

When transmitting an audio signal, e.g. in the case of radio transmission, cable transmission, satellite transmission and recording devices, it is known practice for the analogue audio signal to be converted into a digital audio signal of a specific resolution, transmitted in said form and converted upon reproduction back into an analogue signal. Through digital transmission, a greater signal-to-noise ratio is achieved, particularly upon reproduction.

The bandwidth needed to transmit such a signal is substantially determined by the number of samples to be transmitted per unit of time and by the resolution.

In practice, there is a need to keep the bandwidth required for transmission as low as possible in order to make do with a narrow-band channel or to enable as many audio signals as possible to be transmitted simultaneously through one channel. The required bandwidth may in itself be reduced by reducing the samples or the number of bits per sample.

Said measure, however, generally leads to a deterioration in the quality of reproduction. In a known method (DE-OS-35 06 912), to improve the quality of reproduction, the digital audio signal is transformed in chronologically successive portions into a short-time spectrum which, for each of the time portions, e.g. 20 ms, represents the spectral components of the signal. In the short-time spectrum, components which are not discerned by the listener, i.e. are irrelevant in terms of communication, are generally easier to locate on the basis of psychoacoustic properties than in the time domain. Upon transmission, said components are weighted less or completely left out. By virtue of said measure, a considerable portion of the otherwise necessary data may be dropped during transmission so that the average bit rate may be substantially reduced.

In EP-A-0 251 028, signals are divided into sub-blocks and expanded as a function of the signal energy within the sub-blocks prior to transmission.

Experience has shown, however, that in the case of amplitude variations within such a portion, particularly in the case of signals starting from silence part-way into a block, said signals are superimposed by

interference after transmission. The reason why this is discernible is that the interference occurs also before the start of the signals and so is inadequately masked.

Said interference may occur, for example, as a result of quantization noise being superimposed on the short-time spectrum and resulting, after retransformation, also in noise components arising within the entire block in the time domain.

To reduce said interference, each block is divided into sub-blocks and the amplitude jumps of the signals from one sub-block to the next are detected. In the case of amplitude jumps of more than 20 dB, the signals in the preceding block are compressed and then expanded after retransformation.

During said process, the following problems may arise:

- a) The detection of real jumps is unreliable.
- b) Only those energy changes in predetermined sub-blocks which exceed a specific amount are assessed as jumps.
- c) Uncertainties in the energy detection process may lead to unintentional signal peaking in the accentuated signal portion.

The object of the invention is to improve the method described initially in such a way that given signal variations within a block, such as are customary in great variety in natural audio signals, the signal-to-noise ratio after transmission is improved.

Said object is achieved by the features of claim 1.

In principle, in every sub-block of the blocks into which the audio signal is divided by the time window, signal compression is effected. Said compression is no longer made dependent on whether real jumps of the threshold values are being exceeded. It is therefore possible to reduce the effects and interference during the subsequent signal processing, regardless of their origin. Through complementary expansion of the signals

in the blocks after transmission, the original signal ratios are restored, with the signal-to-noise ratio being improved in respect of weak signals.

The invention is described hereinafter with reference to the drawings. The drawings show:

- Fig.1 the time sequence of the method according to the invention;
- Fig.2 windowing of the signal to achieve the successive blocks;
- Figs.3a-d the view of a signal commencing part-way into a block, in its original state and after retransformation;
- Fig.4 the division into sub-blocks;
- Fig.5 a window function, with which the blocks are superimposed;
- Fig.6 the energy characteristic within a block, with and without compression;
- Fig.7 a view of the non-emphasized and emphasized signal upon a signal jump given non-overlapping windows for energy calculation of the sub-blocks;
- Fig.8 a view similar to Fig.7 but with 50% overlapping windows for energy calculation of the sub-blocks;
- Figs.9-11 block diagrams for effecting signal compression or expansion;
- Fig.12 a transmission function of an adaptive filter;
- Fig.13 a block diagram with process steps for the transmitting end;
- Fig.14 a block diagram with process steps for the receiving end;
- Fig.15 a graph illustrating the accentuation factors of the sub-blocks of a block;
- Fig.16 a graph similar to Fig.15 but in which the accentuation factors are set after a signal jump to 1;
- Fig.18 a switching example for a transmitting-end adaptive filter; and
- Fig.19 a switching example for a receiving-end adaptive filter.

