

DOWNMIX LIMITING

Cross-Reference to Related Applications

This application claims priority to United States Patent Provisional Application No. 61/413,237, filed 12 November 2010, hereby incorporated by reference in its entirety.

Technical Field

The invention disclosed herein generally relates to analogue or digital audio signal processing technique. More particularly, it relates to downmixing of a number of audio signals into a smaller number of audio signals.

Technical Background

As used herein, downmixing refers to the operation of deriving N output audio signals (or channels) from information encoded by M input audio signals (or channels), where $1 \leq N < M$. Common expectations on high-quality downmixing include low information loss, compatible dialogue levels and high psychoacoustic fidelity between the input and output signals.

Downmixing frequently includes combining two signals into one, be it by waveform addition, transform-coefficient addition, weighted averaging or the like. While stereo-to-mono downmixing may be expressed by the simple relationship

$$y_1 = \frac{x_1 + x_2}{\sqrt{2}}, \quad (1)$$

general M-to-N downmixing may be written in matrix form as:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix}. \quad (2)$$

Here, the *relative weight distribution* between input channels contributing to a given output channel y_k , as expressed by downmix coefficients $a_{k1}, \dots, a_{kM}$, may follow from artistic considerations or may be related to the spatial layout of the reproducing audio sources. After fixing the relative ratios of the downmix coefficients, the *gain* of the downmixing may be determined by other concerns, notably energy conservation in cases where one input channel contributes to several output channels. In other situations, the priority may be to maintain a consistent dialogue level. This requirement makes it possible to join audio sections seamlessly together although they have been obtained by different types of mixing or encoding.

A difficulty frequently encountered in downmixing, whether the gain has been chosen by energy conservation or in response to a dialogue-level requirement, is that an output signal exceeds its permitted range. To avoid clipping the output signal or damaging the reproducing audio equipment, a common practice in the art is to reduce the gain, either locally – at or around a point in time where out-of-range values would otherwise be produced – or globally. Supposing that output signal y_k is out of range, the overall gain may be limited as per

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \gamma \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix}, \quad (3)$$

where $0 < \gamma < 1$ is a limiting factor. One may also reduce only the gain of the signals contributing to y_k , by

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & & \vdots \\ a_{k-1,1} & \cdots & a_{k-1,M} \\ \gamma a_{k,1} & \cdots & \gamma a_{k,M} \\ a_{k+1,1} & \cdots & a_{k+1,M} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix}. \quad (4)$$

Irrespective of how limiting factors are applied, the requirements of meeting the dialogue level and performing the limiting in a psychoacoustically unnoticeable manner are clearly contradictory. Limiting the gain more locally favours the consistency of the dialogue level but leads to more sudden and more perceptible gain changes. Similarly, performing the limiting over an extended time period improves one problem but worsens the other. Hence, there is need for improved downmixing techniques.

Summary

To overcome, alleviate or at least mitigate one or more of the problems associated with the prior art, it is an object of the present invention to provide techniques for downmixing audio streams in a psychoacoustically less noticeable fashion. A particular object of the invention is to provide downmixing techniques that enable a consistent dialogue level while avoiding clipping the output signal(s). Another particular object of the invention is to provide downmixing techniques having these general properties and being suitable for preserving dynamic, temporal and/or spatial properties of the audio.

The invention achieves at least one of these objects by providing a method, a mixing system and a computer-program product in accordance with the independent claims. The dependent claims define advantageous embodiments of the invention.

In a first aspect, the invention provides a method of downmixing a plurality of
5 input audio signals, which carry input data, into at least one output audio signal. The mixing properties of the method are dependent on maximal downmix coefficients, at least one in-range condition on the output audio signal(s), and a partition of the input signals into subgroups. The method includes deriving downmix coefficients from the maximal downmix coefficients by downscaling all maximal downmix coefficients be-
10 longing to the same subgroup by a common limiting factor in order to meet the in-range condition(s). The downmix coefficients thus derived are suitable for downmixing the input signals.

In a second aspect, the invention provides a mixing system adapted to perform the method of the first aspect. In a third aspect, the invention provides a com-
15 puter-program product for causing a programmable computer to carry out the method of the first aspect.

The invention teaches that a common limiting factor be applied to all downmix coefficients controlling the contributions of the input signals in a subgroup out of at least two subgroups. By this latitude in limiting different input signals to different ex-
20 tents, relatively more perceptible signals can be limited relatively less. This makes it easier to combine a consistent dialogue level with discreet transitions between signal portions with and without gain limiting.

With reference to the appended claims, it is noted that each of the signals may be either analogue (continuous-valued) or digital (discrete-valued). A “sub-
25 group” may include one input signal or several input signals. An “in-range condition” on a signal may refer to an upper bound on the signal, a lower bound on the signal or a requirement for the signal to remain in an interval having a lower and an upper bound. An in-range condition may apply to a particular time segment, a set of time segments or may be global, applying to the entire signal without restriction. It is un-
30 derstood that the terms “in-range condition” and “non-clip condition” may be used interchangeably in this disclosure, as may the terms “limiting factor” and “gain limiting factor”. The limiting factor for each subgroup is determined on the basis of not only the maximal downmix coefficients assigned to the input signals as such, but

also on the basis of the input data carried by the input signals. Finally, it is noted that the downmixing operation itself, that is, forming linear combinations of the input signals to obtain output signals, may be carried out by techniques that are *per se* known in the art.

