

SYSTEMS AND METHODS FOR COVARIANCE SMOOTHING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/881,825, filed 1 August 2019, and U.S. Provisional Patent Application No. 63/057,533
5 filed 28 July 2020, which are incorporated herein by reference.

TECHNICAL FIELD

[0002] The present disclosure relates to improvements for the processing of signals. In particular, this disclosure relates to processing audio signals to improve covariance smoothing for improved processing.

10 BACKGROUND

[0003] An aspect of audio signal processing includes presenting multi-channel audio to a listener such that the listener can determine virtual spatial locations for the audio, giving an immersive experience to the listener. An early implementation of this is stereo, where the listener can spatially determine the “direction” a sound comes from.

15 [0004] More recent developments in this technology use inter-channel dependencies of a multi-channel system to present a more fully immersive sound experience. This can include the use of a covariance matrix of the audio channels.

SUMMARY

[0005] Various signal processing systems and methods are disclosed herein. Some such
20 systems and methods may involve smoothing the covariance values for a band over successive frames.

[0006] In some examples, a system and method for smoothing an estimate of covariance matrix for a sequence of signal frames in a frequency band, the method comprising: for the signal frames, comparing an effective count of bins in the frequency band to a desired count of
25 bins for the frequency band; computing a forgetting factor for the band as a ratio of the effective count to the desired count; and if the effective count of bins in the frequency band is smaller than the desired count, generating a current estimate of a value of the covariance matrix value for a present frame using a previously generated estimate of a value of the

covariance matrix value for a previous frame relative to the present frame and the forgetting factor; wherein the comparing, computing, and generating are performed by a system including one or more computer processors. The smoothed covariance matrix can be used for improved further processing of the signal by reducing artifacts caused by rough transitions in the matrix.

[0007] In some such examples, the system and method may involve calculating the effective number of bins as a sum of filterbank response values for the frequency band.

[0008] In some such examples, the generating uses a first order filter.

[0009] In some such examples, the first order filter includes a difference between the value for the present frame and the previously generated estimate for the previous frame, the difference weighted by the forgetting factor.

[0010] In some such examples, the system and method also includes comparing the forgetting factor with a maximum forgetting factor; and setting the forgetting factor to the minimum of the calculated forgetting factor and the maximum permitted forgetting factor; wherein the comparing and setting are performed before the generating the current estimate of the value of the covariance matrix for the present frame.

[0011] In some such examples, the system and method also includes detecting whether a transient occurs within a monitored frame; and resetting the smoothing in response to detecting that the transient occurs by setting a current estimate of a value of the covariance matrix value for the monitored frame to an original value of a covariance matrix value for the monitored frame without use of the forgetting factor for the monitored frame. The detecting can be performed by using aspects of a ducking decorrelator.

[0012] In some such examples, the system and method also includes storing the current estimate of the covariance matrix of the present frame; and resampling the stored current estimate over frequency bands, converting from one banding structure to another banding structure.

[0013] The system and method can be a part of an audio signal decoder.

[0014] Some or all of the methods described herein may be performed by one or more devices according to instructions (e.g. software) stored on one or more non-transitory media. Such

non-transitory media may include memory devices such as those described herein, including but not limited to random access memory (RAM) devices, read-only memory (ROM) devices, etc. Accordingly, various innovative aspects of the subject matter described in this disclosure may be implemented in a non-transitory medium having software stored thereon. The software may, for example, be executable by one or more components of a control system such as those disclosed herein. The software may, for example, include instructions for performing one or more of the methods disclosed herein.

[0015] At least some aspects of the present disclosure may be implemented via an apparatus or apparatuses. For example, one or more devices may be configured for performing, at least in part, the methods disclosed herein. In some implementations, an apparatus may include an interface system and a control system. The interface system may include one or more network interfaces, one or more interfaces between the control system and memory system, one or more interfaces between the control system and another device and/or one or more external device interfaces. The control system may include at least one of a general-purpose single- or multi-chip processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, or discrete hardware components. Accordingly, in some implementations the control system may include one or more processors and one or more non-transitory storage media operatively coupled to one or more processors.

[0016] Details of one or more implementations of the subject matter described in this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages will become apparent from the description, the drawings, and the claims. Note that the relative dimensions of the following figures may not be drawn to scale. Like reference numbers and designations in the various drawings generally indicate like elements, but different reference numbers do not necessarily designate different elements between different drawings.

BRIEF DESCRIPTION OF DRAWINGS

[0017] FIG. 1 illustrates an example of a signal channel separated into samples, frames, bins, and bands.

[0018] FIG. 2 illustrates a flowchart for an example method to smooth a signal covariance

matrix using the smoothing algorithm.

[0019] FIG. 3 illustrates a flowchart (partial) for an example method to smooth a signal covariance matrix using the smoothing algorithm with maximum forgetting factor.

5 [0020] FIG. 4 illustrates a flowchart (partial) for an example method to smooth a signal covariance matrix using the smoothing algorithm with transient detection.

[0021] FIG. 5 illustrates an example encoder/decoder system with the smoothing algorithm in the encoder.

[0022] FIG. 6 illustrates an example encoder/decoder system with the smoothing algorithm in the decoder.

10 [0023] FIG. 7 illustrates an example encoder/decoder system with the smoothing algorithm as part of a DTX.

DETAILED DESCRIPTION

15 [0024] The present disclosure describes methods and systems for smoothing the values of an estimated matrix (e.g. a covariance matrix) over time, including cases where the estimate is calculated from a small number of samples, such that the un-smoothed variation of those values could cause issues, such as noticeable artifacts in the output (e.g. audio).

