

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2017/161313 A1

(43) International Publication Date

21 September 2017 (21.09.2017)

WIPO | PCT

(51) International Patent Classification:

*G10L 19/008* (2013.01) *G10L 19/24* (2013.01)  
*G10L 19/04* (2013.01) *G10L 19/02* (2013.01)

(21) International Application Number:

PCT/US2017/023032

(22) International Filing Date:

17 March 2017 (17.03.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/310,626 18 March 2016 (18.03.2016) US  
15/460,928 16 March 2017 (16.03.2017) US

(71) Applicant: **QUALCOMM INCORPORATED** [US/US];  
ATTN: International IP Administration, 5775 Morehouse  
Drive, San Diego, California 92121-1714 (US).

(72) Inventors: **ATTI, Venkatraman S.**; 5775 Morehouse  
Drive, San Diego, California 92121-1714 (US).  
**CHEBIYYAM, Venkata Subrahmanyam Chandra  
Sekhar**; 5775 Morehouse Drive, San Diego, California  
92121-1714 (US).

(74) Agent: **TOLER, Jeffrey G.**; 8500 Bluffstone Cove, Suite  
A201, Austin, Texas 78759 (US).

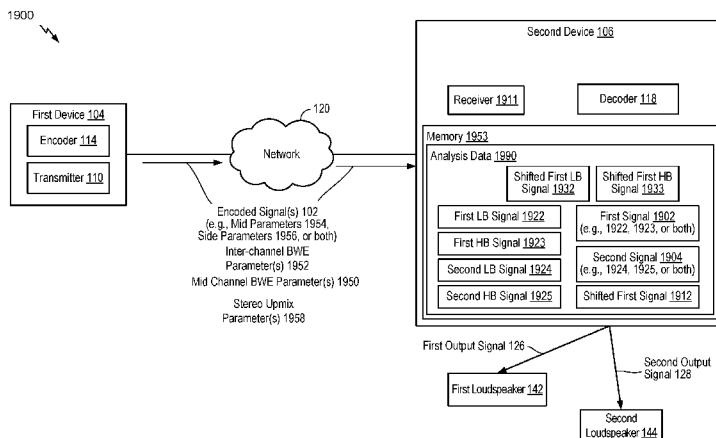
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: AUDIO SIGNAL DECODING



(57) Abstract: An apparatus includes a receiver configured to receive at least one encoded signal that includes inter-channel band-width extension (BWE) parameters. The device also includes a decoder configured to generate a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal. The decoder is also configured to generate, based on the mid channel time-domain high-band signal and the inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal. The decoder is further configured to generate a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal, and to generate a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal. The decoder is also configured to generate a modified target channel signal by modifying the target channel signal based on a temporal mismatch value.

WO 2017/161313 A1

## AUDIO SIGNAL DECODING

### *I. Claim of Priority*

[0001] The present application claims the benefit of priority from the commonly owned U.S. Provisional Patent Application No. 62/310,626, filed March 18, 2016, entitled “AUDIO SIGNAL DECODING,” and U.S. Non-Provisional Patent Application No. 15/460,928, filed March 16, 2017, entitled “AUDIO SIGNAL DECODING,” the contents of each of the aforementioned applications are expressly incorporated herein by reference in their entirety.

### *II. Field*

[0002] The present disclosure is generally related to decoding audio signals.

### *III. Description of Related Art*

[0003] Advances in technology have resulted in smaller and more powerful computing devices. For example, there currently exist a variety of portable personal computing devices, including wireless telephones such as mobile and smart phones, tablets and laptop computers that are small, lightweight, and easily carried by users. These devices can communicate voice and data packets over wireless networks. Further, many such devices incorporate additional functionality such as a digital still camera, a digital video camera, a digital recorder, and an audio file player. Also, such devices can process executable instructions, including software applications, such as a web browser application, that can be used to access the Internet. As such, these devices can include significant computing capabilities.

[0004] A computing device may include multiple microphones to receive audio signals. Generally, a sound source is closer to a first microphone than to a second microphone of the multiple microphones. Accordingly, a second audio signal received from the second microphone may be delayed relative to a first audio signal received from the first microphone. In stereo-encoding, audio signals from the microphones may be encoded to generate a mid channel signal and one or more side channel signals. The mid channel signal may correspond to a sum of the first audio signal and the second audio signal. A

- 2 -

side channel signal may correspond to a difference between the first audio signal and the second audio signal. The first audio signal may not be temporally aligned with the second audio signal because of the delay in receiving the second audio signal relative to the first audio signal. The misalignment (or “temporal offset”) of the first audio signal relative to the second audio signal may result in the side channel signal having high entropy (e.g., the side channel signal may not be maximally decorrelated). Because of the high entropy of the side channel signal, a greater number of bits may be needed to encode the side channel signal.

**[0005]** Additionally, different frame types may cause the computing device to generate different temporal offsets or shift estimates. For example, the computing device may determine that a voiced frame of the first audio signal is offset by a corresponding voiced frame in the second audio signal by a particular amount. However, due to a relatively high amount of noise, the computing device may determine that a transition frame (or unvoiced frame) of the first audio signal is offset by a corresponding transition frame (or corresponding unvoiced frame) of the second audio signal by a different amount. Variations in the shift estimates may cause sample repetition and artifact skipping at frame boundaries. Additionally, variation in shift estimates may result in higher side channel energies, which may reduce coding efficiency.

#### ***IV. Summary***

**[0006]** According to one implementation of the techniques disclosed herein, an apparatus includes a receiver configured to receive at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters. The device also includes a decoder configured to generate a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal. The decoder is also configured to generate, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal. The decoder is further configured to generate a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal. The decoder is also configured to generate a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal. The

- 3 -

decoder is further configured to generate a modified target channel signal by modifying the target channel signal based on a temporal mismatch value. In an example implementation of the techniques disclosed herein, the receiver may be configured to receive the temporal mismatch value. It should be noted that in some implementations of the techniques disclosed herein, the target channel signal may be based on the second channel time-domain high-band signal and the second channel low-band signal, and the reference channel signal may be based on the first channel time-domain high-band signal and the first channel low-band signal. In some implementations of the techniques disclosed herein, the target channel signal and the reference channel signal may vary from frame to frame based on a high-band reference channel indicator. For example, for a first frame, based on a first value of the high-band reference channel indicator, the target channel signal may be based on the second channel time-domain high-band signal and the second channel low-band signal, and the reference channel signal may be based on the first channel time-domain high-band signal and the first channel low-band signal. For a second frame, based on a second value of the high-band reference channel indicator, the target channel signal may be based on the first channel time-domain high-band signal and the first channel low-band signal, and the reference channel signal may be based on the second channel time-domain high-band signal and the second channel low-band signal.

**[0007]** According to another implementation of the techniques disclosed herein, a method of communication includes receiving, at a device, at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters. The method also includes generating, at the device, a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal. The method further includes generating, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal. The method also includes generating, at the device, a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal. The method further includes generating, at the device, a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-

- 4 -

band signal. The method also includes generating, at the device, a modified target channel signal by modifying the target channel signal based on a temporal mismatch value. In an example implementation of the techniques disclosed herein, the receiver may be configured to receive the temporal mismatch value

**[0008]** According to another implementation of the techniques disclosed herein, a computer-readable storage device stores instructions that, when executed by a processor, cause the processor to perform operations including receiving at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters. The operations also include generating a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal. The operations further include generating, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal. The operations also include generating a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal. The operations further include generating a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal. The operations also include generating a modified target channel signal by modifying the target channel signal based on a temporal mismatch value.

**[0009]** According to another implementation of the techniques disclosed herein, an apparatus includes a receiver configured to receive at least one encoded signal. The device also includes a decoder configured to generate a first signal and a second signal based on the at least one encoded signal. The decoder is also configured to generate a shifted first signal by time-shifting first samples of the first signal relative to second samples of the second signal by an amount that is based on a shift value. The decoder is further configured to generate a first output signal based on the shifted first signal and to generate a second output signal based on the second signal.

**[0010]** According to another implementation of the techniques disclosed herein, a method of communication includes receiving, at a device, at least one encoded signal. The method also includes generating, at the device, a plurality of high-band signals

- 5 -

based on the at least one encoded signal. The method further includes generating, independently of the plurality of high-band signals, a plurality of low-band signals based on the at least one encoded signal.

[0011] According to another implementation of the techniques disclosed herein, a computer-readable storage device stores instructions that, when executed by a processor, cause the processor to perform operations including receiving a shift value and at least one encoded signal. The operations also include generating a plurality of high-band signals based on the at least one encoded signal and generating a plurality of low-band signals based on the at least one encoded signal and independently of the plurality of high-band signals. The operations also include generating a first signal based on a first low-band signal of the plurality of low-band signals, a first high-band signal of the plurality of high-band signals, or both. The operations also include generating a second signal based on a second low-band signal of the plurality of low-band signals, a second high-band signal of the plurality of high-band signals, or both. The operations also include generating a shifted first signal by time-shifting first samples of the first signal relative to second samples of the second signal by an amount that is based on the shift value. The operations further include generating a first output signal based on the shifted first signal and generating a second output signal based on the second signal.

[0012] According to another implementation of the techniques disclosed herein, an apparatus includes means for receiving at least one encoded signal. The apparatus also includes means for generating a first output signal based on a shifted first signal and a second output signal based on a second signal. The shifted first signal is generated by time-shifting first samples of a first signal relative to second samples of the second signal by an amount that is based on a shift value. The first signal and the second signal are based on the at least one encoded signal.

#### ***V. Brief Description of the Drawings***

[0013] FIG. 1 is a block diagram of a particular illustrative example of a system that includes a device operable to encode multiple audio signals;

- 6 -

**[0014]** FIG. 2 is a diagram illustrating another example of a system that includes the device of FIG. 1;

**[0015]** FIG. 3 is a diagram illustrating particular examples of samples that may be encoded by the device of FIG. 1;

**[0016]** FIG. 4 is a diagram illustrating particular examples of samples that may be encoded by the device of FIG. 1;

**[0017]** FIG. 5 is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0018]** FIG. 6 is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0019]** FIG. 7 is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0020]** FIG. 8 is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0021]** FIG. 9A is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0022]** FIG. 9B is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0023]** FIG. 9C is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0024]** FIG. 10A is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0025]** FIG. 10B is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0026]** FIG. 11 is a diagram illustrating another example of a system operable to encode

- 7 -

multiple audio signals;

**[0027]** FIG. 12 is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0028]** FIG. 13 is a flow chart illustrating a particular method of encoding multiple audio signals;

**[0029]** FIG. 14 is a diagram illustrating another example of a system operable to encode multiple audio signals;

**[0030]** FIG. 15 depicts graphs illustrating comparison values for voiced frames, transition frames, and unvoiced frames;

**[0031]** FIG. 16 is a flow chart illustrating a method of estimating a temporal offset between audio captured at multiple microphones;

**[0032]** FIG. 17 is a diagram for selectively expanding a search range for comparison values used for shift estimation;

**[0033]** FIG. 18 is depicts graphs illustrating selective expansion of a search range for comparison values used for shift estimation;

**[0034]** FIG. 19 includes a system that is operable to decode audio signals using non-causal shifting;

**[0035]** FIG. 20 illustrates a diagram of a first implementation of a decoder;

**[0036]** FIG. 21 illustrates a diagram of a second implementation of a decoder;

**[0037]** FIG. 22 illustrates a diagram of a third implementation of a decoder;

**[0038]** FIG. 23 illustrates a diagram of a fourth implementation of a decoder;

**[0039]** FIG. 24 is a flowchart of a method for decoding audio signals;

**[0040]** FIG. 25 is a flowchart of another method for decoding audio signals;

- 8 -

[0041] FIG. 26 is a flowchart of another method for decoding audio signals; and

[0042] FIG. 27 is a block diagram of a particular illustrative example of a device that is operable to perform the techniques described with respect to FIGS. 1-26.

#### ***VI. Detailed Description***

[0043] Systems and devices operable to encode multiple audio signals are disclosed. A device may include an encoder configured to encode the multiple audio signals. The multiple audio signals may be captured concurrently in time using multiple recording devices, e.g., multiple microphones. In some examples, the multiple audio signals (or multi-channel audio) may be synthetically (e.g., artificially) generated by multiplexing several audio channels that are recorded at the same time or at different times. As illustrative examples, the concurrent recording or multiplexing of the audio channels may result in a 2-channel configuration (i.e., Stereo: Left and Right), a 5.1 channel configuration (Left, Right, Center, Left Surround, Right Surround, and the low frequency emphasis (LFE) channels), a 7.1 channel configuration, a 7.1+4 channel configuration, a 22.2 channel configuration, or a N-channel configuration.

[0044] Audio capture devices in teleconference rooms (or telepresence rooms) may include multiple microphones that acquire spatial audio. The spatial audio may include speech as well as background audio that is encoded and transmitted. The speech/audio from a given source (e.g., a talker) may arrive at the multiple microphones at different times depending on how the microphones are arranged as well as where the source (e.g., the talker) is located with respect to the microphones and room dimensions. For example, a sound source (e.g., a talker) may be closer to a first microphone associated with the device than to a second microphone associated with the device. Thus, a sound emitted from the sound source may reach the first microphone earlier in time than the second microphone. The device may receive a first audio signal via the first microphone and may receive a second audio signal via the second microphone.

[0045] Mid-side (MS) coding and parametric stereo (PS) coding are stereo coding techniques that may provide improved efficiency over the dual-mono coding techniques. In dual-mono coding, the Left (L) channel (or signal) and the Right (R) channel (or

- 9 -

signal) are independently coded without making use of inter-channel correlation. MS coding reduces the redundancy between a correlated L/R channel-pair by transforming the Left channel and the Right channel to a sum-channel and a difference-channel (e.g., a side channel) prior to coding. The sum signal and the difference signal are waveform coded in MS coding. Relatively more bits are spent on the sum signal than on the side signal. PS coding reduces redundancy in each sub-band by transforming the L/R signals into a sum signal and a set of side parameters. The side parameters may indicate an inter-channel intensity difference (IID), an inter-channel phase difference (IPD), an inter-channel time difference (ITD), etc. The sum signal is waveform coded and transmitted along with the side parameters. In a hybrid system, the side-channel may be waveform coded in the lower bands (e.g., less than 2 kilohertz (kHz)) and PS coded in the upper bands (e.g., greater than or equal to 2 kHz) where the inter-channel phase preservation is perceptually less critical.

**[0046]** The MS coding and the PS coding may be done in either the frequency domain or in the sub-band domain. In some examples, the Left channel and the Right channel may be uncorrelated. For example, the Left channel and the Right channel may include uncorrelated synthetic signals. When the Left channel and the Right channel are uncorrelated, the coding efficiency of the MS coding, the PS coding, or both, may approach the coding efficiency of the dual-mono coding.

**[0047]** Depending on a recording configuration, there may be a temporal shift between a Left channel and a Right channel, as well as other spatial effects such as echo and room reverberation. If the temporal shift and phase mismatch between the channels are not compensated, the sum channel and the difference channel may contain comparable energies reducing the coding-gains associated with MS or PS techniques. The reduction in the coding-gains may be based on the amount of temporal (or phase) shift. The comparable energies of the sum signal and the difference signal may limit the usage of MS coding in certain frames where the channels are temporally shifted but are highly correlated. In stereo coding, a Mid channel (e.g., a sum channel) and a Side channel (e.g., a difference channel) may be generated based on the following Formula:

$$M = (L+R)/2, \quad S = (L-R)/2, \quad \text{Formula 1}$$

- 10 -

[0048] where M corresponds to the Mid channel, S corresponds to the Side channel, L corresponds to the Left channel, and R corresponds to the Right channel.

[0049] In some cases, the Mid channel and the Side channel may be generated based on the following Formula:

$$M=c(L+R), \quad S=c(L-R), \quad \text{Formula 2}$$

[0050] where c corresponds to a complex value which is frequency dependent.

Generating the Mid channel and the Side channel based on Formula 1 or Formula 2 may be referred to as performing a “downmixing” algorithm. A reverse process of generating the Left channel and the Right channel from the Mid channel and the Side channel based on Formula 1 or Formula 2 may be referred to as performing an “upmixing” algorithm.

[0051] An ad-hoc approach used to choose between MS coding or dual-mono coding for a particular frame may include generating a mid signal and a side signal, calculating energies of the mid signal and the side signal, and determining whether to perform MS coding based on the energies. For example, MS coding may be performed in response to determining that the ratio of energies of the side signal and the mid signal is less than a threshold. To illustrate, if a Right channel is shifted by at least a first time (e.g., about 0.001 seconds or 48 samples at 48 kHz), a first energy of the mid signal (corresponding to a sum of the left signal and the right signal) may be comparable to a second energy of the side signal (corresponding to a difference between the left signal and the right signal) for voiced speech frames. When the first energy is comparable to the second energy, a higher number of bits may be used to encode the Side channel, thereby reducing coding efficiency of MS coding relative to dual-mono coding. Dual-mono coding may thus be used when the first energy is comparable to the second energy (e.g., when the ratio of the first energy and the second energy is greater than or equal to the threshold). In an alternative approach, the decision between MS coding and dual-mono coding for a particular frame may be made based on a comparison of a threshold and normalized cross-correlation values of the Left channel and the Right channel.

[0052] In some examples, the encoder may determine a temporal shift value (or a

- 11 -

temporal mismatch value) indicative of a shift (or a temporal mismatch) of the first audio signal relative to the second audio signal. The shift value may correspond to an amount of temporal delay between receipt of the first audio signal at the first microphone and receipt of the second audio signal at the second microphone. Furthermore, the encoder may determine the shift value on a frame-by-frame basis, e.g., based on each 20 milliseconds (ms) speech/audio frame. For example, the shift value may correspond to an amount of time that a second frame of the second audio signal is delayed with respect to a first frame of the first audio signal. Alternatively, the shift value may correspond to an amount of time that the first frame of the first audio signal is delayed with respect to the second frame of the second audio signal.

**[0053]** When the sound source is closer to the first microphone than to the second microphone, frames of the second audio signal may be delayed relative to frames of the first audio signal. In this case, the first audio signal may be referred to as the “reference audio signal” or “reference channel” and the delayed second audio signal may be referred to as the “target audio signal” or “target channel”. Alternatively, when the sound source is closer to the second microphone than to the first microphone, frames of the first audio signal may be delayed relative to frames of the second audio signal. In this case, the second audio signal may be referred to as the reference audio signal or reference channel and the delayed first audio signal may be referred to as the target audio signal or target channel.

**[0054]** Depending on where the sound sources (e.g., talkers) are located in a conference or telepresence room or how the sound source (e.g., talker) position changes relative to the microphones, the reference channel and the target channel may change from one frame to another; similarly, the temporal delay value may also change from one frame to another. However, in some implementations, the shift value may always be positive to indicate an amount of delay of the “target” channel relative to the “reference” channel. Furthermore, the shift value may correspond to a “non-causal shift” value by which the delayed target channel is “pulled back” in time such that the target channel is aligned (e.g., maximally aligned) with the “reference” channel. The down mix algorithm to determine the mid channel and the side channel may be performed on the reference channel and the non-causal shifted target channel.

- 12 -

**[0055]** The encoder may determine the shift value based on the reference audio channel and a plurality of shift values applied to the target audio channel. For example, a first frame of the reference audio channel, X, may be received at a first time (m<sub>1</sub>). A first particular frame of the target audio channel, Y, may be received at a second time (n<sub>1</sub>) corresponding to a first shift value, e.g., shift1 = n<sub>1</sub> - m<sub>1</sub>. Further, a second frame of the reference audio channel may be received at a third time (m<sub>2</sub>). A second particular frame of the target audio channel may be received at a fourth time (n<sub>2</sub>) corresponding to a second shift value, e.g., shift2 = n<sub>2</sub> - m<sub>2</sub>.

**[0056]** The device may perform a framing or a buffering algorithm to generate a frame (e.g., 20 ms samples) at a first sampling rate (e.g., 32 kHz sampling rate (i.e., 640 samples per frame)). The encoder may, in response to determining that a first frame of the first audio signal and a second frame of the second audio signal arrive at the same time at the device, estimate a shift value (e.g., shift1) as equal to zero samples. A Left channel (e.g., corresponding to the first audio signal) and a Right channel (e.g., corresponding to the second audio signal) may be temporally aligned. In some cases, the Left channel and the Right channel, even when aligned, may differ in energy due to various reasons (e.g., microphone calibration).

**[0057]** In some examples, the Left channel and the Right channel may be temporally not aligned due to various reasons (e.g., a sound source, such as a talker, may be closer to one of the microphones than another and the two microphones may be greater than a threshold (e.g., 1-20 centimeters) distance apart). A location of the sound source relative to the microphones may introduce different delays in the Left channel and the Right channel. In addition, there may be a gain difference, an energy difference, or a level difference between the Left channel and the Right channel.

**[0058]** In some examples, a time of arrival of audio signals at the microphones from multiple sound sources (e.g., talkers) may vary when the multiple talkers are alternatively talking (e.g., without overlap). In such a case, the encoder may dynamically adjust a temporal shift value based on the talker to identify the reference channel. In some other examples, the multiple talkers may be talking at the same time, which may result in varying temporal shift values depending on who is the loudest

talker, closest to the microphone, etc.

**[0059]** In some examples, the first audio signal and second audio signal may be synthesized or artificially generated when the two signals potentially show less (e.g., no) correlation. It should be understood that the examples described herein are illustrative and may be instructive in determining a relationship between the first audio signal and the second audio signal in similar or different situations.

**[0060]** The encoder may generate comparison values (e.g., difference values or cross-correlation values) based on a comparison of a first frame of the first audio signal and a plurality of frames of the second audio signal. Each frame of the plurality of frames may correspond to a particular shift value. The encoder may generate a first estimated shift value based on the comparison values. For example, the first estimated shift value may correspond to a comparison value indicating a higher temporal-similarity (or lower difference) between the first frame of the first audio signal and a corresponding first frame of the second audio signal.

**[0061]** The encoder may determine the final shift value by refining, in multiple stages, a series of estimated shift values. For example, the encoder may first estimate a “tentative” shift value based on comparison values generated from stereo pre-processed and re-sampled versions of the first audio signal and the second audio signal. The encoder may generate interpolated comparison values associated with shift values proximate to the estimated “tentative” shift value. The encoder may determine a second estimated “interpolated” shift value based on the interpolated comparison values. For example, the second estimated “interpolated” shift value may correspond to a particular interpolated comparison value that indicates a higher temporal-similarity (or lower difference) than the remaining interpolated comparison values and the first estimated “tentative” shift value. If the second estimated “interpolated” shift value of the current frame (e.g., the first frame of the first audio signal) is different than a final shift value of a previous frame (e.g., a frame of the first audio signal that precedes the first frame), then the “interpolated” shift value of the current frame is further “amended” to improve the temporal-similarity between the first audio signal and the shifted second audio signal. In particular, a third estimated “amended” shift value may correspond to a more

accurate measure of temporal-similarity by searching around the second estimated “interpolated” shift value of the current frame and the final estimated shift value of the previous frame. The third estimated “amended” shift value is further conditioned to estimate the final shift value by limiting any spurious changes in the shift value between frames and further controlled to not switch from a negative shift value to a positive shift value (or vice versa) in two successive (or consecutive) frames as described herein.

**[0062]** In some examples, the encoder may refrain from switching between a positive shift value and a negative shift value or vice-versa in consecutive frames or in adjacent frames. For example, the encoder may set the final shift value to a particular value (e.g., 0) indicating no temporal-shift based on the estimated “interpolated” or “amended” shift value of the first frame and a corresponding estimated “interpolated” or “amended” or final shift value in a particular frame that precedes the first frame. To illustrate, the encoder may set the final shift value of the current frame (e.g., the first frame) to indicate no temporal-shift, i.e.,  $\text{shift1} = 0$ , in response to determining that one of the estimated “tentative” or “interpolated” or “amended” shift value of the current frame is positive and the other of the estimated “tentative” or “interpolated” or “amended” or “final” estimated shift value of the previous frame (e.g., the frame preceding the first frame) is negative. Alternatively, the encoder may also set the final shift value of the current frame (e.g., the first frame) to indicate no temporal-shift, i.e.,  $\text{shift1} = 0$ , in response to determining that one of the estimated “tentative” or “interpolated” or “amended” shift value of the current frame is negative and the other of the estimated “tentative” or “interpolated” or “amended” or “final” estimated shift value of the previous frame (e.g., the frame preceding the first frame) is positive.

**[0063]** The encoder may select a frame of the first audio signal or the second audio signal as a “reference” or “target” based on the shift value. For example, in response to determining that the final shift value is positive, the encoder may generate a reference channel or signal indicator having a first value (e.g., 0) indicating that the first audio signal is a “reference” signal and that the second audio signal is the “target” signal. Alternatively, in response to determining that the final shift value is negative, the encoder may generate the reference channel or signal indicator having a second value (e.g., 1) indicating that the second audio signal is the “reference” signal and that the first

audio signal is the “target” signal.

**[0064]** The encoder may estimate a relative gain (e.g., a relative gain parameter) associated with the reference signal and the non-causal shifted target signal. For example, in response to determining that the final shift value is positive, the encoder may estimate a gain value to normalize or equalize the amplitude or power levels of the first audio signal relative to the second audio signal that is offset by the non-causal shift value (e.g., an absolute value of the final shift value). Alternatively, in response to determining that the final shift value is negative, the encoder may estimate a gain value to normalize or equalize the amplitude or power levels of the non-causal shifted first audio signal relative to the second audio signal. In some examples, the encoder may estimate a gain value to normalize or equalize the amplitude or power levels of the “reference” signal relative to the non-causal shifted “target” signal. In other examples, the encoder may estimate the gain value (e.g., a relative gain value) based on the reference signal relative to the target signal (e.g., the unshifted target signal).

**[0065]** The encoder may generate at least one encoded signal (e.g., a mid signal, a side signal, or both) based on the reference signal, the target signal, the non-causal shift value, and the relative gain parameter. The side signal may correspond to a difference between first samples of the first frame of the first audio signal and selected samples of a selected frame of the second audio signal. The encoder may select the selected frame based on the final shift value. Fewer bits may be used to encode the side channel signal because of reduced difference between the first samples and the selected samples as compared to other samples of the second audio signal that correspond to a frame of the second audio signal that is received by the device at the same time as the first frame. A transmitter of the device may transmit the at least one encoded signal, the non-causal shift value, the relative gain parameter, the reference channel or signal indicator, or a combination thereof.

**[0066]** The encoder may generate at least one encoded signal (e.g., a mid signal, a side signal, or both) based on the reference signal, the target signal, the non-causal shift value, the relative gain parameter, low band parameters of a particular frame of the first audio signal, high band parameters of the particular frame, or a combination thereof.

The particular frame may precede the first frame. Certain low band parameters, high band parameters, or a combination thereof, from one or more preceding frames may be used to encode a mid signal, a side signal, or both, of the first frame. Encoding the mid signal, the side signal, or both, based on the low band parameters, the high band parameters, or a combination thereof, may improve estimates of the non-causal shift value and inter-channel relative gain parameter. The low band parameters, the high band parameters, or a combination thereof, may include a pitch parameter, a voicing parameter, a coder type parameter, a low-band energy parameter, a high-band energy parameter, a tilt parameter, a pitch gain parameter, a FCB gain parameter, a coding mode parameter, a voice activity parameter, a noise estimate parameter, a signal-to-noise ratio parameter, a formants parameter, a speech/music decision parameter, the non-causal shift, the inter-channel gain parameter, or a combination thereof. A transmitter of the device may transmit the at least one encoded signal, the non-causal shift value, the relative gain parameter, the reference channel (or signal) indicator, or a combination thereof.

**[0067]** Referring to FIG. 1, a particular illustrative example of a system is disclosed and generally designated 100. The system 100 includes a first device 104 communicatively coupled, via a network 120, to a second device 106. The network 120 may include one or more wireless networks, one or more wired networks, or a combination thereof.

**[0068]** The first device 104 may include an encoder 114, a transmitter 110, one or more input interfaces 112, or a combination thereof. A first input interface of the input interfaces 112 may be coupled to a first microphone 146. A second input interface of the input interface(s) 112 may be coupled to a second microphone 148. The encoder 114 may include a temporal equalizer 108 and may be configured to down mix and encode multiple audio signals, as described herein. The first device 104 may also include a memory 153 configured to store analysis data 190. The second device 106 may include a decoder 118. The decoder 118 may include a temporal balancer 124 that is configured to upmix and render the multiple channels. The second device 106 may be coupled to a first loudspeaker 142, a second loudspeaker 144, or both.

**[0069]** During operation, the first device 104 may receive a first audio signal 130 via

the first input interface from the first microphone 146 and may receive a second audio signal 132 via the second input interface from the second microphone 148. The first audio signal 130 may correspond to one of a right channel signal or a left channel signal. The second audio signal 132 may correspond to the other of the right channel signal or the left channel signal. A sound source 152 (e.g., a user, a speaker, ambient noise, a musical instrument, etc.) may be closer to the first microphone 146 than to the second microphone 148. Accordingly, an audio signal from the sound source 152 may be received at the input interface(s) 112 via the first microphone 146 at an earlier time than via the second microphone 148. This natural delay in the multi-channel signal acquisition through the multiple microphones may introduce a temporal shift between the first audio signal 130 and the second audio signal 132.

**[0070]** The temporal equalizer 108 may be configured to estimate a temporal offset between audio captured at the microphones 146, 148. The temporal offset may be estimated based on a delay between a first frame of the first audio signal 130 and a second frame of the second audio signal 132, where the second frame includes substantially similar content as the first frame. For example, the temporal equalizer 108 may determine a cross-correlation between the first frame and the second frame. The cross-correlation may measure the similarity of the two frames as a function of the lag of one frame relative to the other. Based on the cross-correlation, the temporal equalizer 108 may determine the delay (e.g., lag) between the first frame and the second frame. The temporal equalizer 108 may estimate the temporal offset between the first audio signal 130 and the second audio signal 132 based on the delay and historical delay data.

**[0071]** The historical data may include delays between frames captured from the first microphone 146 and corresponding frames captured from the second microphone 148. For example, the temporal equalizer 108 may determine a cross-correlation (e.g., a lag) between previous frames associated with the first audio signal 130 and corresponding frames associated with the second audio signal 132. Each lag may be represented by a “comparison value”. That is, a comparison value may indicate a time shift ( $k$ ) between a frame of the first audio signal 130 and a corresponding frame of the second audio signal 132. According to one implementation, the comparison values for previous

frames may be stored at the memory 153. A smoother 192 of the temporal equalizer 108 may “smooth” (or average) comparison values over a long-term set of frames and use the long-term smoothed comparison values for estimating a temporal offset (e.g., “shift”) between the first audio signal 130 and the second audio signal 132.

[0072] To illustrate, if  $CompVal_N(k)$  represents the comparison value at a shift of  $k$  for the frame N, the frame N may have comparison values from  $k=T\_MIN$  (a minimum shift) to  $k=T\_MAX$  (a maximum shift). The smoothing may be performed such that a long-term comparison value  $CompVal_{LT_N}(k)$  is represented by  $CompVal_{LT_N}(k) = f(CompVal_N(k), CompVal_{N-1}(k), CompVal_{LT_{N-2}}(k), \dots)$ . The function  $f$  in the above equation may be a function of all (or a subset) of past comparison values at the shift (k). An alternative representation of the long-term comparison value

$CompVal_{LT_N}(k)$  may be  $CompVal_{LT_N}(k) = g(CompVal_N(k), CompVal_{N-1}(k), CompVal_{N-2}(k), \dots)$ . The functions  $f$  or  $g$  may be simple finite impulse response (FIR) filters or infinite impulse response (IIR) filters, respectively. For example, the function  $g$  may be a single tap IIR filter such that the long-term comparison value  $CompVal_{LT_N}(k)$  is represented by  $CompVal_{LT_N}(k) = (1 - \alpha) * CompVal_N(k) + (\alpha) * CompVal_{LT_{N-1}}(k)$ , where  $\alpha \in (0, 1.0)$ . Thus, the long-term comparison value  $CompVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous comparison value  $CompVal_N(k)$  at frame N and the long-term comparison values  $CompVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases. In a particular aspect, the function  $f$  may be a L-tap FIR filter such that the long-term comparison value  $CompVal_{LT_N}(k)$  is represented by  $CompVal_{LT_N}(k) = (\alpha_1) * CompVal_N(k) + (\alpha_2) * CompVal_{N-1}(k) + \dots + (\alpha_L) * CompVal_{N-L+1}(k)$ , where  $\alpha_1, \alpha_2, \dots$ , and  $\alpha_L$  correspond to weights. In a particular aspect, each of the  $\alpha_1, \alpha_2, \dots$ , and  $\alpha_L \in (0, 1.0)$ , and one of the  $\alpha_1, \alpha_2, \dots$ , and  $\alpha_L$  may be the same as or distinct from another of the  $\alpha_1, \alpha_2, \dots$ , and  $\alpha_L$ . Thus, the long-term comparison value  $CompVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous comparison value  $CompVal_N(k)$  at frame N and the comparison values  $CompVal_{N-i}(k)$  over the previous (L-1) frames.

**[0073]** The smoothing techniques described above may substantially normalize the shift estimate between voiced frames, unvoiced frames, and transition frames. Normalized shift estimates may reduce sample repetition and artifact skipping at frame boundaries. Additionally, normalized shift estimates may result in reduced side channel energies, which may improve coding efficiency.

**[0074]** The temporal equalizer 108 may determine a final shift value 116 (e.g., a non-causal shift value) indicative of the shift (e.g., a non-causal shift) of the first audio signal 130 (e.g., “target”) relative to the second audio signal 132 (e.g., “reference”). The final shift value 116 may be based on the instantaneous comparison value  $CompVal_N(k)$  and the long-term comparison  $CompVal_{LT_{N-1}}(k)$ . For example, the smoothing operation described above may be performed on a tentative shift value, on an interpolated shift value, on an amended shift value, or a combination thereof, as described with respect to FIG. 5. The final shift value 116 may be based on the tentative shift value, the interpolated shift value, and the amended shift value, as described with respect to FIG. 5. A first value (e.g., a positive value) of the final shift value 116 may indicate that the second audio signal 132 is delayed relative to the first audio signal 130. A second value (e.g., a negative value) of the final shift value 116 may indicate that the first audio signal 130 is delayed relative to the second audio signal 132. A third value (e.g., 0) of the final shift value 116 may indicate no delay between the first audio signal 130 and the second audio signal 132.

**[0075]** In some implementations, the third value (e.g., 0) of the final shift value 116 may indicate that delay between the first audio signal 130 and the second audio signal 132 has switched sign. For example, a first particular frame of the first audio signal 130 may precede the first frame. The first particular frame and a second particular frame of the second audio signal 132 may correspond to the same sound emitted by the sound source 152. The delay between the first audio signal 130 and the second audio signal 132 may switch from having the first particular frame delayed with respect to the second particular frame to having the second frame delayed with respect to the first frame. Alternatively, the delay between the first audio signal 130 and the second audio signal 132 may switch from having the second particular frame delayed with respect to the first particular frame to having the first frame delayed with respect to the second

- 20 -

frame. The temporal equalizer 108 may set the final shift value 116 to indicate the third value (e.g., 0) in response to determining that the delay between the first audio signal 130 and the second audio signal 132 has switched sign.

[0076] The temporal equalizer 108 may generate a reference signal indicator 164 based on the final shift value 116. For example, the temporal equalizer 108 may, in response to determining that the final shift value 116 indicates a first value (e.g., a positive value), generate the reference signal indicator 164 to have a first value (e.g., 0) indicating that the first audio signal 130 is a “reference” signal. The temporal equalizer 108 may determine that the second audio signal 132 corresponds to a “target” signal in response to determining that the final shift value 116 indicates the first value (e.g., a positive value). Alternatively, the temporal equalizer 108 may, in response to determining that the final shift value 116 indicates a second value (e.g., a negative value), generate the reference signal indicator 164 to have a second value (e.g., 1) indicating that the second audio signal 132 is the “reference” signal. The temporal equalizer 108 may determine that the first audio signal 130 corresponds to the “target” signal in response to determining that the final shift value 116 indicates the second value (e.g., a negative value). The temporal equalizer 108 may, in response to determining that the final shift value 116 indicates a third value (e.g., 0), generate the reference signal indicator 164 to have a first value (e.g., 0) indicating that the first audio signal 130 is a “reference” signal. The temporal equalizer 108 may determine that the second audio signal 132 corresponds to a “target” signal in response to determining that the final shift value 116 indicates the third value (e.g., 0). Alternatively, the temporal equalizer 108 may, in response to determining that the final shift value 116 indicates the third value (e.g., 0), generate the reference signal indicator 164 to have a second value (e.g., 1) indicating that the second audio signal 132 is a “reference” signal. The temporal equalizer 108 may determine that the first audio signal 130 corresponds to a “target” signal in response to determining that the final shift value 116 indicates the third value (e.g., 0). In some implementations, the temporal equalizer 108 may, in response to determining that the final shift value 116 indicates a third value (e.g., 0), leave the reference signal indicator 164 unchanged. For example, the reference signal indicator 164 may be the same as a reference signal indicator corresponding to the first

- 21 -

particular frame of the first audio signal 130. The temporal equalizer 108 may generate a non-causal shift value 162 indicating an absolute value of the final shift value 116.

