

For the transmission of an audio signal, e.g. for broadcast transmission, cable transmission, satellite links and recording devices, it is known from WO 81/02234 and WO 83/03935 for the analog audio signal to be converted into a digital audio signal with a particular resolution (bits per sample), to be transmitted in this form and to be converted back into an analog signal when reproduced. By means of the digital transmission a greater signal/noise ratio is obtained particularly at the reproduction stage.

In WO 81/02234 it is furthermore disclosed that prior to the transmission the signal is converted into a signal $A(f)$ representing the short-time spectrum and components of said signal are on the basis of psychoacoustic laws (physiological properties of the human ear) weighted differently in their representational accuracy during the coding (spectrum-weighting-stage 3 and data reduction-stage B) of the analog signal to be transmitted.

The band width required for the transmission of such a signal is determined in the main by the number of samples to be transmitted per unit of time and by the resolution (bits per sample).

In practice there is a requirement to keep the band width required for the transmission as small as possible, so as to be able to make do with a narrow-band channel or to transmit as many audio signals as possible simultaneously via an existing channel. The necessary band width may be reduced in principle by a reduction in the samples or the number of bits per sample. This measure usually however results in deterioration at the reproduction stage.

The aim of the invention is to create a method for transmitting a digital audio signal which makes do with a smaller mean bit rate per sample without the quality being noticeably affected at the reproduction stage.

Said aim is achieved by the invention described in claim 1. Advantageous developments of the invention are described in the sub-claims.

The invention is based on the following findings and considerations. The digital audio signal in the time domain, which signal thus represents successively in time the individual samples of the analog signal, is transformed in sections succeeding one another in time into a short-time spectrum, which represents in each case for the short time, e.g. 20 ms, the spectral components of the signal. In the short time spectrum it is generally possible, on the basis of psychoacoustic findings, for components which are not detected by the listener, i.e. are irrelevant in a communications sense, to be detected better than in the time domain. Consequently such components are according to the invention weighted less strongly or omitted completely during the transmission. By means of said measure a considerable portion of the bits required in theory may be dropped during the trans-

mission, so that the mean bit rate for a particular number of samples may be reduced considerably.

Irrelevant components of the spectrum of this kind are e.g. components which do not exceed a particular amplitude relative to a maximum amplitude in close proximity in the spectrum. Such components may preferably be omitted during the transmission without noticeable impairment of the reproduction quality. Components of this kind are in fact not detected by the listener on the basis of psychoacoustic laws due to a so-called masking effect. For a masking effect of this kind the position in time between the components and their mutual frequency and amplitude position are critical.

Since certain time constants are determinative for the masking effect, the signal is divided up by time windows into blocks succeeding one another in time with a duration of the order of magnitude of 20 ms, since within a period of this kind the masking thresholds do not decrease significantly. The signal of such a block is then transformed in each case into the short-time spectrum and checked during the coding for irrelevant components.

The human ear forms frequency bands, so-called groups, which play a significant role in the masking effect. The subdivision into frequency groups and the masking are known and described on p. 46 ff. in E. Zwicker, Psychoakustik, Springer-Verlag, Berlin, Heidelberg, New York, 1982. Each of said frequency groups is evaluated individually during the coding and checked for irrelevant components. The frequency resolution is moreover so fine that the narrowest frequency group with a width of some 100 Hz is still detected. For the conversion of the signal out of the time domain into the frequency domain a Fourier transform, e.g. a so-called Fast Fourier transform, is preferably used.

Further material features and advantages of the solution according to the invention follow from the exemplifying embodiment, which is explained below with reference to the drawing. Here

- Fig. 1 shows the sequence in time of the method according to the invention,
- Fig. 2 the windowing of the signal in order to obtain the successive blocks,
- Fig. 3 amplitude characteristic curves used for the windowing,
- Fig. 4 subdivision of the frequency band of the short-time spectrum into individual frequency ranges and the amplitude evaluation of the components and
- Fig. 5 an additional treatment of the signal for a specific signal content.

In Figure 1 the analog signal $a(t)$, which represents an audio signal such as e.g. speech or music, is converted in the analog-to-digital convertor 1 into a corresponding digital audio signal. In stage 2 there takes place by means of time windows succeeding and overlapping one another in time a so-called windowing of said

signal. The signal is in so doing subdivided into blocks succeeding one another in time and each 20 ms in duration, in such a way that in each case the signal of a block may be further processed separately on its own. In stage 3 a pre-processing of the signal takes place, the significance of which will be explained later. In stage 4 the digital signal of a time window or block is in each case converted into a frequency spectrum by means of a transform. At the output of stage 4 there is therefore present during the blocks succeeding one another in time in each case a signal which represents for the duration of a time window or block the spectral components of the signal over the entire frequency band. Stage 4 therefore produces the conversion of the signal in the time domain into the signal representing the spectrum in the frequency domain.