In Fig.1, the analogue signal $a(t)$, which is an audio signal such as, for example, speech or music, is converted by the process step shown at 1 into a corresponding digital signal. In the advantageous but not absolutely essential process step shown at 2, the signal is subjected to pre-emphasis, with preferably the higher-frequency spectral components of the audio signal being accentuated compared to the lower-frequency components. Said measure is intended to compensate the effect whereby signals of low

frequencies are poorly masked by signals of high frequencies. In order nevertheless to achieve masking, accentuation has to be effected all the more intensely, the further away the useful frequency lies.

In the process step shown at 3, the signal is windowed, thereby producing chronologically successive blocks with signal portions. The duration of said blocks is preferably 20 ms. The signals of a block may later be subjected separately to further processing.

At 4, the blocks are subdivided by weighting with a window function into sub-blocks. A sub-division into 10-20 sub-blocks is advantageous.

In the next process step designated 5, the signals within each of the sub-blocks are constantly amplified or attenuated, with a plurality of process steps being implemented, as is explained below.

To determine the amplification factors or attenuation factors, the mean outputs are derived from the signal portions corresponding to sub-blocks as detection portions. Dimensions (energy : portion duration) are then formed for the mean outputs in the detection portions. Said dimensions are related for the detection portions of a block to the dimension of the detection portion having the greatest dimension. The amplification factors are then determined from the dimensions, with it being possible to select proportional or quantized amplification factors with different quantization factors.

It is advantageous to select lower quantization stage levels for low amplification factors than for higher ones. The maximum amplification factor is preferably restricted to a value of 40 dB. On completion of said measure, roughly the same signal-to-noise ratio with regard to the useful energy and the interference energy arising as a result of coding is achieved in all of the sub-blocks.

In the process step designated 6, the signal of the block is then converted by transformation into a frequency spectrum. At 7, a coding according to psychoacoustic considerations is effected. This means that spectral components, which upon reproduction are not discerned anyway particularly

on account of masking effects, are weighted less or left out during coding. 8 denotes the process step of transmission or storage, after which the previously described process steps may be substantially reversed in reverse order.

Thus, at 9 there is first a decoding process, in which the transmitted or stored signal is converted back into a frequency spectrum. In the process step denoted 10, the short-time spectrum is translated back into the time domain. At 11, on the basis of the transmitted amplification factors, the signals in the blocks and sub-blocks are complementarily equalized. At 12, the sub-blocks are combined into blocks and at 13 continuous digital signals are reformed from the blocks. On completion at 14 of a de-emphasis organized to complement the pre-emphasis at 2, a digital-to-analogue conversion is effected at 15, after which the analogue signal $b(t)$ may finally be picked off. The analogue signal $b(t)$ is not identical to the analogue signal $a(t)$ because, during coding, spectral components have been weighted differently or suppressed. The difference between the analogue signals $b(t)$ and $a(t)$ is however such as to be imperceptible to the listener upon reproduction.

Figs. 2 a, b, c show the windowing of the signal a' , as a result of which chronologically successive blocks 17, 18 and 19 are produced. This may be effected in that the signal a' is multiplied by an amplitude characteristic 16. In a preferred embodiment, the characteristic initially has a sinusoidal shape over a quarter period, then a constant value and finally a cosinusoidal shape. The continuous shape at the ends is to prevent a very broad frequency spectrum arising upon subsequent transformation. Furthermore, the windows are so placed that an overlapping of the blocks occurs. In the following, however, for greater clarity, rectangular amplitude characteristics are assumed.

The windows have a duration of 20 ms. Said time has proved advantageous in order, on the one hand, to obtain a sufficiently finely resolved short-time spectrum which upon subsequent coding contains enough spectral lines to enable an effective data reduction to be carried out and, on the other hand, upon changes in the signal to adjust the amplification relatively quickly to the varied state and hence utilize the psychoacoustic pre- and

post-masking effect.

In practice, however, cases are possible where a signal suddenly starts from silence and said start occurs within a block, e.g. in the second half of the block. Such a case is shown in Fig.3a. The transformed signal is shown in Fig.3b. As a result of quantization errors during coding, an interference spectrum is superimposed on the spectrum shown in Fig.3b, thereby producing the spectrum shown in Fig.3c. After retransformation, said interference spectrum influences the signal characteristic, not from the start of the signal onwards, but from the very start of the block, as Fig.3d shows. Since the pre-masking effect is lower than the post-masking effect, said interference may become audible. By appropriate compression within the block prior to transformation and transmission, and by expansion after transmission and retransformation, the signal-to-noise ratio may be substantially improved.