[0077] The temporal equalizer 108 may generate a gain parameter 160 (e.g., a codec gain parameter) based on samples of the “target” signal and based on samples of the “reference” signal. For example, the temporal equalizer 108 may select samples of the second audio signal 132 based on the non-causal shift value 162. Alternatively, the temporal equalizer 108 may select samples of the second audio signal 132 independent of the non-causal shift value 162. The temporal equalizer 108 may, in response to determining that the first audio signal 130 is the reference signal, determine the gain parameter 160 of the selected samples based on the first samples of the first frame of the first audio signal 130. Alternatively, the temporal equalizer 108 may, in response to determining that the second audio signal 132 is the reference signal, determine the gain parameter 160 of the first samples based on the selected samples. As an example, the gain parameter 160 may be based on one of the following Equations:

$$g_D = \frac{\sum_{n=0}^{N-N_1} Ref(n) Targ(n+N_1)}{\sum_{n=0}^{N-N_1} Targ^2(n+N_1)}, \quad \text{Equation 1a}$$

$$g_D = \frac{\sum_{n=0}^{N-N_1} |Ref(n)|}{\sum_{n=0}^{N-N_1} |Targ(n+N_1)|}, \quad \text{Equation 1b}$$

$$g_D = \frac{\sum_{n=0}^N Ref(n) Targ(n)}{\sum_{n=0}^N Targ^2(n)}, \quad \text{Equation 1c}$$

$$g_D = \frac{\sum_{n=0}^N |Ref(n)|}{\sum_{n=0}^N |Targ(n)|}, \quad \text{Equation 1d}$$

$$g_D = \frac{\sum_{n=0}^{N-N_1} Ref(n) Targ(n)}{\sum_{n=0}^N Ref^2(n)}, \quad \text{Equation 1e}$$

$$g_D = \frac{\sum_{n=0}^{N-N_1} |Targ(n)|}{\sum_{n=0}^N |Ref(n)|}, \quad \text{Equation 1f}$$

[0078] where  $g_D$  corresponds to the relative gain parameter 160 for down mix processing,  $Ref(n)$  corresponds to samples of the “reference” signal,  $N_1$  corresponds to the non-causal shift value 162 of the first frame, and  $Targ(n + N_1)$  corresponds to samples of the “target” signal. The gain parameter 160 ( $g_D$ ) may be modified, e.g., based on one of the Equations 1a – 1f, to incorporate long term smoothing/hysteresis

- 22 -

logic to avoid large jumps in gain between frames. When the target signal includes the first audio signal 130, the first samples may include samples of the target signal and the selected samples may include samples of the reference signal. When the target signal includes the second audio signal 132, the first samples may include samples of the reference signal, and the selected samples may include samples of the target signal.

**[0079]** In some implementations, the temporal equalizer 108 may generate the gain parameter 160 based on treating the first audio signal 130 as a reference signal and treating the second audio signal 132 as a target signal, irrespective of the reference signal indicator 164. For example, the temporal equalizer 108 may generate the gain parameter 160 based on one of the Equations 1a-1f where  $Ref(n)$  corresponds to samples (e.g., the first samples) of the first audio signal 130 and  $Targ(n+N_1)$  corresponds to samples (e.g., the selected samples) of the second audio signal 132. In alternate implementations, the temporal equalizer 108 may generate the gain parameter 160 based on treating the second audio signal 132 as a reference signal and treating the first audio signal 130 as a target signal, irrespective of the reference signal indicator 164. For example, the temporal equalizer 108 may generate the gain parameter 160 based on one of the Equations 1a-1f where  $Ref(n)$  corresponds to samples (e.g., the selected samples) of the second audio signal 132 and  $Targ(n+N_1)$  corresponds to samples (e.g., the first samples) of the first audio signal 130.

**[0080]** The temporal equalizer 108 may generate one or more encoded signals 102 (e.g., a mid channel signal, a side channel signal, or both) based on the first samples, the selected samples, and the relative gain parameter 160 for down mix processing. For example, the temporal equalizer 108 may generate the mid signal based on one of the following Equations:

$$M = Ref(n) + g_D Targ(n + N_1), \quad \text{Equation 2a}$$

$$M = Ref(n) + Targ(n + N_1), \quad \text{Equation 2b}$$

**[0081]** where  $M$  corresponds to the mid channel signal,  $g_D$  corresponds to the relative gain parameter 160 for downmix processing,  $Ref(n)$  corresponds to samples of the “reference” signal,  $N_1$  corresponds to the non-causal shift value 162 of the first frame,

- 23 -

and  $Targ(n + N_1)$  corresponds to samples of the “target” signal.

**[0082]** The temporal equalizer 108 may generate the side channel signal based on one of the following Equations:

$$S = Ref(n) - g_D Targ(n + N_1), \quad \text{Equation 3a}$$

$$S = g_D Ref(n) - Targ(n + N_1), \quad \text{Equation 3b}$$

**[0083]** where  $S$  corresponds to the side channel signal,  $g_D$  corresponds to the relative gain parameter 160 for downmix processing,  $Ref(n)$  corresponds to samples of the “reference” signal,  $N_1$  corresponds to the non-causal shift value 162 of the first frame, and  $Targ(n + N_1)$  corresponds to samples of the “target” signal.

**[0084]** The transmitter 110 may transmit the encoded signals 102 (e.g., the mid channel signal, the side channel signal, or both), the reference signal indicator 164, the non-causal shift value 162, the gain parameter 160, or a combination thereof, via the network 120, to the second device 106. In some implementations, the transmitter 110 may store the encoded signals 102 (e.g., the mid channel signal, the side channel signal, or both), the reference signal indicator 164, the non-causal shift value 162, the gain parameter 160, or a combination thereof, at a device of the network 120 or a local device for further processing or decoding later.

**[0085]** The decoder 118 may decode the encoded signals 102. The temporal balancer 124 may perform upmixing to generate a first output signal 126 (e.g., corresponding to first audio signal 130), a second output signal 128 (e.g., corresponding to the second audio signal 132), or both. The second device 106 may output the first output signal 126 via the first loudspeaker 142. The second device 106 may output the second output signal 128 via the second loudspeaker 144.

**[0086]** The system 100 may thus enable the temporal equalizer 108 to encode the side channel signal using fewer bits than the mid signal. The first samples of the first frame of the first audio signal 130 and selected samples of the second audio signal 132 may correspond to the same sound emitted by the sound source 152 and hence a difference between the first samples and the selected samples may be lower than between the first

- 24 -

samples and other samples of the second audio signal 132. The side channel signal may correspond to the difference between the first samples and the selected samples.

**[0087]** Referring to FIG. 2, a particular illustrative implementation of a system is disclosed and generally designated 200. The system 200 includes a first device 204 coupled, via the network 120, to the second device 106. The first device 204 may correspond to the first device 104 of FIG. 1. The system 200 differs from the system 100 of FIG. 1 in that the first device 204 is coupled to more than two microphones. For example, the first device 204 may be coupled to the first microphone 146, an Nth microphone 248, and one or more additional microphones (e.g., the second microphone 148 of FIG. 1). The second device 106 may be coupled to the first loudspeaker 142, a Yth loudspeaker 244, one or more additional speakers (e.g., the second loudspeaker 144), or a combination thereof. The first device 204 may include an encoder 214. The encoder 214 may correspond to the encoder 114 of FIG. 1. The encoder 214 may include one or more temporal equalizers 208. For example, the temporal equalizer(s) 208 may include the temporal equalizer 108 of FIG. 1.

**[0088]** During operation, the first device 204 may receive more than two audio signals. For example, the first device 204 may receive the first audio signal 130 via the first microphone 146, an Nth audio signal 232 via the Nth microphone 248, and one or more additional audio signals (e.g., the second audio signal 132) via the additional microphones (e.g., the second microphone 148).

**[0089]** The temporal equalizer(s) 208 may generate one or more reference signal indicators 264, final shift values 216, non-causal shift values 262, gain parameters 260, encoded signals 202, or a combination thereof. For example, the temporal equalizer(s) 208 may determine that the first audio signal 130 is a reference signal and that each of the Nth audio signal 232 and the additional audio signals is a target signal. The temporal equalizer(s) 208 may generate the reference signal indicator 164, the final shift values 216, the non-causal shift values 262, the gain parameters 260, and the encoded signals 202 corresponding to the first audio signal 130 and each of the Nth audio signal 232 and the additional audio signals.

**[0090]** The reference signal indicators 264 may include the reference signal indicator

- 25 -

164. The final shift values 216 may include the final shift value 116 indicative of a shift of the second audio signal 132 relative to the first audio signal 130, a second final shift value indicative of a shift of the Nth audio signal 232 relative to the first audio signal 130, or both. The non-causal shift values 262 may include the non-causal shift value 162 corresponding to an absolute value of the final shift value 116, a second non-causal shift value corresponding to an absolute value of the second final shift value, or both. The gain parameters 260 may include the gain parameter 160 of selected samples of the second audio signal 132, a second gain parameter of selected samples of the Nth audio signal 232, or both. The encoded signals 202 may include at least one of the encoded signals 102. For example, the encoded signals 202 may include the side channel signal corresponding to first samples of the first audio signal 130 and selected samples of the second audio signal 132, a second side channel corresponding to the first samples and selected samples of the Nth audio signal 232, or both. The encoded signals 202 may include a mid channel signal corresponding to the first samples, the selected samples of the second audio signal 132, and the selected samples of the Nth audio signal 232.

[0091] In some implementations, the temporal equalizer(s) 208 may determine multiple reference signals and corresponding target signals, as described with reference to FIG. 15. For example, the reference signal indicators 264 may include a reference signal indicator corresponding to each pair of reference signal and target signal. To illustrate, the reference signal indicators 264 may include the reference signal indicator 164 corresponding to the first audio signal 130 and the second audio signal 132. The final shift values 216 may include a final shift value corresponding to each pair of reference signal and target signal. For example, the final shift values 216 may include the final shift value 116 corresponding to the first audio signal 130 and the second audio signal 132. The non-causal shift values 262 may include a non-causal shift value corresponding to each pair of reference signal and target signal. For example, the non-causal shift values 262 may include the non-causal shift value 162 corresponding to the first audio signal 130 and the second audio signal 132. The gain parameters 260 may include a gain parameter corresponding to each pair of reference signal and target signal. For example, the gain parameters 260 may include the gain parameter 160 corresponding to the first audio signal 130 and the second audio signal 132. The

encoded signals 202 may include a mid channel signal and a side channel signal corresponding to each pair of reference signal and target signal. For example, the encoded signals 202 may include the encoded signals 102 corresponding to the first audio signal 130 and the second audio signal 132.

**[0092]** The transmitter 110 may transmit the reference signal indicators 264, the non-causal shift values 262, the gain parameters 260, the encoded signals 202, or a combination thereof, via the network 120, to the second device 106. The decoder 118 may generate one or more output signals based on the reference signal indicators 264, the non-causal shift values 262, the gain parameters 260, the encoded signals 202, or a combination thereof. For example, the decoder 118 may output a first output signal 226 via the first loudspeaker 142, a Yth output signal 228 via the Yth loudspeaker 244, one or more additional output signals (e.g., the second output signal 128) via one or more additional loudspeakers (e.g., the second loudspeaker 144), or a combination thereof. In another implementation, the transmitter 110 may refrain from transmitting the reference signal indicators 264, and the decoder 118 may generate the reference signal indicators 264 based on the final shift values 216 (of the current frame) and final shift values of previous frames.

**[0093]** The system 200 may thus enable the temporal equalizer(s) 208 to encode more than two audio signals. For example, the encoded signals 202 may include multiple side channel signals that are encoded using fewer bits than corresponding mid channels by generating the side channel signals based on the non-causal shift values 262.

**[0094]** Referring to FIG. 3, illustrative examples of samples are shown and generally designated 300. At least a subset of the samples 300 may be encoded by the first device 104, as described herein.

**[0095]** The samples 300 may include first samples 320 corresponding to the first audio signal 130, second samples 350 corresponding to the second audio signal 132, or both. The first samples 320 may include a sample 322, a sample 324, a sample 326, a sample 328, a sample 330, a sample 332, a sample 334, a sample 336, one or more additional samples, or a combination thereof. The second samples 350 may include a sample 352, a sample 354, a sample 356, a sample 358, a sample 360, a sample 362, a sample 364, a

sample 366, one or more additional samples, or a combination thereof.

**[0096]** The first audio signal 130 may correspond to a plurality of frames (e.g., a frame 302, a frame 304, a frame 306, or a combination thereof). Each of the plurality of frames may correspond to a subset of samples (e.g., corresponding to 20 ms, such as 640 samples at 32 kHz or 960 samples at 48 kHz) of the first samples 320. For example, the frame 302 may correspond to the sample 322, the sample 324, one or more additional samples, or a combination thereof. The frame 304 may correspond to the sample 326, the sample 328, the sample 330, the sample 332, one or more additional samples, or a combination thereof. The frame 306 may correspond to the sample 334, the sample 336, one or more additional samples, or a combination thereof.

**[0097]** The sample 322 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 352. The sample 324 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 354. The sample 326 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 356. The sample 328 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 358. The sample 330 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 360. The sample 332 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 362. The sample 334 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 364. The sample 336 may be received at the input interface(s) 112 of FIG. 1 at approximately the same time as the sample 366.

**[0098]** A first value (e.g., a positive value) of the final shift value 116 may indicate that the second audio signal 132 is delayed relative to the first audio signal 130. For example, a first value (e.g., +X ms or +Y samples, where X and Y include positive real numbers) of the final shift value 116 may indicate that the frame 304 (e.g., the samples 326-332) correspond to the samples 358-364. The samples 326-332 and the samples 358-364 may correspond to the same sound emitted from the sound source 152. The samples 358-364 may correspond to a frame 344 of the second audio signal 132. Illustration of samples with cross-hatching in one or more of FIGS. 1-15 may indicate

that the samples correspond to the same sound. For example, the samples 326-332 and the samples 358-364 are illustrated with cross-hatching in FIG. 3 to indicate that the samples 326-332 (e.g., the frame 304) and the samples 358-364 (e.g., the frame 344) correspond to the same sound emitted from the sound source 152.

**[0099]** It should be understood that a temporal offset of Y samples, as shown in FIG. 3, is illustrative. For example, the temporal offset may correspond to a number of samples, Y, that is greater than or equal to 0. In a first case where the temporal offset  $Y = 0$  samples, the samples 326-332 (e.g., corresponding to the frame 304) and the samples 356-362 (e.g., corresponding to the frame 344) may show high similarity without any frame offset. In a second case where the temporal offset  $Y = 2$  samples, the frame 304 and frame 344 may be offset by 2 samples. In this case, the first audio signal 130 may be received prior to the second audio signal 132 at the input interface(s) 112 by  $Y = 2$  samples or  $X = (2/F_s)$  ms, where  $F_s$  corresponds to the sample rate in kHz. In some cases, the temporal offset, Y, may include a non-integer value, e.g.,  $Y = 1.6$  samples corresponding to  $X = 0.05$  ms at 32 kHz.

**[0100]** The temporal equalizer 108 of FIG. 1 may generate the encoded signals 102 by encoding the samples 326-332 and the samples 358-364, as described with reference to FIG. 1. The temporal equalizer 108 may determine that the first audio signal 130 corresponds to a reference signal and that the second audio signal 132 corresponds to a target signal.

**[0101]** Referring to FIG. 4, illustrative examples of samples are shown and generally designated as 400. The samples 400 differ from the samples 300 in that the first audio signal 130 is delayed relative to the second audio signal 132.

**[0102]** A second value (e.g., a negative value) of the final shift value 116 may indicate that the first audio signal 130 is delayed relative to the second audio signal 132. For example, the second value (e.g.,  $-X$  ms or  $-Y$  samples, where X and Y include positive real numbers) of the final shift value 116 may indicate that the frame 304 (e.g., the samples 326-332) correspond to the samples 354-360. The samples 354-360 may correspond to the frame 344 of the second audio signal 132. The samples 354-360 (e.g., the frame 344) and the samples 326-332 (e.g., the frame 304) may correspond to the

same sound emitted from the sound source 152.

**[0103]** It should be understood that a temporal offset of -Y samples, as shown in FIG. 4, is illustrative. For example, the temporal offset may correspond to a number of samples, -Y, that is less than or equal to 0. In a first case where the temporal offset Y = 0 samples, the samples 326-332 (e.g., corresponding to the frame 304) and the samples 356-362 (e.g., corresponding to the frame 344) may show high similarity without any frame offset. In a second case where the temporal offset Y = -6 samples, the frame 304 and frame 344 may be offset by 6 samples. In this case, the first audio signal 130 may be received subsequent to the second audio signal 132 at the input interface(s) 112 by Y = -6 samples or X = (-6/Fs) ms, where Fs corresponds to the sample rate in kHz. In some cases, the temporal offset, Y, may include a non-integer value, e.g., Y = -3.2 samples corresponding to X = -0.1 ms at 32 kHz.

**[0104]** The temporal equalizer 108 of FIG. 1 may generate the encoded signals 102 by encoding the samples 354-360 and the samples 326-332, as described with reference to FIG. 1. The temporal equalizer 108 may determine that the second audio signal 132 corresponds to a reference signal and that the first audio signal 130 corresponds to a target signal. In particular, the temporal equalizer 108 may estimate the non-causal shift value 162 from the final shift value 116, as described with reference to FIG. 5. The temporal equalizer 108 may identify (e.g., designate) one of the first audio signal 130 or the second audio signal 132 as a reference signal and the other of the first audio signal 130 or the second audio signal 132 as a target signal based on a sign of the final shift value 116.

**[0105]** Referring to FIG. 5, an illustrative example of a system is shown and generally designated 500. The system 500 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 500. The temporal equalizer 108 may include a resampler 504, a signal comparator 506, an interpolator 510, a shift refiner 511, a shift change analyzer 512, an absolute shift generator 513, a reference signal designator 508, a gain parameter generator 514, a signal generator 516, or a combination thereof.

**[0106]** During operation, the resampler 504 may generate one or more resampled

- 30 -

signals, as further described with reference to FIG. 6. For example, the resampler 504 may generate a first resampled signal 530 by resampling (e.g., downsampling or upsampling) the first audio signal 130 based on a resampling (e.g., downsampling or upsampling) factor (D) (e.g.,  $\geq 1$ ). The resampler 504 may generate a second resampled signal 532 by resampling the second audio signal 132 based on the resampling factor (D). The resampler 504 may provide the first resampled signal 530, the second resampled signal 532, or both, to the signal comparator 506.

**[0107]** The signal comparator 506 may generate comparison values 534 (e.g., difference values, similarity values, coherence values, or cross-correlation values), a tentative shift value 536, or both, as further described with reference to FIG. 7. For example, the signal comparator 506 may generate the comparison values 534 based on the first resampled signal 530 and a plurality of shift values applied to the second resampled signal 532, as further described with reference to FIG. 7. The signal comparator 506 may determine the tentative shift value 536 based on the comparison values 534, as further described with reference to FIG. 7. According to one implementation, the signal comparator 506 may retrieve comparison values for previous frames of the resampled signals 530, 532 and may modify the comparison values 534 based on a long-term smoothing operation using the comparison values for previous frames. For example, the comparison values 534 may include the long-term comparison value  $CompVal_{LT_N}(k)$  for a current frame (N) and may be represented by  $CompVal_{LT_N}(k) = (1 - \alpha) * CompVal_N(k) + (\alpha) * CompVal_{LT_{N-1}}(k)$ , where  $\alpha \in (0, 1.0)$ . Thus, the long-term comparison value  $CompVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous comparison value  $CompVal_N(k)$  at frame N and the long-term comparison values  $CompVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases.

**[0108]** The first resampled signal 530 may include fewer samples or more samples than the first audio signal 130. The second resampled signal 532 may include fewer samples or more samples than the second audio signal 132. Determining the comparison values 534 based on the fewer samples of the resampled signals (e.g., the first resampled signal 530 and the second resampled signal 532) may use fewer resources (e.g., time, number

of operations, or both) than on samples of the original signals (e.g., the first audio signal 130 and the second audio signal 132). Determining the comparison values 534 based on the more samples of the resampled signals (e.g., the first resampled signal 530 and the second resampled signal 532) may increase precision than on samples of the original signals (e.g., the first audio signal 130 and the second audio signal 132). The signal comparator 506 may provide the comparison values 534, the tentative shift value 536, or both, to the interpolator 510.

**[0109]** The interpolator 510 may extend the tentative shift value 536. For example, the interpolator 510 may generate an interpolated shift value 538, as further described with reference to FIG. 8. For example, the interpolator 510 may generate interpolated comparison values corresponding to shift values that are proximate to the tentative shift value 536 by interpolating the comparison values 534. The interpolator 510 may determine the interpolated shift value 538 based on the interpolated comparison values and the comparison values 534. The comparison values 534 may be based on a coarser granularity of the shift values. For example, the comparison values 534 may be based on a first subset of a set of shift values so that a difference between a first shift value of the first subset and each second shift value of the first subset is greater than or equal to a threshold (e.g.,  $\geq 1$ ). The threshold may be based on the resampling factor (D).

**[0110]** The interpolated comparison values may be based on a finer granularity of shift values that are proximate to the resampled tentative shift value 536. For example, the interpolated comparison values may be based on a second subset of the set of shift values so that a difference between a highest shift value of the second subset and the resampled tentative shift value 536 is less than the threshold (e.g.,  $\geq 1$ ), and a difference between a lowest shift value of the second subset and the resampled tentative shift value 536 is less than the threshold. Determining the comparison values 534 based on the coarser granularity (e.g., the first subset) of the set of shift values may use fewer resources (e.g., time, operations, or both) than determining the comparison values 534 based on a finer granularity (e.g., all) of the set of shift values. Determining the interpolated comparison values corresponding to the second subset of shift values may extend the tentative shift value 536 based on a finer granularity of a smaller set of shift values that are proximate to the tentative shift value 536 without determining

comparison values corresponding to each shift value of the set of shift values. Thus, determining the tentative shift value 536 based on the first subset of shift values and determining the interpolated shift value 538 based on the interpolated comparison values may balance resource usage and refinement of the estimated shift value. The interpolator 510 may provide the interpolated shift value 538 to the shift refiner 511.

**[0111]** According to one implementation, the interpolator 510 may retrieve interpolated shift values for previous frames and may modify the interpolated shift value 538 based on a long-term smoothing operation using the interpolated shift values for previous frames. For example, the interpolated shift value 538 may include a long-term interpolated shift value  $InterVal_{LT_N}(k)$  for a current frame (N) and may be represented by  $InterVal_{LT_N}(k) = (1 - \alpha) * InterVal_N(k) + (\alpha) * InterVal_{LT_{N-1}}(k)$ , where  $\alpha \in (0, 1.0)$ . Thus, the long-term interpolated shift value  $InterVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous interpolated shift value  $InterVal_N(k)$  at frame N and the long-term interpolated shift values  $InterVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases.

**[0112]** The shift refiner 511 may generate an amended shift value 540 by refining the interpolated shift value 538, as further described with reference to FIGS. 9A-9C. For example, the shift refiner 511 may determine whether the interpolated shift value 538 indicates that a change in a shift between the first audio signal 130 and the second audio signal 132 is greater than a shift change threshold, as further described with reference to FIG. 9A. The change in the shift may be indicated by a difference between the interpolated shift value 538 and a first shift value associated with the frame 302 of FIG. 3. The shift refiner 511 may, in response to determining that the difference is less than or equal to the threshold, set the amended shift value 540 to the interpolated shift value 538. Alternatively, the shift refiner 511 may, in response to determining that the difference is greater than the threshold, determine a plurality of shift values that correspond to a difference that is less than or equal to the shift change threshold, as further described with reference to FIG. 9A. The shift refiner 511 may determine comparison values based on the first audio signal 130 and the plurality of shift values

applied to the second audio signal 132. The shift refiner 511 may determine the amended shift value 540 based on the comparison values, as further described with reference to FIG. 9A. For example, the shift refiner 511 may select a shift value of the plurality of shift values based on the comparison values and the interpolated shift value 538, as further described with reference to FIG. 9A. The shift refiner 511 may set the amended shift value 540 to indicate the selected shift value. A non-zero difference between the first shift value corresponding to the frame 302 and the interpolated shift value 538 may indicate that some samples of the second audio signal 132 correspond to both frames (e.g., the frame 302 and the frame 304). For example, some samples of the second audio signal 132 may be duplicated during encoding. Alternatively, the non-zero difference may indicate that some samples of the second audio signal 132 correspond to neither the frame 302 nor the frame 304. For example, some samples of the second audio signal 132 may be lost during encoding. Setting the amended shift value 540 to one of the plurality of shift values may prevent a large change in shifts between consecutive (or adjacent) frames, thereby reducing an amount of sample loss or sample duplication during encoding. The shift refiner 511 may provide the amended shift value 540 to the shift change analyzer 512.

[0113] According to one implementation, the shift refiner may retrieve amended shift values for previous frames and may modify the amended shift value 540 based on a long-term smoothing operation using the amended shift values for previous frames. For example, the amended shift value 540 may include a long-term amended shift value  $AmendVal_{LT_N}(k)$  for a current frame (N) and may be represented by

$$AmendVal_{LT_N}(k) = (1 - \alpha) * AmendVal_N(k) + (\alpha) * AmendVal_{LT_{N-1}}(k),$$
 where  $\alpha \in (0, 1.0)$ . Thus, the long-term amended shift value  $AmendVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous amended shift value  $AmendVal_N(k)$  at frame N and the long-term amended shift values  $AmendVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases.

[0114] In some implementations, the shift refiner 511 may adjust the interpolated shift value 538, as described with reference to FIG. 9B. The shift refiner 511 may determine

the amended shift value 540 based on the adjusted interpolated shift value 538. In some implementations, the shift refiner 511 may determine the amended shift value 540 as described with reference to FIG. 9C.

**[0115]** The shift change analyzer 512 may determine whether the amended shift value 540 indicates a switch or reverse in timing between the first audio signal 130 and the second audio signal 132, as described with reference to FIG. 1. In particular, a reverse or a switch in timing may indicate that, for the frame 302, the first audio signal 130 is received at the input interface(s) 112 prior to the second audio signal 132, and, for a subsequent frame (e.g., the frame 304 or the frame 306), the second audio signal 132 is received at the input interface(s) prior to the first audio signal 130. Alternatively, a reverse or a switch in timing may indicate that, for the frame 302, the second audio signal 132 is received at the input interface(s) 112 prior to the first audio signal 130, and, for a subsequent frame (e.g., the frame 304 or the frame 306), the first audio signal 130 is received at the input interface(s) prior to the second audio signal 132. In other words, a switch or reverse in timing may indicate that a final shift value corresponding to the frame 302 has a first sign that is distinct from a second sign of the amended shift value 540 corresponding to the frame 304 (e.g., a positive to negative transition or vice-versa). The shift change analyzer 512 may determine whether delay between the first audio signal 130 and the second audio signal 132 has switched sign based on the amended shift value 540 and the first shift value associated with the frame 302, as further described with reference to FIG. 10A. The shift change analyzer 512 may, in response to determining that the delay between the first audio signal 130 and the second audio signal 132 has switched sign, set the final shift value 116 to a value (e.g., 0) indicating no time shift. Alternatively, the shift change analyzer 512 may set the final shift value 116 to the amended shift value 540 in response to determining that the delay between the first audio signal 130 and the second audio signal 132 has not switched sign, as further described with reference to FIG. 10A. The shift change analyzer 512 may generate an estimated shift value by refining the amended shift value 540, as further described with reference to FIGS. 10A,11. The shift change analyzer 512 may set the final shift value 116 to the estimated shift value. Setting the final shift value 116 to indicate no time shift may reduce distortion at a decoder by refraining from

time shifting the first audio signal 130 and the second audio signal 132 in opposite directions for consecutive (or adjacent) frames of the first audio signal 130. The shift change analyzer 512 may provide the final shift value 116 to the reference signal designator 508, to the absolute shift generator 513, or both. In some implementations, the shift change analyzer 512 may determine the final shift value 116 as described with reference to FIG. 10B.

**[0116]** The absolute shift generator 513 may generate the non-causal shift value 162 by applying an absolute function to the final shift value 116. The absolute shift generator 513 may provide the non-causal shift value 162 to the gain parameter generator 514.

**[0117]** The reference signal designator 508 may generate the reference signal indicator 164, as further described with reference to FIGS. 12-13. For example, the reference signal indicator 164 may have a first value indicating that the first audio signal 130 is a reference signal or a second value indicating that the second audio signal 132 is the reference signal. The reference signal designator 508 may provide the reference signal indicator 164 to the gain parameter generator 514.

**[0118]** The gain parameter generator 514 may select samples of the target signal (e.g., the second audio signal 132) based on the non-causal shift value 162. To illustrate, the gain parameter generator 514 may select the samples 358-364 in response to determining that the non-causal shift value 162 has a first value (e.g., +X ms or +Y samples, where X and Y include positive real numbers). The gain parameter generator 514 may select the samples 354-360 in response to determining that the non-causal shift value 162 has a second value (e.g., -X ms or -Y samples). The gain parameter generator 514 may select the samples 356-362 in response to determining that the non-causal shift value 162 has a value (e.g., 0) indicating no time shift.

**[0119]** The gain parameter generator 514 may determine whether the first audio signal 130 is the reference signal or the second audio signal 132 is the reference signal based on the reference signal indicator 164. The gain parameter generator 514 may generate the gain parameter 160 based on the samples 326-332 of the frame 304 and the selected samples (e.g., the samples 354-360, the samples 356-362, or the samples 358-364) of the second audio signal 132, as described with reference to FIG. 1. For example, the

gain parameter generator 514 may generate the gain parameter 160 based on one or more of Equation 1a - Equation 1f, where  $g_D$  corresponds to the gain parameter 160,  $Ref(n)$  corresponds to samples of the reference signal, and  $Targ(n+N_1)$  corresponds to samples of the target signal. To illustrate,  $Ref(n)$  may correspond to the samples 326-332 of the frame 304 and  $Targ(n+N_1)$  may correspond to the samples 358-364 of the frame 344 when the non-causal shift value 162 has a first value (e.g.,  $+X$  ms or  $+Y$  samples, where  $X$  and  $Y$  include positive real numbers). In some implementations,  $Ref(n)$  may correspond to samples of the first audio signal 130 and  $Targ(n+N_1)$  may correspond to samples of the second audio signal 132, as described with reference to FIG. 1. In alternate implementations,  $Ref(n)$  may correspond to samples of the second audio signal 132 and  $Targ(n+N_1)$  may correspond to samples of the first audio signal 130, as described with reference to FIG. 1.

**[0120]** The gain parameter generator 514 may provide the gain parameter 160, the reference signal indicator 164, the non-causal shift value 162, or a combination thereof, to the signal generator 516. The signal generator 516 may generate the encoded signals 102, as described with reference to FIG. 1. For examples, the encoded signals 102 may include a first encoded signal frame 564 (e.g., a mid channel frame), a second encoded signal frame 566 (e.g., a side channel frame), or both. The signal generator 516 may generate the first encoded signal frame 564 based on Equation 2a or Equation 2b, where  $M$  corresponds to the first encoded signal frame 564,  $g_D$  corresponds to the gain parameter 160,  $Ref(n)$  corresponds to samples of the reference signal, and  $Targ(n+N_1)$  corresponds to samples of the target signal. The signal generator 516 may generate the second encoded signal frame 566 based on Equation 3a or Equation 3b, where  $S$  corresponds to the second encoded signal frame 566,  $g_D$  corresponds to the gain parameter 160,  $Ref(n)$  corresponds to samples of the reference signal, and  $Targ(n+N_1)$  corresponds to samples of the target signal.

**[0121]** The temporal equalizer 108 may store the first resampled signal 530, the second resampled signal 532, the comparison values 534, the tentative shift value 536, the interpolated shift value 538, the amended shift value 540, the non-causal shift value 162, the reference signal indicator 164, the final shift value 116, the gain parameter 160, the first encoded signal frame 564, the second encoded signal frame 566, or a

combination thereof, in the memory 153. For example, the analysis data 190 may include the first resampled signal 530, the second resampled signal 532, the comparison values 534, the tentative shift value 536, the interpolated shift value 538, the amended shift value 540, the non-causal shift value 162, the reference signal indicator 164, the final shift value 116, the gain parameter 160, the first encoded signal frame 564, the second encoded signal frame 566, or a combination thereof.

**[0122]** The smoothing techniques described above may substantially normalize the shift estimate between voiced frames, unvoiced frames, and transition frames. Normalized shift estimates may reduce sample repetition and artifact skipping at frame boundaries. Additionally, normalized shift estimates may result in reduced side channel energies, which may improve coding efficiency.

**[0123]** Referring to FIG. 6, an illustrative example of a system is shown and generally designated 600. The system 600 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 600.

**[0124]** The resampler 504 may generate first samples 620 of the first resampled signal 530 by resampling (e.g., downsampling or upsampling) the first audio signal 130 of FIG. 1. The resampler 504 may generate second samples 650 of the second resampled signal 532 by resampling (e.g., downsampling or upsampling) the second audio signal 132 of FIG. 1.

**[0125]** The first audio signal 130 may be sampled at a first sample rate (Fs) to generate the first samples 320 of FIG. 3. The first sample rate (Fs) may correspond to a first rate (e.g., 16 kilohertz (kHz)) associated with wideband (WB) bandwidth, a second rate (e.g., 32 kHz) associated with super wideband (SWB) bandwidth, a third rate (e.g., 48 kHz) associated with full band (FB) bandwidth, or another rate. The second audio signal 132 may be sampled at the first sample rate (Fs) to generate the second samples 350 of FIG. 3.

**[0126]** In some implementations, the resampler 504 may pre-process the first audio signal 130 (or the second audio signal 132) prior to resampling the first audio signal 130

(or the second audio signal 132). The resampler 504 may pre-process the first audio signal 130 (or the second audio signal 132) by filtering the first audio signal 130 (or the second audio signal 132) based on an infinite impulse response (IIR) filter (e.g., a first order IIR filter). The IIR filter may be based on the following Equation:

$$H_{pre}(z) = 1/(1 - \alpha z^{-1}), \quad \text{Equation 4}$$

[0127] where  $\alpha$  is positive, such as 0.68 or 0.72. Performing the de-emphasis prior to resampling may reduce effects, such as aliasing, signal conditioning, or both. The first audio signal 130 (e.g., the pre-processed first audio signal 130) and the second audio signal 132 (e.g., the pre- processed second audio signal 132) may be resampled based on a resampling factor (D). The resampling factor (D) may be based on the first sample rate (Fs) (e.g., D = Fs/8, D=2Fs, etc.).

[0128] In alternate implementations, the first audio signal 130 and the second audio signal 132 may be low-pass filtered or decimated using an anti-aliasing filter prior to resampling. The decimation filter may be based on the resampling factor (D). In a particular example, the resampler 504 may select a decimation filter with a first cut-off frequency (e.g.,  $\pi/D$  or  $\pi/4$ ) in response to determining that the first sample rate (Fs) corresponds to a particular rate (e.g., 32 kHz). Reducing aliasing by de-emphasizing multiple signals (e.g., the first audio signal 130 and the second audio signal 132) may be computationally less expensive than applying a decimation filter to the multiple signals.

[0129] The first samples 620 may include a sample 622, a sample 624, a sample 626, a sample 628, a sample 630, a sample 632, a sample 634, a sample 636, one or more additional samples, or a combination thereof. The first samples 620 may include a subset (e.g., 1/8th) of the first samples 320 of FIG. 3. The sample 622, the sample 624, one or more additional samples, or a combination thereof, may correspond to the frame 302. The sample 626, the sample 628, the sample 630, the sample 632, one or more additional samples, or a combination thereof, may correspond to the frame 304. The sample 634, the sample 636, one or more additional samples, or a combination thereof, may correspond to the frame 306.

[0130] The second samples 650 may include a sample 652, a sample 654, a sample 656,

a sample 658, a sample 660, a sample 662, a sample 664, a sample 668, one or more additional samples, or a combination thereof. The second samples 650 may include a subset (e.g., 1/8th) of the second samples 350 of FIG. 3. The samples 654-660 may correspond to the samples 354-360. For example, the samples 654-660 may include a subset (e.g., 1/8th) of the samples 354-360. The samples 656-662 may correspond to the samples 356-362. For example, the samples 656-662 may include a subset (e.g., 1/8th) of the samples 356-362. The samples 658-664 may correspond to the samples 358-364. For example, the samples 658-664 may include a subset (e.g., 1/8th) of the samples 358-364. In some implementations, the resampling factor may correspond to a first value (e.g., 1) where samples 622-636 and samples 652-668 of FIG. 6 may be similar to samples 322-336 and samples 352-366 of FIG. 3, respectively.