5 With the exception of non-local in-range conditions, non-local smoothing processes (see below) or similar measures being applied, the invention includes both real-time and offline embodiments, e.g., processing on a file-to-file basis.

 In one embodiment, at least one subgroup comprises two or more input signals. Since a common limiting factor is used to downscale downmixing coefficients
10 for all these input signals, significant relationships between several input signals may be preserved under downmixing. Hence, perceived dynamical, temporal, timbral and/or spatial impressions which are conveyed by the input signals as a whole are only affected to a limited extent by downmixing in accordance with this embodiment.

 In further developments of the preceding embodiment, the input signals cor-
15 respond to spatially related audio channels, such as left and right channels; left, centre and right channels; left and right wide channels; left and right centre channels; and left, centre and right surround channels.

 In one embodiment, the downmix coefficients are maintained as large as possible. This favours a consistent dialogue level. For example, if the in-range condition
20 is a non-strict inequality, the limiting factors may be set equal or close to their upper values (or 'sharp' values, or 'tight' values, or 'exact' values), that is, values which yield equality in the in-range condition. Preferably, the downmix coefficients should not differ more than 20 % from the values determined from the upper bounds, more preferably not more than 10 % and most preferably not more than 5 %. In embodi-
25 ments which further include smoothing of the downmix coefficients (see below), it is preferable to impose one of the above conditions on the values which the downmix coefficients have before smoothing.

 In one embodiment, the output signal is partitioned into time segments. The time segments may have equal or unequal length; they may be the result of sam-
30 pling of analogue data, transform-based processing of a signal or may result from some similar process. A time segment may consist of a number of samples. Alternatively, a time segment may consist of a number of blocks, which each comprise a number of samples. The input signal may be partitioned into similar or different time

segments, or may be non-partitioned. A method according to this embodiment may attempt to satisfy the in-range condition in each time segment separately, in view of the input data relating to this time segment. The method may be configured to satisfy the in-range condition in all time segments or in some time segments. For slowly
5 varying input signals, the latter option may reduce the computational load at limited quality decrease since not all time segments need be considered.

In a variation suitable for providing downmixing into several output signals, the method may be configured to satisfy the in-range condition in separate time segments, however for all output signals jointly. This may preserve the perceived spatial
10 balance of the output signals.

Embodiments for providing output signals partitioned into time segments may advantageously be combined with smoothing (or regularisation). As one example, the values of a particular downmix coefficient obtained for different time segments may be treated as a (time) sequence and may be subjected to a smoothing operation. The smoothed downmix coefficients may be used in the downmixing operation
15 in place of the non-smoothed downmix coefficients. One or several selected downmix coefficients or all downmix coefficients may undergo smoothing; these processes may operate in parallel to one another. Those skilled in the art will realise that smoothing a limiting factor for a particular subgroup will yield the same result as
20 smoothing the downmix coefficients acting on the input signals in this subgroup; therefore, while both these approaches fall within the scope of the invention, this disclosure need not describe both in detail.

The smoothing may be carried out by any suitable process known *per se* in the art. Preferably, the smoothing is governed by an upper bound on the rate of
25 change. After smoothing in this manner, an isolated value in the sequence of segment-wise values will be surrounded by a downward and an upward ramp of moderately changing values, so that an abrupt change is avoided. The ramps may be characterised by constant increase or decrease, on a linear or logarithmic scale, such as the dB scale. Hence, by adjusting downmix coefficient values so that one
30 obtains a smoothed downmix coefficient in which the increase or decrease rate (in absolute values) is not too large, gradual and hence less perceptible transitions between gain limited and non-limited portions of the downmixed signals may be obtained. Another preferable option is to carry out the smoothing by adjusting the

downmix coefficients by either reducing or maintaining the original values. Increasing the original downmix coefficients should be avoided, as an in-range condition may then no longer be satisfied.

In one embodiment, at least one subgroup of input signals is associated with a lower bound on the limiting factor used to determine the downmix coefficients acting on the input signals in that subgroup. The bound is an a priori bound in the sense that this embodiment of the invention attempts to satisfy the in-range condition on the output signal by looking for solutions above the lower bound only. This ensures that the contribution from the concerned subgroup will not become arbitrarily small.

In a further development of the preceding embodiment, a primary and a secondary subgroup are associated with different lower (a priori) bounds on their respective limiting factors. The lower bound associated with the primary subgroup is greater than or equal to the lower bound associated with the secondary subgroup. This may be used to define a relative balance between the subgroups. For instance, the primary subgroup may be given relatively greater psychoacoustic importance than the secondary subgroup.

In another embodiment, the search for limiting factor values by which to satisfy the in-range condition may be configured to favour the primary group. In particular, a method according to this embodiment may be configured to search for limiting-factor values that satisfy the in-range condition where the primary-subgroup limiting factor is equal to or near an upper bound on the limiting factor for the primary subgroup.