[0025] The present disclosure also describes (among other things) the use of transient detection to prevent over-smoothing the values, thereby preventing adding unwanted artifacts from the smoothing.

20 [0026] The term “smoothing” as used herein refers to reducing the variation of a value of sequential frames of a signal. For example, reducing the variation of covariance values over sequential frames.

[0027] The term “covariance” as used herein refers to the measure of the joint variability of two signals. For example, the covariance of the energies of the signals.

25 [0028] A “covariance matrix” as used herein refers to a matrix of covariances of multiple signals. “Banded covariance matrix” refers to a covariance matrix where several adjacent frequency bins of the per-bin covariance matrix are combined into a band, such that the

covariance can be represented by significantly fewer numbers (equal to the number of bands). For example, an input with 960 frequency bins can become 12 frequency bands with the 960 bins distributed, equally or unequally, among the 12 bands. For the examples provided herein, the “covariance matrix” is taken to be a banded covariance matrix.

- 5 **[0029]** An “estimated matrix” as used herein refers to a sample covariance matrix. As used herein, “covariance matrix” and “smoothed matrix” includes estimated matrices.

[0030] The term “complex advanced coupling” (CACPL) as used herein refers to a method of stereo encoding, whereby left and right signals are combined into mid $(L+R)/2$, and side $(L-R)/2$ signals, and then the side signal is then predicted from the mid.

- 10 **[0031]** The term “smoothing algorithm” as used herein refers to smoothing a covariance matrix in accordance with the present disclosure, in particular, smoothing based on an effective bin count of a band. The term “smoothed matrix” as used herein refers to a covariance matrix modified by the smoothing algorithm.

- 15 **[0032]** The term “effective bin count” as used herein refers to either an actual count of the number of bins in a band or an estimation of the count of the effective bins in a band. An example of an estimation of the count of the effective bins in a band includes summing the filterbank response values for that band.

[0033] The term “forgetting factor” as used herein refers to a weighting value that modifies the significance of a part of a function.

- 20 **[0034]** The term “DTX” as used herein refers to an encoder module that handles discontinuous transmission events.

- 25 **[0035]** A solution to the problem of smoothing the covariance of input channels is described herein. This improves the technology of signal processing, for example for voice and audio services, such as ones providing users with an immersive experience (e.g. virtual reality, augmented reality).

[0036] In digital signal processing as shown in FIG. 1, an input signal (channel) is divided in the time domain into frames 101. A sequence of successive frames can be grouped together into a “frame sequence” 102. In the frequency domain, the signal can be divided up into frequency intervals called bins 103. A given bin at a given frame is called a sample 106.

The bins can be grouped together in a frequency band 104/105. This allows statistical averaging over a given band for a given windowed frame/frame sequence (a “window” being a weighting function applied to a frame or frame sequence). One usage of this averaging is for estimating the covariance between audio channels.

- 5 [0037] Not all bands necessarily contain the same number of bins. For example, lower frequency bands 104 might have a lower number of bins per band than higher frequency bands 105. This means that, for statistical analysis, the lower frequency bands require a larger window (more frames) per bin to produce meaningful statistical results than a higher frequency band would require. A given bin at a given frame can have a signal energy
10 associated with it, and several bin values can be averaged over a band and/or window.

Covariance Matrix Example

- [0038] Taking complex advanced coupling (CACPL) as an example, the parameters α (complex prediction coefficient) and β (real decorrelation coefficient) may be calculated from the elements of a (2x2) mid-side covariance matrix Σ_{MS} . This matrix is calculated
15 over a series of analysis bands $1 \leq b \leq B$.

$$\Sigma_{MS}(b) = \begin{bmatrix} R_{MM}(b) & R_{MS}(b) \\ R_{SM}(b) & R_{SS}(b) \end{bmatrix} \quad \text{eq. 1}$$

$$\alpha(b) = \frac{R_{SM}(b)}{R_{MM}(b)}, \quad \beta(b) = \sqrt{\frac{R_{SS}(b)}{R_{MM}(b)} - \alpha(b)^2}$$

- where terms in the form $R_{XY}(b) = X(b)Y(b)^*$ are the covariance of signals X and Y calculated over a frequency band b . Superscript $*$ denotes the complex conjugate transpose of the value. The example above is a 2 x 2 covariance matrix. In various implementations, any number of
20 input channels are possible (for example, a four-channel system with a 4 x 4 covariance matrix). The example above is for mid (M) and side (S) signals, but one skilled in the art can extrapolate the above for other signal types.

Smoothing Function and Forgetting Factor

- 25 [0039] Generally, the smoothed matrix can be calculated using a low-pass filter designed to meet the smoothing requirements. In the case of the smoothing algorithm, the smoothing requirements are such that previous estimates are used to artificially increase the number

of frequency samples (bins) used to generate the current estimate of a covariance matrix. In some embodiments, calculating the smoothed matrix $\tilde{A}$ from an input covariance matrix A over a frame sequence uses a first order auto-regressive low pass filter that uses a weighted sum of past and present frames' estimated matrix values:

$$\tilde{A}[n] = \tilde{A}[n-1] + \lambda(A[n] - \tilde{A}[n-1]) \quad \text{eq. 2}$$

where λ is a forgetting factor, or an update rate, i.e. how much emphasis is placed on previous estimation data and n is the frame number. In some embodiments, this only has meaning for the frames after the first frame, as there is no value for $A[0]$. In some embodiments, $A[0]$ is given the value of 0, resulting in a smoothing of $A[1]$. In some embodiments, $A[0]$ is given the value of $A[1]$, resulting in no smoothing of $A[1]$. The algorithm in equation 2 is an example of an algorithm that can be used for the smoothing algorithm, but other equations can also be used. For example, while equation 2 shows a first order low pass filter, a higher order filter design can also be used. The important factors of the algorithm (smoothing function) are the looking-back aspect of using previously smoothed results and the forgetting factor to give weight to the influence of those results.