To said end, each block 16, 17 ... is subdivided into sub-blocks, as shown in Fig.4. Said sub-blocks 19, 20, 21... have, apart from at the block edges, identical time dimensions in such a way that each overlaps half of the other. At the block edges there is an overlap with sub-blocks 18 of half the time dimension. In said overlapping rectangular sub-blocks, the mean signal outputs are determined (energy in the time portion divided by the dimension of the time portion).

In a next step, as Fig.5 reveals, the initially rectangular sub-blocks 19, 20, 21... are weighted with $\cos^2$ window functions 22. The time portions at the block edges, which have only half the time dimension of the other sub-blocks, are weighted with a $\cos^2$ half-window 23. The overlapping weighting functions sum at each instant of the signal block to 1.

Fig.6 shows how the signals in the sub-blocks 19, 20, 21... in accordance with the detected mean outputs, as illustrated by the unbroken lines, are amplified or attenuated in such a way that the mean outputs in the sub-blocks 19, 20, 21... become approximately equal, as illustrated by the dashed lines. For greater clarity, the blocks here are illustrated not by dashed lines. The amplification and attenuation of the signals does not alter the ratio between the block useful energy and the block interference

energy arising as a result of coding. On the other hand, the effect achieved by said measure is that there is now the same signal-to-noise ratio in all of the sub-blocks. As regards the signal-to-noise ratio, therefore, the same effect is achieved as if from the start blocks had been selected through windowing which corresponded to the size of the sub-blocks. The previously mentioned drawbacks of shorter blocks are however avoided.

The overlapping sub-blocks are advantageously, for psychoacoustic reasons, dimensioned to a time dimension of around 2-4 ms. Given blocks with around 1000 samples and a sampling frequency of 44.1 kHz, this corresponds to the formation of around 10-20 sub-blocks. It is further advantageous for psychoacoustic reasons to limit the signal amplification to a maximum value of, for example, 40 dB.

It is sufficient to quantize the amplification factors, with it being possible to effect the quantization relatively coarsely in order to limit the additional data required for the quantization stages. Quantization may be effected in such a way that lower quantization stage levels are selected for smaller amplification factors than for larger ones. Quantization is so dimensioned that the mean output in the accentuated sub-block does not exceed the mean output in the sub-block with the highest detected mean output, i.e. the reference block. In said manner it is possible even to obtain a gain in the ratio of the block useful energy to the block interference energy. In such a case, the signal-to-noise ratio of all sub-blocks is however no longer identical but merely approximately identical.

When only the sub-blocks, in which compression of the signal occurs, are weighted by overlapping window functions but not the sub-blocks, which are used to determine the mean signal outputs for calculating the amplification factors, specific signal jumps may give rise to excessively high amplification factors.

Said case is illustrated in Fig.7 for an ideal rectangular jump. The non-accentuated signal characteristic is designated 26, the accentuated signal characteristic is designated 27. The lower-case letters a_0 to a_n stand for the accentuation factors, also known as amplification factors. The

excessive increase occurs when the jump flank and the sub-block boundary do not coincide.

To keep the excessive increase low, according to a development determination of the mean signal outputs is effected likewise in blocks with a 50% overlap, but in this case with rectangular windows. They correspond directly to the sub-blocks in which the signals are amplified. The outcome of said measure is shown in Fig.8 for the same signal jump. 26 again denotes the non-accentuated signal characteristic, 28 this time denotes the accentuated, altered signal characteristic.

When the hitherto described method is applied to the complete audio signal, the amplification factors are then correct only for the energy-rich spectral components since the latter substantially determine the factors. In audio signals, the spectral components up to around 3 kHz are almost always the richest in energy.

If the method presents the greatest representational accuracy for the energy-rich spectral components up to around 3 kHz, then signal jumps at higher frequencies with lower energy fractions may lead to greater inaccuracies during coding, this resulting in possibly audible interference.

According to a first alternative, the signal may be subjected prior to transmission and coding to pre-emphasis and after transmission and decoding to de-emphasis.

Figs.9 to 11 show block diagrams for effecting a compression or expansion together with pre-emphasis or de-emphasis. The circuits each contain a controllable amplifier 24 and a filter 25.