**[0131]** The resampler 504 may store the first samples 620, the second samples 650, or both, in the memory 153. For example, the analysis data 190 may include the first samples 620, the second samples 650, or both.

**[0132]** Referring to FIG. 7, an illustrative example of a system is shown and generally designated 700. The system 700 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 700.

**[0133]** The memory 153 may store a plurality of shift values 760. The shift values 760 may include a first shift value 764 (e.g., -X ms or -Y samples, where X and Y include positive real numbers), a second shift value 766 (e.g., +X ms or +Y samples, where X and Y include positive real numbers), or both. The shift values 760 may range from a lower shift value (e.g., a minimum shift value, T\_MIN) to a higher shift value (e.g., a maximum shift value, T\_MAX). The shift values 760 may indicate an expected temporal shift (e.g., a maximum expected temporal shift) between the first audio signal 130 and the second audio signal 132.

**[0134]** During operation, the signal comparator 506 may determine the comparison values 534 based on the first samples 620 and the shift values 760 applied to the second samples 650. For example, the samples 626-632 may correspond to a first time (t). To illustrate, the input interface(s) 112 of FIG. 1 may receive the samples 626-632

- 40 -

corresponding to the frame 304 at approximately the first time (t). The first shift value 764 (e.g., -X ms or -Y samples, where X and Y include positive real numbers) may correspond to a second time (t-1).

**[0135]** The samples 654-660 may correspond to the second time (t-1). For example, the input interface(s) 112 may receive the samples 654-660 at approximately the second time (t-1). The signal comparator 506 may determine a first comparison value 714 (e.g., a difference value or a cross-correlation value) corresponding to the first shift value 764 based on the samples 626-632 and the samples 654-660. For example, the first comparison value 714 may correspond to an absolute value of cross-correlation of the samples 626-632 and the samples 654-660. As another example, the first comparison value 714 may indicate a difference between the samples 626-632 and the samples 654-660.

**[0136]** The second shift value 766 (e.g., +X ms or +Y samples, where X and Y include positive real numbers) may correspond to a third time (t+1). The samples 658-664 may correspond to the third time (t+1). For example, the input interface(s) 112 may receive the samples 658-664 at approximately the third time (t+1). The signal comparator 506 may determine a second comparison value 716 (e.g., a difference value or a cross-correlation value) corresponding to the second shift value 766 based on the samples 626-632 and the samples 658-664. For example, the second comparison value 716 may correspond to an absolute value of cross-correlation of the samples 626-632 and the samples 658-664. As another example, the second comparison value 716 may indicate a difference between the samples 626-632 and the samples 658-664. The signal comparator 506 may store the comparison values 534 in the memory 153. For example, the analysis data 190 may include the comparison values 534.

**[0137]** The signal comparator 506 may identify a selected comparison value 736 of the comparison values 534 that has a higher (or lower) value than other values of the comparison values 534. For example, the signal comparator 506 may select the second comparison value 716 as the selected comparison value 736 in response to determining that the second comparison value 716 is greater than or equal to the first comparison value 714. In some implementations, the comparison values 534 may correspond to

- 41 -

cross-correlation values. The signal comparator 506 may, in response to determining that the second comparison value 716 is greater than the first comparison value 714, determine that the samples 626-632 have a higher correlation with the samples 658-664 than with the samples 654-660. The signal comparator 506 may select the second comparison value 716 that indicates the higher correlation as the selected comparison value 736. In other implementations, the comparison values 534 may correspond to difference values. The signal comparator 506 may, in response to determining that the second comparison value 716 is lower than the first comparison value 714, determine that the samples 626-632 have a greater similarity with (e.g., a lower difference to) the samples 658-664 than the samples 654-660. The signal comparator 506 may select the second comparison value 716 that indicates a lower difference as the selected comparison value 736.

**[0138]** The selected comparison value 736 may indicate a higher correlation (or a lower difference) than the other values of the comparison values 534. The signal comparator 506 may identify the tentative shift value 536 of the shift values 760 that correspond to the selected comparison value 736. For example, the signal comparator 506 may identify the second shift value 766 as the tentative shift value 536 in response to determining that the second shift value 766 corresponds to the selected comparison value 736 (e.g., the second comparison value 716).

**[0139]** The signal comparator 506 may determine the selected comparison value 736 based on the following Equation:

$$\maxXCorr = \max(|\sum_{k=-K}^K w(n)l'(n) * w(n+k)r'(n+k)|), \quad \text{Equation 5}$$

**[0140]** where  $\maxXCorr$  corresponds to the selected comparison value 736 and  $k$  corresponds to a shift value.  $w(n)*l'$  corresponds to de-emphasized, resampled, and windowed first audio signal 130, and  $w(n)*r'$  corresponds to de-emphasized, resampled, and windowed second audio signal 132. For example,  $w(n)*l'$  may correspond to the samples 626-632,  $w(n-1)*r'$  may correspond to the samples 654-660,  $w(n)*r'$  may correspond to the samples 656-662, and  $w(n+1)*r'$  may correspond to the samples 658-664.  $-K$  may correspond to a lower shift value (e.g., a minimum shift value) of the shift

- 42 -

values 760, and K may correspond to a higher shift value (e.g., a maximum shift value) of the shift values 760. In Equation 5,  $w(n)*l'$  corresponds to the first audio signal 130 independently of whether the first audio signal 130 corresponds to a right (r) channel signal or a left (l) channel signal. In Equation 5,  $w(n)*r'$  corresponds to the second audio signal 132 independently of whether the second audio signal 132 corresponds to the right (r) channel signal or the left (l) channel signal.

[0141] The signal comparator 506 may determine the tentative shift value 536 based on the following Equation:

$$T = \underset{k}{\operatorname{argmax}} (|\sum_{k=-K}^K w(n)l'(n) * w(n+k)r'(n+k)|), \quad \text{Equation 6}$$

[0142] where T corresponds to the tentative shift value 536.

[0143] The signal comparator 506 may map the tentative shift value 536 from the resampled samples to the original samples based on the resampling factor (D) of FIG. 6. For example, the signal comparator 506 may update the tentative shift value 536 based on the resampling factor (D). To illustrate, the signal comparator 506 may set the tentative shift value 536 to a product (e.g., 12) of the tentative shift value 536 (e.g., 3) and the resampling factor (D) (e.g., 4).

[0144] Referring to FIG. 8, an illustrative example of a system is shown and generally designated 800. The system 800 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 800. The memory 153 may be configured to store shift values 860. The shift values 860 may include a first shift value 864, a second shift value 866, or both.

[0145] During operation, the interpolator 510 may generate the shift values 860 proximate to the tentative shift value 536 (e.g., 12), as described herein. Mapped shift values may correspond to the shift values 760 mapped from the resampled samples to the original samples based on the resampling factor (D). For example, a first mapped shift value of the mapped shift values may correspond to a product of the first shift value 764 and the resampling factor (D). A difference between a first mapped shift

- 43 -

value of the mapped shift values and each second mapped shift value of the mapped shift values may be greater than or equal to a threshold value (e.g., the resampling factor (D), such as 4). The shift values 860 may have finer granularity than the shift values 760. For example, a difference between a lower value (e.g., a minimum value) of the shift values 860 and the tentative shift value 536 may be less than the threshold value (e.g., 4). The threshold value may correspond to the resampling factor (D) of FIG. 6. The shift values 860 may range from a first value (e.g., the tentative shift value 536 – (the threshold value-1)) to a second value (e.g., the tentative shift value 536 + (threshold value-1)).

**[0146]** The interpolator 510 may generate interpolated comparison values 816 corresponding to the shift values 860 by performing interpolation on the comparison values 534, as described herein. Comparison values corresponding to one or more of the shift values 860 may be excluded from the comparison values 534 because of the lower granularity of the comparison values 534. Using the interpolated comparison values 816 may enable searching of interpolated comparison values corresponding to the one or more of the shift values 860 to determine whether an interpolated comparison value corresponding to a particular shift value proximate to the tentative shift value 536 indicates a higher correlation (or lower difference) than the second comparison value 716 of FIG. 7.

**[0147]** FIG. 8 includes a graph 820 illustrating examples of the interpolated comparison values 816 and the comparison values 534 (e.g., cross-correlation values). The interpolator 510 may perform the interpolation based on a hanning windowed sinc interpolation, IIR filter based interpolation, spline interpolation, another form of signal interpolation, or a combination thereof. For example, the interpolator 510 may perform the hanning windowed sinc interpolation based on the following Equation:

$$R(k)_{32kHz} = \sum_{i=-4}^4 R(\hat{t}_{N2} - i)_{8kHz} * b(3i + t), \quad \text{Equation 7}$$

**[0148]** where  $t = k - \hat{t}_{N2}$ ,  $b$  corresponds to a windowed sinc function,  $\hat{t}_{N2}$  corresponds to the tentative shift value 536.  $R(\hat{t}_{N2}-i)_{8kHz}$  may correspond to a particular comparison value of the comparison values 534. For example,  $R(\hat{t}_{N2}-i)_{8kHz}$  may indicate a first comparison value of the comparison values 534 that corresponds to a first shift value

- 44 -

(e.g., 8) when i corresponds to 4.  $R(\hat{t}_{N2-i})_{8\text{kHz}}$  may indicate the second comparison value 716 that corresponds to the tentative shift value 536 (e.g., 12) when i corresponds to 0.  $R(\hat{t}_{N2-i})_{8\text{kHz}}$  may indicate a third comparison value of the comparison values 534 that corresponds to a third shift value (e.g., 16) when i corresponds to -4.

**[0149]**  $R(k)_{32\text{kHz}}$  may correspond to a particular interpolated value of the interpolated comparison values 816. Each interpolated value of the interpolated comparison values 816 may correspond to a sum of a product of the windowed sinc function (b) and each of the first comparison value, the second comparison value 716, and the third comparison value. For example, the interpolator 510 may determine a first product of the windowed sinc function (b) and the first comparison value, a second product of the windowed sinc function (b) and the second comparison value 716, and a third product of the windowed sinc function (b) and the third comparison value. The interpolator 510 may determine a particular interpolated value based on a sum of the first product, the second product, and the third product. A first interpolated value of the interpolated comparison values 816 may correspond to a first shift value (e.g., 9). The windowed sinc function (b) may have a first value corresponding to the first shift value. A second interpolated value of the interpolated comparison values 816 may correspond to a second shift value (e.g., 10). The windowed sinc function (b) may have a second value corresponding to the second shift value. The first value of the windowed sinc function (b) may be distinct from the second value. The first interpolated value may thus be distinct from the second interpolated value.

**[0150]** In Equation 7, 8 kHz may correspond to a first rate of the comparison values 534. For example, the first rate may indicate a number (e.g., 8) of comparison values corresponding to a frame (e.g., the frame 304 of FIG. 3) that are included in the comparison values 534. 32 kHz may correspond to a second rate of the interpolated comparison values 816. For example, the second rate may indicate a number (e.g., 32) of interpolated comparison values corresponding to a frame (e.g., the frame 304 of FIG. 3) that are included in the interpolated comparison values 816.

**[0151]** The interpolator 510 may select an interpolated comparison value 838 (e.g., a maximum value or a minimum value) of the interpolated comparison values 816. The

interpolator 510 may select a shift value (e.g., 14) of the shift values 860 that corresponds to the interpolated comparison value 838. The interpolator 510 may generate the interpolated shift value 538 indicating the selected shift value (e.g., the second shift value 866).

**[0152]** Using a coarse approach to determine the tentative shift value 536 and searching around the tentative shift value 536 to determine the interpolated shift value 538 may reduce search complexity without compromising search efficiency or accuracy.

**[0153]** Referring to FIG. 9A, an illustrative example of a system is shown and generally designated 900. The system 900 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 900. The system 900 may include the memory 153, a shift refiner 911, or both. The memory 153 may be configured to store a first shift value 962 corresponding to the frame 302. For example, the analysis data 190 may include the first shift value 962. The first shift value 962 may correspond to a tentative shift value, an interpolated shift value, an amended shift value, a final shift value, or a non-causal shift value associated with the frame 302. The frame 302 may precede the frame 304 in the first audio signal 130. The shift refiner 911 may correspond to the shift refiner 511 of FIG. 1.

**[0154]** FIG. 9A also includes a flow chart of an illustrative method of operation generally designated 920. The method 920 may be performed by the temporal equalizer 108, the encoder 114, the first device 104 of FIG. 1, the temporal equalizer(s) 208, the encoder 214, the first device 204 of FIG. 2, the shift refiner 511 of FIG. 5, the shift refiner 911, or a combination thereof.

**[0155]** The method 920 includes determining whether an absolute value of a difference between the first shift value 962 and the interpolated shift value 538 is greater than a first threshold, at 901. For example, the shift refiner 911 may determine whether an absolute value of a difference between the first shift value 962 and the interpolated shift value 538 is greater than a first threshold (e.g., a shift change threshold).

**[0156]** The method 920 also includes, in response to determining that the absolute value

is less than or equal to the first threshold, at 901, setting the amended shift value 540 to indicate the interpolated shift value 538, at 902. For example, the shift refiner 911 may, in response to determining that the absolute value is less than or equal to the shift change threshold, set the amended shift value 540 to indicate the interpolated shift value 538. In some implementations, the shift change threshold may have a first value (e.g., 0) indicating that the amended shift value 540 is to be set to the interpolated shift value 538 when the first shift value 962 is equal to the interpolated shift value 538. In alternate implementations, the shift change threshold may have a second value (e.g.,  $\geq 1$ ) indicating that the amended shift value 540 is to be set to the interpolated shift value 538, at 902, with a greater degree of freedom. For example, the amended shift value 540 may be set to the interpolated shift value 538 for a range of differences between the first shift value 962 and the interpolated shift value 538. To illustrate, the amended shift value 540 may be set to the interpolated shift value 538 when an absolute value of a difference (e.g., -2, -1, 0, 1, 2) between the first shift value 962 and the interpolated shift value 538 is less than or equal to the shift change threshold (e.g., 2).

**[0157]** The method 920 further includes, in response to determining that the absolute value is greater than the first threshold, at 901, determining whether the first shift value 962 is greater than the interpolated shift value 538, at 904. For example, the shift refiner 911 may, in response to determining that the absolute value is greater than the shift change threshold, determine whether the first shift value 962 is greater than the interpolated shift value 538.

**[0158]** The method 920 also includes, in response to determining that the first shift value 962 is greater than the interpolated shift value 538, at 904, setting a lower shift value 930 to a difference between the first shift value 962 and a second threshold, and setting a greater shift value 932 to the first shift value 962, at 906. For example, the shift refiner 911 may, in response to determining that the first shift value 962 (e.g., 20) is greater than the interpolated shift value 538 (e.g., 14), set the lower shift value 930 (e.g., 17) to a difference between the first shift value 962 (e.g., 20) and a second threshold (e.g., 3). Additionally, or in the alternative, the shift refiner 911 may, in response to determining that the first shift value 962 is greater than the interpolated shift value 538, set the greater shift value 932 (e.g., 20) to the first shift value 962. The

second threshold may be based on the difference between the first shift value 962 and the interpolated shift value 538. In some implementations, the lower shift value 930 may be set to a difference between the interpolated shift value 538 offset and a threshold (e.g., the second threshold) and the greater shift value 932 may be set to a difference between the first shift value 962 and a threshold (e.g., the second threshold).

**[0159]** The method 920 further includes, in response to determining that the first shift value 962 is less than or equal to the interpolated shift value 538, at 904, setting the lower shift value 930 to the first shift value 962, and setting a greater shift value 932 to a sum of the first shift value 962 and a third threshold, at 910. For example, the shift refiner 911 may, in response to determining that the first shift value 962 (e.g., 10) is less than or equal to the interpolated shift value 538 (e.g., 14), set the lower shift value 930 to the first shift value 962 (e.g., 10). Additionally, or in the alternative, the shift refiner 911 may, in response to determining that the first shift value 962 is less than or equal to the interpolated shift value 538, set the greater shift value 932 (e.g., 13) to a sum of the first shift value 962 (e.g., 10) and a third threshold (e.g., 3). The third threshold may be based on the difference between the first shift value 962 and the interpolated shift value 538. In some implementations, the lower shift value 930 may be set to a difference between the first shift value 962 offset and a threshold (e.g., the third threshold) and the greater shift value 932 may be set to a difference between the interpolated shift value 538 and a threshold (e.g., the third threshold).

**[0160]** The method 920 also includes determining comparison values 916 based on the first audio signal 130 and shift values 960 applied to the second audio signal 132, at 908. For example, the shift refiner 911 (or the signal comparator 506) may generate the comparison values 916, as described with reference to FIG. 7, based on the first audio signal 130 and the shift values 960 applied to the second audio signal 132. To illustrate, the shift values 960 may range from the lower shift value 930 (e.g., 17) to the greater shift value 932 (e.g., 20). The shift refiner 911 (or the signal comparator 506) may generate a particular comparison value of the comparison values 916 based on the samples 326-332 and a particular subset of the second samples 350. The particular subset of the second samples 350 may correspond to a particular shift value (e.g., 17) of the shift values 960. The particular comparison value may indicate a difference (or a

- 48 -

correlation) between the samples 326-332 and the particular subset of the second samples 350.

**[0161]** The method 920 further includes determining the amended shift value 540 based on the comparison values 916 generated based on the first audio signal 130 and the second audio signal 132, at 912. For example, the shift refiner 911 may determine the amended shift value 540 based on the comparison values 916. To illustrate, in a first case, when the comparison values 916 correspond to cross-correlation values, the shift refiner 911 may determine that the interpolated comparison value 838 of FIG. 8 corresponding to the interpolated shift value 538 is greater than or equal to a highest comparison value of the comparison values 916. Alternatively, when the comparison values 916 correspond to difference values, the shift refiner 911 may determine that the interpolated comparison value 838 is less than or equal to a lowest comparison value of the comparison values 916. In this case, the shift refiner 911 may, in response to determining that the first shift value 962 (e.g., 20) is greater than the interpolated shift value 538 (e.g., 14), set the amended shift value 540 to the lower shift value 930 (e.g., 17). Alternatively, the shift refiner 911 may, in response to determining that the first shift value 962 (e.g., 10) is less than or equal to the interpolated shift value 538 (e.g., 14), set the amended shift value 540 to the greater shift value 932 (e.g., 13).

**[0162]** In a second case, when the comparison values 916 correspond to cross-correlation values, the shift refiner 911 may determine that the interpolated comparison value 838 is less than the highest comparison value of the comparison values 916 and may set the amended shift value 540 to a particular shift value (e.g., 18) of the shift values 960 that corresponds to the highest comparison value. Alternatively, when the comparison values 916 correspond to difference values, the shift refiner 911 may determine that the interpolated comparison value 838 is greater than the lowest comparison value of the comparison values 916 and may set the amended shift value 540 to a particular shift value (e.g., 18) of the shift values 960 that corresponds to the lowest comparison value.

**[0163]** The comparison values 916 may be generated based on the first audio signal 130, the second audio signal 132, and the shift values 960. The amended shift value

540 may be generated based on comparison values 916 using a similar procedure as performed by the signal comparator 506, as described with reference to FIG. 7.

**[0164]** The method 920 may thus enable the shift refiner 911 to limit a change in a shift value associated with consecutive (or adjacent) frames. The reduced change in the shift value may reduce sample loss or sample duplication during encoding.

**[0165]** Referring to FIG. 9B, an illustrative example of a system is shown and generally designated 950. The system 950 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 950. The system 950 may include the memory 153, the shift refiner 511, or both. The shift refiner 511 may include an interpolated shift adjuster 958. The interpolated shift adjuster 958 may be configured to selectively adjust the interpolated shift value 538 based on the first shift value 962, as described herein. The shift refiner 511 may determine the amended shift value 540 based on the interpolated shift value 538 (e.g., the adjusted interpolated shift value 538), as described with reference to FIGS. 9A, 9C.

**[0166]** FIG. 9B also includes a flow chart of an illustrative method of operation generally designated 951. The method 951 may be performed by the temporal equalizer 108, the encoder 114, the first device 104 of FIG. 1, the temporal equalizer(s) 208, the encoder 214, the first device 204 of FIG. 2, the shift refiner 511 of FIG. 5, the shift refiner 911 of FIG. 9A, the interpolated shift adjuster 958, or a combination thereof.

**[0167]** The method 951 includes generating an offset 957 based on a difference between the first shift value 962 and an unconstrained interpolated shift value 956, at 952. For example, the interpolated shift adjuster 958 may generate the offset 957 based on a difference between the first shift value 962 and an unconstrained interpolated shift value 956. The unconstrained interpolated shift value 956 may correspond to the interpolated shift value 538 (e.g., prior to adjustment by the interpolated shift adjuster 958). The interpolated shift adjuster 958 may store the unconstrained interpolated shift value 956 in the memory 153. For example, the analysis data 190 may include the unconstrained interpolated shift value 956.

- 50 -

**[0168]** The method 951 also includes determining whether an absolute value of the offset 957 is greater than a threshold, at 953. For example, the interpolated shift adjuster 958 may determine whether an absolute value of the offset 957 satisfies a threshold. The threshold may correspond to an interpolated shift limitation MAX\_SHIFT\_CHANGE (e.g., 4).

**[0169]** The method 951 includes, in response to determining that the absolute value of the offset 957 is greater than the threshold, at 953, setting the interpolated shift value 538 based on the first shift value 962, a sign of the offset 957, and the threshold, at 954. For example, the interpolated shift adjuster 958 may in response to determining that the absolute value of the offset 957 fails to satisfy (e.g., is greater than) the threshold, constrain the interpolated shift value 538. To illustrate, the interpolated shift adjuster 958 may adjust the interpolated shift value 538 based on the first shift value 962, a sign (e.g., +1 or -1) of the offset 957, and the threshold (e.g., the interpolated shift value 538 = the first shift value 962 + sign (the offset 957) \* Threshold).

**[0170]** The method 951 includes, in response to determining that the absolute value of the offset 957 is less than or equal to the threshold, at 953, set the interpolated shift value 538 to the unconstrained interpolated shift value 956, at 955. For example, the interpolated shift adjuster 958 may in response to determining that the absolute value of the offset 957 satisfies (e.g., is less than or equal to) the threshold, refrain from changing the interpolated shift value 538.

**[0171]** The method 951 may thus enable constraining the interpolated shift value 538 such that a change in the interpolated shift value 538 relative to the first shift value 962 satisfies an interpolation shift limitation.

**[0172]** Referring to FIG. 9C, an illustrative example of a system is shown and generally designated 970. The system 970 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 970. The system 970 may include the memory 153, a shift refiner 921, or both. The shift refiner 921 may correspond to the shift refiner 511 of FIG. 5.

- 51 -

**[0173]** FIG. 9C also includes a flow chart of an illustrative method of operation generally designated 971. The method 971 may be performed by the temporal equalizer 108, the encoder 114, the first device 104 of FIG. 1, the temporal equalizer(s) 208, the encoder 214, the first device 204 of FIG. 2, the shift refiner 511 of FIG. 5, the shift refiner 911 of FIG. 9A, the shift refiner 921, or a combination thereof.

**[0174]** The method 971 includes determining whether a difference between the first shift value 962 and the interpolated shift value 538 is non-zero, at 972. For example, the shift refiner 921 may determine whether a difference between the first shift value 962 and the interpolated shift value 538 is non-zero.

**[0175]** The method 971 includes, in response to determining that the difference between the first shift value 962 and the interpolated shift value 538 is zero, at 972, setting the amended shift value 540 to the interpolated shift value 538, at 973. For example, the shift refiner 921 may, in response to determining that the difference between the first shift value 962 and the interpolated shift value 538 is zero, determine the amended shift value 540 based on the interpolated shift value 538 (e.g., the amended shift value 540 = the interpolated shift value 538).

**[0176]** The method 971 includes, in response to determining that the difference between the first shift value 962 and the interpolated shift value 538 is non-zero, at 972, determining whether an absolute value of the offset 957 is greater than a threshold, at 975. For example, the shift refiner 921 may, in response to determining that the difference between the first shift value 962 and the interpolated shift value 538 is non-zero, determine whether an absolute value of the offset 957 is greater than a threshold. The offset 957 may correspond to a difference between the first shift value 962 and the unconstrained interpolated shift value 956, as described with reference to FIG. 9B. The threshold may correspond to an interpolated shift limitation MAX\_SHIFT\_CHANGE (e.g., 4).

**[0177]** The method 971 includes, in response to determining that a difference between the first shift value 962 and the interpolated shift value 538 is non-zero, at 972, or determining that the absolute value of the offset 957 is less than or equal to the threshold, at 975, setting the lower shift value 930 to a difference between a first

threshold and a minimum of the first shift value 962 and the interpolated shift value 538, and setting the greater shift value 932 to a sum of a second threshold and a maximum of the first shift value 962 and the interpolated shift value 538, at 976. For example, the shift refiner 921 may, in response to determining that the absolute value of the offset 957 is less than or equal to the threshold, determine the lower shift value 930 based on a difference between a first threshold and a minimum of the first shift value 962 and the interpolated shift value 538. The shift refiner 921 may also determine the greater shift value 932 based on a sum of a second threshold and a maximum of the first shift value 962 and the interpolated shift value 538.

**[0178]** The method 971 also includes generating the comparison values 916 based on the first audio signal 130 and the shift values 960 applied to the second audio signal 132, at 977. For example, the shift refiner 921 (or the signal comparator 506) may generate the comparison values 916, as described with reference to FIG. 7, based on the first audio signal 130 and the shift values 960 applied to the second audio signal 132. The shift values 960 may range from the lower shift value 930 to the greater shift value 932. The method 971 may proceed to 979.

**[0179]** The method 971 includes, in response to determining that the absolute value of the offset 957 is greater than the threshold, at 975, generating a comparison value 915 based on the first audio signal 130 and the unconstrained interpolated shift value 956 applied to the second audio signal 132, at 978. For example, the shift refiner 921 (or the signal comparator 506) may generate the comparison value 915, as described with reference to FIG. 7, based on the first audio signal 130 and the unconstrained interpolated shift value 956 applied to the second audio signal 132.

**[0180]** The method 971 also includes determining the amended shift value 540 based on the comparison values 916, the comparison value 915, or a combination thereof, at 979. For example, the shift refiner 921 may determine the amended shift value 540 based on the comparison values 916, the comparison value 915, or a combination thereof, as described with reference to FIG. 9A. In some implementations, the shift refiner 921 may determine the amended shift value 540 based on a comparison of the comparison value 915 and the comparison values 916 to avoid local maxima due to shift variation.

**[0181]** In some cases, an inherent pitch of the first audio signal 130, the first resampled signal 530, the second audio signal 132, the second resampled signal 532, or a combination thereof, may interfere with the shift estimation process. In such cases, pitch de-emphasis or pitch filtering may be performed to reduce the interference due to pitch and to improve reliability of shift estimation between multiple channels. In some cases, background noise may be present in the first audio signal 130, the first resampled signal 530, the second audio signal 132, the second resampled signal 532, or a combination thereof, that may interfere with the shift estimation process. In such cases, noise suppression or noise cancellation may be used to improve reliability of shift estimation between multiple channels.

**[0182]** Referring to FIG. 10A, an illustrative example of a system is shown and generally designated 1000. The system 1000 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 1000.

**[0183]** FIG. 10A also includes a flow chart of an illustrative method of operation generally designated 1020. The method 1020 may be performed by the shift change analyzer 512, the temporal equalizer 108, the encoder 114, the first device 104, or a combination thereof.

**[0184]** The method 1020 includes determining whether the first shift value 962 is equal to 0, at 1001. For example, the shift change analyzer 512 may determine whether the first shift value 962 corresponding to the frame 302 has a first value (e.g., 0) indicating no time shift. The method 1020 includes, in response to determining that the first shift value 962 is equal to 0, at 1001, proceeding to 1010.

**[0185]** The method 1020 includes, in response to determining that the first shift value 962 is non-zero, at 1001, determining whether the first shift value 962 is greater than 0, at 1002. For example, the shift change analyzer 512 may determine whether the first shift value 962 corresponding to the frame 302 has a first value (e.g., a positive value) indicating that the second audio signal 132 is delayed in time relative to the first audio signal 130.

- 54 -

**[0186]** The method 1020 includes, in response to determining that the first shift value 962 is greater than 0, at 1002, determining whether the amended shift value 540 is less than 0, at 1004. For example, the shift change analyzer 512 may, in response to determining that the first shift value 962 has the first value (e.g., a positive value), determine whether the amended shift value 540 has a second value (e.g., a negative value) indicating that the first audio signal 130 is delayed in time relative to the second audio signal 132. The method 1020 includes, in response to determining that the amended shift value 540 is less than 0, at 1004, proceeding to 1008. The method 1020 includes, in response to determining that the amended shift value 540 is greater than or equal to 0, at 1004, proceeding to 1010.

**[0187]** The method 1020 includes, in response to determining that the first shift value 962 is less than 0, at 1002, determining whether the amended shift value 540 is greater than 0, at 1006. For example, the shift change analyzer 512 may in response to determining that the first shift value 962 has the second value (e.g., a negative value), determine whether the amended shift value 540 has a first value (e.g., a positive value) indicating that the second audio signal 132 is delayed in time with respect to the first audio signal 130. The method 1020 includes, in response to determining that the amended shift value 540 is greater than 0, at 1006, proceeding to 1008. The method 1020 includes, in response to determining that the amended shift value 540 is less than or equal to 0, at 1006, proceeding to 1010.

**[0188]** The method 1020 includes setting the final shift value 116 to 0, at 1008. For example, the shift change analyzer 512 may set the final shift value 116 to a particular value (e.g., 0) that indicates no time shift.

**[0189]** The method 1020 includes determining whether the first shift value 962 is equal to the amended shift value 540, at 1010. For example, the shift change analyzer 512 may determine whether the first shift value 962 and the amended shift value 540 indicate the same time delay between the first audio signal 130 and the second audio signal 132.

**[0190]** The method 1020 includes, in response to determining that the first shift value 962 is equal to the amended shift value 540, at 1010, setting the final shift value 116 to

the amended shift value 540, at 1012. For example, the shift change analyzer 512 may set the final shift value 116 to the amended shift value 540.

**[0191]** The method 1020 includes, in response to determining that the first shift value 962 is not equal to the amended shift value 540, at 1010, generating an estimated shift value 1072, at 1014. For example, the shift change analyzer 512 may determine the estimated shift value 1072 by refining the amended shift value 540, as further described with reference to FIG. 11.

**[0192]** The method 1020 includes setting the final shift value 116 to the estimated shift value 1072, at 1016. For example, the shift change analyzer 512 may set the final shift value 116 to the estimated shift value 1072.

**[0193]** In some implementations, the shift change analyzer 512 may set the non-causal shift value 162 to indicate the second estimated shift value in response to determining that the delay between the first audio signal 130 and the second audio signal 132 did not switch. For example, the shift change analyzer 512 may set the non-causal shift value 162 to indicate the amended shift value 540 in response to determining that the first shift value 962 is equal to 0, 1001, that the amended shift value 540 is greater than or equal to 0, at 1004, or that the amended shift value 540 is less than or equal to 0, at 1006.

**[0194]** The shift change analyzer 512 may thus set the non-causal shift value 162 to indicate no time shift in response to determining that delay between the first audio signal 130 and the second audio signal 132 switched between the frame 302 and the frame 304 of FIG. 3. Preventing the non-causal shift value 162 from switching directions (e.g., positive to negative or negative to positive) between consecutive frames may reduce distortion in down mix signal generation at the encoder 114, avoid use of additional delay for upmix synthesis at a decoder, or both.

**[0195]** Referring to FIG. 10B, an illustrative example of a system is shown and generally designated 1030. The system 1030 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 1030.

- 56 -

**[0196]** FIG. 10B also includes a flow chart of an illustrative method of operation generally designated 1031. The method 1031 may be performed by the shift change analyzer 512, the temporal equalizer 108, the encoder 114, the first device 104, or a combination thereof.

**[0197]** The method 1031 includes determining whether the first shift value 962 is greater than zero and the amended shift value 540 is less than zero, at 1032. For example, the shift change analyzer 512 may determine whether the first shift value 962 is greater than zero and whether the amended shift value 540 is less than zero.

**[0198]** The method 1031 includes, in response to determining that the first shift value 962 is greater than zero and that the amended shift value 540 is less than zero, at 1032, setting the final shift value 116 to zero, at 1033. For example, the shift change analyzer 512 may, in response to determining that the first shift value 962 is greater than zero and that the amended shift value 540 is less than zero, set the final shift value 116 to a first value (e.g., 0) that indicates no time shift.

**[0199]** The method 1031 includes, in response to determining that the first shift value 962 is less than or equal to zero or that the amended shift value 540 is greater than or equal to zero, at 1032, determining whether the first shift value 962 is less than zero and whether the amended shift value 540 is greater than zero, at 1034. For example, the shift change analyzer 512 may, in response to determining that the first shift value 962 is less than or equal to zero or that the amended shift value 540 is greater than or equal to zero, determine whether the first shift value 962 is less than zero and whether the amended shift value 540 is greater than zero.

**[0200]** The method 1031 includes, in response to determining that the first shift value 962 is less than zero and that the amended shift value 540 is greater than zero, proceeding to 1033. The method 1031 includes, in response to determining that the first shift value 962 is greater than or equal to zero or that the amended shift value 540 is less than or equal to zero, setting the final shift value 116 to the amended shift value 540, at 1035. For example, the shift change analyzer 512 may, in response to determining that the first shift value 962 is greater than or equal to zero or that the amended shift value 540 is less than or equal to zero, set the final shift value 116 to the amended shift value

540.

**[0201]** Referring to FIG. 11, an illustrative example of a system is shown and generally designated 1100. The system 1100 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 1100. FIG. 11 also includes a flow chart illustrating a method of operation that is generally designated 1120. The method 1120 may be performed by the shift change analyzer 512, the temporal equalizer 108, the encoder 114, the first device 104, or a combination thereof. The method 1120 may correspond to the step 1014 of FIG. 10A.

**[0202]** The method 1120 includes determining whether the first shift value 962 is greater than the amended shift value 540, at 1104. For example, the shift change analyzer 512 may determine whether the first shift value 962 is greater than the amended shift value 540.

**[0203]** The method 1120 also includes, in response to determining that the first shift value 962 is greater than the amended shift value 540, at 1104, setting a first shift value 1130 to a difference between the amended shift value 540 and a first offset, and setting a second shift value 1132 to a sum of the first shift value 962 and the first offset, at 1106. For example, the shift change analyzer 512 may, in response to determining that the first shift value 962 (e.g., 20) is greater than the amended shift value 540 (e.g., 18), determine the first shift value 1130 (e.g., 17) based on the amended shift value 540 (e.g., amended shift value 540 – a first offset). Alternatively, or in addition, the shift change analyzer 512 may determine the second shift value 1132 (e.g., 21) based on the first shift value 962 (e.g., the first shift value 962 + the first offset). The method 1120 may proceed to 1108.

**[0204]** The method 1120 further includes, in response to determining that the first shift value 962 is less than or equal to the amended shift value 540, at 1104, setting the first shift value 1130 to a difference between the first shift value 962 and a second offset, and setting the second shift value 1132 to a sum of the amended shift value 540 and the second offset. For example, the shift change analyzer 512 may, in response to determining that the first shift value 962 (e.g., 10) is less than or equal to the amended

shift value 540 (e.g., 12), determine the first shift value 1130 (e.g., 9) based on the first shift value 962 (e.g., first shift value 962 – a second offset). Alternatively, or in addition, the shift change analyzer 512 may determine the second shift value 1132 (e.g., 13) based on the amended shift value 540 (e.g., the amended shift value 540 + the second offset). The first offset (e.g., 2) may be distinct from the second offset (e.g., 3). In some implementations, the first offset may be the same as the second offset. A higher value of the first offset, the second offset, or both, may improve a search range.

**[0205]** The method 1120 also includes generating comparison values 1140 based on the first audio signal 130 and shift values 1160 applied to the second audio signal 132, at 1108. For example, the shift change analyzer 512 may generate the comparison values 1140, as described with reference to FIG. 7, based on the first audio signal 130 and the shift values 1160 applied to the second audio signal 132. To illustrate, the shift values 1160 may range from the first shift value 1130 (e.g., 17) to the second shift value 1132 (e.g., 21). The shift change analyzer 512 may generate a particular comparison value of the comparison values 1140 based on the samples 326-332 and a particular subset of the second samples 350. The particular subset of the second samples 350 may correspond to a particular shift value (e.g., 17) of the shift values 1160. The particular comparison value may indicate a difference (or a correlation) between the samples 326-332 and the particular subset of the second samples 350.