Forgetting Factor

[0040] The effect of the forgetting factor is that, as the smoothing is applied over successive frames, the effect of previous frames becomes less and less impactful on the smoothing of the frame being smoothed (adjusted).

[0041] When the forgetting factor in the equation 1 is one ($\lambda = 1$), no smoothing occurs and it effectively acts as an all-pass filter. When $0 < \lambda < 1$ the equation acts as a low pass filter. The lower λ places more emphasis on the old covariance data, while a higher λ takes more of the new covariance into account. A forgetting factor over one (e.g. $1 < \lambda < 2$) implements as a high pass filter. In some embodiments, a maximum permissible forgetting factor λ_{max} is implemented. This maximum value will determine the behavior of the algorithm once the bins/band values become large. In some embodiments, $\lambda_{max} < 1$ will always implement some smoothing in every band, regardless of what the calculated forgetting factor is; and $\lambda_{max} = 1$ will only apply the smoothing function to bands with less bins than the desired N_{min} , leaving larger bands unsmoothed.

[0042] In some of those embodiments, the forgetting factor for a particular band λ_b is calculated as the minimum of the maximum permitted forgetting factor λ_{max} and the ratio of the effective number of bins in the band N_b and the minimum number of bins N_{min} that are determined to give a good statistical estimate based on the window size.

$$5 \quad \lambda_b = \min \left(\lambda_{max}, \frac{N_b}{N_{min}} \right) \quad \text{eq. 3}$$