In a variation to the preceding embodiment, upper and lower bounds may be defined for the respective limiting factors for the primary subgroup and the secondary subgroup. A method according to this embodiment is configured to initially look for solutions including the primary-subgroup limiting factor being equal to its upper bound. The secondary-subgroup limiting factor is varied between its upper and lower bound. Then, if no solution to the in-range condition is found, the method looks for solutions including the secondary-subgroup limiting factor being equal to its lower bound. The primary-subgroup limiting factor is varied between its upper and lower bound. Put differently, the method initially sets both limiting factors equal to their maximal values (which will best preserve a consistent dialogue level) and then decreases them in a selective fashion until a pair of limiting factors is found by which

the in-range condition is satisfied. The selective decreasing includes initially decreasing the secondary-subgroup limiting factor to its lower bound and then, if needed, decreasing also the primary-subgroup limiting factor. Advantageously, this ensures that the primary channels, which may be defined as the perceptually more important ones, are affected by gain limiting as little as possible.

With reference to the above embodiments wherein a primary and a secondary subgroup are distinguished, the primary subgroup may include signals corresponding to channels that are more important from a psychoacoustic point of view. These include channels intended for playback by audio sources located in a half space in front of a listener; the secondary group may then collect the remaining channels, particularly those intended for playback behind or to the sides of the listener. By another model, the primary channels may be those intended for playback by audio sources located at substantially the same height as a listener (or a listener's ears) and/or propagating substantially horizontally; the secondary group may then contain the remaining channels, for reproduction at other heights or/and propagating non-horizontally. As still another option, the primary subgroup may be composed of channels to be reproduced in the front half space and at substantially the same height as the listener.

In one embodiment, at least one of the subgroups is associated with an upper bound on the limiting factor for that subgroup. In embodiments where several subgroups are assigned an upper bound on their limiting factor and the method is configured to search for largest possible limiting factor values as solutions, the combination of both limiting factors being equal to their upper bounds is an admissible solution. In this situation, it is preferable to set the upper bounds equal, so that the proportions, as expressed by the predefined maximal downmix coefficients, between input signal from *different* subgroups are preserved under downmixing.

One embodiment is configured to provide at least two output audio signals corresponding to spatially related channels. Such spatially related channels may belong to one of the following channel groups or a combination of these: front, surround, rear surround, direct surround, wide, centre, side, high, vertical high. The invention teaches to derive one limiting factor for each subgroup in order to satisfy in-range conditions for all output channels jointly. This may translate the perceived spatial balance of the input signals into a corresponding balance of the output signals,

and may thus avoid undesirable drift of the perceived location of an audio source and similar problems. In one particular embodiment, the determination of a common limiting factor may happen in two substeps. Firstly, downmix coefficients are determined, as products of the maximal downmix coefficients and preliminary limiting factors, which satisfy the in-range condition on each of the (spatially related) output signals which are derived from input signals in the concerned subgroup. Secondly, the limiting factor to be applied to this subgroup is obtained by extracting the minimum of all preliminary limiting factors derived for said output signals in the first substep.

In one embodiment, an encoding system is adapted to receive a plurality of audio signals, to downmix these into at least one downmix signal in accordance with the invention and to encode the downmix signal(s) as a bit stream.

In one embodiment, a decoding system is adapted to receive a bitstream which encodes audio signals and a downmix specification generated in accordance with the invention. The downmix specification may include downmix coefficients and/or a partition of the signals into subgroups. The decoder is further adapted to downmix the audio signals into at least one downmix signal in accordance with the downmix specification, e.g., by applying the downmix coefficients.

In one embodiment, a decoding system may include an input port, a decoder and a mixer. The decoding system is adapted to decode and downmix a signal in accordance with a specification generated in accordance with the invention. As seen above, the invention teaches that downmix coefficients are downscaled in order to meet an in-range condition by a multiplicative limiting factor that is common within each subgroup of signals. This will imply that ratios of coefficients to be applied to signals in one subgroup are constant, while ratios of coefficients to be applied to signals in different subgroups are variable. Here, the terms "constant" and "variable" refer to the possible variation between different sets of downmix coefficients. For instance, one set of downmix coefficients may be computed for each time segment. However, as the invention teaches, the downmixing system will preserve certain ratios between the downmix coefficients within such sets. Because some of the ratios are variable, the decoding system may be adapted to limit relatively more perceptible signals (e.g., in a primary subgroup) relatively less. This makes it easier to combine a consistent dialogue level with discreet transitions between signal portions with and without gain limiting. If a subgroup contains two or more signals, the decoding sys-

tem may preserve significant relationships between these signals under its combined decoding and downmixing, so that perceived dynamical, temporal, timbral and/or spatial impressions which are conveyed by the input signals as a whole are only affected to a small extent

5 It is noted that the invention relates to all possible combinations of features recited in the claims.

Brief Description of the Drawings

The present invention will now be described in more detail with reference to the accompanying drawings, on which:

10 Figure 1 is a generalised block diagram of a portion of a mixing system according to an embodiment;

Figure 2 is a graph illustrating the selection of mixing factors for a primary and a secondary subgroup according to an embodiment;

15 Figure 3 are two graphs illustrating the selection of admissible intervals for limiting factors on the basis of maximal downmix coefficients according to an embodiment;

Figure 4 is a generalised block diagram of a mixing system according to an embodiment; and

Figure 5 illustrates a smoothing process forming part of an embodiment.