**[0206]** The method 1120 further includes determining the estimated shift value 1072 based on the comparison values 1140, at 1112. For example, the shift change analyzer 512 may, when the comparison values 1140 correspond to cross-correlation values, select a highest comparison value of the comparison values 1140 as the estimated shift value 1072. Alternatively, the shift change analyzer 512 may, when the comparison values 1140 correspond to difference values, select a lowest comparison value of the comparison values 1140 as the estimated shift value 1072.

**[0207]** The method 1120 may thus enable the shift change analyzer 512 to generate the estimated shift value 1072 by refining the amended shift value 540. For example, the shift change analyzer 512 may determine the comparison values 1140 based on original samples and may select the estimated shift value 1072 corresponding to a comparison

value of the comparison values 1140 that indicates a highest correlation (or lowest difference).

**[0208]** Referring to FIG. 12, an illustrative example of a system is shown and generally designated 1200. The system 1200 may correspond to the system 100 of FIG. 1. For example, the system 100, the first device 104 of FIG. 1, or both, may include one or more components of the system 1200. FIG. 12 also includes a flow chart illustrating a method of operation that is generally designated 1220. The method 1220 may be performed by the reference signal designator 508, the temporal equalizer 108, the encoder 114, the first device 104, or a combination thereof.

**[0209]** The method 1220 includes determining whether the final shift value 116 is equal to 0, at 1202. For example, the reference signal designator 508 may determine whether the final shift value 116 has a particular value (e.g., 0) indicating no time shift.

**[0210]** The method 1220 includes, in response to determining that the final shift value 116 is equal to 0, at 1202, leaving the reference signal indicator 164 unchanged, at 1204. For example, the reference signal designator 508 may, in response to determining that the final shift value 116 has the particular value (e.g., 0) indicating no time shift, leave the reference signal indicator 164 unchanged. To illustrate, the reference signal indicator 164 may indicate that the same audio signal (e.g., the first audio signal 130 or the second audio signal 132) is a reference signal associated with the frame 304 as with the frame 302.

**[0211]** The method 1220 includes, in response to determining that the final shift value 116 is non-zero, at 1202, determining whether the final shift value 116 is greater than 0, at 1206. For example, the reference signal designator 508 may, in response to determining that the final shift value 116 has a particular value (e.g., a non-zero value) indicating a time shift, determine whether the final shift value 116 has a first value (e.g., a positive value) indicating that the second audio signal 132 is delayed relative to the first audio signal 130 or a second value (e.g., a negative value) indicating that the first audio signal 130 is delayed relative to the second audio signal 132.

**[0212]** The method 1220 includes, in response to determining that the final shift value

- 60 -

116 has the first value (e.g., a positive value), set the reference signal indicator 164 to have a first value (e.g., 0) indicating that the first audio signal 130 is a reference signal, at 1208. For example, the reference signal designator 508 may, in response to determining that the final shift value 116 has the first value (e.g., a positive value), set the reference signal indicator 164 to a first value (e.g., 0) indicating that the first audio signal 130 is a reference signal. The reference signal designator 508 may, in response to determining that the final shift value 116 has the first value (e.g., the positive value), determine that the second audio signal 132 corresponds to a target signal.

**[0213]** The method 1220 includes, in response to determining that the final shift value 116 has the second value (e.g., a negative value), set the reference signal indicator 164 to have a second value (e.g., 1) indicating that the second audio signal 132 is a reference signal, at 1210. For example, the reference signal designator 508 may, in response to determining that the final shift value 116 has the second value (e.g., a negative value) indicating that the first audio signal 130 is delayed relative to the second audio signal 132, set the reference signal indicator 164 to a second value (e.g., 1) indicating that the second audio signal 132 is a reference signal. The reference signal designator 508 may, in response to determining that the final shift value 116 has the second value (e.g., the negative value), determine that the first audio signal 130 corresponds to a target signal.

**[0214]** The reference signal designator 508 may provide the reference signal indicator 164 to the gain parameter generator 514. The gain parameter generator 514 may determine a gain parameter (e.g., a gain parameter 160) of a target signal based on a reference signal, as described with reference to FIG. 5.

**[0215]** A target signal may be delayed in time relative to a reference signal. The reference signal indicator 164 may indicate whether the first audio signal 130 or the second audio signal 132 corresponds to the reference signal. The reference signal indicator 164 may indicate whether the gain parameter 160 corresponds to the first audio signal 130 or the second audio signal 132.

**[0216]** Referring to FIG. 13, a flow chart illustrating a particular method of operation is shown and generally designated 1300. The method 1300 may be performed by the reference signal designator 508, the temporal equalizer 108, the encoder 114, the first

device 104, or a combination thereof.

**[0217]** The method 1300 includes determining whether the final shift value 116 is greater than or equal to zero, at 1302. For example, the reference signal designator 508 may determine whether the final shift value 116 is greater than or equal to zero. The method 1300 also includes, in response to determining that the final shift value 116 is greater than or equal to zero, at 1302, proceeding to 1208. The method 1300 further includes, in response to determining that the final shift value 116 is less than zero, at 1302, proceeding to 1210. The method 1300 differs from the method 1220 of FIG. 12 in that, in response to determining that the final shift value 116 has a particular value (e.g., 0) indicating no time shift, the reference signal indicator 164 is set to a first value (e.g., 0) indicating that the first audio signal 130 corresponds to a reference signal. In some implementations, the reference signal designator 508 may perform the method 1220. In other implementations, the reference signal designator 508 may perform the method 1300.

**[0218]** The method 1300 may thus enable setting the reference signal indicator 164 to a particular value (e.g., 0) indicating that the first audio signal 130 corresponds to a reference signal when the final shift value 116 indicates no time shift independently of whether the first audio signal 130 corresponds to the reference signal for the frame 302.

**[0219]** Referring to FIG. 14, an illustrative example of a system is shown and generally designated 1400. The system 1400 includes the signal comparator 506 of FIG. 5, the interpolator 510 of FIG. 5, the shift refiner 511 of FIG. 5, and the shift change analyzer 512 of FIG. 5.

**[0220]** The signal comparator 506 may generate the comparison values 534 (e.g., difference values, similarity values, coherence values, or cross-correlation values), the tentative shift value 536, or both. For example, the signal comparator 506 may generate the comparison values 534 based on the first resampled signal 530 and a plurality of shift values 1450 applied to the second resampled signal 532. The signal comparator 506 may determine the tentative shift value 536 based on the comparison values 534. The signal comparator 506 includes a smoother 1410 configured to retrieve comparison values for previous frames of the resampled signals 530, 532 and may modify the

comparison values 534 based on a long-term smoothing operation using the comparison values for previous frames. For example, the comparison values 534 may include the long-term comparison value  $CompVal_{LT_N}(k)$  for a current frame (N) and may be represented by  $CompVal_{LT_N}(k) = (1 - \alpha) * CompVal_N(k), +(\alpha) * CompVal_{LT_{N-1}}(k)$ , where  $\alpha \in (0, 1.0)$ . Thus, the long-term comparison value  $CompVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous comparison value  $CompVal_N(k)$  at frame N and the long-term comparison values  $CompVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases. The signal comparator 506 may provide the comparison values 534, the tentative shift value 536, or both, to the interpolator 510.

**[0221]** The interpolator 510 may extend the tentative shift value 536 to generate the interpolated shift value 538. For example, the interpolator 510 may generate interpolated comparison values corresponding to shift values that are proximate to the tentative shift value 536 by interpolating the comparison values 534. The interpolator 510 may determine the interpolated shift value 538 based on the interpolated comparison values and the comparison values 534. The comparison values 534 may be based on a coarser granularity of the shift values. The interpolated comparison values may be based on a finer granularity of shift values that are proximate to the resampled tentative shift value 536. Determining the comparison values 534 based on the coarser granularity (e.g., the first subset) of the set of shift values may use fewer resources (e.g., time, operations, or both) than determining the comparison values 534 based on a finer granularity (e.g., all) of the set of shift values. Determining the interpolated comparison values corresponding to the second subset of shift values may extend the tentative shift value 536 based on a finer granularity of a smaller set of shift values that are proximate to the tentative shift value 536 without determining comparison values corresponding to each shift value of the set of shift values. Thus, determining the tentative shift value 536 based on the first subset of shift values and determining the interpolated shift value 538 based on the interpolated comparison values may balance resource usage and refinement of the estimated shift value. The interpolator 510 may provide the interpolated shift value 538 to the shift refiner 511.

**[0222]** The interpolator 510 includes a smoother 1420 configured to retrieve interpolated shift values for previous frames and may modify the interpolated shift value 538 based on a long-term smoothing operation using the interpolated shift values for previous frames. For example, the interpolated shift value 538 may include a long-term interpolated shift value  $InterVal_{LT_N}(k)$  for a current frame (N) and may be represented by  $InterVal_{LT_N}(k) = (1 - \alpha) * InterVal_N(k) + (\alpha) * InterVal_{LT_{N-1}}(k)$ , where  $\alpha \in (0, 1.0)$ . Thus, the long-term interpolated shift value  $InterVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous interpolated shift value  $InterVal_N(k)$  at frame N and the long-term interpolated shift values  $InterVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases.

**[0223]** The shift refiner 511 may generate the amended shift value 540 by refining the interpolated shift value 538. For example, the shift refiner 511 may determine whether the interpolated shift value 538 indicates that a change in a shift between the first audio signal 130 and the second audio signal 132 is greater than a shift change threshold. The change in the shift may be indicated by a difference between the interpolated shift value 538 and a first shift value associated with the frame 302 of FIG. 3. The shift refiner 511 may, in response to determining that the difference is less than or equal to the threshold, set the amended shift value 540 to the interpolated shift value 538. Alternatively, the shift refiner 511 may, in response to determining that the difference is greater than the threshold, determine a plurality of shift values that correspond to a difference that is less than or equal to the shift change threshold. The shift refiner 511 may determine comparison values based on the first audio signal 130 and the plurality of shift values applied to the second audio signal 132. The shift refiner 511 may determine the amended shift value 540 based on the comparison values. For example, the shift refiner 511 may select a shift value of the plurality of shift values based on the comparison values and the interpolated shift value 538. The shift refiner 511 may set the amended shift value 540 to indicate the selected shift value. A non-zero difference between the first shift value corresponding to the frame 302 and the interpolated shift value 538 may indicate that some samples of the second audio signal 132 correspond to both frames (e.g., the frame 302 and the frame 304). For example, some samples of the second

audio signal 132 may be duplicated during encoding. Alternatively, the non-zero difference may indicate that some samples of the second audio signal 132 correspond to neither the frame 302 nor the frame 304. For example, some samples of the second audio signal 132 may be lost during encoding. Setting the amended shift value 540 to one of the plurality of shift values may prevent a large change in shifts between consecutive (or adjacent) frames, thereby reducing an amount of sample loss or sample duplication during encoding. The shift refiner 511 may provide the amended shift value 540 to the shift change analyzer 512.

**[0224]** The shift refiner 511 includes a smoother 1430 configured to retrieve amended shift values for previous frames and may modify the amended shift value 540 based on a long-term smoothing operation using the amended shift values for previous frames. For example, the amended shift value 540 may include a long-term amended shift value  $AmendVal_{LT_N}(k)$  for a current frame (N) and may be represented by

$$AmendVal_{LT_N}(k) = (1 - \alpha) * AmendVal_N(k) + (\alpha) * AmendVal_{LT_{N-1}}(k), \text{ where } \alpha \in (0, 1.0).$$

Thus, the long-term amended shift value  $AmendVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous amended shift value  $AmendVal_N(k)$  at frame N and the long-term amended shift values  $AmendVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases.

**[0225]** The shift change analyzer 512 may determine whether the amended shift value 540 indicates a switch or reverse in timing between the first audio signal 130 and the second audio signal 132. The shift change analyzer 512 may determine whether the delay between the first audio signal 130 and the second audio signal 132 has switched sign based on the amended shift value 540 and the first shift value associated with the frame 302. The shift change analyzer 512 may, in response to determining that the delay between the first audio signal 130 and the second audio signal 132 has switched sign, set the final shift value 116 to a value (e.g., 0) indicating no time shift. Alternatively, the shift change analyzer 512 may set the final shift value 116 to the amended shift value 540 in response to determining that the delay between the first audio signal 130 and the second audio signal 132 has not switched sign.

**[0226]** The shift change analyzer 512 may generate an estimated shift value by refining the amended shift value 540. The shift change analyzer 512 may set the final shift value 116 to the estimated shift value. Setting the final shift value 116 to indicate no time shift may reduce distortion at a decoder by refraining from time shifting the first audio signal 130 and the second audio signal 132 in opposite directions for consecutive (or adjacent) frames of the first audio signal 130. The shift change analyzer 512 may provide the final shift value 116 to the absolute shift generator 513. The absolute shift generator 513 may generate the non-causal shift value 162 by applying an absolute function to the final shift value 116.

**[0227]** The smoothing techniques described above may substantially normalize the shift estimate between voiced frames, unvoiced frames, and transition frames. Normalized shift estimates may reduce sample repetition and artifact skipping at frame boundaries. Additionally, normalized shift estimates may result in reduced side channel energies, which may improve coding efficiency.

**[0228]** As described with respect to FIG. 14, smoothing may be performed at the signal comparator 506, the interpolator 510, the shift refiner 511, or a combination thereof. If the interpolated shift is consistently different from the tentative shift at an input sampling rate (FSin), smoothing of the interpolated shift value 538 may be performed in addition to smoothing of the comparison values 534 or in alternative to smoothing of the comparison values 534. During estimation of the interpolated shift value 538, the interpolation process may be performed on smoothed long-term comparison values generated at the signal comparator 506, on un-smoothed comparison values generated at the signal comparator 506, or on a weighted mixture of interpolated smoothed comparison values and interpolated un-smoothed comparison values. If smoothing is performed at the interpolator 510, the interpolation may be extended to be performed at the proximity of multiple samples in addition to the tentative shift estimated in a current frame. For example, interpolation may be performed in proximity to a previous frame's shift (e.g., one or more of the previous tentative shift, the previous interpolated shift, the previous amended shift, or the previous final shift) and in proximity to the current frame's tentative shift. As a result, smoothing may be performed on additional samples for the interpolated shift values which may improve the interpolated shift estimate.

- 66 -

**[0229]** Referring to FIG. 15, graphs illustrating comparison values for voiced frames, transition frames, and unvoiced frames are shown. According to FIG. 15, the graph 1502 illustrates comparison values (e.g., cross-correlation values) for a voiced frame processed without using the long-term smoothing techniques described, the graph 1504 illustrates comparison values for a transition frame processed without using the long-term smoothing techniques described, and the graph 1506 illustrates comparison values for an unvoiced frame processed without using the long-term smoothing techniques described.

**[0230]** The cross-correlation represented in each graph 1502, 1504, 1506 may be substantially different. For example, the graph 1502 illustrates that a peak cross-correlation between a voiced frame captured by the first microphone 146 of FIG. 1 and a corresponding voiced frame captured by the second microphone 148 of FIG. 1 occurs at approximately a 17 sample shift. However, the graph 1504 illustrates that a peak cross-correlation between a transition frame captured by the first microphone 146 and a corresponding transition frame captured by the second microphone 148 occurs at approximately a 4 sample shift. Moreover, the graph 1506 illustrates that a peak cross-correlation between an unvoiced frame captured by the first microphone 146 and a corresponding unvoiced frame captured by the second microphone 148 occurs at approximately a -3 sample shift. Thus, the shift estimate may be inaccurate for transition frames and unvoiced frames due to a relatively high level of noise.

**[0231]** According to FIG. 15, the graph 1512 illustrates comparison values (e.g., cross-correlation values) for a voiced frame processed using the long-term smoothing techniques described, the graph 1514 illustrates comparison values for a transition frame processed using the long-term smoothing techniques described, and the graph 1516 illustrates comparison values for an unvoiced frame processed using the long-term smoothing techniques described. The cross-correlation values in each graph 1512, 1514, 1516 may be substantially similar. For example, each graph 1512, 1514, 1516 illustrates that a peak cross-correlation between a frame captured by the first microphone 146 of FIG. 1 and a corresponding frame captured by the second microphone 148 of FIG. 1 occurs at approximately a 17 sample shift. Thus, the shift estimate for transition frames (illustrated by the graph 1514) and unvoiced frames

(illustrated by the graph 1516) may be relatively accurate (or similar) to the shift estimate of the voiced frame in spite of noise.

**[0232]** The comparison value long-term smoothing process described with respect to FIG. 15 may be applied when the comparison values are estimated on the same shift ranges in each frame. The smoothing logic (e.g., the smoothers 1410, 1420, 1430) may be performed prior to estimation of a shift between the channels based on generated comparison values. For example, the smoothing may be performed prior to estimation of either the tentative shift, the estimation of interpolated shift, or the amended shift. To reduce adaptation of comparison values during silent portions (or background noise which may cause drift in the shift estimation), the comparison values may be smoothed based on a higher time-constant (e.g.,  $\alpha = 0.995$ ); otherwise the smoothing may be based on  $\alpha = 0.9$ . The determination whether to adjust the comparison values may be based on whether the background energy or long-term energy is below a threshold.

**[0233]** Referring to FIG. 16, a flow chart illustrating a particular method of operation is shown and generally designated 1600. The method 1600 may be performed by the temporal equalizer 108, the encoder 114, the first device 104 of FIG. 1, or a combination thereof.

**[0234]** The method 1600 includes capturing a first audio signal at a first microphone, at 1602. The first audio signal may include a first frame. For example, referring to FIG. 1, the first microphone 146 may capture the first audio signal 130. The first audio signal 130 may include a first frame.

**[0235]** A second audio signal may be captured at a second microphone, at 1604. The second audio signal may include a second frame, and the second frame may have substantially similar content as the first frame. For example, referring to FIG. 1, the second microphone 148 may capture the second audio signal 132. The second audio signal 132 may include a second frame, and the second frame may have substantially similar content as the first frame. The first frame and the second frames may be one of voiced frames, transition frames, or unvoiced frames.

**[0236]** A delay between the first frame and the second frame may be estimated, at 1606.

For example, referring to FIG. 1, the temporal equalizer 108 may determine a cross-correlation between the first frame and the second frame. A temporal offset between the first audio signal and the second audio signal may be estimated based on the delay based on historical delay data, at 1608. For example, referring to FIG. 1, the temporal equalizer 108 may estimate a temporal offset between audio captured at the microphones 146, 148. The temporal offset may be estimated based on a delay between a first frame of the first audio signal 130 and a second frame of the second audio signal 132, where the second frame includes substantially similar content as the first frame. For example, the temporal equalizer 108 may use a cross-correlation function to estimate the delay between the first frame and the second frame. The cross-correlation function may be used to measure the similarity of the two frames as a function of the lag of one frame relative to the other. Based on the cross-correlation function, the temporal equalizer 108 may determine the delay (e.g., lag) between the first frame and the second frame. The temporal equalizer 108 may estimate the temporal offset between the first audio signal 130 and the second audio signal 132 based on the delay and historical delay data.

[0237] The historical data may include delays between frames captured from the first microphone 146 and corresponding frames captured from the second microphone 148. For example, the temporal equalizer 108 may determine a cross-correlation (e.g., a lag) between previous frames associated with the first audio signal 130 and corresponding frames associated with the second audio signal 132. Each lag may be represented by a “comparison value”. That is, a comparison value may indicate a time shift (k) between a frame of the first audio signal 130 and a corresponding frame of the second audio signal 132. According to one implementation, the comparison values for previous frames may be stored at the memory 153. A smoother 192 of the temporal equalizer 108 may “smooth” (or average) comparison values over a long-term set of frames and used the long-term smoothed comparison values for estimating a temporal offset (e.g., “shift”) between the first audio signal 130 and the second audio signal 132.

[0238] Thus, the historical delay data may be generated based on smoothed comparison values associated with the first audio signal 130 and the second audio signal 132. For example, the method 1600 may include smoothing comparison values associated with

the first audio signal 130 and the second audio signal 132 to generate the historical delay data. The smoothed comparison values may be based on frames of the first audio signal 130 generated earlier in time than the first frame and based on frames of the second audio signal 132 generated earlier in time than the second frame. According to one implementation, the method 1600 may include temporally shifting the second frame by the temporal offset.

[0239] To illustrate, if  $CompVal_N(k)$  represents the comparison value at a shift of  $k$  for the frame N, the frame N may have comparison values from  $k=T\_MIN$  (a minimum shift) to  $k=T\_MAX$  (a maximum shift). The smoothing may be performed such that a long-term comparison value  $CompVal_{LT_N}(k)$  is represented by  $CompVal_{LT_N}(k) = f(CompVal_N(k), CompVal_{N-1}(k), CompVal_{LT_{N-2}}(k), \dots)$ . The function  $f$  in the above equation may be a function of all (or a subset) of past comparison values at the shift (k). An alternative representation of the may be  $CompVal_{LT_N}(k) = g(CompVal_N(k), CompVal_{N-1}(k), CompVal_{N-2}(k), \dots)$ . The functions  $f$  or  $g$  may be simple finite impulse response (FIR) filters or infinite impulse response (IIR) filters, respectively. For example, the function  $g$  may be a single tap IIR filter such that the long-term comparison value  $CompVal_{LT_N}(k)$  is represented by  $CompVal_{LT_N}(k) = (1 - \alpha) * CompVal_N(k) + (\alpha) * CompVal_{LT_{N-1}}(k)$ , where  $\alpha \in (0, 1.0)$ . Thus, the long-term comparison value  $CompVal_{LT_N}(k)$  may be based on a weighted mixture of the instantaneous comparison value  $CompVal_N(k)$  at frame N and the long-term comparison values  $CompVal_{LT_{N-1}}(k)$  for one or more previous frames. As the value of  $\alpha$  increases, the amount of smoothing in the long-term comparison value increases.

[0240] According to one implementation, the method 1600 may include adjusting a range of comparison values that are used to estimate the delay between the first frame and the second frame, as described in greater detail with respect to FIGS. 17-18. The delay may be associated with a comparison value in the range of comparison values having a highest cross-correlation. Adjusting the range may include determining whether comparison values at a boundary of the range are monotonically increasing and expanding the boundary in response to a determination that the comparison values at the boundary are monotonically increasing. The boundary may include a left boundary or a

right boundary.

**[0241]** The method 1600 of FIG. 16 may substantially normalize the shift estimate between voiced frames, unvoiced frames, and transition frames. Normalized shift estimates may reduce sample repetition and artifact skipping at frame boundaries. Additionally, normalized shift estimates may result in reduced side channel energies, which may improve coding efficiency.

**[0242]** Referring to FIG. 17, a process diagram 1700 for selectively expanding a search range for comparison values used for shift estimation is shown. For example, the process diagram 1700 may be used to expand the search range for comparison values based on comparison values generated for a current frame, comparison values generated for past frames, or a combination thereof.

**[0243]** According to the process diagram 1700, a detector may be configured to determine whether the comparison values in the vicinity of a right boundary or left boundary is increasing or decreasing. The search range boundaries for future comparison value generation may be pushed outward to accommodate more shift values based on the determination. For example, the search range boundaries may be pushed outward for comparison values in subsequent frames or comparison values in a same frame when comparison values are regenerated. The detector may initiate search boundary extension based on the comparison values generated for a current frame or based on comparison values generated for one or more previous frames.

**[0244]** At 1702, the detector may determine whether comparison values at the right boundary are monotonically increasing. As a non-limiting example, the search range may extend from -20 to 20 (e.g., from 20 sample shifts in the negative direction to 20 samples shifts in the positive direction). As used herein, a shift in the negative direction corresponds to a first signal, such as the first audio signal 130 of FIG. 1, being a reference signal and a second signal, such as the second audio signal 132 of FIG. 1, being a target signal. A shift in the positive direction corresponds to the first signal being the target signal and the second signal being the reference signal.

**[0245]** If the comparison values at the right boundary are monotonically increasing, at

1702, the detector may adjust the right boundary outwards to increase the search range, at 1704. To illustrate, if comparison value at sample shift 19 has a particular value and the comparison value at sample shift 20 has a higher value, the detector may extend the search range in the positive direction. As a non-limiting example, the detector may extend the search range from -20 to 25. The detector may extend the search range in increments of one sample, two samples, three samples, etc. According to one implementation, the determination at 1702 may be performed by detecting comparison values at a plurality of samples towards the right boundary to reduce the likelihood of expanding the search range based on a spurious jump at the right boundary.

**[0246]** If the comparison values at the right boundary are not monotonically increasing, at 1702, the detector may determine whether the comparison values at the left boundary are monotonically increasing, at 1706. If the comparison values at the left boundary are monotonically increasing, at 1706, the detector may adjust the left boundary outwards to increase the search range, at 1708. To illustrate, if comparison value at sample shift -19 has a particular value and the comparison value at sample shift -20 has a higher value, the detector may extend the search range in the negative direction. As a non-limiting example, the detector may extend the search range from -25 to 20. The detector may extend the search range in increments of one sample, two samples, three samples, etc. According to one implementation, the determination at 1702 may be performed by detecting comparison values at a plurality of samples towards the left boundary to reduce the likelihood of expanding the search range based on a spurious jump at the left boundary. If the comparison values at the left boundary are not monotonically increasing, at 1706, the detector may leave the search range unchanged, at 1710.

**[0247]** Thus, the process diagram 1700 of FIG. 17 may initiate search range modification for future frames. For example, if the past three consecutive frames are detected to be monotonically increasing in the comparison values over the last ten shift values before the threshold (e.g., increasing from sample shift 10 to sample shift 20 or increasing from sample shift -10 to sample shift -20), the search range may be increased outwards by a particular number of samples. This outward increase of the search range may be continuously implemented for future frames until the comparison value at the boundary is no longer monotonically increasing. Increasing the search

range based on comparison values for previous frames may reduce the likelihood that the “true shift” might lay very close to the search range’s boundary but just outside the search range. Reducing this likelihood may result in improved side channel energy minimization and channel coding.

**[0248]** Referring to FIG. 18, graphs illustrating selective expansion of a search range for comparison values used for shift estimation is shown. The graphs may operate in conjunction with the data in Table 1.

Frame	Is current frame's correlation monotonically increasing at left boundary?	No. of consecutive frames with monotonically increasing left boundary	Is current frame's correlation monotonically increasing at right boundary?	No. of consecutive frames with monotonically increasing right boundary	Action to take	Best Boundary range	Estimated shift
i-2	No	0	Yes	1	Leave future search range unchanged	[-20, 20]	-12
i-1	No	0	Yes	2	Leave future search range unchanged	[-20, 20]	-12
i	No	0	Yes	3	Push the future right boundary outward	[-20, 20]	-12
i+1	No	0	Yes	4	Push the future right boundary outward	[-23, 23]	-12
i+2	No	0	Yes	5	Push the future right boundary outward	[-26, 26]	26
i+3	No	0	No	0	Leave future search range unchanged	[-29, 29]	27
i+4	No	1	No	1	Leave future search range unchanged	[-29, 29]	27

**Table 1: Selective Search Range Expansion Data**

**[0249]** According to Table 1, the detector may expand the search range if a particular boundary increases at three or more consecutive frames. The first graph 1802 illustrates comparison values for frame i-2. According to the first graph 1802, the left boundary is not monotonically increasing and the right boundary is monotonically increasing for one consecutive frame. As a result, the search range remains unchanged for the next frame (e.g., frame i-1) and the boundary may range from -20 to 20. The second graph 1804 illustrates comparison values for frame i-1. According to the second graph 1804, the left boundary is not monotonically increasing and the right boundary is monotonically increasing for two consecutive frames. As a result, the search range remains unchanged for the next frame (e.g., frame i) and the boundary may range from -20 to 20.

**[0250]** The third graph 1806 illustrates comparison values for frame i. According to the

third graph 1806, the left boundary is not monotonically increasing and the right boundary is monotonically increasing for three consecutive frames. Because the right boundary is monotonically increasing for three or more consecutive frames, the search range for the next frame (e.g., frame  $i+1$ ) may be expanded and the boundary for the next frame may range from -23 to 23. The fourth graph 1808 illustrates comparison values for frame  $i+1$ . According to the fourth graph 1808, the left boundary is not monotonically increasing and the right boundary is monotonically increasing for four consecutive frames. Because the right boundary is monotonically increasing for three or more consecutive frames, the search range for the next frame (e.g., frame  $i+2$ ) may be expanded and the boundary for the next frame may range from -26 to 26. The fifth graph 1810 illustrates comparison values for frame  $i+2$ . According to the fifth graph 1810, the left boundary is not monotonically increasing and the right boundary is monotonically increasing for five consecutive frames. Because the right boundary is monotonically increasing for three or more consecutive frames, the search range for the next frame (e.g., frame  $i+3$ ) may be expanded and the boundary for the next frame may range from -29 to 29.

**[0251]** The sixth graph 1812 illustrates comparison values for frame  $i+3$ . According to the sixth graph 1812, the left boundary is not monotonically increasing and the right boundary is not monotonically increasing. As a result, the search range remains unchanged for the next frame (e.g., frame  $i+4$ ) and the boundary may range from -29 to 29. The seventh graph 1814 illustrates comparison values for frame  $i+4$ . According to the seventh graph 1814, the left boundary is not monotonically increasing and the right boundary is monotonically increasing for one consecutive frame. As a result, the search range remains unchanged for the next frame and the boundary may range from -29 to 29.

**[0252]** According to FIG. 18, the left boundary is expanded along with the right boundary. In alternative implementations, the left boundary may be pushed inwards to compensate for the outward push of the right boundary to maintain a constant number of shift values on which the comparison values are estimated for each frame. In another implementation, the left boundary may remain constant when the detector indicates that the right boundary is to be expanded outwards.

**[0253]** According to one implementation, when the detector indicates a particular boundary is to be expanded outwards, the amount of samples that the particular boundary is expanded outward may be determined based on the comparison values. For example, when the detector determines that the right boundary is to be expanded outwards based on the comparison values, a new set of comparison values may be generated on a wider shift search range and the detector may use the newly generated comparison values and the existing comparison values to determine the final search range. To illustrate, for frame  $i+1$ , a set of comparison values on a wider range of shifts ranging from -30 to 30 may be generated. The final search range may be limited based on the comparison values generated in the wider search range.

**[0254]** Although the examples in FIG. 18 indicate that the right boundary may be extended outwards, similar analogous functions may be performed to extend the left boundary outwards if the detector determines that the left boundary is to be extended. According to some implementations, absolute limitations on the search range may be utilized to prevent the search range for indefinitely increasing or decreasing. As a non-limiting example, the absolute value of the search range may not be permitted to increase above 8.75 milliseconds (e.g., the look-ahead of the CODEC).

**[0255]** Referring to FIG. 19, a system 1900 for decoding audio signals is shown. The system 1900 includes the first device 104, the second device 106, and the network 120 of FIG. 1.

**[0256]** As described with respect to FIG. 1, the first device 104 may transmit at least one encoded signal (e.g., the encoded signals 102) to the second device 106 via the network 120. The encoded signals 102 may include mid channel bandwidth extension (BWE) parameters 1950, mid channel parameters 1954, side channel parameters 1956, inter-channel BWE parameters 1952, stereo upmix parameters 1958, or a combination thereof. According to one implementation, the mid channel BWE parameters 1950 may include mid channel high-band linear predictive coding (LPC) parameters, a set of gain parameters, or both. According to one implementation, the inter-channel BWE parameters 1952 may include a set of adjustment gain parameters, an adjustment spectral shape parameter, a high-band reference channel indicator, or a combination

thereof. The high-band reference channel indicator may be the same as or distinct from the reference signal indicator 164 of FIG. 1.

**[0257]** The second device 106 includes the decoder 118, a receiver 1911, and a memory 1953. The memory 1953 may include analysis data 1990. The receiver 1911 may be configured to receive the encoded signals 102 (e.g., a bitstream) from the first device 104 and may provide the encoded signals 102 (e.g., the bitstream) to the decoder 118. Different implementations of the decoder 118 are described with respect to FIGS. 20-23. It should be understood that the implementations of the decoder 118 described with respect to FIGS. 20-23 are merely for illustrative purposes and are not to be considered limiting. The decoder 118 may be configured to generate the first output signal 126 and the second output signal 128 based on the encoded signals 102. The first output signal 126 and the second output signal 128 may be provided to the first loudspeaker 142 and the second loudspeaker 144, respectively.

**[0258]** The decoder 118 may generate a plurality of low-band (LB) signals based on the encoded signals 102 and may generate a plurality of high-band (HB) signals based on the encoded signals 102. The plurality of low-band signals may include a first LB signal 1922 and a second LB signal 1924. The plurality of high-band signals may include a first HB signal 1923 and a second HB signal 1925. Generation of the first LB signal 1922 and the second LB signal 1924 is described in greater detail with respect to FIGS. 20-23. According to one implementation, the plurality of high-band signals may be generated independently of the plurality of low-band signals. In some implementations, the plurality of high-band signals may be generated based on stereo inter-channel bandwidth extension (ICBWE) HB upmix processing, and the plurality of low-band signals may be generated based on stereo LB upmix processing. The stereo LB upmix processing may be based on MS to left-right (LR) conversion in the time-domain or in the frequency-domain. Generation of the first HB signal 1923 and the second HB signal 1925 is described in greater detail with respect to FIGS. 20-23.

**[0259]** The decoder 118 may be configured to generate a first signal 1902 by combining the first LB signal 1922 of the plurality of low-band signals and the first HB signal 1923 of the plurality of high-band signals. The decoder 118 may also be configured to

generate a second signal 1904 by combining the second LB signal 1924 of the plurality of low-band signals and the second HB signal 1925 of the plurality of high-band signals. The second output signal 128 may correspond to the second signal 1904. The decoder 118 may be configured to generate the first output signal 126 by shifting the first signal 1902. For example, the decoder 118 may time-shift first samples of the first signal 1902 relative to second samples of the second signal 1904 by an amount that is based on the non-causal shift value 162 to generate a shifted first signal 1912. In other implementations, the decoder 118 may shift based on other shift values described herein, such as the first shift value 962 of FIG. 9, the amended shift value 540 of FIG. 5, the interpolated shift value 538 of FIG. 5, etc. Thus, with respect to the decoder 118, it should be understood that the non-causal shift value 162 may include other shift values described herein. The first output signal 126 may correspond to the shifted first signal 1912.

**[0260]** According to one implementation, the decoder 118 may generate a shifted first HB signal 1933 by time-shifting the first HB signal 1923 of the plurality of high-band signals relative to the second HB signal 1925 of the plurality of high-band signals by an amount that is based on the non-causal shift value 162. In other implementations, the decoder 118 may shift based on other shift values described herein, such as the first shift value 962 of FIG. 9, the amended shift value 540 of FIG. 5, the interpolated shift value 538 of FIG. 5, etc. The decoder 118 may generate a shifted first LB signal 1932 by shifting the first LB signal 1922 based on the non-causal shift value 162, described in greater detail with respect to FIG. 20. The first output signal 126 may be generated by combining the shifted first LB signal 1932 and the shifted first HB signal 1933. The second output signal 128 may be generated by combining the second LB signal 1924 and the second HB signal 1925. It should be noted that in other implementations (e.g., the implementations described with respect to FIGS. 21-23), the low-band and high-band signals may be combined, and the combined signal may be shifted.

**[0261]** For ease of description and illustration, additional operations of the decoder 118 are described with respect to FIGS. 20-26. The system 1900 of FIG. 19 may enable integration of the inter-channel BWE parameters 1952 with target channel shifting, a sequence of upmix techniques, and shift compensation techniques, as further described

with respect to FIGS. 20-26.

**[0262]** Referring to FIG. 20, a first implementation 2000 of the decoder 118 is shown. According to the first implementation 2000, the decoder 118 includes a mid BWE decoder 2002, a LB mid core decoder 2004, a LB side core decoder 2006, an upmix parameter decoder 2008, an inter-channel BWE spatial balancer 2010, a LB upmixer 2012, a shifter 2016, and a synthesizer 2018.

**[0263]** The mid channel BWE parameters 1950 may be provided to the mid BWE decoder 2002. The mid channel BWE parameters 1950 may include mid channel HB LPC parameters and a set of gain parameters. The mid channel parameters 1954 may be provided to the LB mid core decoder 2004, and the side channel parameters 1956 may be provided to the LB side core decoder 2006. The stereo upmix parameters 1958 may be provided to the upmix parameter decoder 2008.