[0043] In some embodiments N_b is the actual count of bins for the frequency band. In some embodiments, N_b can be calculated from the sum of a particular band's frequency response, e.g. if a band's response is $r = [0.5, 1, 1, 0.5, 0, \dots, 0]$, the effective number of bins $N_b = \text{sum}(r) = 0.5 + 1 + 1 + 0.5 = 3$. In some embodiments, $\lambda_{max} = 1$ such that λ_b stays within a reasonable range, e.g. $0 \leq \lambda_b \leq 1$. This means that smoothing is applied proportionally to small sample estimates, and no smoothing is applied at all to large sample estimates. In some embodiments, $\lambda_{max} < 1$ which forces larger bands to be smoothed to a certain extent regardless of their size (e.g. $\lambda_{max} = 0.9$). In some embodiments, N_{min} can be selected based on the data at hand that produces the best subjective results. In some
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995
 1000
 1005
 1010
 1015
 1020
 1025
 1030
 1035
 1040
 1045
 1050
 1055
 1060
 1065
 1070
 1075
 1080
 1085
 1090
 1095
 1100
 1105
 1110
 1115
 1120
 1125
 1130
 1135
 1140
 1145
 1150
 1155
 1160
 1165
 1170
 1175
 1180
 1185
 1190
 1195
 1200
 1205
 1210
 1215
 1220
 1225
 1230
 1235
 1240
 1245
 1250
 1255
 1260
 1265
 1270
 1275
 1280
 1285
 1290
 1295
 1300
 1305
 1310
 1315
 1320
 1325
 1330
 1335
 1340
 1345
 1350
 1355
 1360
 1365
 1370
 1375
 1380
 1385
 1390
 1395
 1400
 1405
 1410
 1415
 1420
 1425
 1430
 1435
 1440
 1445
 1450
 1455
 1460
 1465
 1470
 1475
 1480
 1485
 1490
 1495
 1500
 1505
 1510
 1515
 1520
 1525
 1530
 1535
 1540
 1545
 1550
 1555
 1560
 1565
 1570
 1575
 1580
 1585
 1590
 1595
 1600
 1605
 1610
 1615
 1620
 1625
 1630
 1635
 1640
 1645
 1650
 1655
 1660
 1665
 1670
 1675
 1680
 1685
 1690
 1695
 1700
 1705
 1710
 1715
 1720
 1725
 1730
 1735
 1740
 1745
 1750
 1755
 1760
 1765
 1770
 1775
 1780
 1785
 1790
 1795
 1800
 1805
 1810
 1815
 1820
 1825
 1830
 1835
 1840
 1845
 1850
 1855
 1860
 1865
 1870
 1875
 1880
 1885
 1890
 1895
 1900
 1905
 1910
 1915
 1920
 1925
 1930
 1935
 1940
 1945
 1950
 1955
 1960
 1965
 1970
 1975
 1980
 1985
 1990
 1995
 2000
 2005
 2010
 2015
 2020
 2025
 2030
 2035
 2040
 2045
 2050
 2055
 2060
 2065
 2070
 2075
 2080
 2085
 2090
 2095
 2100
 2105
 2110
 2115
 2120
 2125
 2130
 2135
 2140
 2145
 2150
 2155
 2160
 2165
 2170
 2175
 2180
 2185
 2190
 2195
 2200
 2205
 2210
 2215
 2220
 2225
 2230
 2235
 2240
 2245
 2250
 2255
 2260
 2265
 2270
 2275
 2280
 2285
 2290
 2295
 2300
 2305
 2310
 2315
 2320
 2325
 2330
 2335
 2340
 2345
 2350
 2355
 2360
 2365
 2370
 2375
 2380
 2385
 2390
 2395
 2400
 2405
 2410
 2415
 2420
 2425
 2430
 2435
 2440
 2445
 2450
 2455
 2460
 2465
 2470
 2475
 2480
 2485
 2490
 2495
 2500
 2505
 2510
 2515
 2520
 2525
 2530
 2535
 2540
 2545
 2550
 2555
 2560
 2565
 2570
 2575
 2580
 2585
 2590
 2595
 2600
 2605
 2610
 2615
 2620
 2625
 2630
 2635
 2640
 2645
 2650
 2655
 2660
 2665
 2670
 2675
 2680
 2685
 2690
 2695
 2700
 2705
 2710
 2715
 2720
 2725
 2730
 2735
 2740
 2745
 2750
 2755
 2760
 2765
 2770
 2775
 2780
 2785
 2790
 2795
 2800
 2805
 2810
 2815
 2820
 2825
 2830
 2835
 2840
 2845
 2850
 2855
 2860
 2865
 2870
 2875
 2880
 2885
 2890
 2895
 2900
 2905
 2910
 2915
 2920
 2925
 2930
 2935
 2940
 2945
 2950
 2955
 2960
 2965
 2970
 2975
 2980
 2985
 2990
 2995
 3000
 3005
 3010
 3015
 3020
 3025
 3030
 3035
 3040
 3045
 3050
 3055
 3060
 3065
 3070
 3075
 3080
 3085
 3090
 3095
 3100
 3105
 3110
 3115
 3120
 3125
 3130
 3135
 3140
 3145
 3150
 3155
 3160
 3165
 3170
 3175
 3180
 3185
 3190
 3195
 3200
 3205
 3210
 3215
 3220
 3225
 3230
 3235
 3240
 3245
 3250
 3255
 3260
 3265
 3270
 3275
 3280
 3285
 3290
 3295
 3300
 3305
 3310
 3315
 3320
 3325
 3330
 3335
 3340
 3345
 3350
 3355
 3360
 3365
 3370
 3375
 3380
 3385
 3390
 3395
 3400
 3405
 3410
 3415
 3420
 3425
 3430
 3435
 3440
 3445
 3450
 3455
 3460
 3465
 3470
 3475
 3480
 3485
 3490
 3495
 3500
 3505
 3510
 3515
 3520
 3525
 3530
 3535
 3540
 3545
 3550
 3555
 3560
 3565
 3570
 3575
 3580
 3585
 3590
 3595
 3600
 3605
 3610
 3615
 3620
 3625
 3630
 3635
 3640
 3645
 3650
 3655
 3660
 3665
 3670
 3675
 3680
 3685
 3690
 3695
 3700
 3705
 3710
 3715
 3720
 3725
 3730
 3735
 3740
 3745
 3750
 3755
 3760
 3765
 3770
 3775
 3780
 3785
 3790
 3795
 3800
 3805
 3810
 3815
 3820
 3825
 3830
 3835
 3840
 3845
 3850
 3855
 3860
 3865
 3870
 3875
 3880
 3885
 3890
 3895
 3900
 3905
 3910
 3915
 3920
 3925
 3930
 3935
 3940
 3945
 3950
 3955
 3960
 3965
 3970
 3975
 3980
 3985
 3990
 3995
 4000
 4005
 4010
 4015
 4020
 4025
 4030
 4035
 4040
 4045
 4050
 4055
 4060
 4065
 4070
 4075
 4080
 4085
 4090
 4095
 4100
 4105
 4110
 4115
 4120
 4125
 4130
 4135
 4140
 4145
 4150
 4155
 4160
 4165
 4170
 4175
 4180
 4185
 4190
 4195
 4200
 4205
 4210
 4215
 4220
 4225
 4230
 4235
 4240
 4245
 4250
 4255
 4260
 4265
 4270
 4275
 4280
 4285
 4290
 4295
 4300
 4305
 4310
 4315
 4320
 4325
 4330
 4335
 4340
 4345
 4350
 4355
 4360
 4365
 4370
 4375
 4380
 4385
 4390
 4395
 4400
 4405
 4410
 4415
 4420
 4425
 4430
 4435
 4440
 4445
 4450
 4455
 4460
 4465
 4470
 4475
 4480
 4485
 4490
 4495
 4500
 4505
 4510
 4515
 4520
 4525
 4530
 4535
 4540
 4545
 4550
 4555
 4560
 4565
 4570
 4575
 4580
 4585
 4590
 4595
 4600
 4605
 4610
 4615
 4620
 4625
 4630
 4635
 4640
 4645
 4650
 4655
 4660
 4665
 4670
 4675
 4680
 4685
 4690
 4695
 4700
 4705
 4710
 4715
 4720
 4725
 4730
 4735
 4740
 4745
 4750
 4755
 4760
 4765
 4770
 4775
 4780
 4785
 4790
 4795
 4800
 4805
 4810
 4815
 4820
 4825
 4830
 4835
 4840
 4845
 4850
 4855
 4860
 4865
 4870
 4875
 4880
 4885
 4890
 4895
 4900
 4905
 4910
 4915
 4920
 4925
 4930
 4935
 4940
 4945
 4950
 4955
 4960
 4965
 4970
 4975
 4980
 4985
 4990
 4995
 5000
 5005
 5010
 5015
 5020
 5025
 5030
 5035
 5040
 5045
 5050
 5055
 5060
 5065
 5070
 5075
 5080
 5085
 5090
 5095
 5100
 5105
 5110
 5115
 5120
 5125
 5130
 5135
 5140
 5145
 5150
 5155
 5160
 5165
 5170
 5175
 5180
 5185
 5190
 5195
 5200
 5205
 5210
 5215
 5220
 5225
 5230
 5235
 5240
 5245
 5250
 5255
 5260
 5265
 5270
 5275
 5280
 5285
 5290
 5295
 5300
 5305
 5310
 5315
 5320
 5325
 5330
 5335
 5340
 5345
 5350
 5355
 5360
 5365
 5370
 5375
 5380
 5385
 5390
 5395
 5400
 5405
 5410
 5415
 5420
 5425
 5430
 5435
 5440
 5445
 5450
 5455
 5460
 5465
 5470
 5475
 5480
 5485
 5490
 5495
 5500
 5505
 5510
 5515
 5520
 5525
 5530
 5535
 5540
 5545
 5550
 5555
 5560
 5565
 5570
 5575
 5580
 5585
 5590
 5595
 5600
 5605
 5610
 5615
 5620
 5625
 5630
 5635
 5640
 5645
 5650
 5655
 5660
 5665
 5670
 5675
 5680
 5685
 5690
 5695
 5700
 5705
 5710
 5715
 5720
 5725
 5730
 5735
 5740
 5745
 5750
 5755
 5760
 5765
 5770
 5775
 5780
 5785
 5790
 5795
 5800
 5805
 5810
 5815
 5820
 5825
 5830
 5835
 5840
 5845
 5850
 5855
 5860
 5865
 5870
 5875
 5880
 5885
 5890
 5895
 5900
 5905
 5910
 5915
 5920
 5925
 5930
 5935
 5940
 5945
 5950
 5955
 5960
 5965
 5970
 5975
 5980
 5985
 5990
 5995
 6000
 6005
 6010
 6015
 6020
 6025
 6030
 6035
 6040
 6045
 6050
 6055
 6060
 6065
 6070
 6075
 6080
 6085
 6090
 6095
 6100
 6105
 6110
 6115
 6120
 6125
 6130
 6135
 6140
 6145
 6150
 6155
 6160
 6165
 6170
 6175
 6180
 6185
 6190
 6195
 6200
 6205
 6210
 6215
 6220
 6225
 6230
 6235
 6240
 6245
 6250
 6255
 6260
 6265
 6270
 6275
 6280
 6285
 6290
 6295
 6300
 6305
 6310
 6315
 6320
 6325
 6330
 6335
 6340
 6345
 6350
 6355
 6360
 6365
 6370
 6375
 6380
 6385
 6390
 6395
 6400
 6405
 6410
 6415
 6420
 6425
 6430
 6435
 6440
 6445
 6450
 6455
 6460
 6465
 6470
 6475
 6480
 6485
 6490
 6495
 6500
 6505
 6510
 6515
 6520
 6525
 6530
 6535
 6540
 6545
 6550
 6555
 6560
 6565
 6570
 6575
 6580
 6585
 6590
 6595
 6600
 6605
 6610
 6615
 6620
 6625
 6630
 6635
 6640
 6645
 6650
 6655
 6