In Fig.9, only the controllable amplifier 24 is disposed in the signal path. The filter 25 is situated in the control signal path. If said circuit through inverse control is used for complementary expansion, the filter 25 may be dispensed with.

In Fig.10, the filter 25 is connected upstream of the controllable amplifier 24. Said refinement requires a filter also during expansion.

In Fig.11, the filter 25 is likewise connected upstream of the controllable amplifier 24 and is in addition itself controllable. The critical frequency may be shifted in such a way that in practically every case masking is achievable. During expansion, the controllable filter 25 and the controllable amplifier 24 are to be inversely controlled.

A further alternative is such that, instead of a fixed pre-emphasis and de-emphasis, a selective-frequency signal accentuation is effected, using an adaptive filter. The transmission function of such an adaptive filter for various amplifications is shown in Fig.12. The filter accentuates only the medium frequencies and disregards the bottom and top frequencies.

Figures 13 and 14 show block diagrams with process steps for effecting the method including adaptive filtering, with Fig.13 relating to the transmitting end and Fig.14 to the receiving end.

At the transmitting end, the input signal is first reduced with the aid of a fixed bandpass filter 29 to the signal components which are relevant for determining the filter coefficients. Said bandpass filter 29 should, in its characteristic, be comparable to the adaptive filter as shown in Fig.12. The output signal of said bandpass 29 is at 30 divided, for example, into blocks having preferably 1024 samples, with the blocks each overlapping by 6.25%. The blocks thereby produced correspond in their time position to the blocks which are later supplied to the adaptive filter. Alternatively, the fixed bandpass filtering may be effected after a corresponding block formation for an individual block. None of the subsequent processing steps are changed thereby.

In 31, the filtered signal blocks are divided by 50% overlapping rectangular windows into, for example, 9 sub-blocks. The mean signal output from said blocks is determined and the maximum of said signal outputs is defined as a reference quantity. Then, in 32, the accentuation factors for each sub-block, which relate to said reference quantity, are calculated. Mathematically, the accentuation factors $f_i = P_{ref}/P_i$, in

which f_i is the factor of the i -th block, P_{ref} the reference quantity and P_i the mean output of the i -th block.

The factors f_i should apply at the supporting points marked with a cross in Fig.15 and are communicated to the receiving end. For complete adaptive filtering, the existing 9 values have to be extrapolated to 1024 in 33 in that, between the supporting points, the factors f_i are softly cross-faded into one another. The corresponding factor characteristic for all 1024 samples of a block is shown for an example in Fig.15 by the solid line when all factors remain unchanged.

Taking psychoacoustic criteria into account, however, the factors may be set to 1 after a signal jump since the post-masking effect is better than the pre-masking effect. Said case is reproduced in Fig.16. Cross-fading is effected in the present case by means of $\cos^2$ functions.

From the illustrated factor characteristics, the filter coefficients a_1 to a_3 for an adaptive filter 35, of the kind illustrated by way of example in Fig.17, may be calculated in 34 for each sample. The result is an adaptive filter, whose step response alters when each sample is read in. Said filter is so dimensioned that it may be varied, say, between the absolute-value function shown in Fig.12. The transmission function for the factor 1 is characterized there by H_{min} and the transmission function for the maximum permitted factor of preferably 10,000 is characterized by H_{max} . The adaptive filter thus calculated is applied to the unfiltered block obtained in 36, which was removed from the continuous signal and has already been weighted with the analysis window. Thus, only the signal components in the medium frequency range are accentuated, while low- and high-frequency signal components remain substantially unchanged. The block thus filtered is supplied to the encoder 37.

The factors f_i may, for example, be logarithmically quantized and transmitted with, in each case, 5 bits. From these, in the receiver, after corresponding extrapolation in 38, are then recovered in 39 the 1024 factors which determine the filter coefficients b_1 to b_3 for an inverse filter 40, an example of which is shown in Fig.18. With said filter, the block after decoding in 41 is filtered in the receiver in order to reverse