Detailed Description of Embodiments

20 Figure 1 shows a portion of a mixing system 100 in accordance with an embodiment of the invention. The system 100 is adapted to satisfy the following in-range condition on the k^{th} output signal:

$$|y_k| \leq \hat{y}_k \quad (5)$$

25 First multipliers 101 and a summer 103 compute the k^{th} output signal on the basis of 1^{st} , 2^{nd} and 4^{th} input signals as per

$$y_k = a_{k1}x_1 + a_{k2}x_2 + a_{k4}x_4,$$

30 where a_{k1} , a_{k2} , a_{k4} are predefined maximal downmix coefficients determining the relative weights of the input signals in the absence of limiting. By a predefined partition, the 1^{st} and 4^{th} input signals belong to a first subgroup, while the 2^{nd} and 3^{rd} input signals belong to a second subgroup. In view of this partition into subgroups, a control-

ler 104 will attempt to satisfy the in-range condition (5) by choosing values of limiting factors $\alpha_1, \alpha_2 > 0$ in

$$y_k = \alpha_1(a_{k1}x_1 + a_{k4}x_4) + \alpha_2 a_{k2}x_2. \quad (6)$$

With reference to figure 1, second multipliers 102 apply the limiting factors α_1, α_2 to the input signals. The controller 104 selects the values of the limiting factors α_1, α_2 in response to the value of the output signal y_k .

With reference now to the whole mixing system 100 discussed above, the action of limiting input signals at downmixing may be expressed as follows in matrix notation. Downmixing without limiting follows a relationship $Y = AX$, where X, Y are input and output signal vectors and

$$A = \begin{bmatrix} a_{11} & \cdots & a_{14} \\ \vdots & & \vdots \\ a_{M1} & \cdots & a_{M4} \end{bmatrix}.$$

Downmixing with limiting follows the equation

$$Y = (\alpha_1 A_1 + \alpha_2 A_2)X$$

with

$$A_1 = \begin{bmatrix} a_{11} & 0 & 0 & a_{14} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & 0 & 0 & a_{M4} \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0 & a_{12} & a_{13} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & a_{M2} & a_{M3} & 0 \end{bmatrix}.$$

Clearly, if one imposes one of the in-range conditions $Y \leq \tilde{Y}$, $\tilde{Y} \leq Y$ and $\tilde{Y} \leq Y \leq \tilde{Y}$, where $\tilde{Y}, \tilde{Y}$ are constant vectors, then the limiting factors α_1, α_2 will be chosen small enough that the in-range conditions on all output signals are satisfied jointly.

The gain limiting according to the invention may be made less perceptible by treating the above subgroups differently. The first subgroup $\{x_1, x_4\}$ may be treated as a primary subgroup, while the second subgroup $\{x_2, x_3\}$ may be treated as a secondary subgroup. For example, the signals in the primary subgroup may correspond to front left and front right signals, which are of primary psychoacoustic significance. Those in the second subgroup may correspond to surround left and surround right, which are intended for playback by non-frontal audio source and therefore carry less significance.

To reflect the unequal significance of the two subgroups, the mixing system 100 according to this embodiment may choose the primary limiting factor from the

interval $L_1 \leq \alpha_1 \leq U_1$ and the secondary limiting factor from the interval $L_2 \leq \alpha_2 \leq U_2$. Suitably, $L_1, L_2 > 0$.

This will now be illustrated by an example in which it is assumed that the upper bounds are equal, which preserves the mixing proportions expressed by the maximal downmixing coefficients where this is possible, and are unity, that is
 5 $U_1 = U_2 = 1$. Further, it is assumed that $\hat{y}_k = 1$.

Clearly, in a situation where $a_{k1}x_1 + a_{k4}x_4 = 0.5$ and $a_{k2}x_2 = 0.4$ in equation (6), no gain limiting is needed, so that the limiting factors can be set to
 10 $(\alpha_1, \alpha_2) = (1, 1)$ and still meet the in-range condition, that is, the maximum downmixing coefficients are applied as downmixing coefficients.