for the current frame, and the forgetting factor. The new (smoothed) matrix formed by these new values is used in further signal processing 206.

[0046] FIG. 3 shows an example modification to the flow for a maximum permitted forgetting factor. As in FIG. 2, a forgetting factor is computed 301 for the band. Additionally, a maximum permitted forgetting factor is determined 302. The values are compared 303, and in response to the calculated factor being less than the maximum permitted factor, then the calculated factor is used in the smoothing 305. If the calculated factor is greater than the maximum permitted factor, the maximum permitted factor is used 304 in the smoothing 305. The example shows the calculated factor being used if the factors are equal (not greater than), but an equivalent flow can be envisioned where the minimum value is used if they are equal.

Smoothing Reset

[0047] In some embodiments, there may be a desire avoid smoothing over transients (sudden changes in signal magnitude) as this may produce unwanted signal distortion/artifacts in the output. In these embodiments, the smoothing can be “reset” at points where transients are detected in the signal.

[0048] The previous time frame’s estimated matrix can be stored to facilitate calculation the smoothed value for the current frame. If a transient is detected in the input signals during that frame, the smoothing function can be set to re-initialize itself.

[0049] When a transient is detected, the past matrix estimate is reset to the current estimate, such that the output of the smoothing filter after a transient is the estimate itself (no change applied). In other words, for the reset frame $\tilde{A}_b[n] = A_b[n]$. After the reset frame, subsequent frames can have the smoothing function applied again until the next reset.

[0050] FIG. 4 shows an example modification to the flow for transient detection. A determination is made 401 if a transient is detected for a given frame. If it is 403, then the new matrix value remains the same as the input value. If not 402, the usual smoothing algorithm is used for that frame. The combination (matrix) of smoothed and non-smoothed (transient) frame values are used for signal processing 404.

[0051] In some embodiments, the smoothing is reset when a transient is detected on any channel. For example, if there are N channels, N transient detectors can be used (one per channel) and if any of them detect a transient, the smoothing is reset or end of signal or end of smoothing (smoothing is turned off).

- 5 [0052] For the example of a stereo input, the channels may be determined to be distinct (or possibly distinct) enough such that only considering transients in the left channel might mean an important transient in the right channel may be improperly smoothed (and vice versa). Therefore, two transient detectors are used (left and right) and either one of these can trigger a smoothing reset of the entire 2x2 matrix.