From stage 4 the signal passes to the coder 5. Here a coding according to psychoacoustic principles takes place. This means that spectral components which are in any case not detected at the reproduction stage in particular by virtue of masking effects are weighted less strongly or omitted during the coding. A processing of this kind of the short-time spectrum is possible e.g. by means of a computer.

The signal coded in this way passes via the transmitter 6 to the transmission channel 7. Due to the reduction achieved in the mean bit rate said transmission channel may accordingly be constructed as a narrow-band one. The transmission channel 7 is followed by the receiver 8, which performs essentially the inverse functions to the transmitter. The signal passes next to a decoder 9, which brings about the decoding in accordance with the coder 5. In stage 10 the spectrum-representing signal obtained in this way in the frequency domain is converted back into a digital signal in the time domain. In stage 11 the signal is assembled again into a uniform continuous digital signal and the pre-processing of stage 3 taken into account. The signal is then fed to the digital-to-analog convertor 12. The convertor 12 again supplies the analog signal $b(t)$. Said signal is not identical to the signal $a(t)$, because in the coder 5 during the coding spectral components have been weighted differently or suppressed. The difference between the analog signals $b(t)$ and $a(t)$ is however such that said difference is not noticed by the listener at the reproduction stage. In the signal, therefore, only irrelevant information, inaudible to the listener, is removed, in order to reduce the required bit rate during the transmission via the transmission channel 7, in particular the decision content is reduced.

Figure 2 A - C show the windowing of the signal in stage 2. The signal $a(t)$ is evaluated by means of a time window with the amplitude characteristic curve WA of $t_1 - t_7$, i.e. the signal values of the signal $a(t)$ as per Figure 2 A are multiplied by the values of the characteristic curve WA as per Figure 2 B. The characteristic curve has from t_1 to t_2 a sinusoidal shape over a quarter cycle and from t_6 to t_7 a cosinusoidal shape over a quarter cycle and between the latter from t_2 to t_6 it has the constant value 1. The steady shape of $t_1 - t_2$ and $t_6 - t_7$ is

essential, since with such a subdivision into blocks succeeding one another in time, the so-called windowing, a jump between zero and a final value would produce a very broad frequency spectrum in the case of a Fourier transform. Multiplication of the signal $a(t)$ with the characteristic curve WA produces the signal A as per Figure 2 C, which is therefore distorted in amplitude in the periods $t_1 - t_2$ and $t_6 - t_7$. Said windowing, i.e. time-selective evaluation, which takes place in Figure 1 in stage 2, is represented by means of analog signals in Figure 2 A - C. In reality the signal $a(t)$ and WA are digital signals, so that from t_1 to t_7 respectively digital signal values are multiplied by each other.

In each case the signal of a time window $t_1 - t_7$, i.e. of a block, is converted as per Figure 1 in stage 4 into the short-time spectrum. Figure 2 D shows a plot of the amplitude versus the frequency f and Figure 2 E a plot of the phase versus the frequency f for such a signal. On the receive side the plot shown in the Figures 2 A - E takes place in reverse sequence. The amplitude and the phase are received as digital signals and converted back into an analog signal $b(t)$ by means of an Inverse Fast Fourier Transform and a digital/analog conversion.

Figure 3 shows the amplitude characteristic curves for the successive time windows on the transmit side and on the receive side. The time windows on the transmit side in stage 2 have the shape WA (A = analysis) as per Figure 3 A. $t_1 - t_2$ and $t_6 - t_7$ are overlapping periods in which two time windows are effective. One time window extends over 1024 samples, the overlapping amounting to some 64 samples, i.e. 6.25% of the duration of the time window. The amplitude characteristic curve shown in Figure 3 A applies in identical manner to the receive side and is plotted in stage 11 (WS = window synthesis). In the period $t_2 - t_6$ the signal traverses without affecting the amplitude, hence with the factor 1, the path shown in Figure 1. A sample during the overlapping period $t_1 - t_2$ is first of all multiplied by the sine in stage 2. Since the same curve is effective on the receive side in stage 11, said sample will be multiplied altogether by $\sin^2(x)$, which would mean a weakening in the amplitude of said sample by the factor $\sin^2(x)$. During the period $t_1 - t_2$, however, the descending, cosinusoidal slope of the preceding time window is effective additionally, so that the same sample is due to the cosinusoidal slope multiplied altogether by $\cos^2(x)$. Because of the linear overlapping of the two signals and the mathematical relationship