Now, if $a_{k1}x_1 + a_{k4}x_4 = 0.8$ and $a_{k2}x_2 = 0.4$ in equation (6), then the in-range condition $|y_k| \leq 1$ is satisfied by limiting factor pairs (α_1, α_2) within the pentagonal area with corners at $(L_1, L_2), (1, L_2), (1, \frac{1}{2}), (\frac{3}{4}, 1)$ and $(L_1, 1)$, as shown in figure 2. For reasons already stated, the gain is preferably not limited more than necessary and
 15 accordingly, the system 100 preferably attempts to find an upper (or 'sharp') solution $y_k = 1$ by selecting limiting factors from the edge segment between $(1, \frac{1}{2})$ and $(\frac{3}{4}, 1)$. Further, it is advantageous to limit secondary input channels rather than primary input channels, and this translates to selecting a pair of limiting factors at the right extreme (highest α_1) on this segment. This leads to the solution $(\alpha_1, \alpha_2) = (1, \frac{1}{2})$, and
 20 the k^{th} output signal will be given by

$$y_k = a_{k1}x_1 + a_{k2}x_2 + \frac{a_{k4}}{2}x_4.$$

However, if $L_2 > \frac{1}{2}$, then the primary limiting factor α_1 will necessarily be less than its upper bound $U_1 = 1$. To favour the primary subgroup over the secondary maximally, the preferred choice of limiting factors is $(\alpha_1, \alpha_2) = (\frac{5}{4} - \frac{L_2}{2}, L_2)$.

25 In variations to this embodiment where the system 100 is configured to search for limiting factors in a different way than described in the example of the preceding paragraph, the primary subgroup may be favoured by being associated with a greater lower bound than the secondary subgroup, that is, $L_1 > L_2$.

In one embodiment, the mixing system 100 may determine suitable upper and
 30 lower bounds on the limiting factors on the basis of the maximal downmix coefficient

cients. If the in-range condition is $-1 \leq Y \leq 1$, a number $W \leq 1$ is given and the bounds are written on the form

$$L_1 = m_P W, L_2 = m_S W, U_1 = U_2 = W, \quad (7)$$

then this embodiment uses

$$m_S = \min \left\{ Q, \frac{1}{W(P+S)} \right\}, m_P = \frac{1}{P} \left(\frac{1}{W} - m_S S \right), \quad (8)$$

where P is the sum of the absolute values of the downmix coefficients applied to the signals in the primary subgroup and S is the sum of the absolute values of the downmix coefficients applied to the signals in the secondary subgroup. By varying the value of constant $0 < Q < 1$, the system's tendency to limit secondary signals rather than primary signals can be made more or less pronounced. In the example discussed above, $P = |a_{k1}| + |a_{k3}|$ and $S = |a_{k2}|$.

In figures 3A and 3B, the dotted areas represent choices (α_1, α_2) of limiting factors that satisfy the double inequality

$$-1 \leq W(m_P P + m_S S) \leq 1,$$

which is what the above in-range condition amounts to in the worst-case situation of all input signals having unity magnitude and of equal signs as the downmix coefficients, that is, for some k , $\alpha_{ki} x_i = |a_{ki}|$ for all i or $\alpha_{ki} x_i = -|a_{ki}|$ for all i . The hashed sub-areas represents choices of limiting factors for which primary signals are limited less than secondary signals. The lower bounds in formulas (7), (8) represent choices of limiting values for which the in-range condition is just satisfied (i.e., satisfied 'sharply') in the worst case. For the purpose of illustration, the constant Q has been set to 1/2. This embodiment is based on the realisation that limiting factors need never be chosen smaller than these values. Having understood this exemplifying embodiment, those skilled in the art will be able to generalise it to other in-range conditions than $-1 \leq Y \leq 1$.

Figure 4 shows a mixing system 400 for downmixing eight audio channels into two channels. It may be argued that the system 400 has a three-layered structure comprising a configuring section 420, a controller (gain limiting section) 440 and a mixing section 460. The configuring section 420 is adapted to determine suitable intervals for limiting factors on the basis of parameters configuring the properties of the system 400. The limiting controller 440 is adapted to determine the values of the downmix coefficients to be applied by the mixing section 460 on the basis of the in-

tervals supplied by the configuring section 420 and further on the basis of certain input data supplied by the mixing section 460. The mixing section 460 is adapted to receive a vector of input audio signals $X = [L_s \ R_s \ C \ LFE \ Ls \ Rs \ Lrs \ Rrs]^T$ and to downmix these into a vector of output audio signals $Y = [L \ R]^T$ by means of
 5 a mixer 462 and using the downmix coefficients.

The mixing system 400 is adapted to handle signals partitioned into time segments. As an example, the signals may be conformal to the digital distribution format described in the paper J.R. Stuart *et al.*, "MLP lossless compression", Meridian Audio Ltd., Huntingdon, England, which is hereby incorporated by reference. In
 10 this distribution format, blocks (or access units) are formed from between 40 and 160 samples, and packets (corresponding to restart intervals) are formed from a fixed number of blocks. A packet, which may consist of 128 blocks and include a restart header, will be regarded as a time segment for the purposes of this example.