**[0264]** The LB mid core decoder 2004 may be configured to generate core parameters 2056 and a mid channel LB signal 2052 based on the mid channel parameters 1954. The core parameters 2056 may include a mid channel LB excitation signal. The core parameters 2056 may be provided to the mid BWE decoder 2002 and to the LB side core decoder 2006. The mid channel LB signal 2052 may be provided to the LB upmixer 2012. The mid BWE decoder 2002 may generate a mid channel HB signal 2054 based on the mid channel BWE parameters 1950 and based on the core parameters 2056 from the LB mid core decoder 2004. In a particular implementation, the mid BWE decoder 2002 may include a time-domain bandwidth extension decoder (or module). The time-domain bandwidth extension decoder (e.g., the mid BWE decoder 2002) may generate the mid channel HB signal 2054. For example, the time-domain bandwidth extension decoder may generate an upsampled mid channel LB excitation signal by upsampling the mid channel LB excitation signal. The time-domain bandwidth extension decoder may apply a function (e.g., a non-linear function or an absolute value function) to the upsampled mid channel LB excitation signal corresponding to the high-band to generate a high-band signal. The time-domain bandwidth extension decoder may filter the high-band signal based on HB LPC parameters (e.g., the mid channel HB LPC parameters) to generate a filtered signal (e.g.,

a LPC synthesized high-band excitation). The mid channel BWE parameters 1950 may include the HB LPC parameters. The time-domain bandwidth extension decoder may generate the mid channel HB signal 2054 by scaling the filtered signal based on subframe gains or frame gain. The mid channel BWE parameters 1950 may include the subframe gains, the frame gain, or a combination thereof.

**[0265]** In an alternative implementation, the mid BWE decoder 2002 may include a frequency-domain bandwidth extension decoder (or module). The frequency-domain bandwidth extension decoder (e.g., the mid BWE decoder 2002) may generate the mid channel HB signal 2054. For example, the frequency-domain bandwidth extension decoder may generate the mid channel HB signal 2054 by scaling the mid channel LB excitation signal based on subframe gains, sub-band gains (subsets of the high-band frequency range), or frame gain. The mid channel BWE parameters 1950 may include the subframe gains, the sub-band gains, the frame gain, or a combination thereof. In some implementations, the mid BWE decoder 2002 is configured to provide the LPC synthesized filtered high-band excitation as an additional input to the inter-channel BWE spatial balancer 2010. The mid channel HB signal 2054 may be provided to the inter-channel BWE spatial balancer 2010.

**[0266]** The inter-channel BWE spatial balancer 2010 may be configured to generate the first HB signal 1923 and the second HB signal 1925 based on the mid channel HB signal 2054 and based on the inter-channel BWE parameters 1952. The inter-channel BWE parameters 1952 may include a set of adjustment gain parameters, a high-band reference channel indicator, adjustment spectral shape parameters, or a combination thereof. In a particular implementation, the inter-channel BWE spatial balancer 2010 may, in response to determining that the set of adjustment gain parameters includes a single adjustment gain parameter and that the adjustment spectral shape parameters are absent from the inter-channel BWE parameters 1952, scale the (decoded) mid channel HB signal 2054 based on the adjustment gain parameter to generate an adjustment gain scaled mid channel HB signal. The inter-channel BWE spatial balancer 2010 may determine, based on the high-band reference channel indicator, whether the adjustment gain scaled mid channel HB signal is designated as the first HB signal 1923 or the second HB signal 1925. For example, the inter-channel BWE spatial balancer 2010

may, in response to determining that the high-band reference channel indicator has a first value, output the adjustment gain scaled mid channel HB signal as the first HB signal 1923. As another example, the inter-channel BWE spatial balancer 2010 may, in response to determining that the high-band reference channel indicator has a second value, output the adjustment gain scaled mid channel HB signal as the second HB signal 1925. The inter-channel BWE spatial balancer 2010 may generate the other of the first HB signal 1923 or the second HB signal 1925 by scaling the mid channel HB signal 2054 by a factor (e.g., 2 – (the adjustment gain parameter)).

**[0267]** The inter-channel BWE spatial balancer 2010 may, in response to determining that the inter-channel BWE parameters 1952 include the adjustment spectral shape parameters, generate (or receive from the mid BWE decoder 2002) a synthesized non-reference signal (e.g., the LPC synthesized high-band excitation). The inter-channel BWE spatial balancer 2010 may include a spectral shape adjuster module. The spectral shape adjuster module (e.g., the inter-channel BWE spatial balancer 2010) may include a spectral shaping filter. The spectral shaping filter may be configured to generate a spectral shape adjusted signal based on the synthesized non-reference signal (e.g., the LPC synthesized high-band excitation) and the adjustment spectral shape parameters. The adjustment spectral shape parameters may correspond to a parameter or coefficient (e.g., “u”) of the spectral shaping filter, where the spectral shaping filter is defined by a function (e.g.,  $H(z) = 1 / (1 - uz^{-1})$ ). The spectral shaping filter may output the spectral shape adjusted signal to a gain adjustment module. The inter-channel BWE spatial balancer 2010 may include the gain adjustment module. The gain adjustment module may be configured to generate a gain adjusted signal by applying a scaling factor to the spectral shape adjusted signal. The scaling factor may be based on the adjustment gain parameter. The inter-channel BWE spatial balancer 2010 may determine, based on a value of the high-band reference channel indicator, whether the gain adjusted signal is designated as the first HB signal 1923 or the second HB signal 1925. For example, the inter-channel BWE spatial balancer 2010 may, in response to determining that the high-band reference channel indicator has a first value, output the gain adjusted signal as the first HB signal 1923. As another example, the inter-channel BWE spatial balancer 2010 may, in response to determining that the high-band reference channel indicator has a

- 80 -

second value, output the gain adjusted signal as the second HB signal 1925. The inter-channel BWE spatial balancer 2010 may generate the other of the first HB signal 1923 or the second HB signal 1925 by scaling the mid channel HB signal 2054 by a factor (e.g., 2 – (the adjustment gain parameter)). The first HB signal 1923 and the second HB signal 1925 may be provided to the shifter 2016.

**[0268]** The LB side core decoder 2006 may be configured to generate a side channel LB signal 2050 based on the side channel parameters 1956 and based on the core parameters 2056. The side channel LB signal 2050 may be provided to the LB upmixer 2012. The mid channel LB signal 2052 and the side channel LB signal 2050 may be sampled at a core frequency. The upmix parameter decoder 2008 may regenerate the gain parameters 160, the non-causal shift value 156, and the reference signal indicator 164 based on the stereo upmix parameters 1958. The gain parameters 160, the non-causal shift value 156, and the reference signal indicator 164 may be provided to the LB upmixer 2012 and to the shifter 2016.

**[0269]** The LB upmixer 2012 may be configured to generate the first LB signal 1922 and the second LB signal 1924 based on the mid channel LB signal 2052 and the side channel LB signal 2050. For example, the LB upmixer 2012 may apply one or more of the gain parameters 160, the non-causal shift value 162, and the reference signal indicator 164 to the signals 2050, 2052 to generate the first LB signal 1922 and the second LB signal 1924. In other implementations, the decoder 118 may shift based on other shift values described herein, such as the first shift value 962 of FIG. 9, the amended shift value 540 of FIG. 5, the interpolated shift value 538 of FIG. 5, etc. The first LB signal 1922 and the second LB signal 1924 may be provided to the shifter 2016. The non-causal shift value 162 may also be provided to the shifter 2016.

**[0270]** The shifter 2016 may be configured to generate the shifted first HB signal 1933 based on the first HB signal 1923, the non-causal shift value 162, the gain parameters 160, the non-causal shift value 162, and the reference signal indicator 164. For example, the shifter 2016 may shift the first HB signal 1923 to generate the shifted first HB signal 1933. To illustrate, the shifter 2016 may, in response to determining that the reference signal indicator 164 indicates that the first HB signal 1921 corresponds to a

target signal, shift the first HB signal 1921 to generate the shifted first HB signal 1933. The shifted first HB signal 1933 may be provided to the synthesizer 2018. The shifter 2016 may also provide the second HB signal 1925 to the synthesizer 2018.

**[0271]** The shifter 2016 may also be configured to generate the shifted first LB signal 1932 based on the first LB signal 1922, the non-causal shift value 162, the gain parameters 160, the non-causal shift value 162, and the reference signal indicator 164. In other implementations, the decoder 118 may shift based on other shift values described herein, such as the first shift value 962 of FIG. 9, the amended shift value 540 of FIG. 5, the interpolated shift value 538 of FIG. 5, etc. The shifter 2016 may shift the first LB signal 1922 to generate the shifted first LB signal 1932. To illustrate, the shifter 2016 may, in response to determining that the reference signal indicator 164 indicates that the first LB signal 1922 corresponds to a target signal, shift the first LB signal 1922 to generate the shifted first LB signal 1932. The shifted first LB signal 1932 may be provided to the synthesizer 2018. The shifter 2016 may also provide the second LB signal 1924 to the synthesizer 2018.

**[0272]** The synthesizer 2018 may be configured to generate the first output signal 126 and the second output signal 128. For example, the synthesizer 2018 may resample and combine the shifted first LB signal 1932 and the shifted first HB signal 1933 to generate the first output signal 126. Additionally, the synthesizer 2018 may resample and combine the second LB signal 1924 and the second HB signal 1925 to generate the second output signal 128. In a particular aspect, the first output signal 126 may correspond to a left output signal and the second output signal 128 may correspond to a right output signal. In an alternative aspect, the first output signal 126 may correspond to a right output signal and the second output signal 128 may correspond to a left output signal.

**[0273]** Thus, the first implementation 2000 of the decoder 118 enables generation the first LB signal 1922 and the second LB signal 1924 independently of generation of the first and second HB signals 1923, 1925. Also, the first implementation 2000 of the decoder 118 shifts the high-band and the low-band individually, and then combines the resultant signals to form a shifted output signal.

**[0274]** Referring to FIG. 21, a second implementation 2100 of the decoder 118 is shown that combines a low-band and a high-band before applying a shift to generate a shifted signal. According to the second implementation 2100, the decoder 118 includes the mid BWE decoder 2002, the LB mid core decoder 2004, the LB side core decoder 2006, the upmix parameter decoder 2008, the inter-channel BWE spatial balancer 2010, a LB resampler 2114, a stereo upmixer 2112, a combiner 2118, and a shifter 2116.

**[0275]** The mid channel BWE parameters 1950 may be provided to the mid BWE decoder 2002. The mid channel BWE parameters 1950 may include mid channel HB LPC parameters and a set of gain parameters. The mid channel parameters 1954 may be provided to the LB mid core decoder 2004, and the side channel parameters 1956 may be provided to the LB side core decoder 2006. The stereo upmix parameters 1958 may be provided to the upmix parameter decoder 2008.

**[0276]** The LB mid core decoder 2004 may be configured to generate core parameters 2056 and the mid channel LB signal 2052 based on the mid channel parameters 1954. The core parameters 2056 may include a mid channel LB excitation signal. The core parameters 2056 may be provided to the mid BWE decoder 2002 and to the LB side core decoder 2006. The mid channel LB signal 2052 may be provided to the LB resampler 2114. The mid BWE decoder 2002 may generate the mid channel HB signal 2054 based on the mid channel BWE parameters 1950 and based on the core parameters 2056 from the LB mid core decoder 2004. The mid channel HB signal 2054 may be provided to the inter-channel BWE spatial balancer 2010.

**[0277]** The inter-channel BWE spatial balancer 2010 may be configured to generate the first HB signal 1923 and the second HB signal 1925 based on the mid channel HB signal 2054, the inter-channel BWE parameters 1952, a non-linear extended harmonic LB excitation, a mid HB synthesis signal, or a combination thereof, as described with reference to FIG. 20. The inter-channel BWE parameters 1952 may include a set of adjustment gain parameters, a high-band reference channel indicator, adjustment spectral shape parameters, or a combination thereof. The first HB signal 1923 and the second HB signal 1925 may be provided to the combiner 2118.

**[0278]** The LB side core decoder 2006 may be configured to generate the side channel

LB signal 2050 based on the side channel parameters 1956 and based on the core parameters 2056. The side channel LB signal 2050 may be provided to the LB resampler 2114. The mid channel LB signal 2052 and the side channel LB signal 2050 may be sampled at a core frequency. The upmix parameter decoder 2008 may regenerate the gain parameters 160, the non-causal shift value 162, and the reference signal indicator 164 based on the stereo upmix parameters 1958. The gain parameters 160, the non-causal shift value 156, and the reference signal indicator 164 may be provided to the stereo upmixer 2112 and to the shifter 2116.

**[0279]** The LB resampler 2114 may be configured to sample the mid channel LB signal 2052 to generate an extended mid channel signal 2152. The extended mid channel signal 2152 may be provided to the stereo upmixer 2112. The LB resampler 2114 may also be configured to sample the side channel LB signal 2050 to generate an extended side channel signal 2150. The extended side channel signal 2150 may also be provided to the stereo upmixer 2112.

**[0280]** The stereo upmixer 2112 may be configured to generate the first LB signal 1922 and the second LB signal 1924 based on the extended mid channel signal 2152 and the extended side channel signal 2150. For example, the stereo upmixer 2112 may apply one or more of the gain parameters 160, the non-causal shift value 162, and the reference signal indicator 164 to the signals 2150, 2152 to generate the first LB signal 1922 and the second LB signal 1924. The first LB signal 1922 and the second LB signal 1924 may be provided to the combiner 2118.

**[0281]** The combiner 2118 may be configured to combine the first HB signal 1923 with the first LB signal 1922 to generate the first signal 1902. The combiner 2118 may also be configured to combine the second HB signal 1925 with the second LB signal 1924 to generate the second signal 1904. The first signal 1902 and the second signal 1904 may be provided to the shifter 2116. The non-causal shift value 162 may also be provided to the shifter 2116. The combiner 2118 may select, based on the high-band reference channel indicator and the inter-channel BWE parameters 1952, the first HB signal 1923 or the second HB signal 1925 to be combined with the first LB signal 1922. Similarly, the combiner 2118 may select, based on the high-band reference channel indicator and

the inter-channel BWE parameters 1952, the other of the first HB signal 1923 or the second HB signal 1925 to be combined with the second LB signal 1924.

**[0282]** The shifter 2116 may also be configured to generate the first output signal 126 and the second output signal 128 based on the first signal 1902 and the second signal 1904, respectively. For example, the shifter 2116 may shift the first signal 1902 by the non-causal shift value 162 to generate the first output signal 126. The first output signal 126 of FIG. 21 may correspond to the shifted first signal 1912 of FIG. 19. The shifter 2116 may also pass the second signal 1904 as the second output signal 128 (e.g., the second signal 1904 of FIG. 19). In some implementations, the shifter 2116 may determine, based on the reference signal indicator 164, the sign of the final shift values 216, or the sign of the final shift value 116, whether to shift the first signal 1902 or the second signal 1904 to compensate for the encoder-side non-causal shifting of one of the channels.

**[0283]** Thus, the second implementation 2100 of the decoder 118 may combine low-band and high-band signals prior to performing a shift that generates a shifted signal (e.g., the first output signal 126).

**[0284]** Referring to FIG. 22, a third implementation 2200 of the decoder 118 is shown. According to the third implementation 2200, the decoder 118 includes the mid BWE decoder 2002, the LB mid core decoder 2004, a side parameter mapper 2220, the upmix parameter decoder 2008, the inter-channel BWE spatial balancer 2010, a LB resampler 2214, a stereo upmixer 2212, the combiner 2118, and the shifter 2116.

**[0285]** The mid channel BWE parameters 1950 may be provided to the mid BWE decoder 2002. The mid channel BWE parameters 1950 may include mid channel HB LPC parameters and a set of gain parameters (e.g., gain shape parameters, gain frame parameters, mix factors, etc). The mid channel parameters 1954 may be provided to the LB mid core decoder 2004, and the side channel parameters 1956 may be provided to the side parameter mapper 2220. The stereo upmix parameters 1958 may be provided to the upmix parameter decoder 2008.

**[0286]** The LB mid core decoder 2004 may be configured to generate core parameters

2056 and the mid channel LB signal 2052 based on the mid channel parameters 1954. The core parameters 2056 may include a mid channel LB excitation signal, a LB voicing factor, or both. The core parameters 2056 may be provided to the mid BWE decoder 2002. The mid channel LB signal 2052 may be provided to the LB resampler 2214. The mid BWE decoder 2002 may generate the mid channel HB signal 2054 based on the mid channel BWE parameters 1950 and based on the core parameters 2056 from the LB mid core decoder 2004. The mid BWE decoder 2002 may also generate a non-linear extended harmonic LB excitation as an intermediate signal. The mid BWE decoder 2002 may perform a high-band LP synthesis of the combined non-linear harmonic LB excitation and shaped white noise to generate the mid HB synthesis signal. The mid BWE decoder 2002 may generate the mid channel HB signal 2054 by applying the gain shape parameter, the gain frame parameters, or a combination thereof, to the mid HB synthesis signal. The mid channel HB signal 2054 may be provided to the inter-channel BWE spatial balancer 2010. The non-linear extended harmonic LB excitation (e.g., the intermediate signal), the mid HB synthesis signal, or both, may also be provided to the inter-channel BWE spatial balancer 2010.

[0287] The inter-channel BWE spatial balancer 2010 may be configured to generate the first HB signal 1923 and the second HB signal 1925 based on the mid channel HB signal 2054, the inter-channel BWE parameters 1952, a non-linear extended harmonic LB excitation, a mid HB synthesis signal, or a combination thereof, as described with reference to FIG. 20. The inter-channel BWE parameters 1952 may include a set of adjustment gain parameters, a high-band reference channel indicator, adjustment spectral shape parameters, or a combination thereof. The first HB signal 1923 and the second HB signal 1925 may be provided to the combiner 2118.

[0288] The LB resampler 2214 may be configured to sample the mid channel LB signal 2052 to generate an extended mid channel signal 2252. The extended mid channel signal 2252 may be provided to the stereo upmixer 2212. The side parameter mapper 2220 may be configured to generate parameters 2256 based on the side channel parameters 1956. The parameters 2256 may be provided to the stereo upmixer 2212. The stereo upmixer 2212 may apply the parameters 2256 to the extended mid channel signal 2252 to generate the first LB signal 1922 and the second LB signal 1924. The

first and second LB signal 1922, 1924 may be provided to the combiner 2118. The combiner 2118 and the shifter 2116 may operate in a substantially similar manner as described with respect to FIG. 21.

**[0289]** The third implementation 2200 of the decoder 118 may combine low-band and high-band signals prior to performing a shift that generates a shifted signal (e.g., the first output signal 126). Additionally, generation of the side channel LB signal 2050 may be bypassed in the third implementation 2200 to reduce an amount of signal processing in comparison to the second implementation 2100.

**[0290]** Referring to FIG. 23, a fourth implementation 2300 of the decoder 118 is shown. According to the fourth implementation 2300, the decoder 118 includes the mid BWE decoder 2002, the LB mid core decoder 2004, the side parameter mapper 2220, the upmix parameter decoder 2008, a mid side generator 2310, a stereo upmixer 2312, the LB resampler 2214, the stereo upmixer 2212, the combiner 2118, and the shifter 2116.

**[0291]** The mid channel BWE parameters 1950 may be provided to the mid BWE decoder 2002. The mid channel BWE parameters 1950 may include mid channel HB LPC parameters and a set of gain parameters. The mid channel parameters 1954 may be provided to the LB mid core decoder 2004, and the side channel parameters 1956 may be provided to the side parameter mapper 2220. The stereo upmix parameters 1958 may be provided to the upmix parameter decoder 2008.

**[0292]** The LB mid core decoder 2004 may be configured to generate core parameters 2056 and the mid channel LB signal 2052 based on the mid channel parameters 1954. The core parameters 2056 may include a mid channel LB excitation signal. The core parameters 2056 may be provided to the mid BWE decoder 2002. The mid channel LB signal 2052 may be provided to the LB resampler 2214. The mid BWE decoder 2002 may generate the mid channel HB signal 2054 based on the mid channel BWE parameters 1950 and based on the core parameters 2056 from the LB mid core decoder 2004. The mid channel HB signal 2054 may be provided to the mid side generator 2310.

**[0293]** The mid side generator 2310 may be configured to generate an adjusted mid

channel signal 2354 and a side channel signal 2350 based on the mid channel HB signal 2054 and the inter-channel BWE parameters 1952. The adjusted mid channel signal 2354 and the side channel signal 2350 may be provided to the stereo upmixer 2312. The stereo upmixer 2312 may generate the first HB signal 1923 and the second HB signal 1925 based on the adjusted mid channel signal 2354 and the side channel signal 2350. The first HB signal 1923 and the second HB signal 1925 may be provided to the combiner 2118.

**[0294]** The side parameter mapper 2220, the upmix parameter decoder 2008, the LB resampler 2214, the stereo upmixer 2212, the combiner 2118, and the shifter 2116 may operate in a substantially similar manner as described with respect to FIGS. 20-22.

**[0295]** The fourth implementation 2300 of the decoder 118 may combine low-band and high-band signals prior to performing a shift that generates a shifted signal (e.g., the first output signal 126).

**[0296]** Referring to FIG. 24, a flowchart of a method 2400 of communication is shown. The method 2400 may be performed by the second device 106 of FIGS. 1 and 19.

**[0297]** The method 2400 includes receiving, at a device, at least one encoded signal, at 2402. For example, referring to FIG. 19, the receiver 1911 may receive the encoded signals 102 from the first device 104 and may provide the encoded signals the decoder 118.

**[0298]** The method 2400 also includes generating, at the device, a first signal and a second signal based on the at least one encoded signal, at 2404. For example, referring to FIG. 19, the decoder 118 may generate the first signal 1902 and the second signal 1904 based on the encoded signals 102. To illustrate, in FIG. 20, the first signal may correspond to the first HB signal 1923 and the second signal may correspond to the second HB signal 1925. Alternatively, in FIG. 19, the first signal may correspond to the first LB signal 1922 and the second signal may correspond to the second LB signal 1924. As another example, in FIGS. 20-23, the first signal and the second signal may correspond to the first signal 1902 and the second signal 1904, respectively.

**[0299]** The method 2400 also includes generating, at the device, a shifted first signal by time-shifting first samples of the first signal relative to second samples of the second signal by an amount that is based on a shift value, at 2406. For example, referring to FIG. 19, the decoder 118 may time-shift first samples of the first signal 1902 relative to second samples of the second signal 1904 by an amount that is based on the non-causal shift value 162 to generate a shifted first signal 1912. In FIG. 20, the shifter 2016 may shift the first HB signal 1923 to generate the shifted first HB signal 1933. Additionally, the shifter 2016 may shift the first LB signal 1922 to generate the shifted first LB signal 1932. In FIGS. 21-23, the shifter 2116 may shift the first signal 1902 to generate the shifted first signal 1912 (e.g., the first output signal 126).

**[0300]** The method 2400 also includes generating, at the device, a first output signal based on the shifted first signal, at 2408. The first output signal may be provided to a first speaker. For example, referring to FIG. 19, the decoder 118 may generate the first output signal 126 based on the shifted first signal 1912. In FIG. 20, the synthesizer 2018 generates the first output signal 126. In FIGS. 21-23, the shifted first signal 1912 may be the first output signal 126.

**[0301]** The method 2400 also includes generating, at the device, a second output signal based on the second signal, at 2410. The second output signal may be provided to a second speaker. For example, referring to FIG. 19, the decoder 118 may generate the second output signal 128 based on the second signal 1904. In FIG. 20, the synthesizer 2018 generates the second output signal 128. In FIGS. 21-23, the second signal 1904 may be the second output signal 128.

**[0302]** According to one implementation, the method 2400 may include generating a plurality of low-band signals 1922, 1924 based on the at least one encoded signal 102. The method 2400 may also include generating, independently of the plurality of low-band signals 1922, 1924, a plurality of high-band signals 1923, 1925 based on the at least one encoded signal 102. The plurality of high-band signals 1923, 1925 may include the first signal 1902 and the second signal 1904. The method 2400 may also include generating the first signal 1902 by combining a first low-band signal 1922 of the plurality of low-band signals 1922, 1924 and a first high-band signal 1923 of the

plurality of high-band signals 1923, 1925. The method 2400 may also include generating the second signal 1904 by combining a second low-band signal 1924 of the plurality of low-band signals 1922, 1924 and a second high-band signal 1925 of the plurality of high-band signals 1923, 1925. The first output signal 126 may correspond to the shifted first signal 1912, and the second output signal 128 may correspond to the second signal 1904.

**[0303]** According to one implementation, the plurality of low-band signals may include the first signal 1902 and the second signal 1904, and the method 2400 may also include generating a shifted first high-band signal 1933 by time-shifting a first high-band signal 1923 of the plurality of high-band signals relative to a second high-band signal 1925 of the plurality of high-band signals by an amount that is based on the non-causal shift value 162. The method 2400 may also include generating the first output signal 126 by combining the shifted first signal 1912 (e.g., the shifted first LB signal 1932) and the shifted first high-band signal 1933, such as illustrated with respect to FIG. 20. The method 2400 may also include generating the second output signal 128 by combining the second signal 1904 (e.g., the second LB signal 1924) and the second high-band signal 1925.

**[0304]** In some implementations, the method 2400 may include generating a first low-band signal 1922, a first high-band signal 1923, a second low-band signal 1924, and a second high-band signal 1925 based on the at least one encoded signal 102. The first signal 1902 may be based on the first low-band signal 1922, the first high-band signal 1923, or both. The second signal 1904 may be based on the second low-band signal 1924, the second high-band signal 1925, or both. To illustrate, the method 2400 may include generating a mid low-band signal (e.g., the mid channel LB signal 2052) based on the at least one encoded signal and generating a side low-band signal (e.g., the side channel LB signal 2050) based on the at least one encoded signal. The first low-band signal (e.g., the first LB signal 1922) and the second low-band signal (e.g., the second LB signal 1924) may be based on the mid low-band signal and the side low-band signal. The first low-band signal and the second low-band signal may be further based on a gain parameter (e.g., the gain parameter 160). The first low-band signal and the second low-band signal may be generated independently of the first high-band signal and the

second high-band signal (e.g., components 2012, 2114, 2112, 2214, 2212 in a low-band processing path are independent from components 2010 in a high-band processing path).

**[0305]** According to one implementation, the method 2400 may include generating a mid low-band signal based on the at least one encoded signal. The method 2400 may also include receiving one or more BWE parameters and generating a mid signal by performing bandwidth extension on the mid low-band signal based on the one or more BWE parameters. The method may also include receiving one or more inter-channel BWE parameters and generating the first high-band signal and the second high-band signal based on a mid signal and the one or more inter-channel BWE parameters.

**[0306]** According to one implementation, the method 2400 may also include generating a mid low-band signal based on the at least one encoded signal. The first signal and the second signal may be based on the mid signal and one or more side parameters.

**[0307]** The method 2400 of FIG. 24 may enable integration of the inter-channel BWE parameters 1952 with target channel shifting, a sequence of upmix techniques, and shift compensation techniques.

**[0308]** Referring to FIG. 25, a flowchart of a method 2500 of communication is shown. The method 2500 may be performed by the second device 106 of FIGS. 1 and 19.

**[0309]** The method 2500 includes receiving, at a device, at least one encoded signal, at 2502. For example, referring to FIG. 19, the receiver 1911 may receive the encoded signals 102 from the first device 104 via the network 120.

**[0310]** The method 2500 also includes generating, at the device, a plurality of high-band signals based on the at least one encoded signal, at 2504. For example, referring to FIG. 19, the decoder 118 may generate the plurality of high-band signals 1923, 1925 based on the encoded signals 102.

**[0311]** The method 2500 also includes generating, independently of the plurality of high-band signals, a plurality of low-band signals based on the at least one encoded signal, at 2506. For example, referring to FIG. 19, the decoder 118 may generate the

plurality of low-band signals 1922, 1924 based on the encoded signals 102. The plurality of low-band signals 1922, 1924 may be generated independently of the plurality of high-band signals 1923, 1925. For example, in FIG. 20, the inter-channel BWE spatial balancer 2010 operates independent of the outputs of the LB upmixer 2012. Likewise, the LB upmixer 2012 operates independent of the outputs of the inter-channel BWE spatial balancer 2010. In FIG. 21, the inter-channel BWE spatial balancer 2010 operates independent of the outputs of the LB resampler 2114 and independent of the outputs of the stereo upmixer 2112, and the LB resampler 2114 and the stereo upmixer 2112 operate independent of the outputs of the inter-channel BWE spatial balancer 2010. Additionally, in FIG. 22, the inter-channel BWE spatial balancer 2010 operates independent of the outputs of the LB resampler 2214 and independent of the outputs of the stereo upmixer 2212, and the LB resampler 2214 and the stereo upmixer 2212 operate independent of the outputs of the inter-channel BWE spatial balancer 2010.

**[0312]** According to one implementation, the method 2500 may include generating a mid low-band signal and a side low-band signal based on the at least one encoded signal. The plurality of low-band signals may be based on the mid low-band signal, the side low-band signal, and a gain parameter.

**[0313]** According to one implementation, the method 2500 may include generating a first signal based on a first low-band signal of the plurality of low-band signals, a first high-band signal of the plurality of high-band signals, or both. The method 2500 may also include generating a second signal based on a second low-band signal of the plurality of low-band signals, a second high-band signal of the plurality of high-band signals, or both. The method 2500 may further include generating a shifted first signal by time-shifting first samples of the first signal relative to second samples of the second signal by an amount that is based on the shift value. The method 2500 may also include generating a first output signal based on the shifted first signal and generating a second output signal based on the second signal.

**[0314]** According to one implementation, the method 2500 may include receiving a shift value and generating a first signal by combining a first low-band signal of the

plurality of low-band signals and a first high-band signal of the plurality of high-band signals. The method 2500 may also include generating a second signal by combining a second low-band signal of the plurality of low-band signals and a second high-band signal of the plurality of high-band signals. The method 2500 may also include generating a shifted first signal by time-shifting first samples of the first signal relative to second samples of the second signal by an amount that is based on the shift value. The method 2500 may also include providing the shifted first signal to a first speaker and providing the second signal to a second speaker.

**[0315]** According to one implementation, the method 2500 may include receiving a shift value and generating a shifted first low-band signal by time-shifting a first low-band signal of the plurality of low-band signals relative to a second low-band signal of the plurality of low-band signals by an amount that is based on the shift value. The method 2500 may also include generating a shifted first high-band signal by time-shifting a first high-band signal of the plurality of high-band signals relative to a second high-band signal of the plurality of high-band signals. The method 2500 may also include generating a shifted first signal by combining the shifted first low-band signal and the shifted first high-band signal. The method 2500 may further include generating a second signal by combining the second low-band signal and the second high-band signal. The method 2500 may also include providing the shifted first signal to a first loudspeaker and providing the second signal to a second loudspeaker.

**[0316]** Referring to FIG. 26, a flowchart of a method 2600 of communication is shown. The method 2600 may be performed by the second device 106 of FIGS. 1 and 19.

**[0317]** The method 2600 includes receiving, at a device, at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters, at 2602. For example, referring to FIG. 19, the receiver 1911 may receive the encoded signals 102 from the first device 104 via the network 120. The encoded signals 102 may include the inter-channel BWE parameters 1952.

**[0318]** The method 2600 also includes generating, at the device, a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal, at 2604. For example, referring to FIG. 20, the decoder 118 may

generate the mid channel HB signal 2054 by performing bandwidth extension based on the encoded signals 102. To illustrate, the encoded signals 102 may include the mid channel parameters 1954, the mid channel BWE parameters 1950, or a combination thereof. The LB mid core decoder 2004 may generate the core parameters 2056 based on the mid channel parameters 1954. The mid BWE decoder 2002 of FIG. 20 may generate the mid channel HB signal 2054 based on the mid channel BWE parameters 1950, the core parameters 2056, or a combination thereof, as described with reference to FIG. 20. With reference to the method 2600, the mid channel HB signal 2054 may also be referred to as the “mid channel time-domain high-band signal.”

**[0319]** The method 2600 further includes generating, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal, at 2606. For example, referring to FIG. 19, the decoder 118 may generate, based on the mid channel HB signal 2054, the mid channel BWE parameters 1950, a non-linear extended harmonic LB excitation, a mid HB synthesis signal, or a combination thereof, the first HB signal 1923 and the second HB signal 1925, as described with reference to FIG. 20. With reference to the method 2600, the first HB signal 1923 may also be referred to as the “first channel time-domain high-band signal” and the second HB signal 1925 may also be referred to as the “second channel time-domain high-band signal.”

**[0320]** The method 2600 also includes generating, at the device, a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal, at 2608. For example, referring to FIG. 21, the decoder 118 may generate the first signal 1902 by combining the first HB signal 1923 and the first LB signal 1922. With reference to the method 2600, the first signal 1902 may also be referred to as the “target channel signal” and the first LB signal 1922 may also be referred to as the “first channel low-band signal.”

**[0321]** The method 2600 further includes generating, at the device, a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal, at 2610. For example, referring to FIG. 21, the decoder 118

may generate the second signal 1904 by combining the second HB signal 1925 and the second LB signal 1924. With reference to the method 2600, the second signal 1904 may also be referred to as the “reference channel signal” and the second LB signal 1924 may also be referred to as the “second channel low-band signal.”

**[0322]** The method 2600 also includes generating, at the device, a modified target channel signal by modifying the target channel signal based on a temporal mismatch value, at 2612. For example, referring to FIG. 21, the decoder 118 may generate the shifted first signal 1912 by modifying the first signal 1902 based on the non-causal shift value 162. With reference to the method 2600, the shifted first signal 1912 may also be referred to as the “modified target channel signal” and the non-causal shift value 162 may also be referred to as the “temporal mismatch value.”

**[0323]** According to one implementation, the method 2600 may include generating, at the device, a mid channel low-band signal and a side channel low-band signal based on the at least one encoded signal. The first channel low-band signal and the second channel low-band signal may be based on the mid channel low-band signal, the side channel low-band signal, and a gain parameter. With reference to the method 2600, the mid channel LB signal 2052 may also be referred to as the “mid channel low-band signal” and the side channel LB signal 2050 may also be referred to as the “side channel low-band signal.”

**[0324]** According to one implementation, the method 2600 may include generating a first output signal based on the modified target channel signal. The method 2600 may also include generating a second output signal based on the reference channel signal. The method 2600 may further include providing the first output signal to a first speaker and providing the second output signal to a second speaker.

**[0325]** According to one implementation, the method 2600 may include receiving the temporal mismatch value at the device. The modified target channel signal may be generated by temporally shifting first samples of the target channel signal relative to second samples of the reference channel signal by an amount that is based on the temporal mismatch value. In some implementations, the temporal shift corresponds to a “causal shift” by which the target channel signal is “pulled forward” in time relative to

the reference channel signal.

**[0326]** According to one implementation, the method 2600 may include generating one or more mapped parameters based on one or more side parameters. The at least one encoded signal may include the one or more side parameters. The method 2600 may also include generating the first channel low-band signal and the second channel low-band signal by applying the one or more side parameters to the mid channel low-band signal. With reference to the method 2600, the parameters 2256 of FIG. 22 may also be referred to as the “mapped parameters.”

**[0327]** The techniques described with respect to FIGS. 19-26 may enable an upmix framework in a multi-channel decoder to decode audio signals with non-causal shifting. According to the techniques, a mid channel is decoded. For example, a low-band mid channel may be decoded for an ACELP core and a high-band mid channel may be decoded using high-band mid BWE. A TCX full band may be decoded for a MDCT frame (along with IGF parameters or other BWE parameters). An inter-channel spatial balancer may be applied to the high-band BWE signal to generate a high-band for a first and second channel based on a tilt, a gain, an ILD, and a reference channel indicator. For an ACELP frame, an LP core signal may be up-sampled using frequency domain or transform domain (e.g., DFT) resampling. Side channel parameters may be applied in the DFT domain on a core mid signal and an upmix may be performed followed by IDFT and windowing. First and second low-band channels may be generated in the time domain at an output sampling frequency. First and second high-band channels may be added to the first and second low-band channels, respectively, in the time domain to generate full-band channels. For a TCX frame or an MDCT frame, the side parameters may be applied to the full band to produce first and second channel outputs. An inverse non-causal shifting may be applied on a target channel to generate a temporal alignment between the channels.

**[0328]** Referring to FIG. 27, a block diagram of a particular illustrative example of a device (e.g., a wireless communication device) is depicted and generally designated 2700. In various implementations, the device 2700 may have fewer or more components than illustrated in FIG. 27. In an illustrative implementation, the device

2700 may correspond to the first device 104 or the second device 106 of FIG. 1. In an illustrative implementation, the device 2700 may perform one or more operations described with reference to systems and methods of FIGS. 1-26.

**[0329]** In a particular implementation, the device 2700 includes a processor 2706 (e.g., a central processing unit (CPU)). The device 2700 may include one or more additional processors 2710 (e.g., one or more digital signal processors (DSPs)). The processors 2710 may include a media (e.g., speech and music) coder-decoder (CODEC) 2708, and an echo canceller 2712. The media CODEC 2708 may include the decoder 118, such as described with respect to FIG. 1, 19, 20, 21, 22, or 23, the encoder 114, or both, of FIG. 1.

**[0330]** The device 2700 may include a memory 2753 and a CODEC 2734. Although the media CODEC 2708 is illustrated as a component of the processors 2710 (e.g., dedicated circuitry and/or executable programming code), in other implementations one or more components of the media CODEC 2708, such as the decoder 118, the encoder 114, or both, may be included in the processor 2706, the CODEC 2734, another processing component, or a combination thereof.