$$\sin^2(x) + \cos^2(x) = 1$$

there is thus obtained also during the overlapping periods the constant amplitude factor 1. This is shown in Figure 3 B by the upper horizontal line in the overlapping periods $t_1 - t_2$ and $t_6 - t_7$. Altogether there is thus obtained continuously for the signal the amplitude transmission value 1, so that no amplitude distortion of the signal takes place.

Figure 4 shows the amplitude plot of the spectral lines of a short-time frequency spectrum, as is obtained at the outset of stage 4 in Figure 1. The whole frequency band of the short-time spectrum $f_1 - f_{15}$ is subdivided into a plurality of frequency groups $f_1 - f_2$, $f_2 - f_4$, $f_4 - f_{12}$, $f_{12} - f_{14}$ and $f_{14} - f_{15}$. In the individual frequency groups the spectral lines are examined and weighted according to psychoacoustic principles. Only the dominant amplitude values are transmitted, irrelevant amplitude values are weighted less strongly or suppressed. The absolute maximum 15 of the whole frequency band is first of all transmitted as an absolute value with 12 - 16 bits. The maxima 14, 16, 17, 18 of the remaining frequency groups are transmitted with an accuracy of 8 bits, i.e. in their relative positive to the absolute maximum 15. The remaining values 20 - 26 of the frequency group $f_4 - f_{12}$ are for their part related to the maximum value 16, i.e. their deviation from the maximum value 16 is transmitted. To this end the amplitude range is, starting from the maximum 16, subdivided into three ranges A1, A2, A3 each of 10 dB and one range A4 for the remainder. Signal values 16, 20, 21 or 22, 23 or 24 or 25, 26, lying in each case in an amplitude range, are transmitted as an identical value. No distinction is therefore drawn between the values 16, 20, 21 and 22, 23 and 25, 26. The amplitude values 25, 26 at the frequencies f_{10} , f_{11} , which fall below the value of 30 dB below the maximum 16, are set at zero. The phase of the values 25, 26 is not transmitted. Said spectral components would because of their close position to the value 16 and their small amplitude in any case no longer be detectable by virtue of the masking effect. In practice the whole frequency band $f_1 - f_{15}$ is divided up into 26 frequency groups, of which only 5 are shown in Figure 4 for the sake of simplicity. Due to the dividing up into the amplitude ranges A1, A2, A3, A4 a total of 2 bits is sufficient for the transmission of the amplitude values 20 - 26 relative to the maximum 16. For each transmitted amplitude value which lies in the ranges A1 to A3, 2 bits are transmitted for the associated phase value.

A considerable reduction in the amount of data required for the transmission is already achieved by the coarse quantization of the amplitude and phase values with 2 bits. An additional saving on bits is also made during the transmission, however, by the dropping of components, namely of the phase values for the amplitude values 25, 26 in the frequency group $f_4 - f_{12}$. Said liberated bits may be used for the transmission of a more detailed amplitude subdivision in the ranges A1 - A3. To this end, for example, each range A1 - A3 is divided into two ranges of 5 dB each. Moreover there is assigned to each frequency value 20 - 24 a bit which indicates whether the amplitude value, e.g. 20, 21, lies within the first 5 dB or the second 5 dB below the maximum 16. The assignment of said bits takes place on the basis of a table 28 which is created on the transmit side and is reconstructable on the receive side. To this end there is placed over the whole plot of the frequency spectrum as per Figure 4 a grid 27 with a gradation in stages of 6 dB. The amplitude values 20 - 24 are therefore assigned to said grid. The table 28 assigns to each amplitude value 20 - 24 a particular position in relation to the maximum 16. The table 28 begins with the lowest frequency

value and shows by means of the line the respective position in relation to the maximum of the corresponding frequency range.

If further free bits are available, a subdivision of the 5 dB ranges into 2.5 dB ranges will be carried out. The subdivision process may be continued indefinitely. The saving on bits and the use of said bits for refinement of the resolution is termed adaptive quantization.