The configuring section 420 includes a unit 421 for receiving a matrix of maximal downmix coefficients
 15

$$dm_{8 \rightarrow 2} = \begin{bmatrix} 1 & 0 & 10^{-3/20} & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 10^{-3/20} & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

and for receiving masking matrices

$$mask_P = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$mask_S = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

20 which define a partition of the input signals into a primary subgroup (L_s, R_s, C , which are intended for playback in front of a listener and at approximate ear level) and a secondary subgroup (Ls, Rs, Lrs, Rrs). A third subgroup containing only the low-frequency effects (LFE) channel will not contribute to any output signals in this mixing system 400. The receiving unit 421 computes the numbers P, S referred to above
 25 and forms masked mixing matrices

$$primary_{8 \rightarrow 2} = mask_P \cdot dm_{8 \rightarrow 2}, \quad secondary_{8 \rightarrow 2} = mask_S \cdot dm_{8 \rightarrow 2},$$

where $\cdot$ denotes element-wise (or Hadamard) matrix multiplication. Since the maximal downmix coefficients are symmetric, the numbers are

$$P = 1 + 10^{-3/20} \text{ and } S = 1 + 1 = 2.$$

The configuring section 420 further comprises units 423, 424, 434 for computing upper and lower bounds on the respective limiting factors for the primary and secondary subgroups. A first unit 423 determines an intermediate value

$$\alpha = \frac{1}{W(P + S)}$$

based on the value of a parameter *maxaudio* determining the in-range condition to be applied, the values of *P*, *S* obtained from the receiving unit 421 and further based on a common upper bound *W* on the primary and secondary limiting factors. The value of the upper bound *mW* may be supplied directly to the first unit 423 as a configuration parameter to the system 400. It may also, as shown in figure 4, be supplied by a converter 422 for calculating the upper bound *W* on the basis of dialogue norm values; as an illustrative example, the upper bound may be given by the relationship

$$W = 10^{(dialnorm_{8ch} - dialnorm_{2ch})/20},$$

where *dialnorm_{8ch}* denotes the dialogue norm pertaining to the 8-channel input representation of the audio and *dialnorm_{2ch}* is the desired dialogue norm in the 2-channel output representation. Returning to the computation of the upper and lower bounds, a second unit 424 is adapted to evaluate, based on α , the variables m_p, m_s given by equations (8). Finally, third and fourth units 425, 426 are adapted to receive m_p, W and m_s, W respectively, and to derive the primary and secondary upper and lower bounds on the limiting factors using equations (7).

Turning now to the controller 440, output channel *L* has an associated limiter 442 for determining what values the primary and secondary limiting factors α_{pL}, α_{sL} are required to have in order to satisfy the in-range condition defined by the parameter *maxaudio*. The limiter 442 determines the values for one time segment at a time and may be configured to carry this out in the manner described previously, favouring the primary input signals over the secondary ones. For a given time segment, the limiter 442 bases its decisions on the in-range parameter *maxaudio*, on the intervals $[L_1, U_1], [L_2, U_2]$ in which the limiter 442 is permitted to choose the limiting factors α_1, α_2 , and further on input signal data for the time segment. In this

embodiment, the input data is supplied from a preliminary mixer 441 to the limiter 442 in the form of signals L_{2P}, L_{2S} given by

$$\begin{bmatrix} L_{2P} \\ R_{2P} \end{bmatrix} = \text{primary}_{8 \rightarrow 2} X \text{ and } \begin{bmatrix} L_{2S} \\ R_{2S} \end{bmatrix} = \text{secondary}_{8 \rightarrow 2} X.$$

The preliminary mixer 441 is communicatively connected to an input port 461 to obtain the input signals X or, possibly, a subset (e.g. not including LFE) sufficient to compute $L_{2P}, L_{2S}, R_{2P}, R_{2S}$. A limiter 443 for the other output channel R is configured in a similar manner as the L limiter 442, except that it receives signals R_{2P}, R_{2S} in lieu of L_{2P}, L_{2S} and outputs α_{PR}, α_{SR} .

Subsequently, to restore the balance between the input channels going to the output channels, the left and right primary limiting factors α_{PL}, α_{PR} are fed to a minimum extractor 444 adapted to return $\alpha_P = \min\{\alpha_{PL}, \alpha_{PR}\}$. Similarly, the left and right secondary limiting factors α_{SL}, α_{SR} are supplied to a further minimum extractor 445 configured to output $\alpha_S = \min\{\alpha_{SL}, \alpha_{SR}\}$.

In this embodiment, smoothing of the time sequence of primary and secondary limiting factors $\alpha_P(n), \alpha_S(n)$, where n is a time-segment index, is performed by regularisers 446, 447 which return smoothed sequences of limiting factors $\tilde{\alpha}_P(n), \tilde{\alpha}_S(n)$. The functioning of the regularisers 446, 447 will be described in more detail below. In this embodiment, the regularisers 446, 447 are assisted by respective buffers 448, 449 enabling the regularisers 446, 447 to operate on more values of the limiting factor than the current one. The buffers 448, 449 may be realised as shift registers.