**[0331]** The device 2700 may include a transceiver 2711 coupled to an antenna 2742. The device 2700 may include a display 2728 coupled to a display controller 2726. One or more speakers 2748 may be coupled to the CODEC 2734. One or more microphones 2746 may be coupled, via the input interface(s) 112, to the CODEC 2734. In a particular aspect, the speakers 2748 may include the first loudspeaker 142, the second loudspeaker 144 of FIG. 1, the Yth loudspeaker 244 of FIG. 2, or a combination thereof. In a particular implementation, the microphones 2746 may include the first microphone 146, the second microphone 148 of FIG. 1, the Nth microphone 248 of FIG. 2, the third microphone 1146, the fourth microphone 1148 of FIG. 11, or a combination thereof. The CODEC 2734 may include a digital-to-analog converter (DAC) 2702 and an analog-to-digital converter (ADC) 2704.

**[0332]** The memory 2753 may include instructions 2760 executable by the processor 2706, the processors 2710, the CODEC 2734, another processing unit of the device 2700, or a combination thereof, to perform one or more operations described with

reference to FIGS. 1-26. The memory 2753 may store the analysis data 190, 1990.

**[0333]** One or more components of the device 2700 may be implemented via dedicated hardware (e.g., circuitry), by a processor executing instructions to perform one or more tasks, or a combination thereof. As an example, the memory 2753 or one or more components of the processor 2706, the processors 2710, and/or the CODEC 2734 may be a memory device, such as a random access memory (RAM), magnetoresistive random access memory (MRAM), spin-torque transfer MRAM (STT-MRAM), flash memory, read-only memory (ROM), programmable read-only memory (PROM), erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), registers, hard disk, a removable disk, or a compact disc read-only memory (CD-ROM). The memory device may include instructions (e.g., the instructions 2760) that, when executed by a computer (e.g., a processor in the CODEC 2734, the processor 2706, and/or the processors 2710), may cause the computer to perform one or more operations described with reference to FIGS. 1-26. As an example, the memory 2753 or the one or more components of the processor 2706, the processors 2710, and/or the CODEC 2734 may be a non-transitory computer-readable medium that includes instructions (e.g., the instructions 2760) that, when executed by a computer (e.g., a processor in the CODEC 2734, the processor 2706, and/or the processors 2710), cause the computer perform one or more operations described with reference to FIGS. 1-26.

**[0334]** In a particular implementation, the device 2700 may be included in a system-in-package or system-on-chip device (e.g., a mobile station modem (MSM)) 2722. In a particular implementation, the processor 2706, the processors 2710, the display controller 2726, the memory 2753, the CODEC 2734, and a transceiver 2711 are included in a system-in-package or the system-on-chip device 2722. In a particular implementation, an input device 2730, such as a touchscreen and/or keypad, and a power supply 2744 are coupled to the system-on-chip device 2722. Moreover, in a particular implementation, as illustrated in FIG. 27, the display 2728, the input device 2730, the speakers 2748, the microphones 2746, the antenna 2742, and the power supply 2744 are external to the system-on-chip device 2722. However, each of the display 2728, the input device 2730, the speakers 2748, the microphones 2746, the

antenna 2742, and the power supply 2744 can be coupled to a component of the system-on-chip device 2722, such as an interface or a controller.

**[0335]** The device 2700 may include a wireless telephone, a mobile communication device, a mobile phone, a smart phone, a cellular phone, a laptop computer, a desktop computer, a computer, a tablet computer, a set top box, a personal digital assistant (PDA), a display device, a television, a gaming console, a music player, a radio, a video player, an entertainment unit, a communication device, a fixed location data unit, a personal media player, a digital video player, a digital video disc (DVD) player, a tuner, a camera, a navigation device, a decoder system, an encoder system, a base station, a vehicle, or any combination thereof.

**[0336]** In a particular implementation, one or more components of the systems described herein and the device 2700 may be integrated into a decoding system or apparatus (e.g., an electronic device, a CODEC, or a processor therein), into an encoding system or apparatus, or both. In other implementations, one or more components of the systems described herein and the device 2700 may be integrated into a wireless communication device (e.g., a wireless telephone), a tablet computer, a desktop computer, a laptop computer, a set top box, a music player, a video player, an entertainment unit, a television, a game console, a navigation device, a communication device, a personal digital assistant (PDA), a fixed location data unit, a personal media player, a base station, a vehicle, or another type of device.

**[0337]** It should be noted that various functions performed by the one or more components of the systems described herein and the device 2700 are described as being performed by certain components or modules. This division of components and modules is for illustration only. In an alternate implementation, a function performed by a particular component or module may be divided amongst multiple components or modules. Moreover, in an alternate implementation, two or more components or modules of the systems described herein may be integrated into a single component or module. Each component or module illustrated in systems described herein may be implemented using hardware (e.g., a field-programmable gate array (FPGA) device, an application-specific integrated circuit (ASIC), a DSP, a controller, etc.), software (e.g.,

instructions executable by a processor), or any combination thereof.

**[0338]** In conjunction with the described implementations, an apparatus includes means for receiving at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters. For example, the means for receiving may include the second device 106 of FIG. 1, the receiver 1911 of FIG. 19, the transceiver 2711 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0339]** The apparatus also includes means for generating a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal. For example, the means for generating the mid channel time-domain high-band signal may include the second device 106, the decoder 118, the temporal balancer 124 of FIG. 1, the mid BWE decoder 2002 of FIG. 20, the speech and music codec 2708, the processors 2710, the CODEC 2734, the processor 2706 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0340]** The apparatus further includes means for generating a first channel time-domain high-band signal and a second channel time-domain high-band signal based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters. For example, the means for generating the first channel time-domain high-band signal and the second channel time-domain high-band signal may include the second device 106, the decoder 118, the temporal balancer 124 of FIG. 1, the inter-channel BWE spatial balancer 2010 of FIG. 20, the stereo upmixer 2312 of FIG. 23, the speech and music codec 2708, the processors 2710, the CODEC 2734, the processor 2706 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0341]** The apparatus also includes means for generating a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal. For example, the means for generating the target channel signal may include the second device 106, the decoder 118, the temporal balancer 124 of FIG. 1, the inter-channel BWE spatial balancer 2010 of FIG. 20, the combiner 2118 of FIG. 21, the speech and music codec 2708, the processors 2710, the CODEC 2734, the processor

- 100 -

2706 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0342]** The apparatus further includes means for generating a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal. For example, the means for generating the reference channel signal may include the second device 106, the decoder 118, the temporal balancer 124 of FIG. 1, the inter-channel BWE spatial balancer 2010 of FIG. 20, the combiner 2118 of FIG. 21, the speech and music codec 2708, the processors 2710, the CODEC 2734, the processor 2706 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0343]** The apparatus also includes means for generating a modified target channel signal by modifying the target channel signal based on a temporal mismatch value. For example, the means for generating the modified target channel signal may include the second device 106, the decoder 118, the temporal balancer 124 of FIG. 1, the inter-channel BWE spatial balancer 2010 of FIG. 20, the shifter 2116 of FIG. 21, the speech and music codec 2708, the processors 2710, the CODEC 2734, the processor 2706 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0344]** Also in conjunction with the described implementations, an apparatus includes means for receiving at least one encoded signal. For example, the means for receiving may include the receiver 1911 of FIG. 19, the transceiver 2711 of FIG. 27, one or more other devices configured to receive the at least one encoded signal, or a combination thereof.

**[0345]** The apparatus may also include means for generating a first output signal based on a shifted first signal and a second output signal based on a second signal. The shifted first signal may be generated by time-shifting first samples of a first signal relative to second samples of the second signal by an amount that is based on a shift value. The first signal and the second signal may be based on the at least one encoded signal. For example, the means for generating may include the decoder 118 of FIG. 19, one or more devices/sensors configured to generate the first output signal and the second output

signal (e.g., a processor executing instructions that are stored at a computer-readable storage device), or a combination thereof.

**[0346]** Those of skill would further appreciate that the various illustrative logical blocks, configurations, modules, circuits, and algorithm steps described in connection with the implementations disclosed herein may be implemented as electronic hardware, computer software executed by a processing device such as a hardware processor, or combinations of both. Various illustrative components, blocks, configurations, modules, circuits, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or executable software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present disclosure.

**[0347]** The steps of a method or algorithm described in connection with the implementations disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may reside in a memory device, such as random access memory (RAM), magnetoresistive random access memory (MRAM), spin-torque transfer MRAM (STT-MRAM), flash memory, read-only memory (ROM), programmable read-only memory (PROM), erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), registers, hard disk, a removable disk, or a compact disc read-only memory (CD-ROM). An exemplary memory device is coupled to the processor such that the processor can read information from, and write information to, the memory device. In the alternative, the memory device may be integral to the processor. The processor and the storage medium may reside in an application-specific integrated circuit (ASIC). The ASIC may reside in a computing device or a user terminal. In the alternative, the processor and the storage medium may reside as discrete components in a computing device or a user terminal.

**[0348]** The previous description of the disclosed implementations is provided to enable a person skilled in the art to make or use the disclosed implementations. Various

- 102 -

modifications to these implementations will be readily apparent to those skilled in the art, and the principles defined herein may be applied to other implementations without departing from the scope of the disclosure. Thus, the present disclosure is not intended to be limited to the implementations shown herein but is to be accorded the widest scope possible consistent with the principles and novel features as defined by the following claims.

**WHAT IS CLAIMED IS:**

1. An apparatus comprising:
  - a receiver configured to receive at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters; and
  - a decoder configured to:
    - generate a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal;
    - generate, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal;
    - generate a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal;
    - generate a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal; and
    - generate a modified target channel signal by modifying the target channel signal based on a temporal mismatch value.
2. The apparatus of claim 1, wherein the one or more inter-channel BWE parameters include a set of adjustment gain parameters, an adjustment spectral shape parameter, or a combination thereof.
3. The apparatus of claim 1, wherein the receiver is further configured to receive one or more BWE parameters, and wherein the decoder is further configured to:
  - generate a mid channel low-band signal based on the at least one encoded signal;
  - and
  - generate the mid channel time-domain high-band signal by performing bandwidth extension on the mid channel low-band signal based on the one or more BWE parameters.
4. The apparatus of claim 3, wherein the BWE parameters include mid channel

- 104 -

high-band linear predictive coding (LPC) parameters, a set of gain parameters, or a combination thereof.

5. The apparatus of claim 3, wherein the decoder includes a time-domain bandwidth extension decoder, and wherein the time-domain bandwidth extension decoder is configured to generate the mid channel time-domain high-band signal based on the BWE parameters.

6. The apparatus of claim 1, wherein the decoder is further configured to: generate, based on the at least one encoded signal, a mid channel low-band signal and a side channel low-band signal; and generate the first channel low-band signal and the second channel low-band signal by upmixing the mid channel low-band signal and the side channel low-band signal.

7. The apparatus of claim 1, wherein the decoder is further configured to: generate a mid channel low-band signal based on the at least one encoded signal; generate one or more mapped parameters based on one or more side parameters, wherein the at least one encoded signal includes the one or more side parameters; and generate the first channel low-band signal and the second channel low-band signal by applying the one or more side parameters to the mid channel low-band signal.

8. The apparatus of claim 1, wherein the decoder is further configured to generate the modified target channel signal by temporally shifting first samples of the target channel signal relative to second samples of the reference channel signal by an amount based on the temporal mismatch value.

- 105 -

9. The apparatus of claim 1, wherein the decoder is further configured to:  
generate a left output signal corresponding to one of the reference channel signal  
or the modified target channel signal; and  
generate a right output signal corresponding to the other of the reference channel  
signal or the modified target channel signal.
10. The apparatus of claim 9, wherein the inter-channel BWE parameters  
include a high-band reference channel indicator, wherein the decoder is further  
configured to determine, based on the high-band reference channel indicator, whether  
the left output signal or the right output signal corresponds to the reference channel  
signal.
11. The apparatus of claim 9, wherein the decoder is further configured to:  
provide the left output signal to a first loudspeaker; and  
provide the right output signal to a second loudspeaker.
12. The apparatus of claim 1, wherein the first channel low-band signal and the  
second channel low-band signal are generated based on stereo low-band upmix  
processing, and wherein the first channel time-domain high-band signal and the second  
channel time-domain high-band signal are generated based on stereo inter-channel  
bandwidth extension high-band upmix processing.
13. The apparatus of claim 1, wherein the decoder is further configured to:  
generate a first output signal based on the reference channel signal;  
generate a second output signal based on the modified target channel signal;  
provide the first output signal to a first speaker; and  
provide the second output signal to a second speaker.
14. The apparatus of claim 1, further comprising an antenna coupled to the  
receiver, wherein the receiver is configured to receive the at least one encoded signal via  
the antenna.
15. The apparatus of claim 1, wherein the receiver and the decoder are

- 106 -

integrated into a mobile communication device.

16. The apparatus of claim 1, wherein the receiver and the decoder are integrated into a base station.

17. A method of communication comprising:  
receiving, at a device, at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters;  
generating, at the device, a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal;  
generating, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal;  
generating, at the device, a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal;  
generating, at the device, a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal; and  
generating, at the device, a modified target channel signal by modifying the target channel signal based on a temporal mismatch value.

18. The method of claim 17, further comprising generating, at the device, a mid channel low-band signal and a side channel low-band signal based on the at least one encoded signal, wherein the first channel low-band signal and the second channel low-band signal are based on the mid channel low-band signal, the side channel low-band signal, and a gain parameter.

19. The method of claim 17, further comprising:  
generating a first output signal based on the modified target channel signal; and  
generating a second output signal based on the reference channel signal.

- 107 -

20. The method of claim 19, further comprising:  
providing the first output signal to a first speaker; and  
providing the second output signal to a second speaker.
21. The method of claim 17, further comprising receiving the temporal mismatch value at the device,  
wherein the modified target channel signal is generated by temporally shifting first samples of the target channel signal relative to second samples of the reference channel signal by an amount that is based on the temporal mismatch value.
22. The method of claim 17, wherein the device comprises a mobile communication device.
23. The method of claim 17, wherein the device comprises a base station.
24. A computer-readable storage device storing instructions that, when executed by a processor, cause the processor to perform operations comprising:  
receiving at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters;  
generating a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal;  
generating, based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters, a first channel time-domain high-band signal and a second channel time-domain high-band signal;  
generating a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal;  
generating a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal; and  
generating a modified target channel signal by modifying the target channel signal based on a temporal mismatch value.

- 108 -

25. The computer-readable storage device of claim 24, wherein the operations further comprise:

- generating a first output signal based on the reference channel signal;
- generating a second output signal based on the modified target channel signal;
- providing the first output signal to a first loudspeaker; and
- providing the second output signal to a second loudspeaker.

26. The computer-readable storage device of claim 24, wherein the operations further comprise:

- receiving one or more BWE parameters; and
- generating a mid channel low-band signal based on the at least one encoded signal,

wherein the mid channel time-domain high-band signal is generated by

- performing bandwidth extension on the mid channel low-band signal based at least in part on the one or more BWE parameters.

27. The computer-readable storage device of claim 26, wherein the one or more BWE parameters include mid channel high-band linear predictive coding (LPC) parameters, a set of gain parameters, or a combination thereof.

28. The computer-readable storage device of claim 24, wherein the one or more inter-channel BWE parameters include a set of adjustment gain parameters, an adjustment spectral shape parameter, or a combination thereof.

29. The computer-readable storage device of claim 24, wherein the operations further comprise generating the modified target channel signal by temporally shifting first samples of the target channel signal relative to second samples of the reference channel signal by an amount that is based on the temporal mismatch value.

- 109 -

30. An apparatus comprising:

means for receiving at least one encoded signal that includes one or more inter-channel bandwidth extension (BWE) parameters;

means for generating a mid channel time-domain high-band signal by performing bandwidth extension based on the at least one encoded signal;

means for generating a first channel time-domain high-band signal and a second channel time-domain high-band signal based on the mid channel time-domain high-band signal and the one or more inter-channel BWE parameters;

means for generating a target channel signal by combining the first channel time-domain high-band signal and a first channel low-band signal;

means for generating a reference channel signal by combining the second channel time-domain high-band signal and a second channel low-band signal; and

means for generating a modified target channel signal by modifying the target channel signal based on a temporal mismatch value.

31. The apparatus of claim 30, wherein the means for receiving the at least one encoded signal, the means for generating the mid channel time-domain high-band signal, the means for generating the first channel time-domain high-band signal and the second channel time-domain high-band signal, the means for generating the target channel signal, the means for generating the reference channel signal, and the means for generating the modified target channel signal are integrated into at least one of a mobile phone, a communication device, a computer, a music player, a video player, an entertainment unit, a navigation device, a personal digital assistant (PDA), a decoder, or a set top box.

- 110 -

32. The apparatus of claim 30, wherein the means for receiving the at least one encoded signal, the means for generating the mid channel time-domain high-band signal, the means for generating the first channel time-domain high-band signal and the second channel time-domain high-band signal, the means for generating the target channel signal, the means for generating the reference channel signal, and the means for generating the modified target channel signal are integrated into a mobile communication device.

33. The apparatus of claim 30, wherein the means for receiving the at least one encoded signal, the means for generating the mid channel time-domain high-band signal, the means for generating the first channel time-domain high-band signal and the second channel time-domain high-band signal, the means for generating the target channel signal, the means for generating the reference channel signal, and the means for generating the modified target channel signal are integrated into a base station.

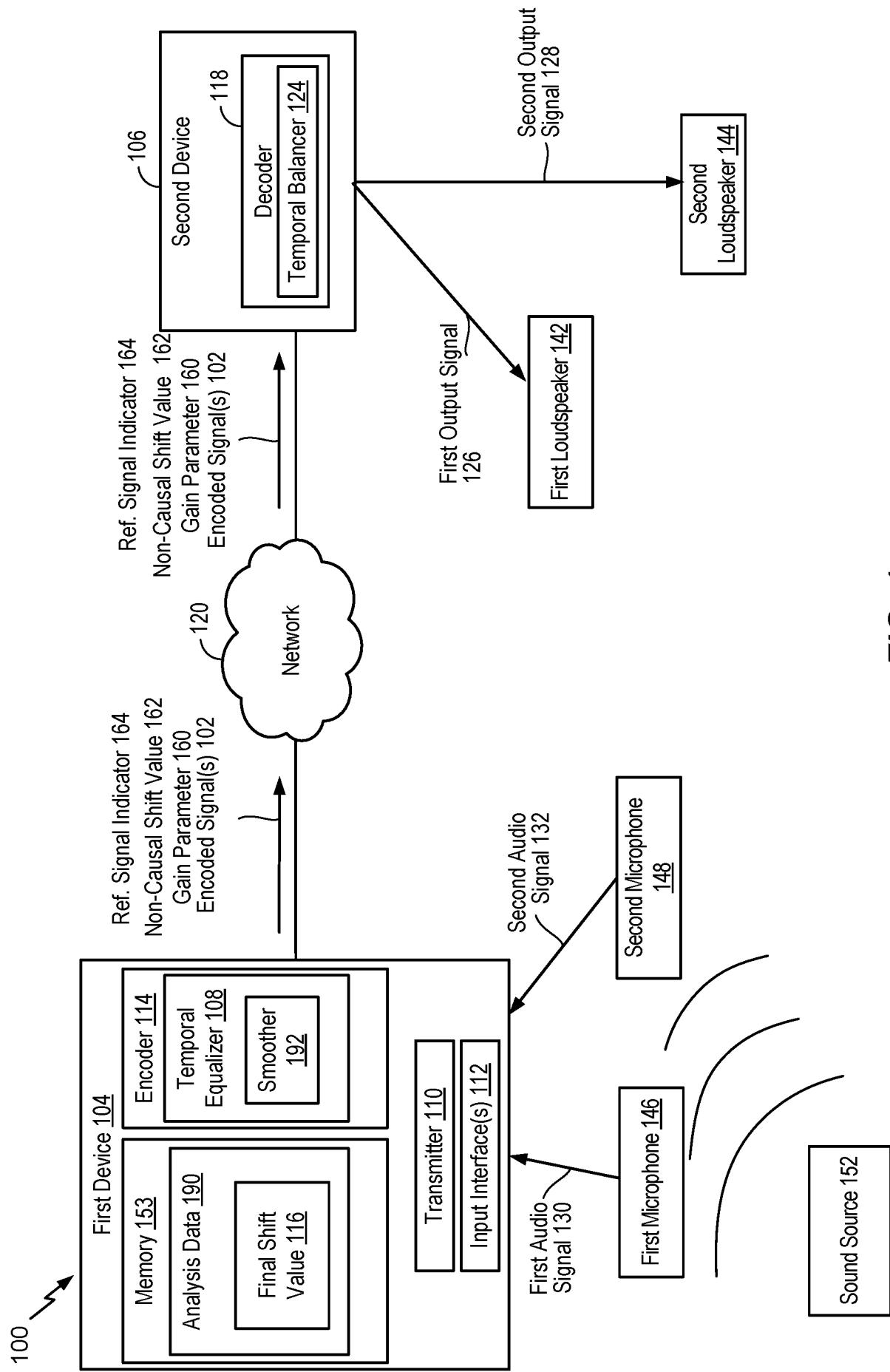


FIG. 1

2/30

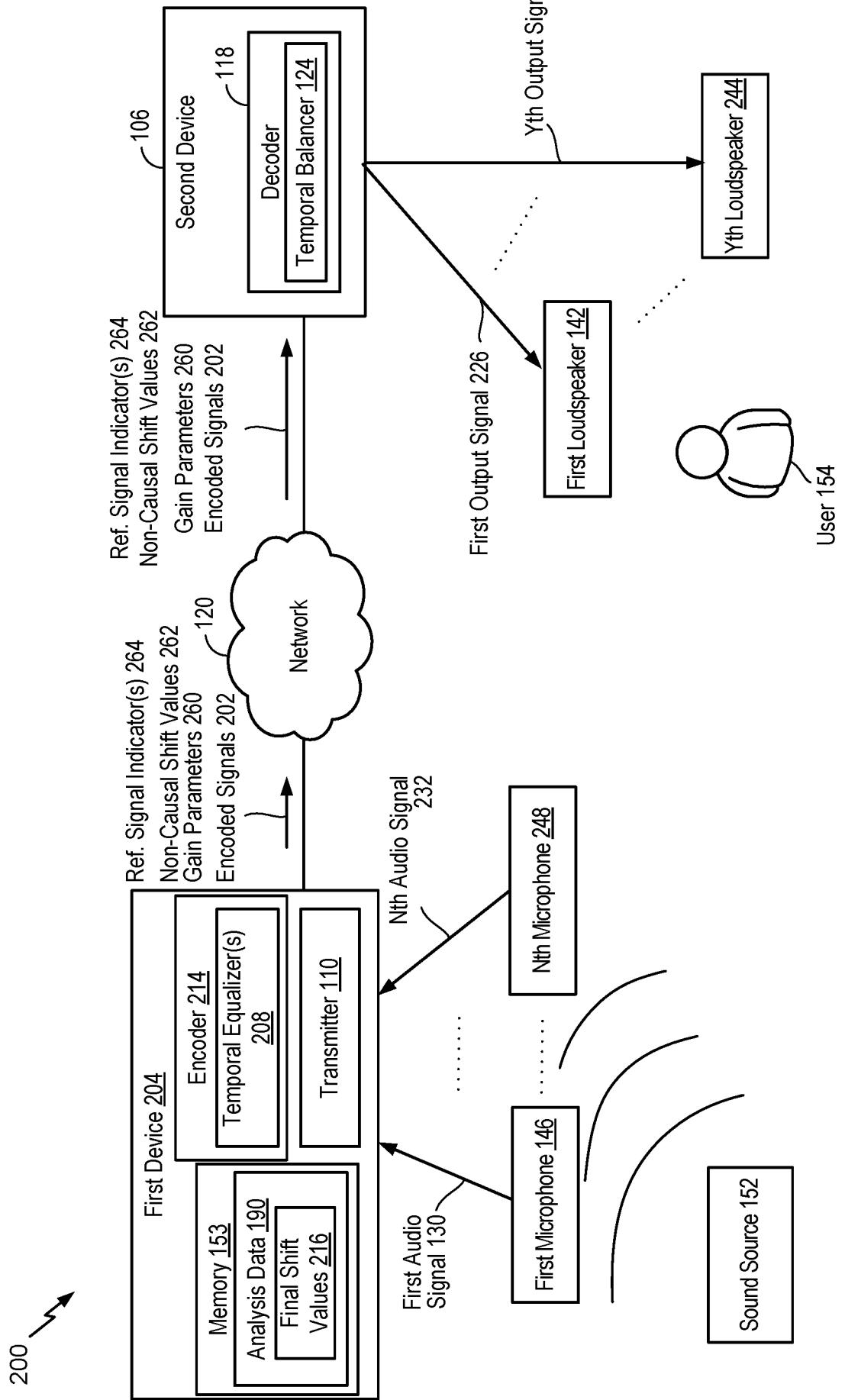
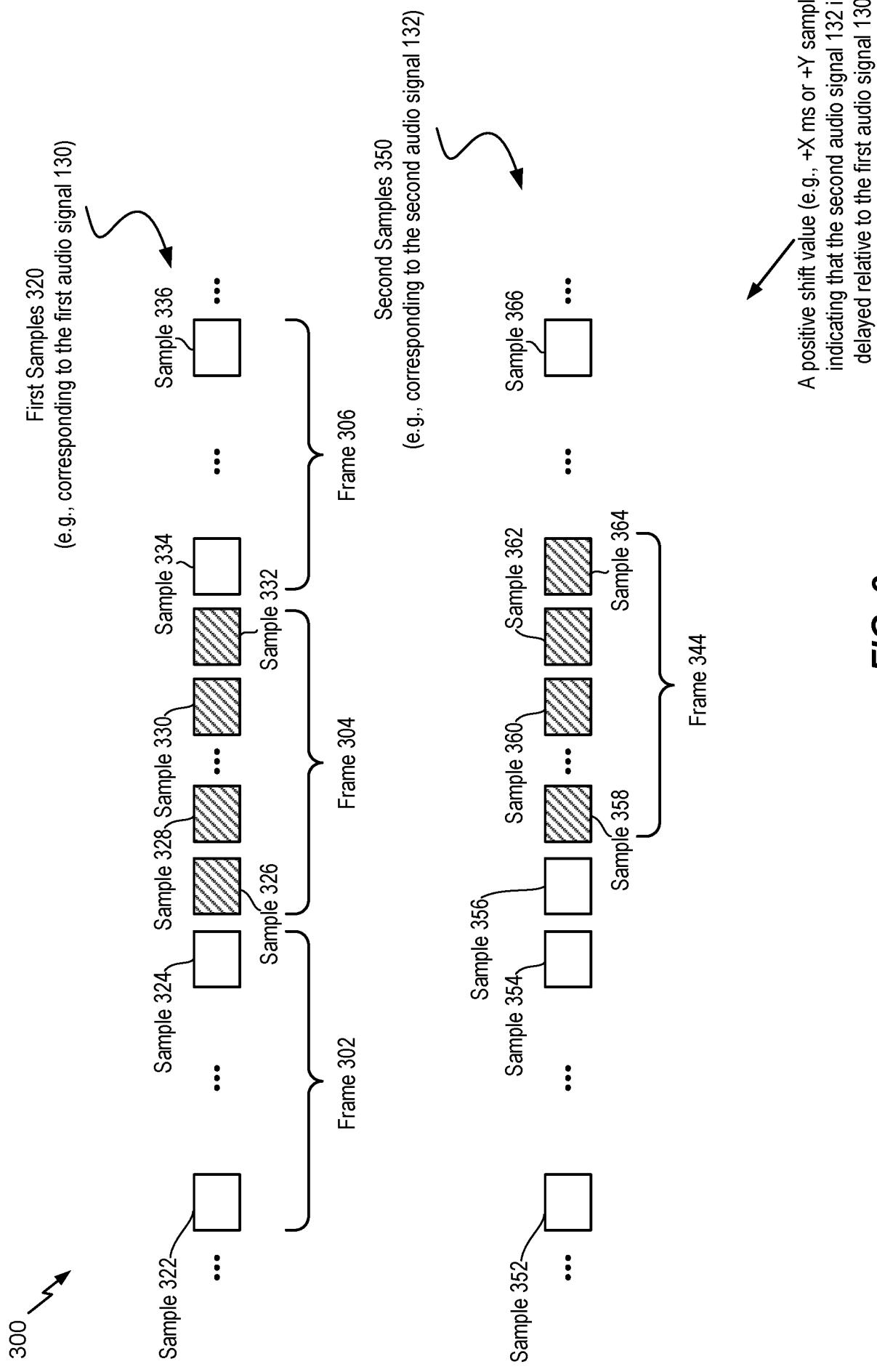
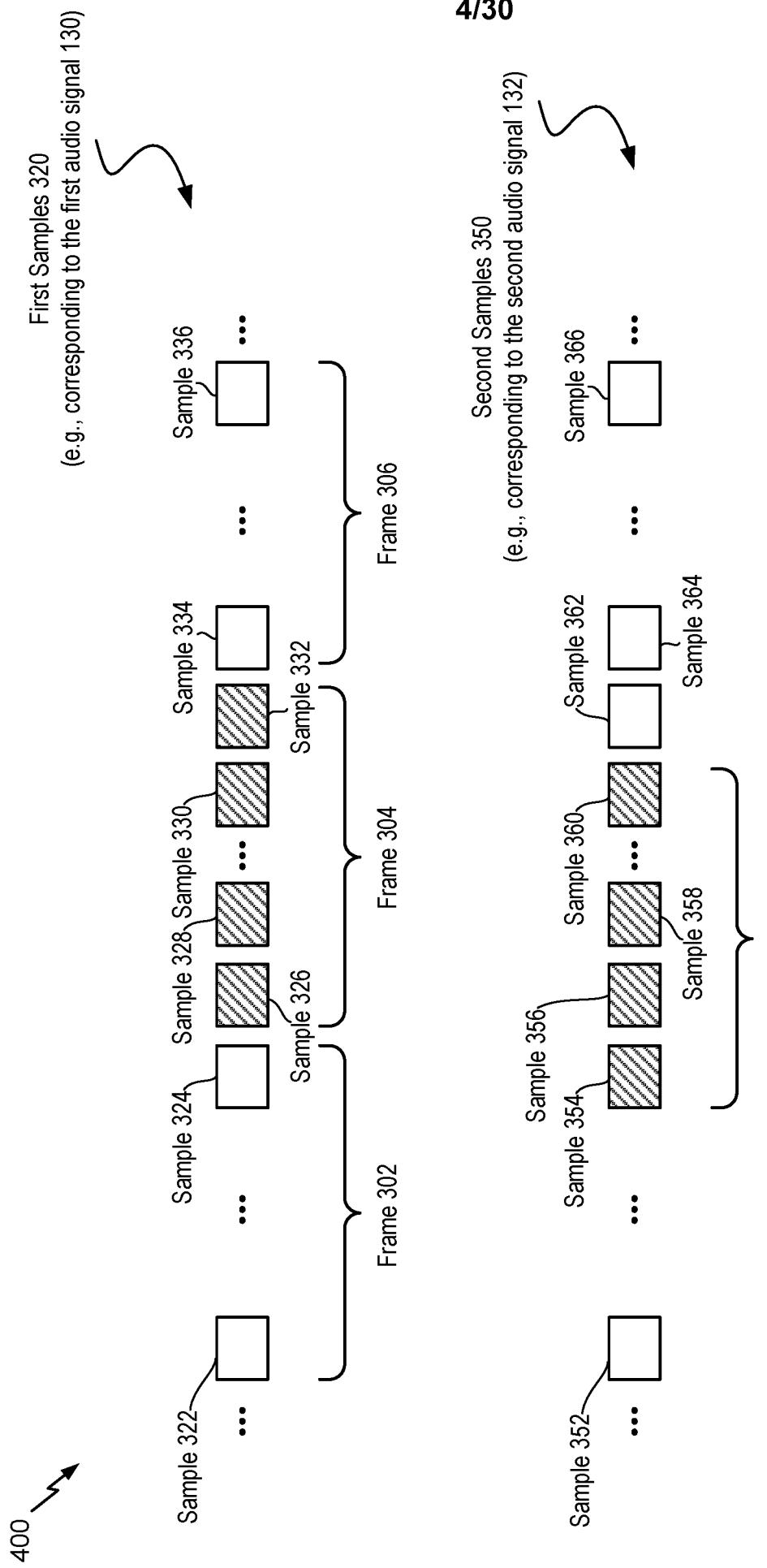