Figure 5 shows the pre-processing of a sudden acoustic event 29 which occurs within a time window $t_1 - t_7$ at the point in time t_9 . Such an acoustic event may be e.g. the striking of a triangle. The pre-processing described takes place in Figure 1 in stage 3. The acoustic event 29 is also preceded by a preshoot between t_8 and t_9 , which is however inaudible due to a preliminary masking. During the conversion into the frequency spectrum in stage 4 in Figure 1 there arises in each case in the frequency domain a signal which indicates the spectral distribution in the window $t_1 - t_7$. Since in the case of said signal the assignment of spectral lines to individual points in time within a time window no longer takes place, the event 29 averaged over the whole time window $t_1 - t_9$ would therefore be as it were blurred. An audible distortion may occur as a result.

In order to prevent said still conceivable defect, a time window $t_1 - t_7$ or else block is subdivided into 32 sub-blocks. The amplitudes of the individual sub-blocks are determined. As soon as an amplitude jump of more than 20 dB occurs between two sub-blocks, produced in Figure 5 by the event 29, an additional measure will be triggered. Said measure consists in the fact that prior to the amplitude jump the signal is by means of a compander process increased on the transmit side and correspondingly reduced again on the receive side. The above-mentioned defects caused by the blurring of the short-time event over the whole time window will thereby be reduced.

Fig. 6 shows an alternative to the transmission of more detailed amplitude subdivisions. A grid 40 is laid over the frequency spectrum. The uppermost limit of said table coincides with the absolute maximum 15. The grid 40 comprises not more than sixteen 6 dB gradations and thus extends over a range of 96 dB. On the basis of said grid 40 a table 41 is created. The table is created on the receive side and is reconstructable on the transmit side. This means that the position of the values 42, 44, 43, 20, 21, 22, 23 and 45 in the table is reconstructable. The table indicates by means of a bit (0 or 1) whether a pertinent value lies in the 5 dB range or in the 6 - 10 dB range of the ranges A1 - A3, A5 - A7. For the more detailed amplitude sub-division only one bit is therefore transmitted for a value. For the value 20 a 0 will be transmitted, in order to indicate that the value 20 lies in the 0 - 5 dB range of A1. The values 42 and 44 will be assigned to the 0 dB line, the values 43, 20, 21 to the 6 dB line, the values 22, 23 to the 12 dB line, the value 45 to the 18 dB line. The bits corresponding to the values will be transmitted in lines starting with the 0 dB line as per table 41. The transmission time is the same for all frequency spectra.

According to the number of bits saved on by means of the coarse quantization of the amplitude and phase values, the same number of additional bits may be transmitted from the table 41. This means that bits from the first line 0 dB will be transmitted with a higher probability than bits from the following lines. The dotted line 46 sets the boundary between the lines 47 and 48. The bit for the value 23 will be transmitted with greater probability than the bit for the value 45, although the magnitude of the value 45 is greater than the magnitude of the value 23.

Claims

1. Method for transmitting an audio signal, in which the analog signal is converted into a digital signal, is prior to the transmission converted into a signal representing the short-time spectrum, and the whole frequency band is divided up into a plurality of frequency groups, signal components in the frequency groups are on the basis of psychoacoustic laws weighted differently in their representational accuracy during the coding of the signal to be transmitted, the code signal is transmitted, and after the transmission decoded, inverse-transformed and converted back into an analog signal, characterised in that the value of the absolute maximum magnitude of the whole frequency band as an absolute quantity and the maximum magnitudes of the remaining frequency groups are divided up into value ranges of the respective transformed magnitude values in the frequency groups starting from the respective maximum magnitude above a threshold and the magnitude values lying in the same value range are transmitted as the same value, that magnitude values below the threshold are set as zero and free bits not required so far for the transmission are used for the more detailed subdivision of the value ranges.
2. Method according to claim 1, characterised in that the conversion is carried out by means of a Fourier transform (FT).
3. Method according to claim 2, characterised in that the transform is a Fast Fourier transform (FFT).
4. Method according to claim 1, characterised in that the short time lies in the order of magnitude of 20 ms.
5. Method according to claim 1, characterised in that components of the signal are weighted less strongly or suppressed for the coding.



According to the number of bits saved on by means of the coarse quantization of the amplitude and phase values, the same number of additional bits may be transmitted from the table 41. This means that bits from the first line 0 dB will be transmitted with a higher probability than bits from the following lines. The dotted line 46 sets the boundary between the lines 47 and 48. The bit for the value 23 will be transmitted with greater probability than the bit for the value 45, although the magnitude of the value 45 is greater than the magnitude of the value 23.