As a final step to be carried out by the controller 440, multipliers 450, 451 and a summer 452 compute, using the smoothed limiting factors and the masked mixing matrices, the following downmix matrix to be applied in the n^{th} time segment:

$$\tilde{\alpha}_P(n) \text{primary}_{8 \rightarrow 2} + \tilde{\alpha}_S(n) \text{primary}_{8 \rightarrow 2}.$$

As has been already mentioned, the mixing section 460 comprises an input port 461 for receiving the input signals X and for supplying these to the preliminary mixer 441. The input port 461 further provides the input signals X to a mixer 461, which is adapted to receive the downmix matrix and to evaluate the equation

$$Y = (\tilde{\alpha}_P(n) \text{primary}_{8 \rightarrow 2} + \tilde{\alpha}_S(n) \text{primary}_{8 \rightarrow 2})X.$$

Figure 5 shows an example of the smoothing provided by one or both of the regularisers 446, 447. Limiting factors before smoothing (upper curve) and after smoothing (lower curve) have been plotted in a semi-logarithmic diagram. The sharp downward peaks in the non-smoothed values, which may be occasioned by high input signal values, correspond to broadened peaks in the smoothed values in order to ensure that a greatest (absolute) rate-of-change condition is satisfied. In this example, the broadening is double sided. Further, both the location and the amplitude of the peak are preserved. It is possible to achieve this by means of a look-ahead filter. For the acceptable rate of change R_{max} [signal units per time segment] and the maximal expected change in signal magnitude A_{max} [signal units] a suitable number of taps is $A_{\text{max}}/R_{\text{max}}$, and the look-ahead period will be approximately the number of taps multiplied by the segment length. In the smoothing, as already noted, it is not advisable to adjust individual segment-wise values of downmix coefficients by increasing them, as this may violate the in-range condition in time segments affected by smoothing.

In an analogue implementation, the regularisers 446, 447 may be realised by rate-limiting filters of the kind exemplified by US3252105, which is hereby incorporated by reference. Such filters are preferably applied in conjunction with appropriate delay lines to ensure sufficient synchronicity of the limiting factors and the input signals to be downmixed. In the embodiment shown in figure 4, a delay line may be arranged between the input port 461 and the mixer 462 and may correspond to the size of buffers 448, 449.

Further embodiments of the present invention will become apparent to a person skilled in the art after studying the description above. Even though the present description and drawings disclose embodiments and examples, the invention is not restricted to these specific examples. Numerous modifications and variations can be made without departing from the scope of the present invention, which is defined by the accompanying claims.

The systems and methods disclosed hereinabove may be implemented as software, firmware, hardware or a combination thereof. In a hardware implementation, the division of tasks between functional units referred to in the above description does not necessarily correspond to the division into physical units; to the contrary, one physical component may have multiple functionalities, and one task may be

carried out by several physical components in cooperation. Certain components or all components may be implemented as software executed by a digital signal processor or microprocessor, or be implemented as hardware or as an application-specific integrated circuit. Such software may be distributed on computer readable
5 media, which may comprise computer storage media (or non-transitory media) and communication media (or transitory media). As is well known to a person skilled in the art, computer storage media includes both volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program mod-
10 ules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a
15 computer. Further, it is well known to the skilled person that communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media.

What is claimed is

1. A method for downmixing a plurality of input audio signals containing input data into at least one output audio signal,

wherein maximal downmix coefficients are predefined, at least one in-range condition on said at least one output audio signal is predefined and the input audio signals are partitioned into predefined subgroups,

the method, being restricted in that, said in-range condition for said at least one output audio signal is an upper bound to said at least one output audio signal, a lower bound to said at least one output audio signal, or said at least one output audio signal is maintained in a range having an upper bound and a lower limit, the method comprising:

determining downmix coefficients as products of said maximal downmix coefficients and a limiting factor which is common within each subgroup in order to satisfy, in view of the input data, an in-range condition on said at least one output signal; and

applying the downmix coefficients to downmix the plurality of input audio signals into at least two output audio signals corresponding with spatially related channels,

wherein said downmixing coefficients are determined as products of all output audio signals, a limiting factor which is common within each subgroup, and said maximal downmix coefficients, in order to commonly satisfy an in-range condition of each of at least two output audio signals corresponding with said spatially related channels,

the step of determining said downmixing coefficients comprising:

determining, for each of the output audio signals to which the input audio signals in a sub-group contribute, a downmix coefficient as a product of the maximal downmix coefficient and a preliminary limiting factor; and

determining a limiting factor common within the subgroup by selecting the minimum of the preliminary limiting factors.

2. The method of claim 1, wherein at least one of said subgroups of input audio signals comprises two or more input audio signals.

3. The method of claim 1, wherein input audio signals in a subgroup correspond to spatially related audio channels.
4. The method of claim 1, wherein the downmix coefficients are determined in such manner that the in-range condition will be satisfied by at most 20 per cent margin.
5. The method of claim 1, wherein the output audio signal is partitioned into time segments, and wherein a segment-wise set of downmix coefficients is determined for each of a plurality of time segments as products of said maximal downmix coefficients and a limiting factor which is common within each subgroup in order to satisfy, independently in view of the input data in this time segment, an upper output-signal bound.
6. The method of claim 5, wherein a segment-wise set of downmix coefficients is determined for each of a plurality of time segments as products of said maximal downmix coefficients and a limiting factor which is common within each subgroup in order to jointly satisfy an in-range condition on each of said at least two output audio signals corresponding with spatially related channels, independently in view of the input data within a time segment.
7. The method of claim 6, comprising the steps of:
 - defining a sequence of segment-wise values of a downmix coefficient from said segment-wise sets of downmix coefficients;
 - smoothing the sequence of segment-wise values of the downmix coefficient;
 - and
 - applying the smoothed segment-wise values to downmix the plurality of input audio signals.
8. The method of claim 7, wherein the sequence of segment-wise values is smoothed by applying an upper rate-of-change bound.