FIG. 2

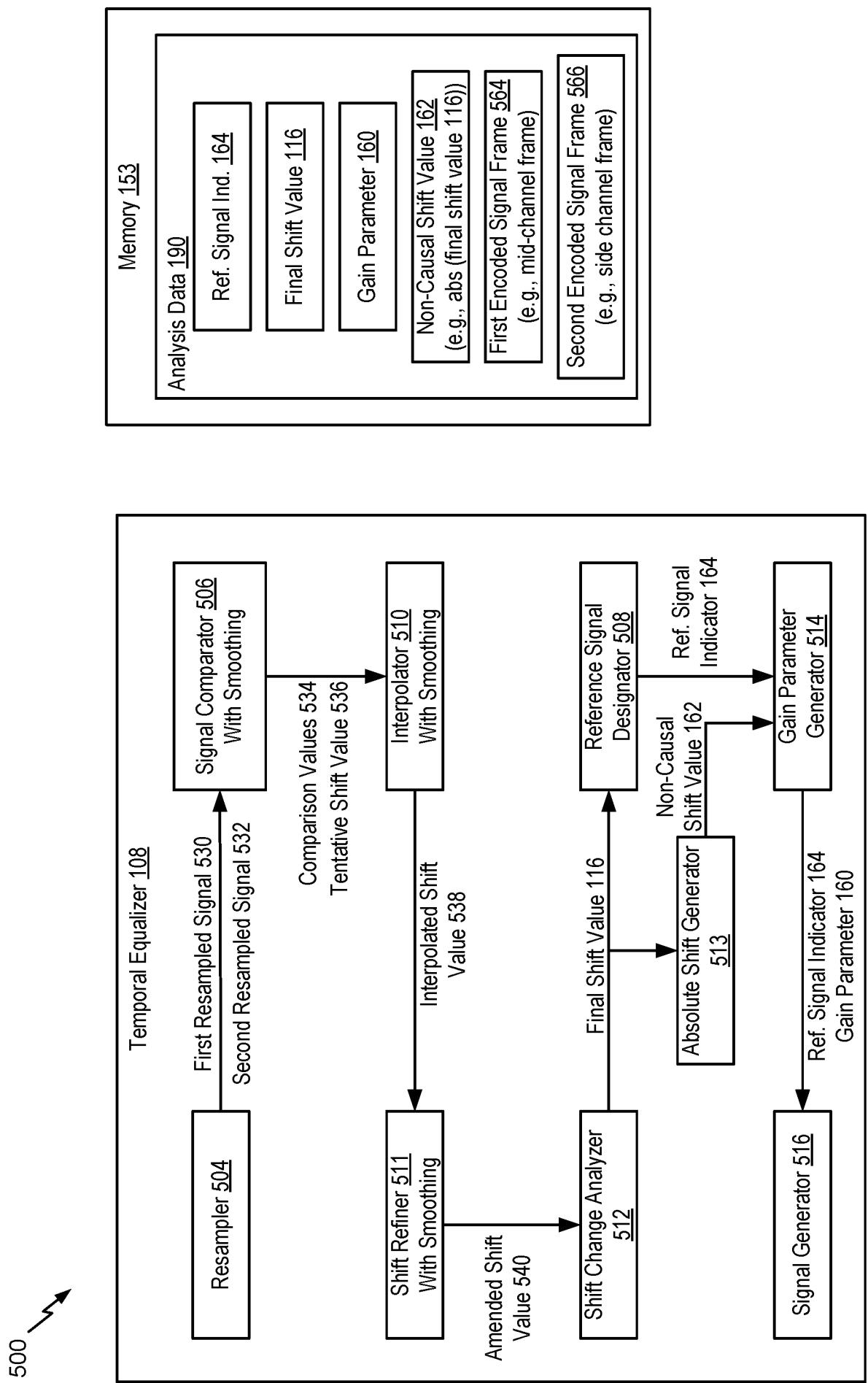
3/30

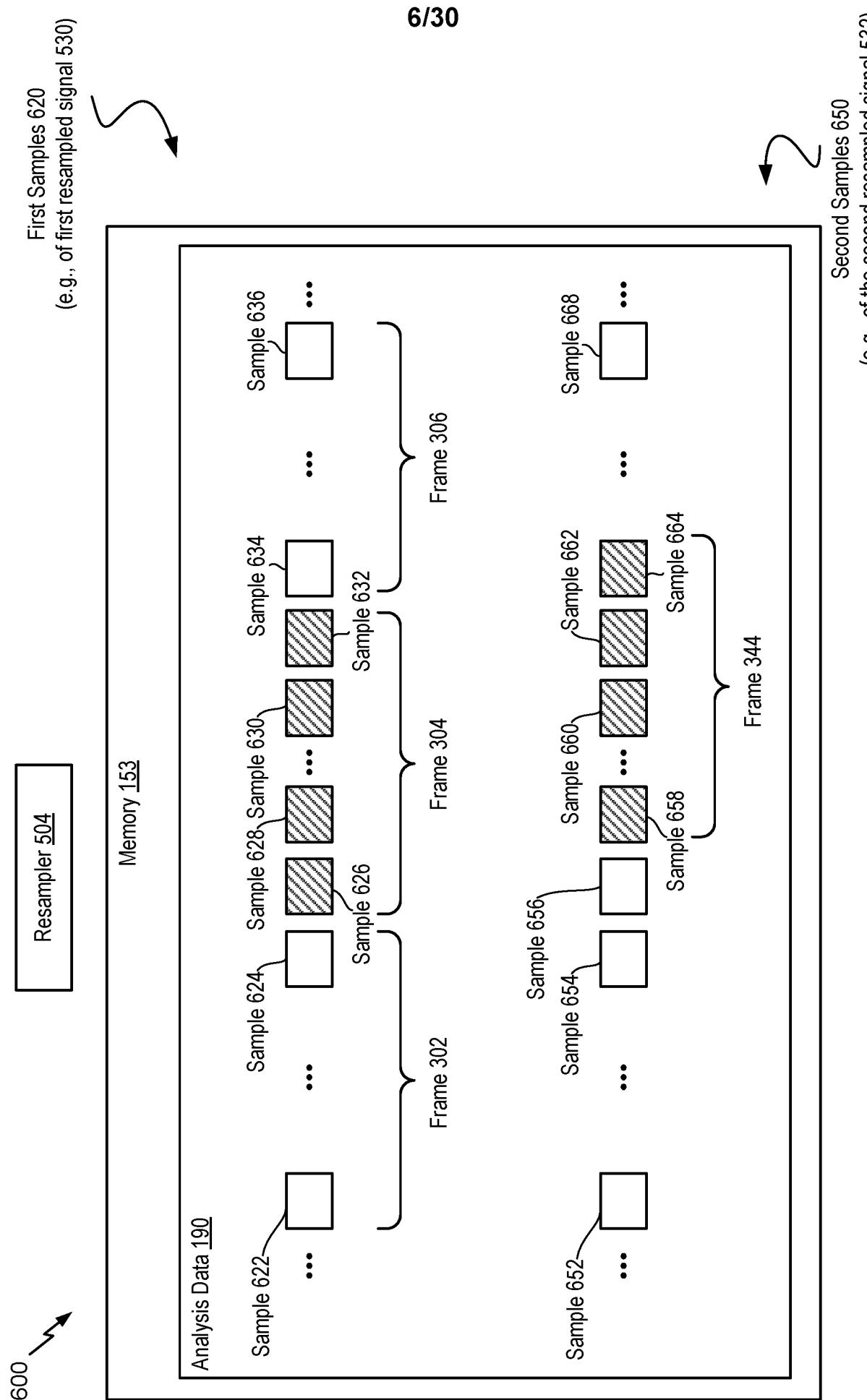
**FIG. 3**

4/30

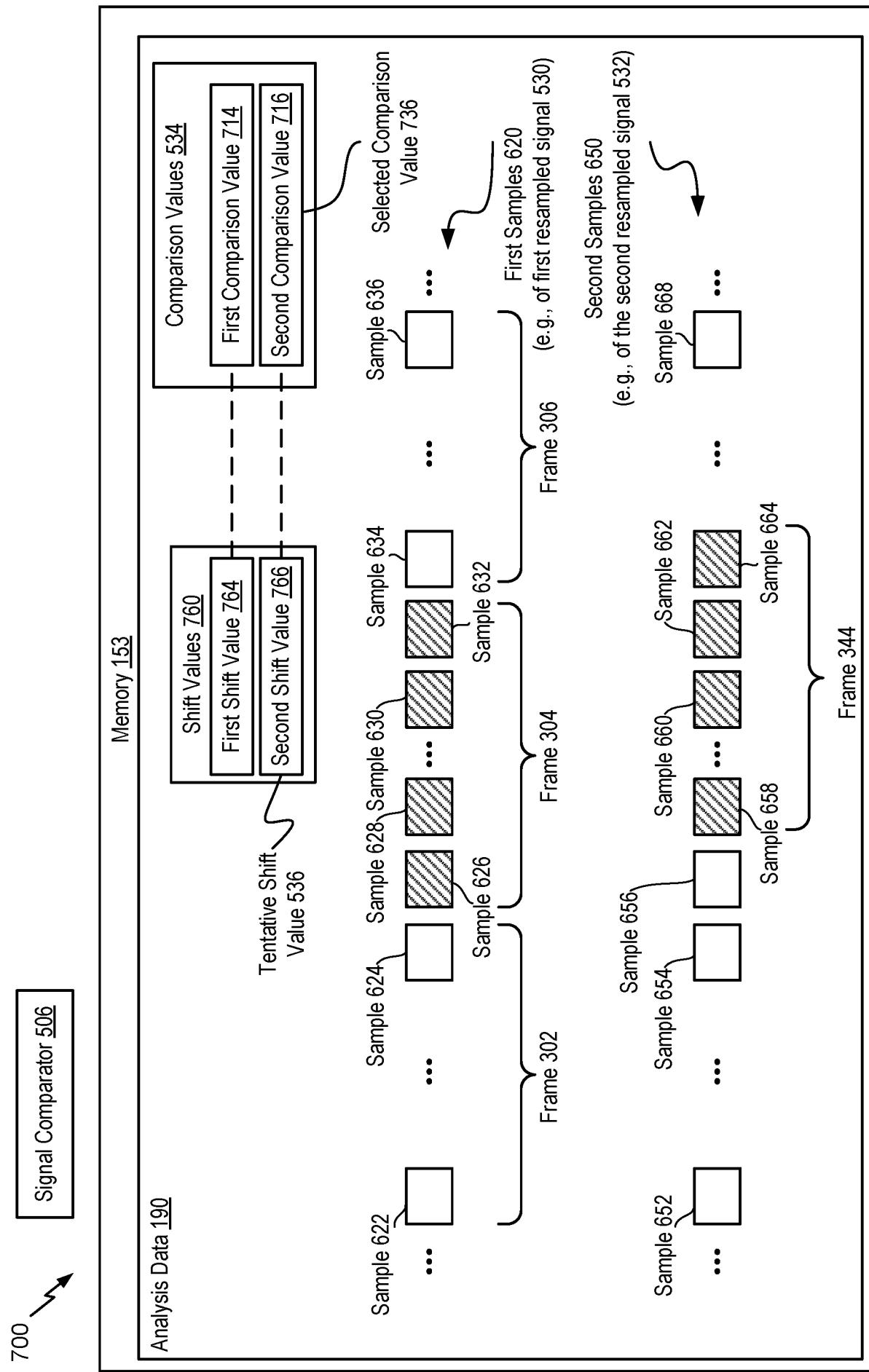
**FIG. 4**

5/30



**FIG. 6**

7/30

**FIG. 7**

8/30

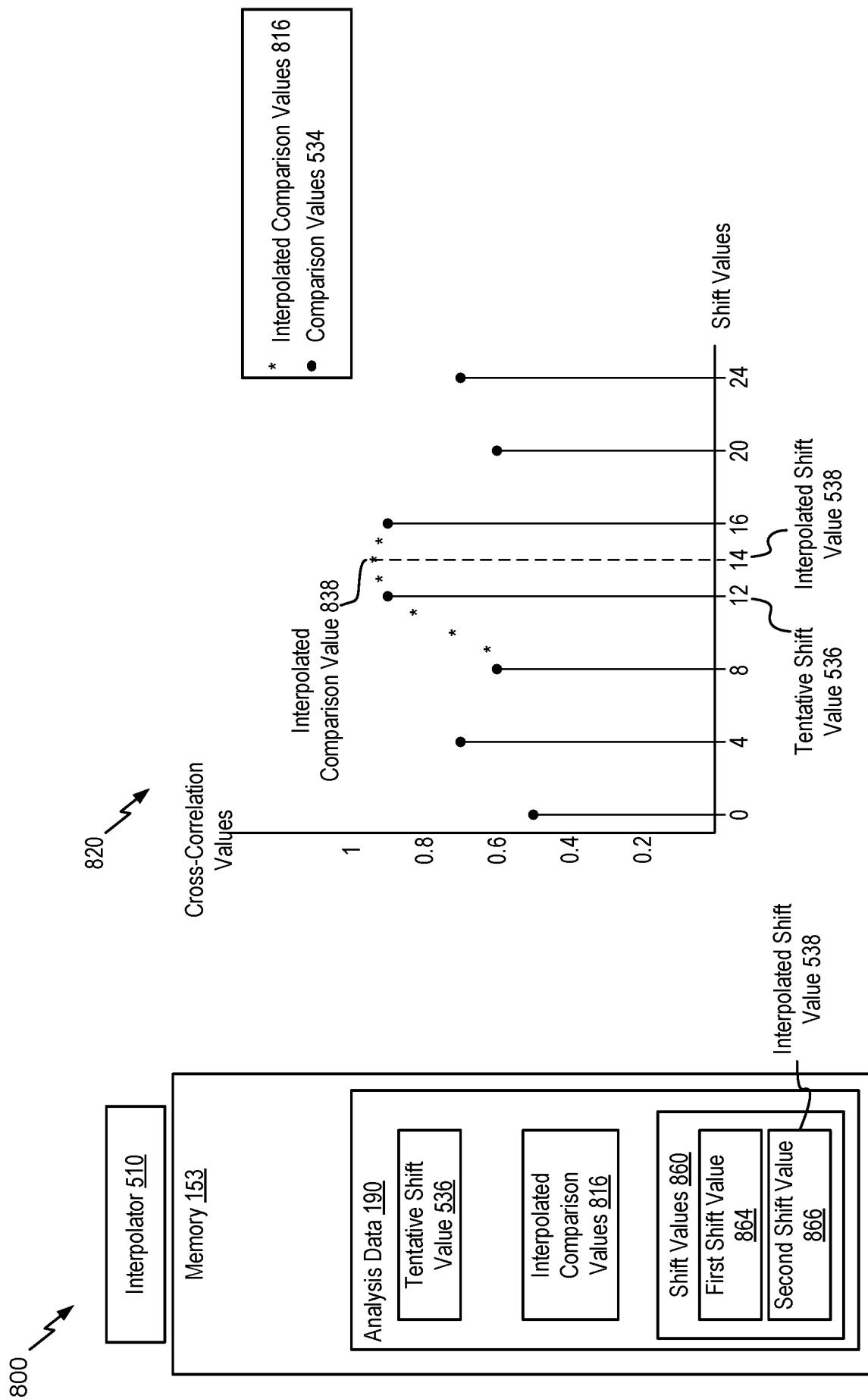
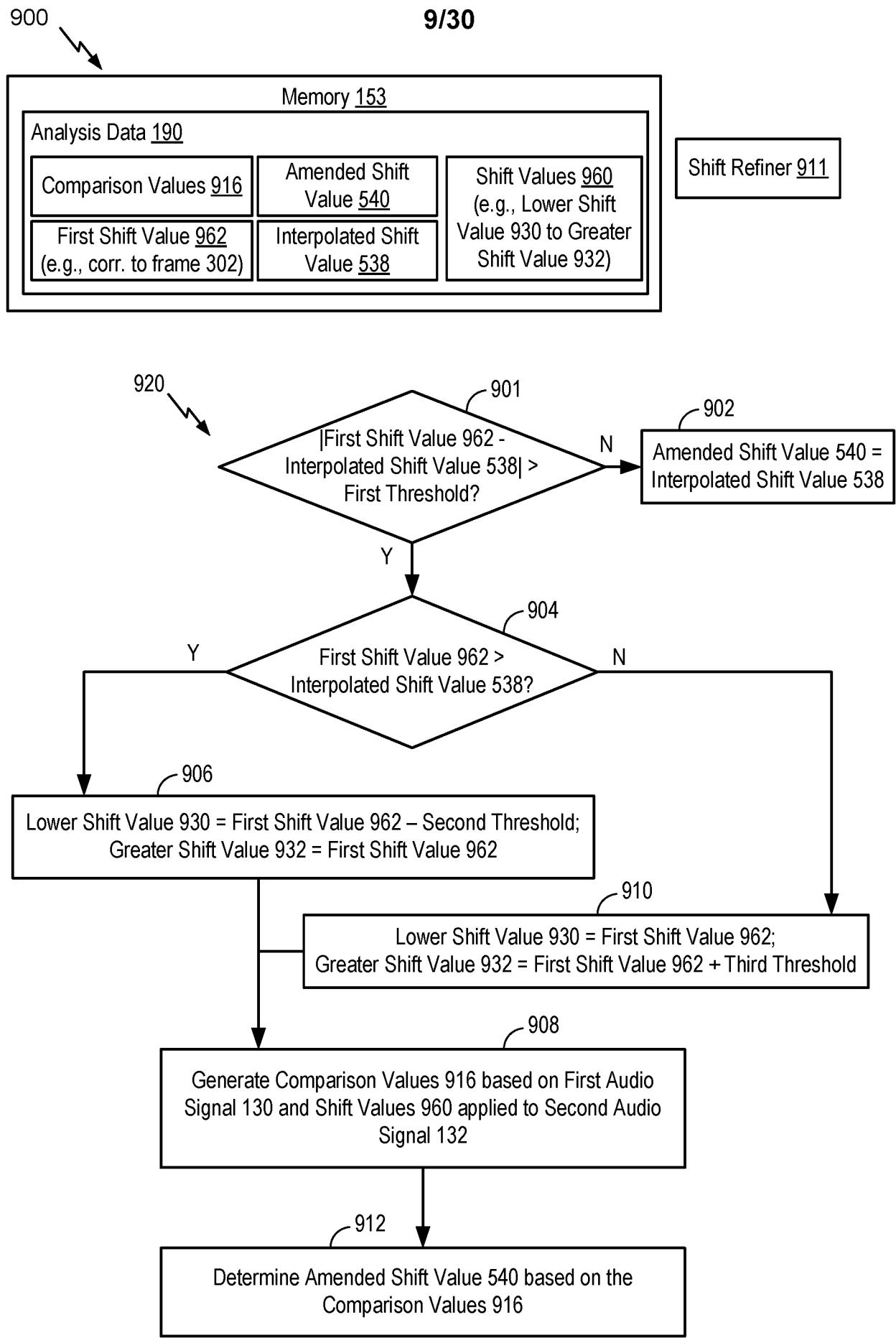
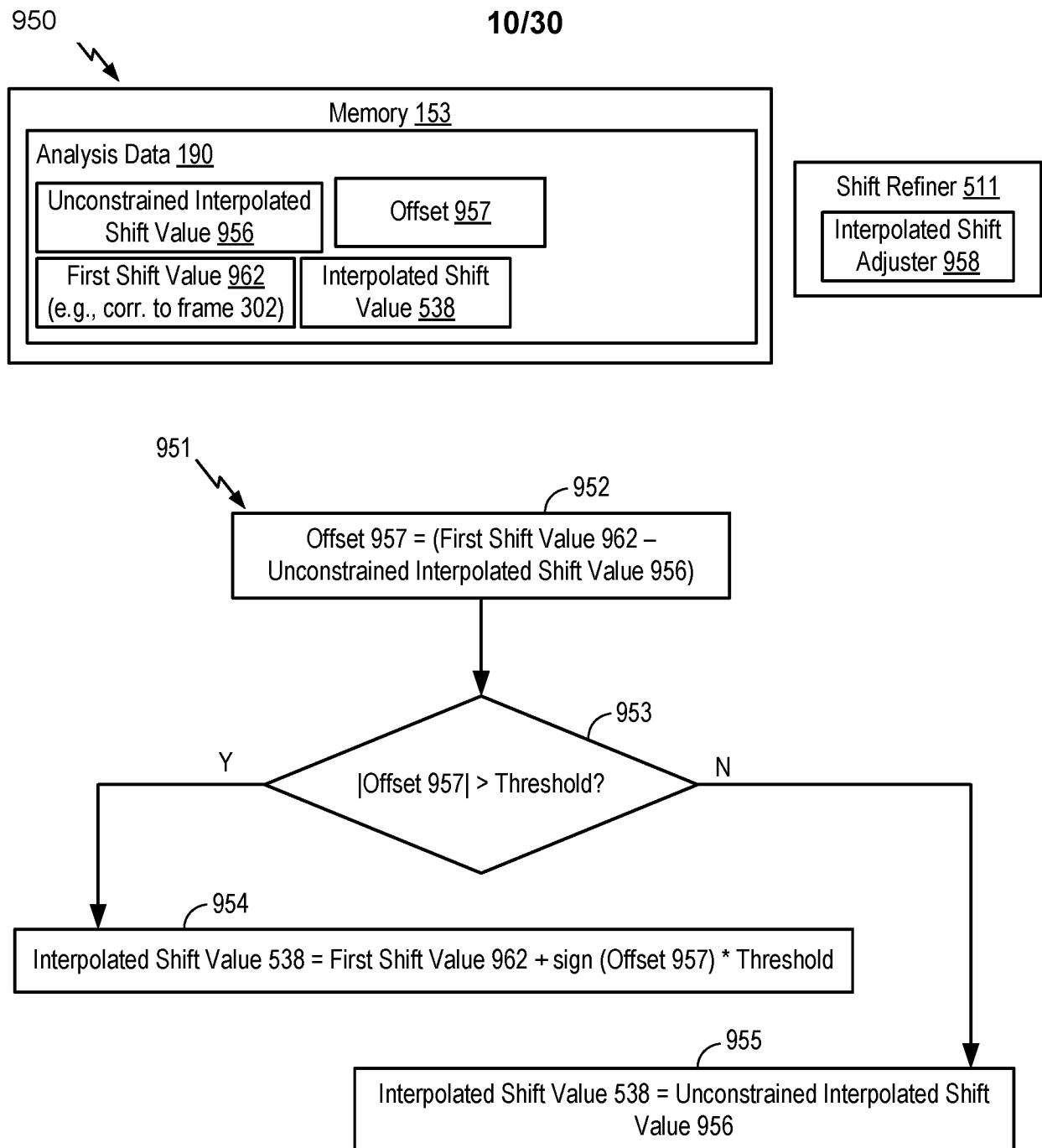
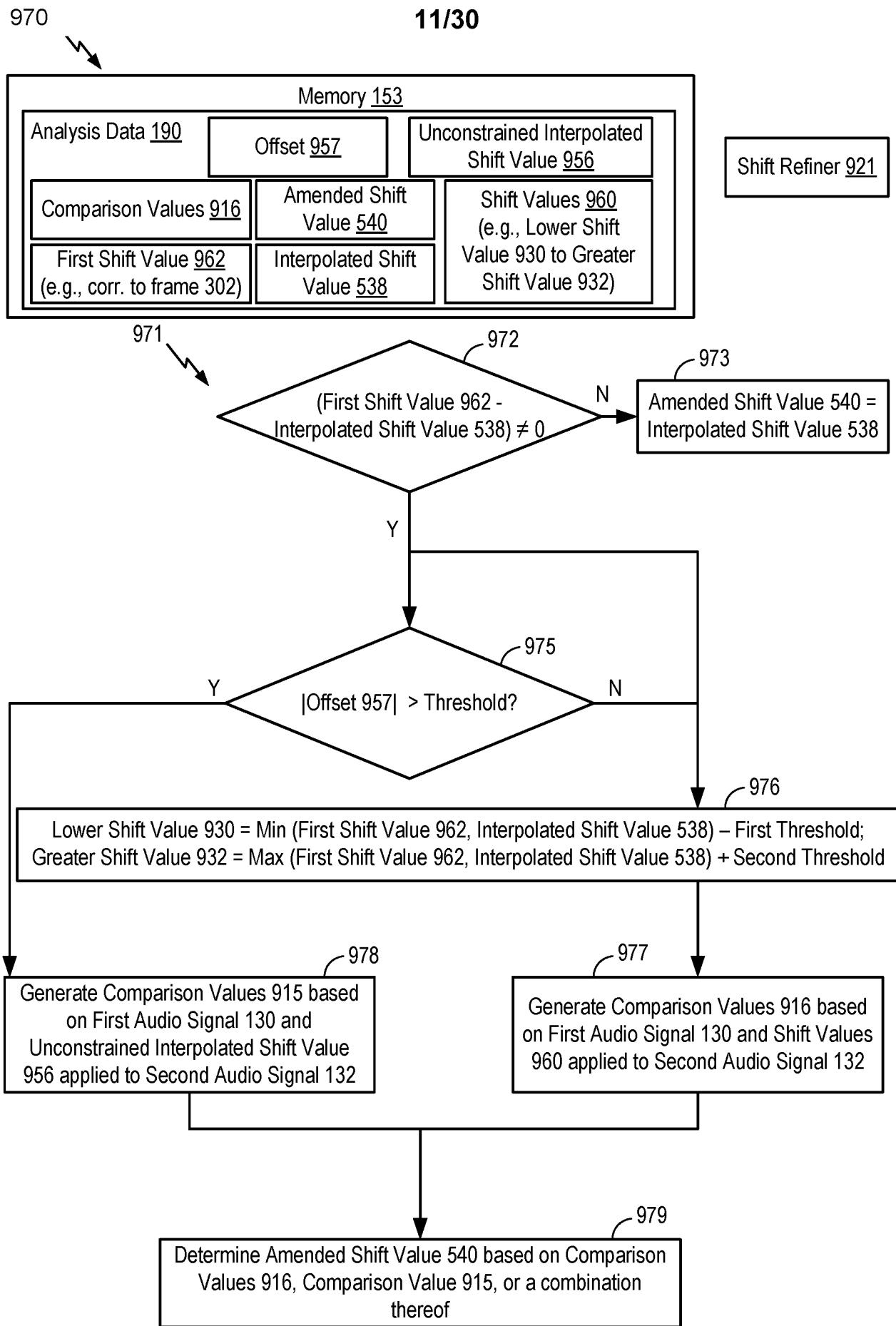