Claims

1. Method for transmitting an audio signal, in which the analog signal is converted into a digital signal, is prior to the transmission converted into a signal representing the short-time spectrum, and the whole frequency band is divided up into a plurality of frequency groups, signal components in the frequency groups are on the basis of psychoacoustic laws weighted differently in their representational accuracy during the coding of the signal to be transmitted, the code signal is transmitted, and after the transmission decoded, inverse-transformed and converted back into an analog signal, characterised in that the value of the absolute maximum magnitude of the whole frequency band as an absolute quantity and the maximum magnitudes of the remaining frequency groups are divided up into value ranges of the respective transformed magnitude values in the frequency groups starting from the respective maximum magnitude above a threshold and the magnitude values lying in the same value range are transmitted as the same value, that magnitude values below the threshold are set as zero and free bits not required so far for the transmission are used for the more detailed subdivision of the value ranges.
2. Method according to claim 1, characterised in that the conversion is carried out by means of a Fourier transform (FT).
3. Method according to claim 2, characterised in that the transform is a Fast Fourier transform (FFT).
4. Method according to claim 1, characterised in that the short time lies in the order of magnitude of 20 ms.
5. Method according to claim 1, characterised in that components of the signal are weighted less strongly or suppressed for the coding.



6. Method according to claim 1, characterised in that components of the signal with decreasing amplitude are weighted less strongly.
7. Method according to claim 1, characterised in that components of the signal below a certain amplitude are suppressed for the coding (25, 26 in Fig. 4).
8. Method according to claim 1, characterised in that components of the signal are weighted less strongly or suppressed as a function of their amplitude and their frequency spacing relative to masking spectral components of greater amplitude.
9. Method according to claim 1, characterised in that the signal is divided up by time windows ($t_1 - t_7$) into blocks succeeding one another in time which are each converted into a signal sequence representing the short-time spectrum.
10. Method according to claim 9, characterised in that the time windows overlap one another in time.
11. Method according to claim 10, characterised in that the overlapping time ($t_1 - t_2$; $t_6 - t_7$) lies in the order of magnitude of 5 - 8%, e.g. 6.25% of the duration of one time window.
12. Method according to claim 9, characterised in that the amplitude characteristic curve (WA, WS) of the time windows is identical on the transmit side and the receive side (Fig. 3).
13. Method according to claim 9, characterised in that the amplitude characteristic curve (WA, WS) of the blocks has a continuously rising or descending shape during the overlapping time ($t_1 - t_2$; $t_6 - t_7$).
14. Method according to claim 12 and/or 13, characterised in that the plot of the amplitude characteristic curve (WA, WS) during the overlapping time is shaped so that during said time the sum of the products of the amplitude characteristic curve (WA, WS) over transmit side and receive side equals 1.



15. Method according to claim 14, characterised in that the descending slope of the amplitude characteristic curve at the end of a time window has a cosinusoidal shape and the rising slope of the characteristic curve at the beginning of the time window has a sinusoidal shape.
16. Method according to claim 1, characterised in that in each case the values of the amplitude maxima (14, 15, 16, 17, 18) of a frequency group are transmitted (Fig. 4).
17. Method according to claim 9, characterised in that each individual block is divided up into a plurality (32) of sub-blocks and whenever an amplitude jump above a particular value (29 dB) occurs from one sub-block to the next, the amplitude values in the sub-blocks coming before in time are raised by means of a compander system for the transmission (Fig. 5).
18. Method according to claim 2, characterised in that the spectrum is represented according to magnitude and phase.
19. Method for transmitting an audio signal, in which the analog signal is converted into a digital signal, is prior to the transmission converted into a signal representing the short-time spectrum, in which the whole frequency band is divided up into a plurality of frequency groups, signal components in the frequency groups are on the basis of psychoacoustic laws weighted differently in their representational accuracy during the coding of the signal to be transmitted, the code signal is transmitted, and after the transmission decoded, inverse-transformed and converted back into an analog signal, characterised in that the transformed magnitude values are divided up into frequency groups starting from the respective maximum magnitude above a threshold in the value range and the magnitude values lying in the same value range are transmitted as the same value, and that magnitude values below the threshold are set as zero.
20. Method according to claim 19, characterised in that the subdivision of the value ranges depends on the number of free bits available.