9. The method of claim 1, wherein at least one subgroup is associated with a lower bound on the limiting factor for that subgroup.

10. The method of claim 9, wherein a primary and secondary subgroup are defined, and a lower bound on the limiting factor associated with the primary subgroup is greater than a lower bound on the limiting factor associated with the secondary subgroup.

11. The method of claim 1, wherein a primary and a secondary subgroup are pre-defined and the primary subgroup is associated with an upper bound on the limiting factor, and

wherein said step of determining downmix coefficients includes setting the upper bound on the limiting factor for the primary subgroup as a value of the limiting factor for the primary subgroup.

12. The method of claim 11, wherein a primary and a secondary subgroup are predefined and each is associated with a respective lower bound and a respective upper bound on the limiting factors ($L_1 \leq a_1 \leq U_1$, $L_2 \leq a_2 \leq U_2$), and

the step of determining said downmixing coefficients comprises the substeps of:

initially attempting to satisfy the in-range condition on said at least one output audio signal in the subspace of limiting factors such that the primary-subgroup limiting factor is equal to its upper bound ($a_1 = U_1$, $L_2 \leq a_2 \leq U_2$); and

further, if the initial attempt fails, attempting to satisfy the in-range condition on said at least one output audio signal in the subspace of limiting factors such that the secondary-subgroup limiting factor is equal to its lower bound ($L_1 \leq a_1 \leq U_1$, $a_2 = L_2$).

13. The method of any one of claims 10, wherein the primary subgroup corresponds to channels from one of the following groups:

- (i) channels for playback by audio sources located in a front half space with respect to a listener, and
- (ii) channels for playback by audio sources located at substantially the same height as a listener; and

the secondary subgroup corresponds to channels other than (i) or (ii).

14. The method of any one of claims 13, wherein the primary subgroup corresponds to channels from one of the following groups:

(iii) front channels,

(iv) center channels, and

(v) wide channels; and

the secondary subgroup corresponds to channels other than (iii), (iv) or (v).

15. The method of claim 1, wherein at least one subgroup is associated with an upper bound on the limiting factor.

16. The method of claim 15, wherein two or more subgroups are associated with a common upper bound on the limiting factor.

17. The method of claim 1, wherein said spatially related channels belong to one of the following channel groups: front, surround, rear surround, wide, center, and side channel.

18. A method for encoding a plurality of audio signals as a bit stream, comprising the steps of:

receiving a plurality of audio signals;

downmixing the audio signals into a downmix signal according to the downmixing method of claim 1; and

encoding the downmix signal as a bit stream.

19. A data carrier storing computer-executable instructions for performing the method of any one of the claim 1 through claim 17 and claim 18.

20. A method for decoding, comprising the steps of:

receiving a bit stream containing a plurality of encoded audio signals and a mixing matrix obtained from the downmixing coefficients determined by any one method of claim 1 to claim 17;

decoding said encoded audio signals to generate decoded audio signals; and
mixing said decoded audio signals into one or more output audio signals
according to said mixing matrix.

21. A data carrier storing computer-executable instructions for performing the
decoding method of claim 20.

22. A mixing system (400), comprising:

an input port (461) for receiving a plurality of input audio signals containing
input data;

a configuring section (420), for receiving
maximal downmix coefficients,

an in-range condition on said at least one output signal, and

a partition of the subgroups of said plurality of input audio signals;

said configuring section (420), being restricted in that, said in-range condition
for said at least one output audio signal is an upper bound to said at least one output
audio signal, a lower bound to said at least one output audio signal, or said at least
one output audio signal is maintained in a range having an upper bound and a lower
limit,

a controller (440) for determining downmix coefficients as products of said
maximal downmixing coefficients and a limiting factor which is common within each
subgroup in order to satisfy, in view of the input data, an in-range condition on said at
least one output audio signal; and

a mixer (462) for applying the downmix coefficients determined by the
controller (440) to downmix said plurality of input audio signals into at least two
spatially related output audio signals;

wherein the controller (440) is adapted to determine the downmix coefficients
as products of said maximal downmixing coefficients and said limiting factor in order
to jointly satisfy the in-range conditions pertinent to each of said output audio signals,
said limiting factor is common within all output audio signals and each subgroup,
said controller (440) comprising:

the means (442, 443) for determining, for each of the output audio signals to
which the input audio signals in a subgroup contribute, a downmix coefficient as a

product of the maximal downmix coefficient and a preliminary limiting factor; and
a minimal extractor (444, 445) determining a limiting factor common within the subgroup by selecting the minimum value of the preliminary limiting factors.

23. A decoding system, comprising:

an input port receiving a bit stream containing a plurality of encoded audio signals and a mixing matrix obtained from the downmixing coefficients determined by any one method of claim 1 to claim 17;

a decoder decoding said encoded audio signals to generate decoded audio signals; and

a mixer mixing said decoded audio signals into one or more output audio signals according to said mixing matrix.