- 10 [0053] In some embodiments, the smoothing is only reset on transients for certain channels. For example, if there are N channels, only M (<N, possibly 1) detectors are used.

- [0054] For the example of a First Order Ambisonics (FOA) input, the first (W) channel can be determined to be the most important compared to the other three (X, Y, Z) and, given the spatial relationships between FOA signals, transients in the latter three channels will likely
15 be reflected in the W channel anyway. Therefore, the system can be set up with a transient detector only on the W channel, triggering a reset of the entire 4x4 covariance matrix when it detects a transient on W.

- [0055] In some embodiments, the reset only resets covariance elements that have experienced the transient. This would mean that a transient in the n^{th} channel would only reset values in
20 the n^{th} row and in the n^{th} column of the covariance matrix (entire row and entire column). This can be performed by having separate transient monitoring on each channel and a detected transient on any given channel would trigger a reset for matrix positions that correspond to that channel's covariance to another channel (and vice versa, and, trivially, to itself).

- 25 [0056] In some embodiments, the reset only occurs on majority/threshold number of channels detecting a transient. For example, in a four channel system, the threshold could be set to trigger a reset only if at least two of the channels report a transient in the same frame.

Transient Detection Example

[0057] An example of a transient detector is the to use to ducking gain property of a ducking decorrelator, which can be used to determine if a transient has occurred in a single input channel. See e.g. US 9,747,909, the contents of which are incorporated by reference herein.

- 5 Sudden decreases in the ducking gain occur at transients. The change in the ducking gain can be compared to a threshold value to determine if a transient occurred. The threshold value can be determined by subjective evaluation of what provides an accurate transient detection.

[0058] The example transient detector works in the following manner:

- 10 **1.** The input signals are high pass filtered to enhance the transients.
- 2.** The absolute value of the high pass filter output signal is then low pass filtered to determine the signal envelope.
- 3.** Two envelopes are generated, fast $e_f[n]$ and slow $e_s[n]$, using two different low pass filters with slightly different cut off frequencies, $\omega_f > \omega_s$.
- 15 **4.** For each time-domain sample n of the signal envelopes within a particular frame, where $1 \leq n \leq N$, if

$$1.1e_s[n] < g_d[n-1]e_f[n] \quad \text{eq. 4}$$

then the ducking gain g_d is decreased to

$$g_d[n] = 1.1 \frac{e_s[n]}{e_f[n]} \quad \text{eq. 5}$$

- 20 otherwise the gain is allowed to rise slowly back to 1, following:

$$g_d[n] = (g_d[n-1] - 1)e^{-\frac{1}{\tau_d F_s}} + 1 \quad \text{eq. 6}$$

where τ_d is the time constant determining how quickly the ducking gain returns to 1.

- 5.** If there has been a large change in the ducking gain over the current frame, it is deemed that a transient has occurred: i.e.

$$d_g[1] - d_g[N] > threshold \quad \text{eq. 7}$$

[0059] In some embodiments, the method to determine a transient's occurrence from the ducking gain is counting the number/fraction of time-domain samples within each frame that the condition above is met. If this value passes a particular threshold, the frame is determined to contain a transient.

[0060] While ducking gain can be used for transient detection for the smoothing algorithm, any method of transient detection known in the art can be utilized.

Resampling (Banding Transition)

[0061] In some embodiments, a band undergoing smoothing can be converted to a new band for the smoothing when the banding of the channel changes. If the input signal banding changes from one frame to the next, in terms of number and/or bandwidths (e.g., because of a filterbank switch at the encoder), the stored (previous) matrix estimate can be resampled to the new banding structure in order to continue with uninterrupted operation of the smoothing mechanism. A conversion matrix between banding structures can be calculated by taking the set of windowing functions in the frequency domain, i.e. the ideal filterbank magnitude responses for the previous banding scheme (i.e., the previous filterbank), as a matrix.

[0062] As an example:

$Bank_1$ is a $(B_1 \times F)$ array containing rows of previous ideal filterbank responses $FR1$, and $Bank_2$ is a $(B_2 \times F)$ array containing the new filterbank ideal responses $FR2$. F is the number of frequency bins.

$$Bank_1 = [FR1_1^T \dots FR1_{b_1}^T \dots FR1_{\{B_1\}B_1}^T]^T$$

$$Bank_2 = [FR2_1^T \dots FR2_{b_2}^T \dots FR2_{\{B_2\}B_2}^T]^T$$

[0063] The elements of the conversion matrix $C_{b_1 \rightarrow b_2}$ that convert a band b_1 from $Bank_1$ to a band b_2 from $Bank_2$:

$$C_{b_1 \rightarrow b_2} = \frac{FR1_{b_1} FR2_{b_2}^*}{\sum_{b=1}^{B_1} FR1_b FR2_b^*} \quad \text{eq.8}$$

creating a $(B_1 \times B_2)$ matrix that can be used to convert the previous banded estimates to the new banding scheme.

Example of Smoothing with Reset

- 5 **[0064]** Given an example of a single band b of a 1x1 covariance matrix $A_b[n]$, where the band consists of 3 bins ($N_b = 3$) but it is determined (subjectively) that at least 8 bins ($N_{min} = 8$) are needed for good statistical analysis. This gives a band forgetting factor of $\lambda_b = 3/8$.