Claims

1. Method of transmitting an audio signal (a'), which is divided (30) by time window (3) into chronologically successive blocks (17, 18, 19), then the signal components contained in the blocks are converted by transformation (6) into a short-time spectrum, then the short-time spectra are coded on the basis of psychoacoustic characteristics (7, 37) and transmitted (8) and after transmission are decoded (9), the short-time spectra are translated by retransformation (10) back into the time domain and finally the blocks in the time domain are joined together (13), with the blocks being subdivided into sub-blocks (4, 31), the signals in the sub-blocks being subjected prior to transformation to compression/expansion (5) and after retransformation to complementary expansion/compression (11), characterized in that the signal as a function of the mean signal outputs in the sub-blocks (4, 31) is amplified and/or attenuated (32) in said sub-blocks during compression/expansion and during expansion/compression is subjected to complementary treatment (40), with lower mean signal outputs being linked with high amplification factors and high mean signal outputs being linked with lower amplification factors.
2. Method according to claim 1, characterized in that the maximum change in the time-dependent amplification and/or attenuation is limited.
3. Method according to claim 1 or 2, characterized in that the subdivision of the blocks into sub-blocks (31) is effected by weighting with window functions (4, 16) which have a constant value over the time portion.
4. Method according to claim 3, characterized in that the window functions (16) overlap and the blocks at their two edges are weighted with window functions of half the time width.
5. Method according to claim 4, characterized in that the blocks are weighted with further "soft" window functions, which have a continuous shape, become zero at the boundaries of the sub-blocks and



sum in the entire block to a constant weighting value, preferably to 1.

6. Method according to one of claims 3 to 5, characterized in that the window functions, apart from at the block edges, are Hanning windows corresponding to $\cos^2$ windows and at the block edges are $\cos^2$ half-windows.
7. Method according to one of claims 3 to 6, characterized in that the amplification factors and/or attenuation factors are derived from the mean outputs in the signal portions corresponding to sub-blocks, as detection portions.
8. Method according to one of claims 3 to 6, characterized in that the detection portions like the sub-blocks are gained by weighting of the blocks with window functions which have a constant value over the time portion.
9. Method according to claim 8, characterized in that the window functions overlap.
10. Method according to one of claims 7 to 9, characterized in that dimensions, e.g. energy divided by portion length, are formed for the mean outputs in the detection portions.
11. Method according to claim 10, characterized in that the dimensions for the detection portions of a block are related to the dimension of a detection portion, preferably the detection portion having the greatest dimension, and that the resulting dimension ratios determine, preferably proportionally, the amplification factors.
12. Method according to claim 11, characterized in that quantized amplification factors are selected.
13. Method according to claim 12, characterized in that the quantization is effected non-uniformly, preferably in such a way that lower quantization stage levels are selected for low amplification factors



than for higher ones.

14. Method according to one of claims 1 to 13, characterized in that the amplification factors are limited to a maximum value of preferably 40 dB.
15. Method according to one of claims 1 to 14, characterized in that the signals are subjected, prior to block formation, to pre-emphasis (24, 25) and, after transmission and joining together of the blocks, to complementary de-emphasis.
16. Method according to claim 15, characterized in that the pre-emphasis accentuates higher-frequency spectral components of the audio signal compared to lower-frequency components.
17. Method according to one of claims 1 to 14, characterized in that the signals, which are used to determine the mean signal outputs in the sub-blocks, are filtered as a function of frequency, preferably with the aid of a bandpass filter (29), and that compression and expansion is likewise effected as a function of frequency.
18. Method according to claim 17, characterized in that from the signals, which are used to determine the mean signal outputs in the sub-blocks, dimensions are obtained for varying an adaptive bandpass, by means of which compression and expansion are effected.
19. Method according to claim 18, characterized in that the adaptive bandpass takes the form of a transverse filter with three coefficients.
20. Method according to claim 18 or 19, characterized in that from the dimensions, which initially for each sub-block are related to the sub-block with the highest signal output, intermediate values are extrapolated for each sample in the sub-blocks and that the adaptive filter for each sample is varied according to the extrapolated intermediate values.



21. Method according to claim 20, characterized in that in the sub-blocks following the sub-block with the highest signal output, the dimensions are set equal to the reference value of said sub-block with the highest signal output, i.e. in said sub-blocks there is no longer any change of the adaptive filter in dependence upon the samples.
22. Method according to claim 20 or 21, characterized in that only the dimensions corresponding to the number of sub-blocks and serving as supporting values are transmitted and possibly the extrapolation specification is additionally transmitted and that at the receiving end the extrapolated intermediate values are reconstructed and inverted to effect expansion.
23. Use of the method according to one of claims 1 to 22 for radio transmission, cable transmission, satellite transmission or in recording devices.