[0065] If the initial banded covariant matrix is (first frame and transients shown in bold):

$$A_b[n] = [\mathbf{0.1}, 0.05, 0.01, 0.05, 0.1, 0, \mathbf{0.9}, 0.1, 0.05, 0.01, 0.05]$$

- 10 and a transient detector with the following output, indicating the first frame and the seventh frame containing transients:

$$T[n] = [\mathbf{1}, 0, 0, 0, 0, 0, \mathbf{1}, 0, 0, 0, 0]$$

The smoothed covariance output would look like:

$$\begin{aligned} \tilde{A}_b[n] \\ = [\mathbf{0.1}, 0.0813, 0.0545, 0.0528, 0.0705, 0.0478, \mathbf{0.9}, 0.6, 0.3937, 0.2498, 0.1749] \end{aligned}$$

- 15 **[0066]** The first frame remains the same in $\tilde{A}[n]$ as it was in $A[n]$ as smoothing does not start until the second frame, due to the transient indicated in the first. If there was no transient in the first frame, the first frame can be smoothed by, in one embodiment, taking the previous frame as having a zero value, or as, in another embodiment, having a value equal to itself; the situation depending on how the algorithm is configured. In some embodiments,
- 20 the first frame is always marked as having a transient detected regardless of any actual detection taking place. The seventh value of $\tilde{A}[n]$ is the same as the one from $A[n]$ due to the smoothing reset.

- [0067]** The second frame uses the values of itself (0.05), the “smoothed” value of the previous frame (0.1), and the forgetting factor (3/8) to calculate the smoothed value of $\tilde{A}[2]$ (approx.
- 25 0.813). The formula being in this example (first order filter):

$$\tilde{A}[2] = \tilde{A}[1] + \lambda(A[2] - \tilde{A}[1]) = 0.1 + (3/8)(0.05 - 0.1) = 0.8125$$

[0068] This is repeated for subsequent frames until it reaches frame 7, which in $T[n]$ is designated as having a transient. For this, $\tilde{A}[7] = A[7] = 0.9$. The subsequent frames are then calculated by the smoothing function again (as if frame 7 was an initial frame).

5 [0069] In some embodiments, the covariant matrix smoothing system can be integrated into an encoder. An example system is provided in FIG. 5. The example encoder 501 generates 502 one or more covariance matrices for a signal 504 which the smoothing algorithm 503 modifies before the signal 504 is sent to the decoder 505.

10 [0070] In some embodiments, the covariant matrix smoothing system can be integrated into a decoder. An example of a decoder with the smoothing algorithm is shown in FIG. 6. The encoder 601 sends a signal 602 (e.g. audio channels) to the decoder 603. The decoder 603 applies the smoothing algorithm 604 before further processing 605 is performed (there may be processing before smoothing as well).

15 [0071] In some embodiments, covariance smoothing may also be used in the encoder in the case of discontinuous transmission (DTX) frames, to stabilize parametrization of spatial comfort noise. In this instance, during DTX frames, there will be no resetting of the covariance smoothing due to transients, and instead it will be reset upon entry/exit of the codec's DTX mode. An example of an encoder with the smoothing algorithm is shown in FIG. 7, where the encoder 701 has a DTX module 702 which incorporates the smoothing algorithm 703.

20 [0072] A number of embodiments of the disclosure have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the present disclosure. Accordingly, other embodiments are within the scope of the following claims.

25 [0073] A computing device implementing the techniques of covariance smoothing can have the following example architecture. Other architectures are possible, including architectures with more or fewer components. In some implementations, the example architecture includes one or more processors (e.g., dual-core Intel® Xeon® Processors), one or more output devices (e.g., LCD), one or more network interfaces, one or more input devices (e.g., mouse, keyboard, touch-sensitive display) and one or more computer-readable mediums (e.g., RAM, ROM, SDRAM, hard disk, optical disk, flash memory, etc.).
30 These components can exchange communications and data over one or more

communication channels (e.g., buses), which can utilize various hardware and software for facilitating the transfer of data and control signals between components.

5 [0074] The term “computer-readable medium” refers to a medium that participates in providing instructions to processor for execution, including without limitation, non-volatile media (e.g., optical or magnetic disks), volatile media (e.g., memory) and transmission media. Transmission media includes, without limitation, coaxial cables, copper wire and fiber optics.

10 [0075] Computer-readable medium can further include operating system (e.g., a Linux® operating system), network communication module, audio interface manager, audio processing manager and live content distributor. Operating system can be multi-user, multiprocessing, multitasking, multithreading, real time, etc. Operating system performs basic tasks, including but not limited to: recognizing input from and providing output to network interfaces 706 and/or devices 708; keeping track and managing files and directories on computer-readable mediums (e.g., memory or a storage device); controlling 15 peripheral devices; and managing traffic on the one or more communication channels. Network communications module includes various components for establishing and maintaining network connections (e.g., software for implementing communication protocols, such as TCP/IP, HTTP, etc.).

20 [0076] Architecture can be implemented in a parallel processing or peer-to-peer infrastructure or on a single device with one or more processors. Software can include multiple software components or can be a single body of code.

25 [0077] The described features can be implemented advantageously in one or more computer programs that are executable on a programmable system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to, a data storage system, at least one input device, and at least one output device. A computer program is a set of instructions that can be used, directly or indirectly, in a computer to perform a certain activity or bring about a certain result. A computer program can be written in any form of programming language (e.g., Objective-C, Java), including compiled or interpreted languages, and it can be deployed in any form, including 30 as a stand-alone program or as a module, component, subroutine, a browser-based web application, or other unit suitable for use in a computing environment.

[0078] Suitable processors for the execution of a program of instructions include, by way of example, both general and special purpose microprocessors, and the sole processor or one of multiple processors or cores, of any kind of computer. Generally, a processor will receive instructions and data from a read-only memory or a random access memory or both.

5 The essential elements of a computer are a processor for executing instructions and one or more memories for storing instructions and data. Generally, a computer will also include, or be operatively coupled to communicate with, one or more mass storage devices for storing data files; such devices include magnetic disks, such as internal hard disks and removable disks; magneto-optical disks; and optical disks. Storage devices suitable for
10 tangibly embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory devices, such as EPROM, EEPROM, and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory can be supplemented by, or incorporated in, ASICs (application-specific integrated circuits).
15

[0079] To provide for interaction with a user, the features can be implemented on a computer having a display device such as a CRT (cathode ray tube) or LCD (liquid crystal display) monitor or a retina display device for displaying information to the user. The computer can have a touch surface input device (e.g., a touch screen) or a keyboard and a pointing device
20 such as a mouse or a trackball by which the user can provide input to the computer. The computer can have a voice input device for receiving voice commands from the user.

[0080] The features can be implemented in a computer system that includes a back-end component, such as a data server, or that includes a middleware component, such as an application server or an Internet server, or that includes a front-end component, such as a
25 client computer having a graphical user interface or an Internet browser, or any combination of them. The components of the system can be connected by any form or medium of digital data communication such as a communication network. Examples of communication networks include, e.g., a LAN, a WAN, and the computers and networks forming the Internet.

30 [0081] The computing system can include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the

respective computers and having a client-server relationship to each other. In some embodiments, a server transmits data (e.g., an HTML page) to a client device (e.g., for purposes of displaying data to and receiving user input from a user interacting with the client device). Data generated at the client device (e.g., a result of the user interaction) can
5 be received from the client device at the server.

[0082] A system of one or more computers can be configured to perform particular actions by virtue of having software, firmware, hardware, or a combination of them installed on the system that in operation causes or cause the system to perform the actions. One or more computer programs can be configured to perform particular actions by virtue of including
10 instructions that, when executed by data processing apparatus, cause the apparatus to perform the actions.

[0083] While this specification contains many specific implementation details, these should not be construed as limitations on the scope of any inventions or of what may be claimed, but rather as descriptions of features specific to particular embodiments of particular
15 inventions. Certain features that are described in this specification in the context of separate embodiments can also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment can also be implemented in multiple embodiments separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain
20 combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

[0084] Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order
25 shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the described program components and systems can
30 generally be integrated together in a single software product or packaged into multiple software products.

[0085] Thus, particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. In some cases, the actions recited in the claims can be performed in a different order and still achieve desirable results. In addition, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In certain implementations, multitasking and parallel processing may be advantageous.

[0086] A number of implementations of the invention have been described. Nevertheless, it will be understood that various modifications can be made without departing from the spirit and scope of the invention.

CLAIMS

What is claimed is:

1. A method of smoothing an input covariance matrix for a sequence of signal frames in a frequency band, the method comprising:

5 for the signal frames, comparing an effective count of bins in the frequency band to a desired count of bins for the frequency band;

 computing a forgetting factor for the band as a ratio of the effective count to the desired count; and

 in response to determining that the effective count of bins in the frequency band is
10 smaller than the desired count, generating a value of a smoothed matrix for a present frame using a previously generated value of the smoothed matrix for a previous frame relative to the present frame and the forgetting factor;

 wherein the comparing, computing, and generating are performed by a system including one or more computer processors.

15 2. The method of claim 1, wherein the effective number of bins is calculated as a sum of filterbank response values for the frequency band.

3. The method of any of claims 1-2, wherein the generating uses a first order filter.

4. The method of claim 3, wherein the first order filter includes a difference between the value of the input covariant matrix for the present frame and the previously generated value of the
20 smoothed matrix for the previous frame, the difference weighted by the forgetting factor.

5. The method of any of claims 1-4, further comprising:

 comparing the forgetting factor with a maximum permitted forgetting factor; and

 setting the forgetting factor to the minimum of the forgetting factor and the maximum permitted forgetting factor;

25 wherein the comparing and setting are performed before the generating the value of the smoothed matrix for the present frame.

6. The method of any of claims 1-5, further comprising:
- detecting whether a transient occurs within a monitored frame; and
 - resetting the smoothing in response to detecting that the transient occurs by setting a value of the smoothed matrix value for the monitored frame to an original value of the input
- 5 matrix value for the monitored frame without use of the forgetting factor.
7. The method of claim 6, wherein the detecting is performed by using aspects of a ducking decorrelator.
8. The method of any of claims 1-7, wherein the detecting is performed on all channels.
9. The method of claim 8, wherein the detecting detects that the transient occurs if any transient
- 10 is detected on any channel for the monitored frame.
10. The method of claim 8, wherein the detecting detects that the transient occurs only if a transient occurs on a particular channel.
11. The method of any of claims 1-10, further comprising:
- storing the smoothed matrix of the present frame; and
- 15 resampling the stored smoothed matrix over frequency bands.
12. A system comprising:
- one or more computer processors; and
 - a non-transitory computer-readable medium storing instructions that, upon execution by the one or more computer processors, cause the one or more computer processors to perform
- 20 operations of any of claims 1-11.
13. A non-transitory computer-readable medium storing instructions that, upon execution by one or more computer processors, cause the one or more computer processors to perform operations of any of claims 1-11.
14. An encoder comprising the system of claim 12.
- 25 15. A decoder comprising the system of claim 12.