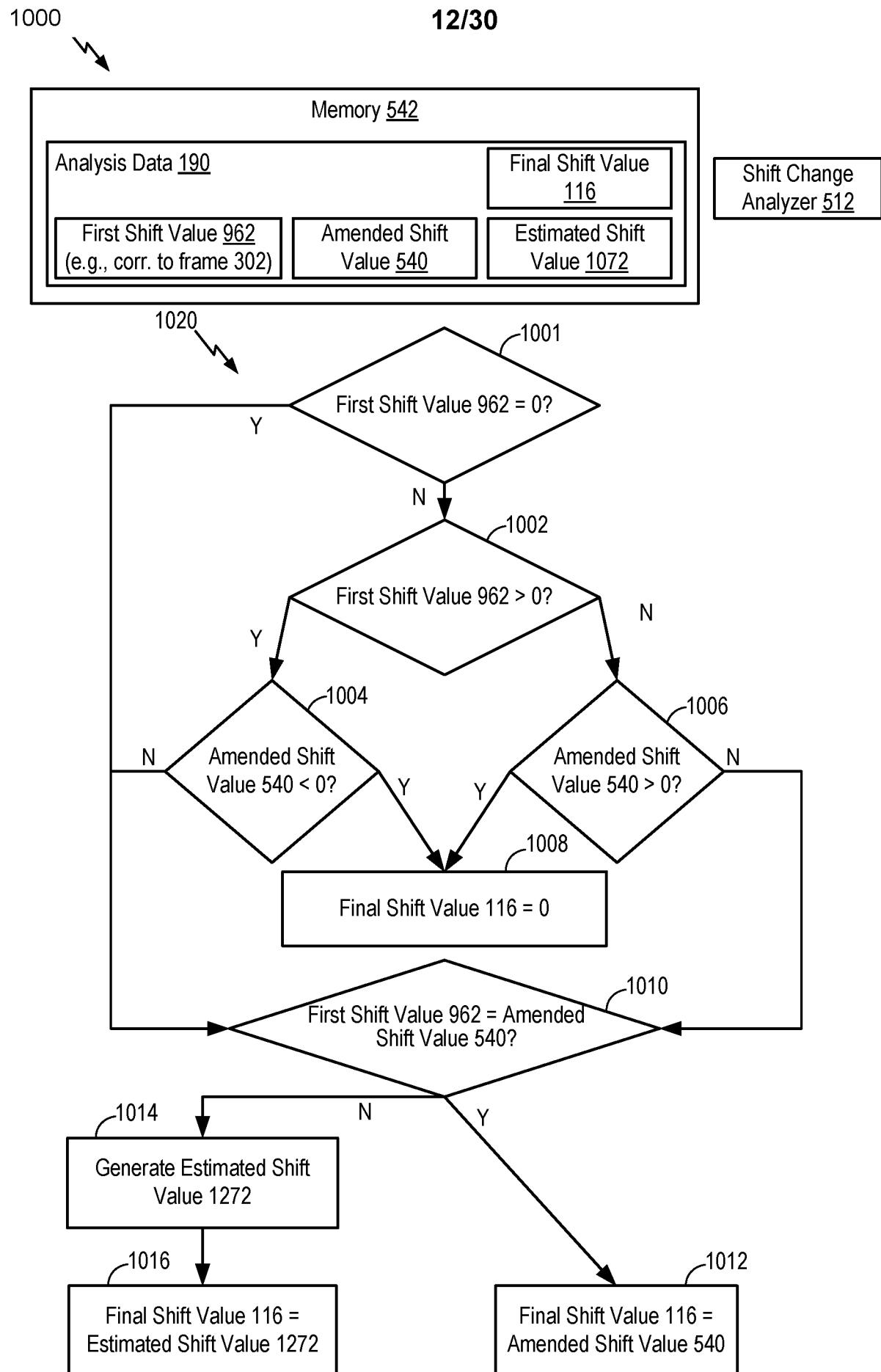
FIG. 8

**FIG. 9A**

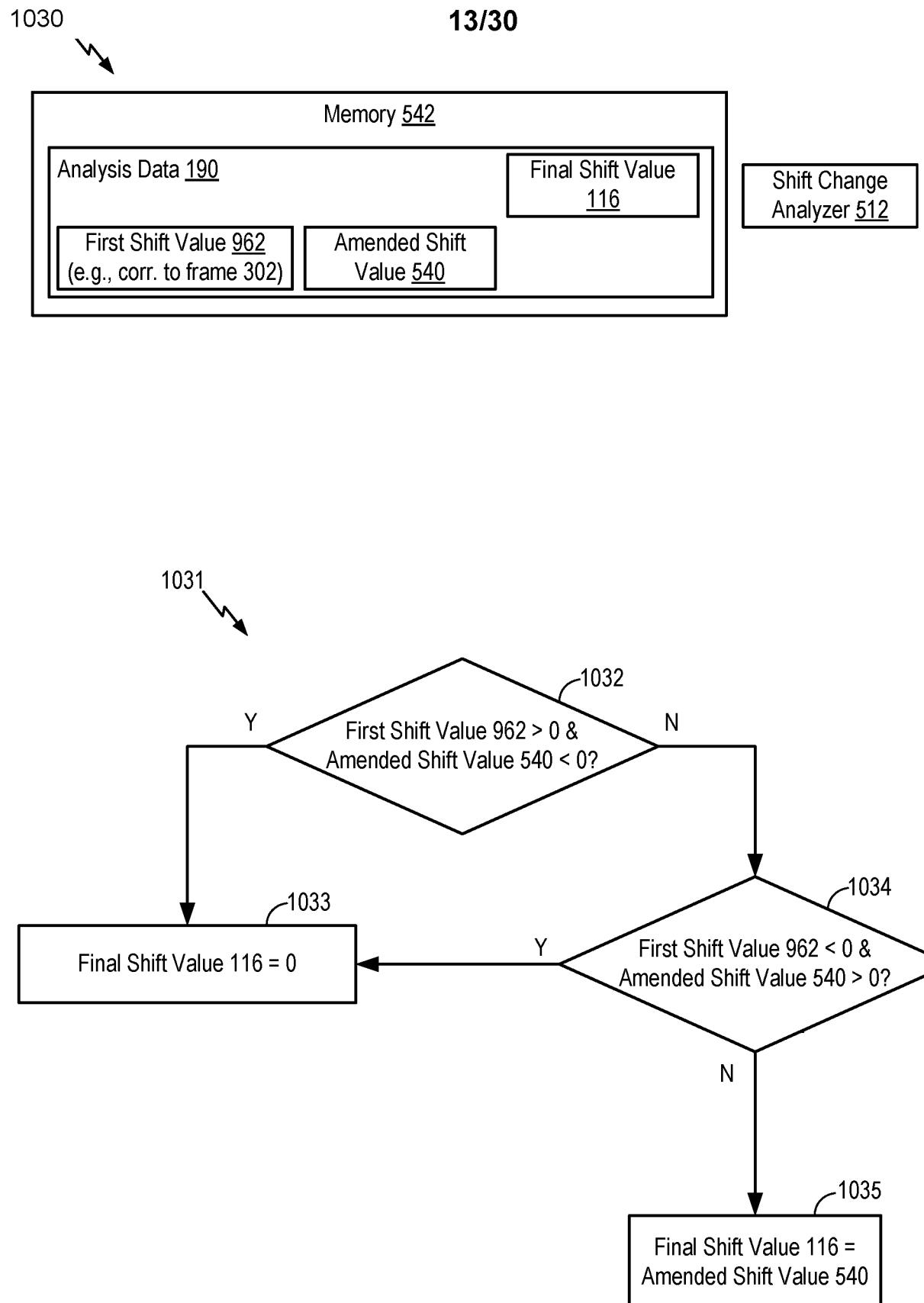


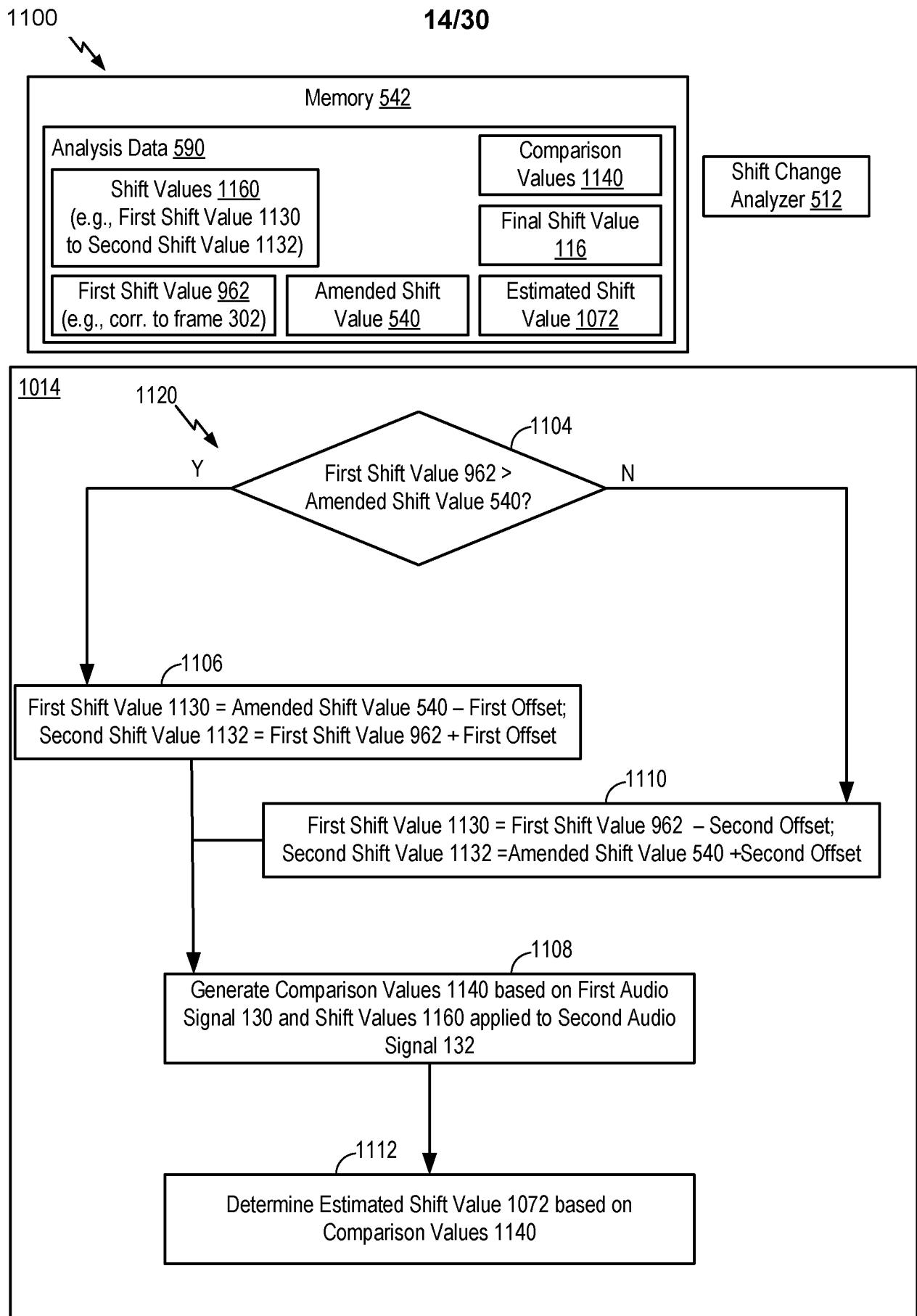
**FIG. 9B**

**FIG.9C**



**FIG. 10A**

**FIG. 10B**

**FIG. 11**

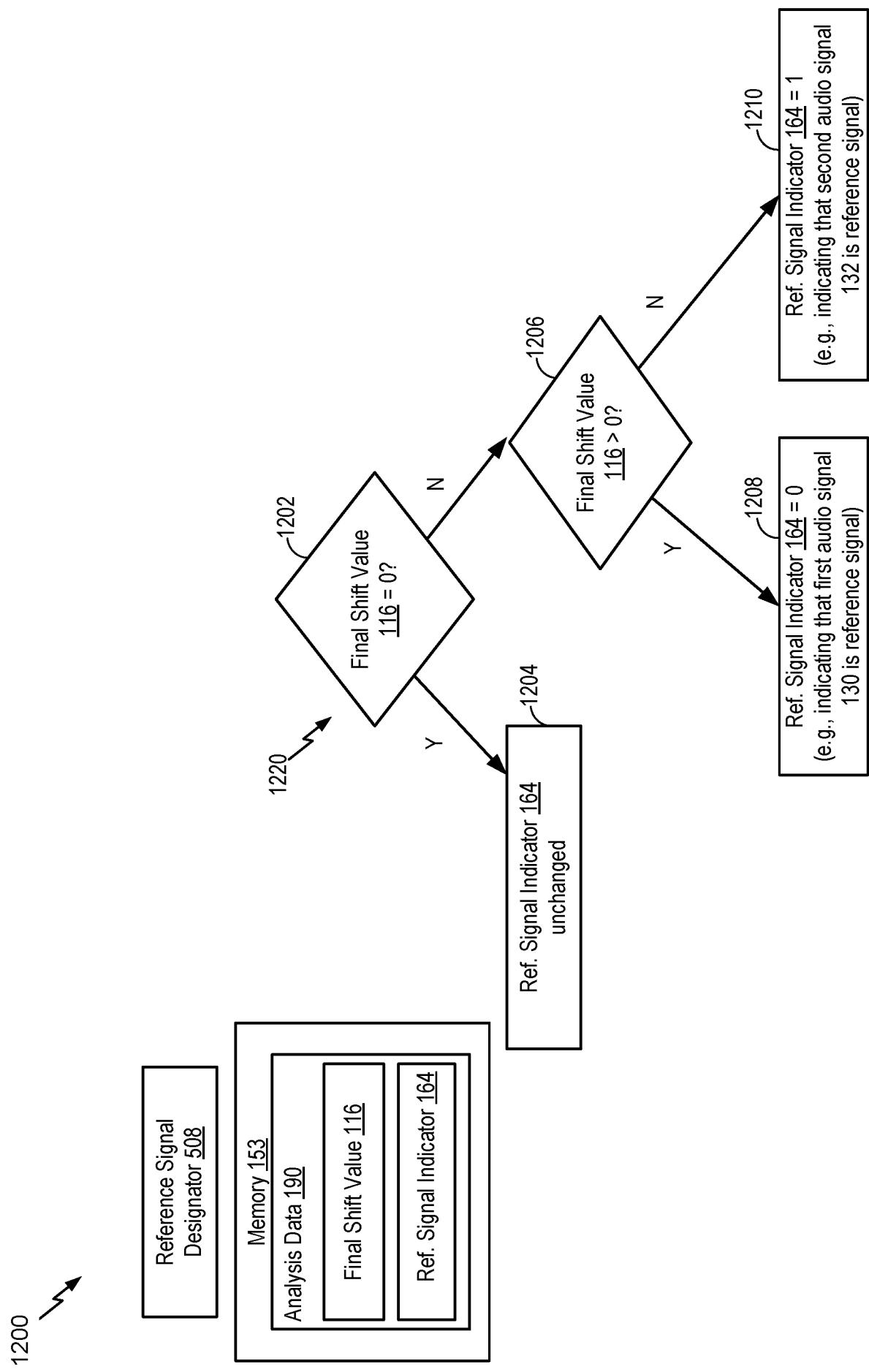


FIG. 12

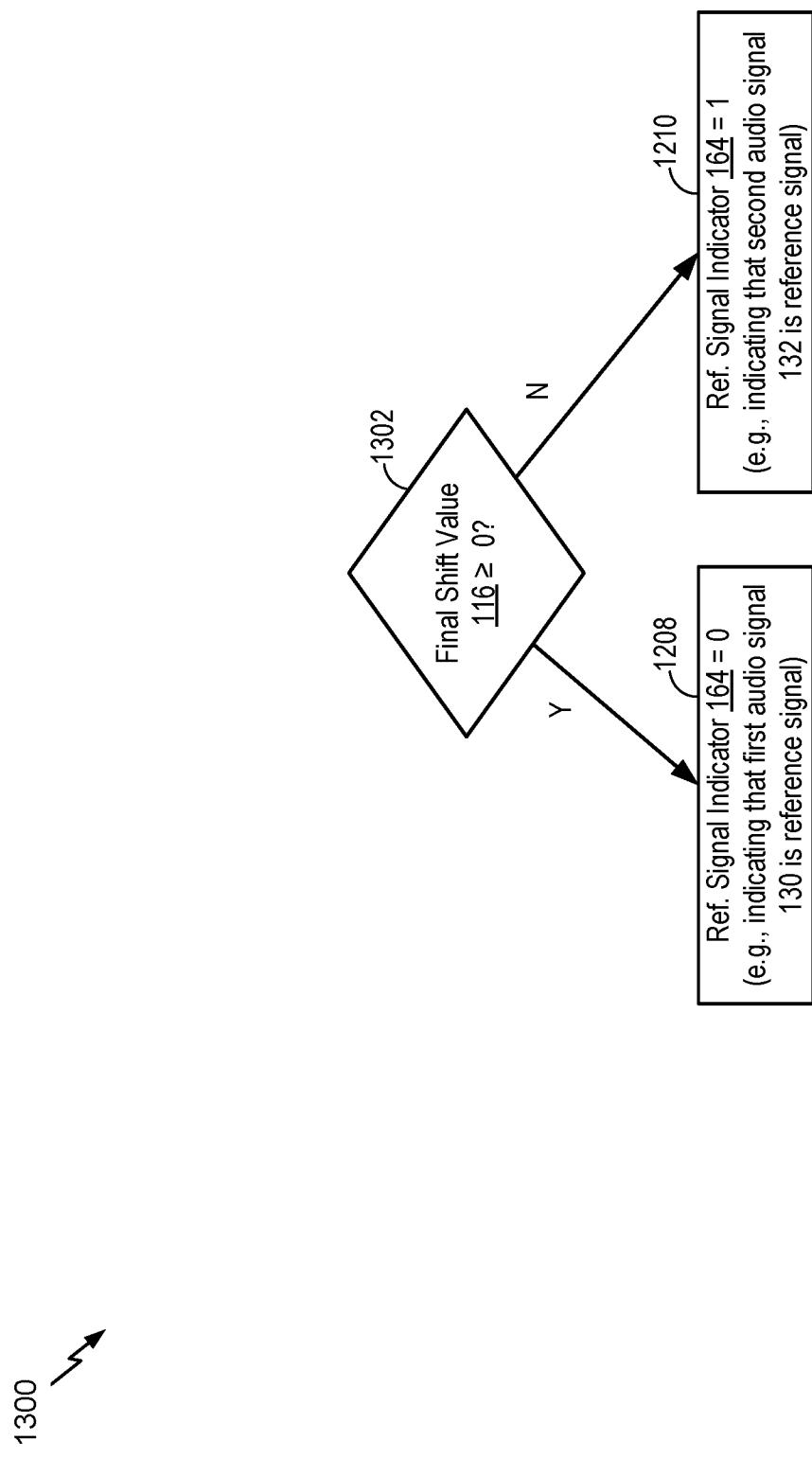


FIG. 13

17/30

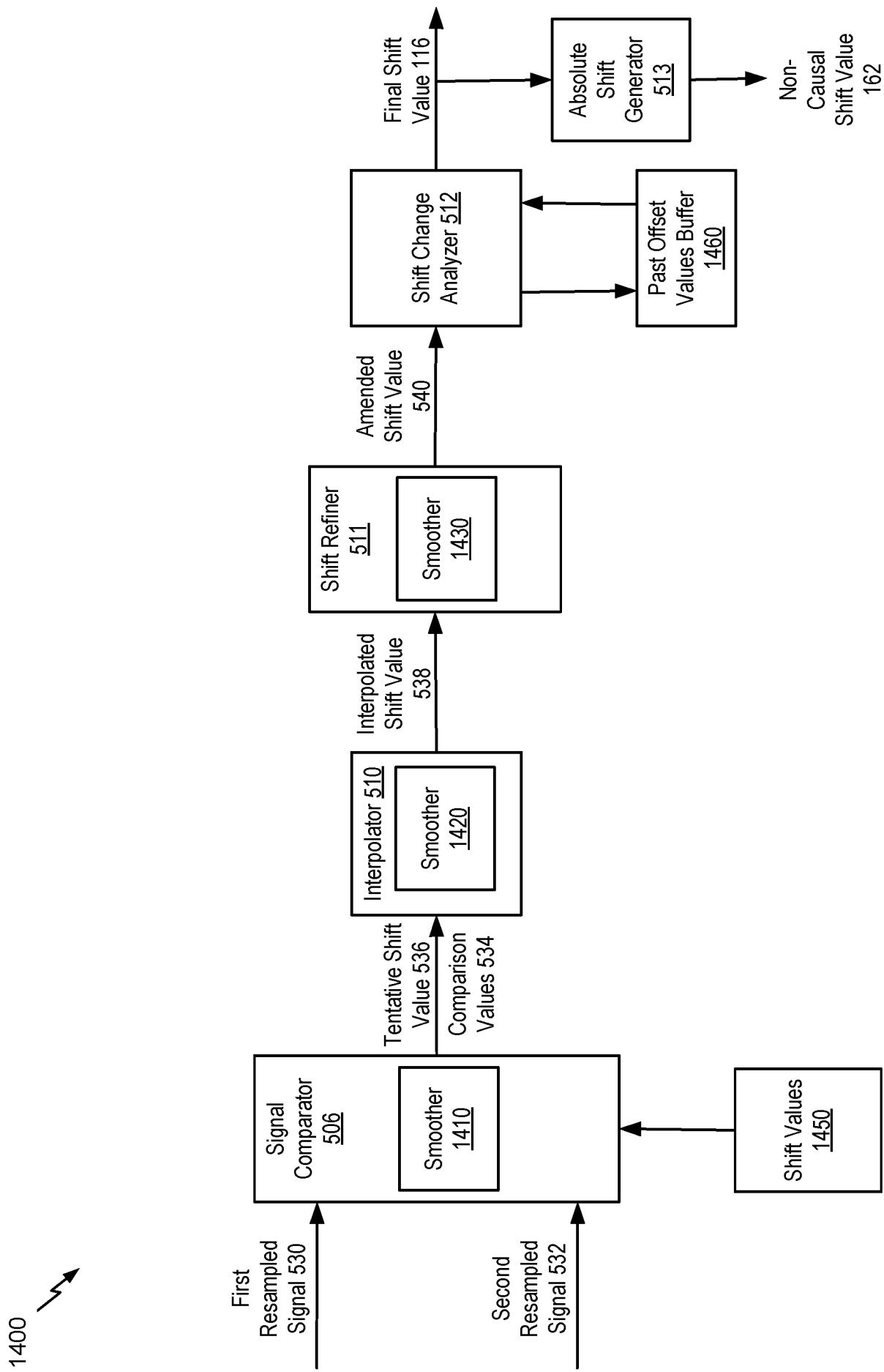


FIG. 14

18/30

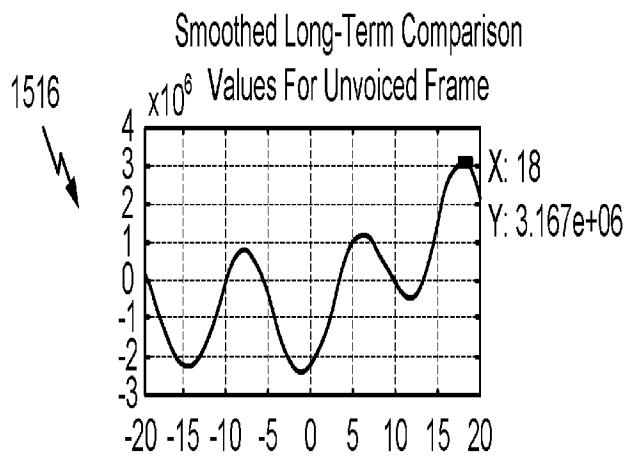
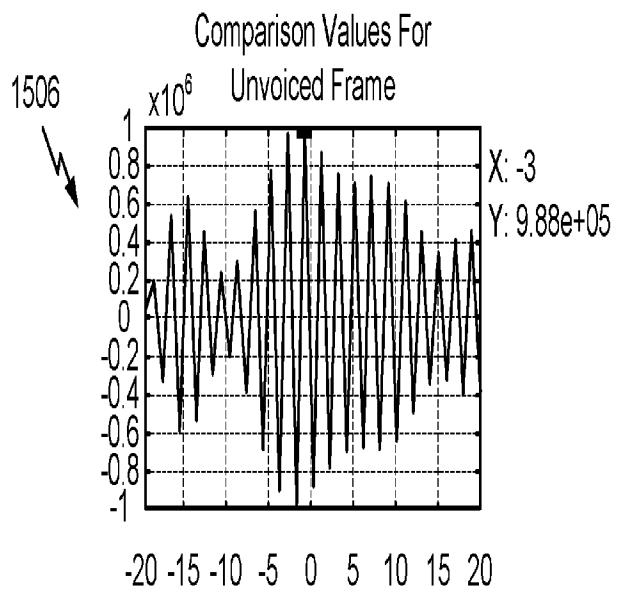
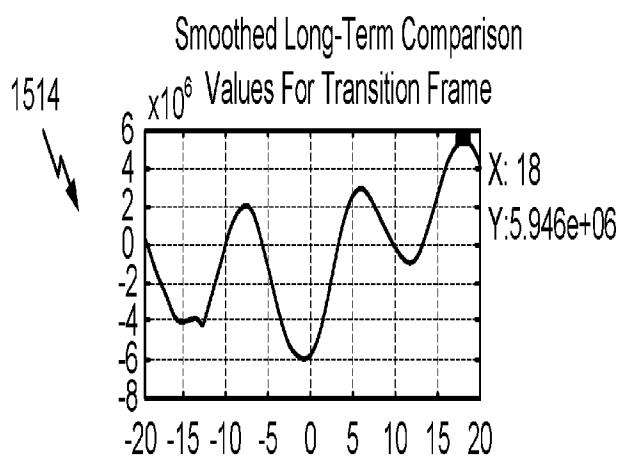
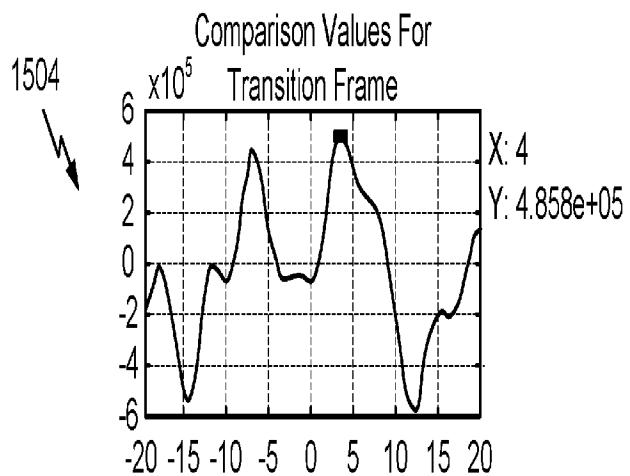
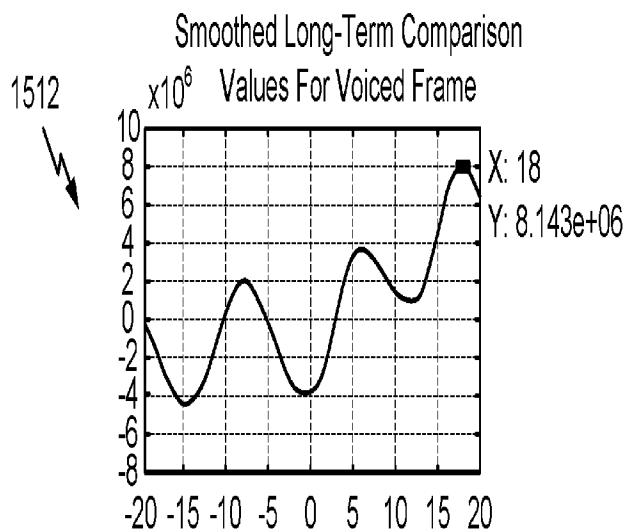
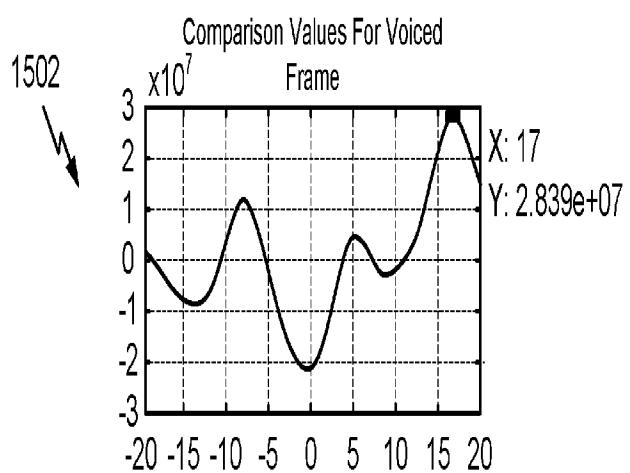
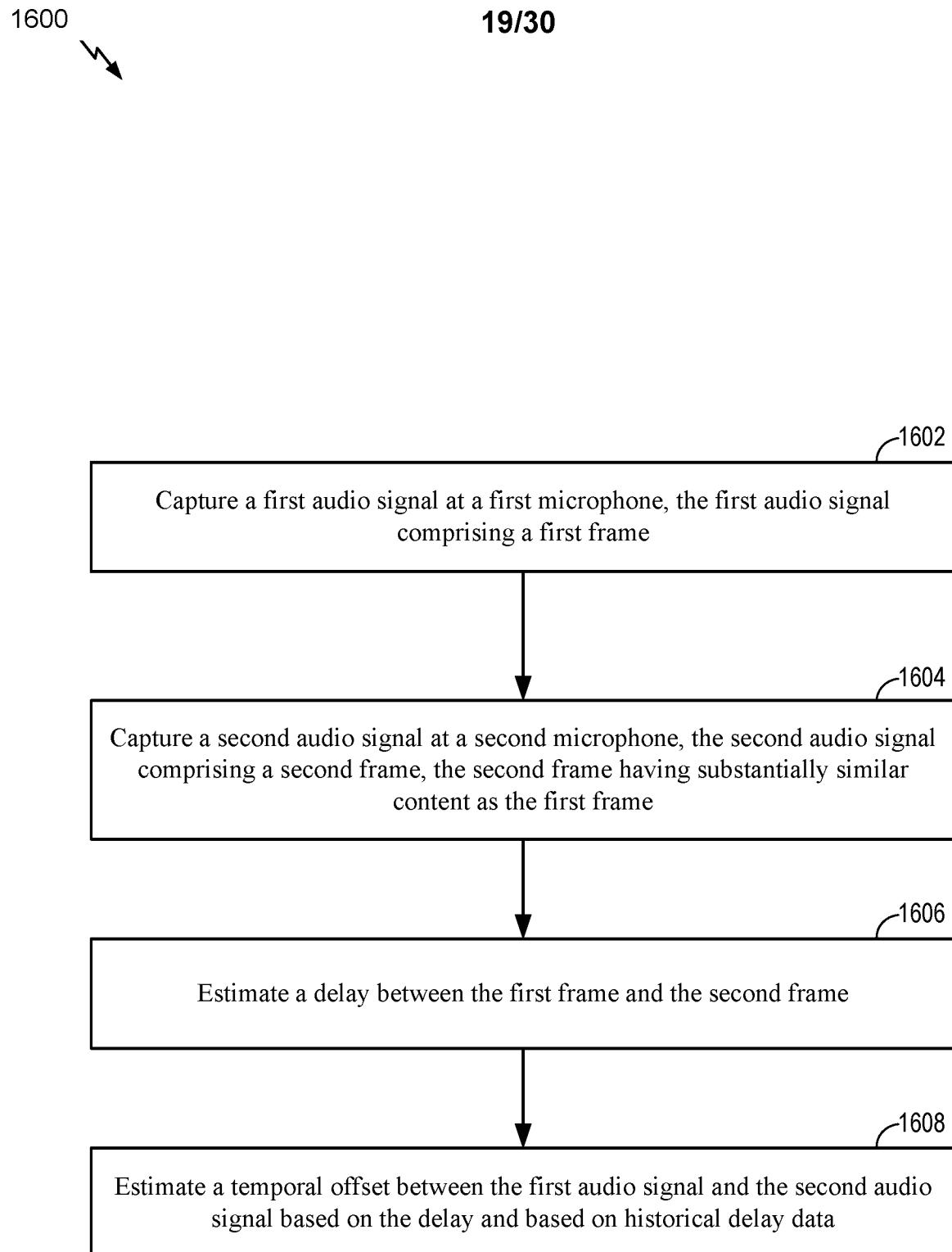


FIG. 15

**FIG. 16**

20/30

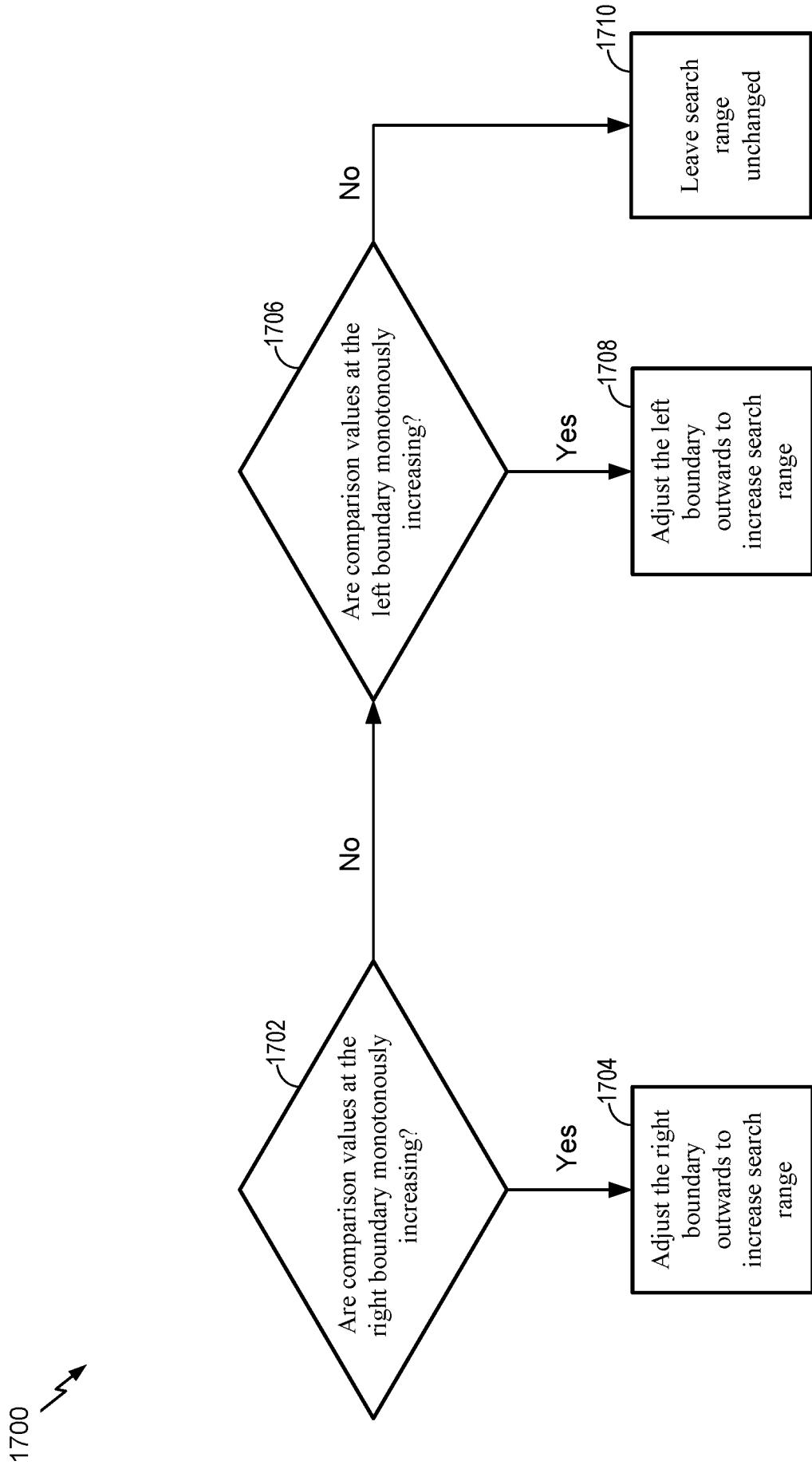
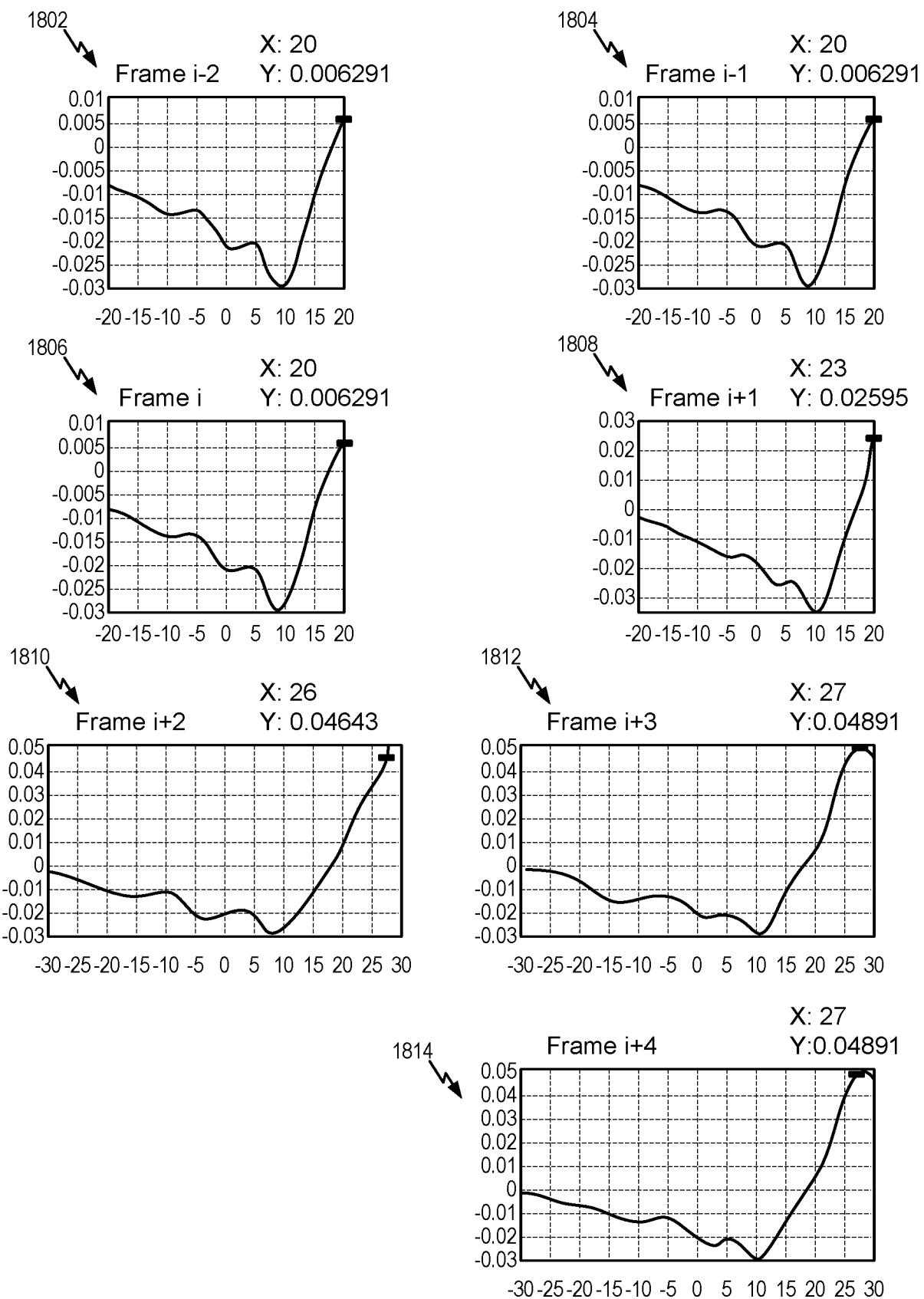


FIG. 17



**FIG. 18**  
SUBSTITUTE SHEET (RULE 26)

22/30

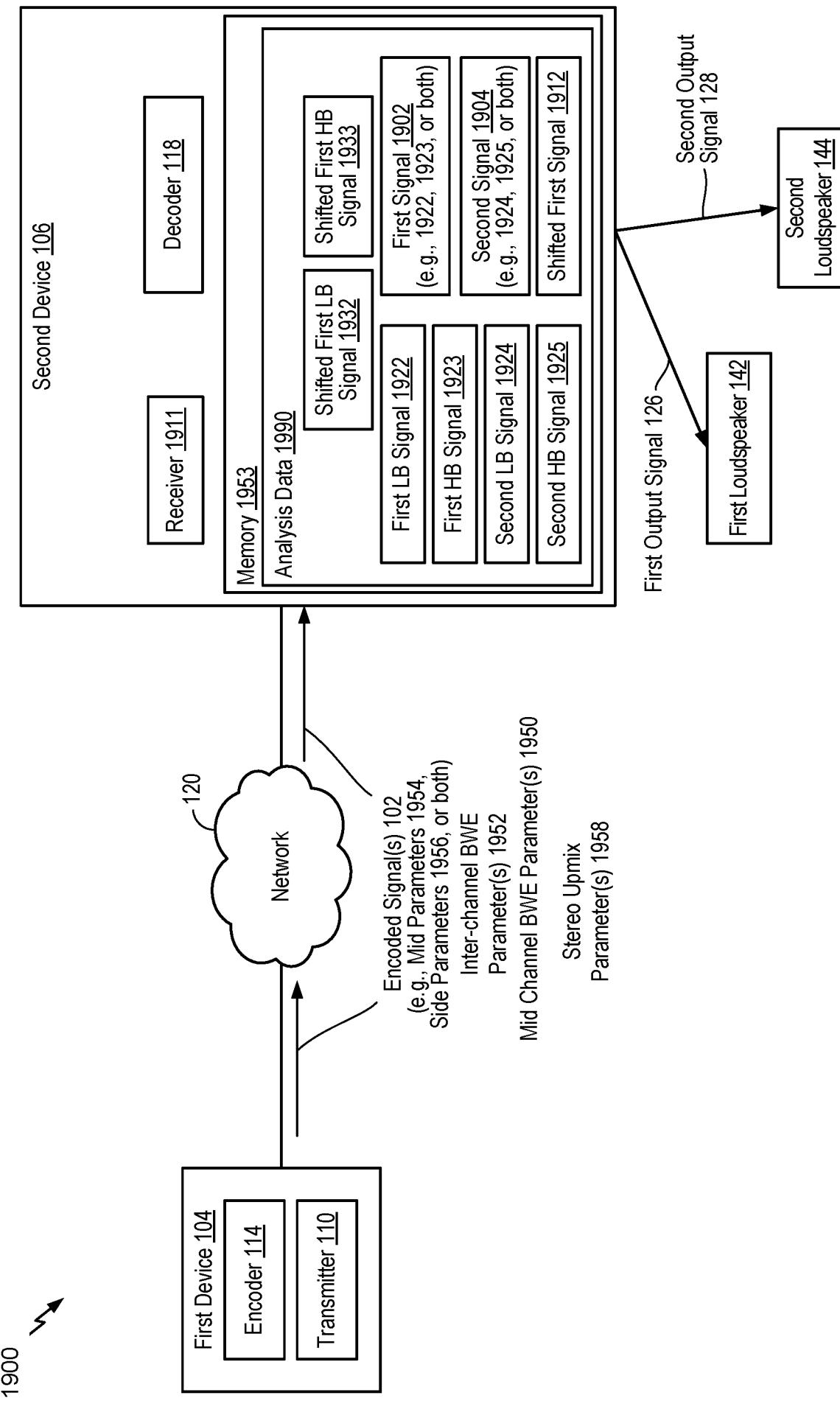


FIG. 19

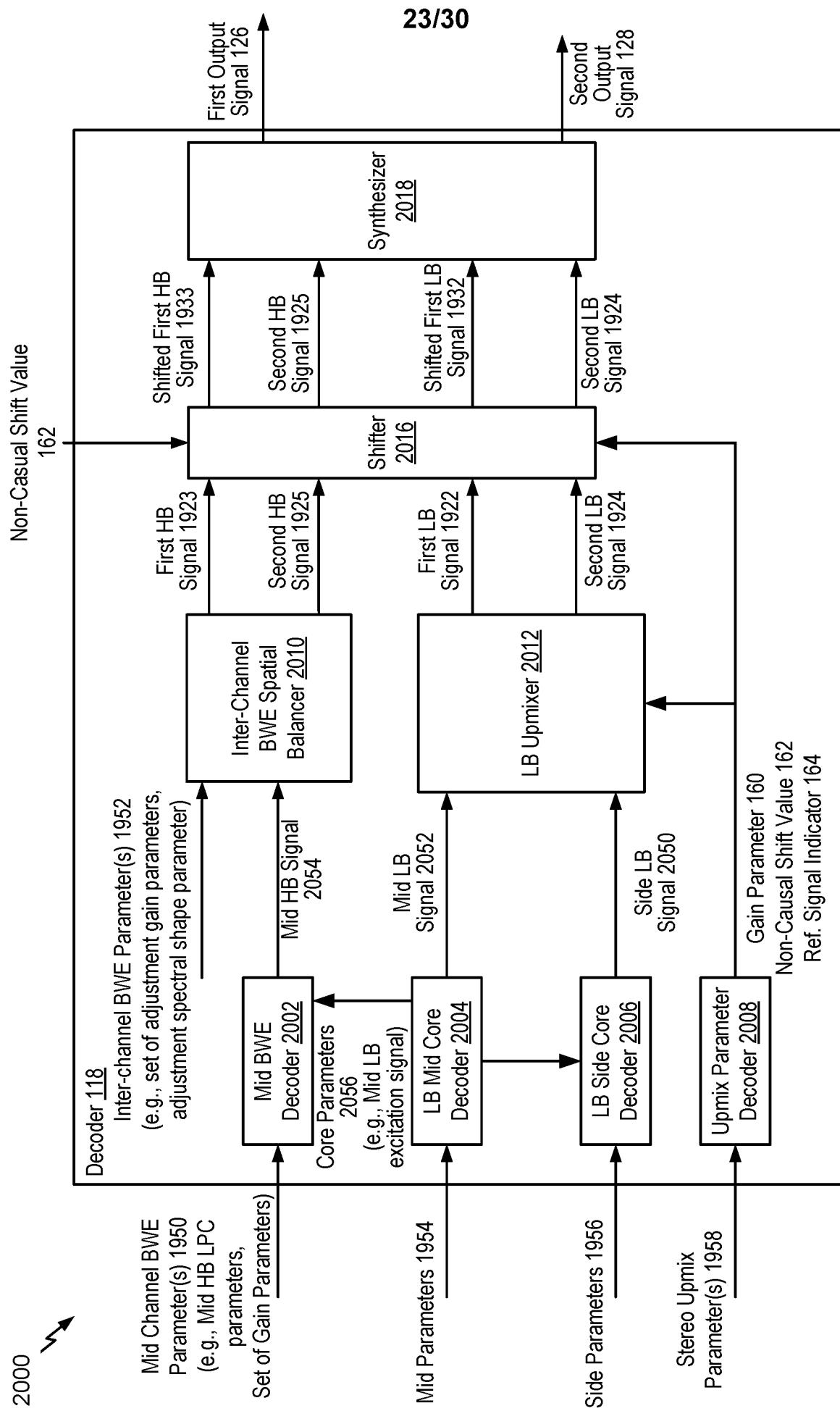


FIG. 20

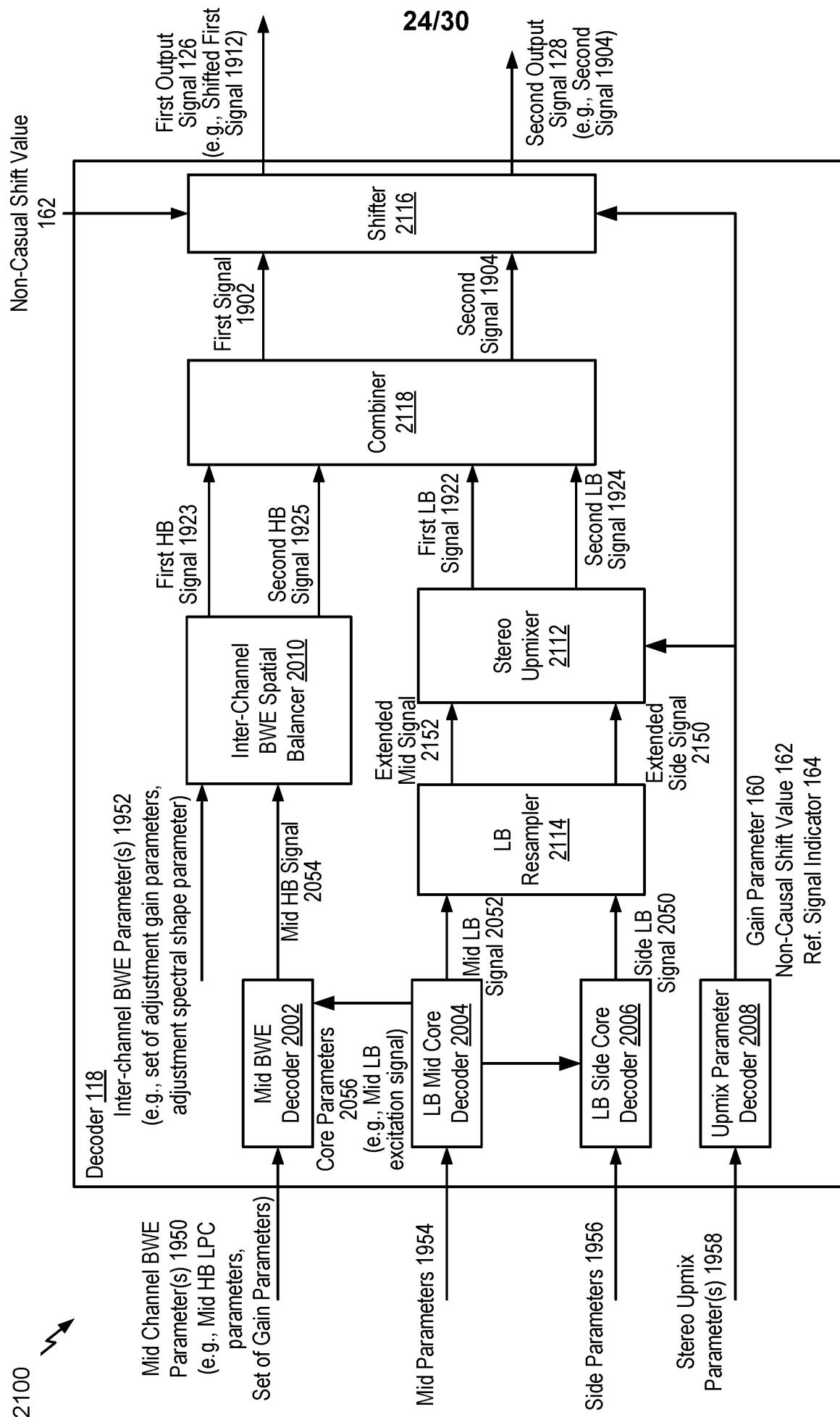
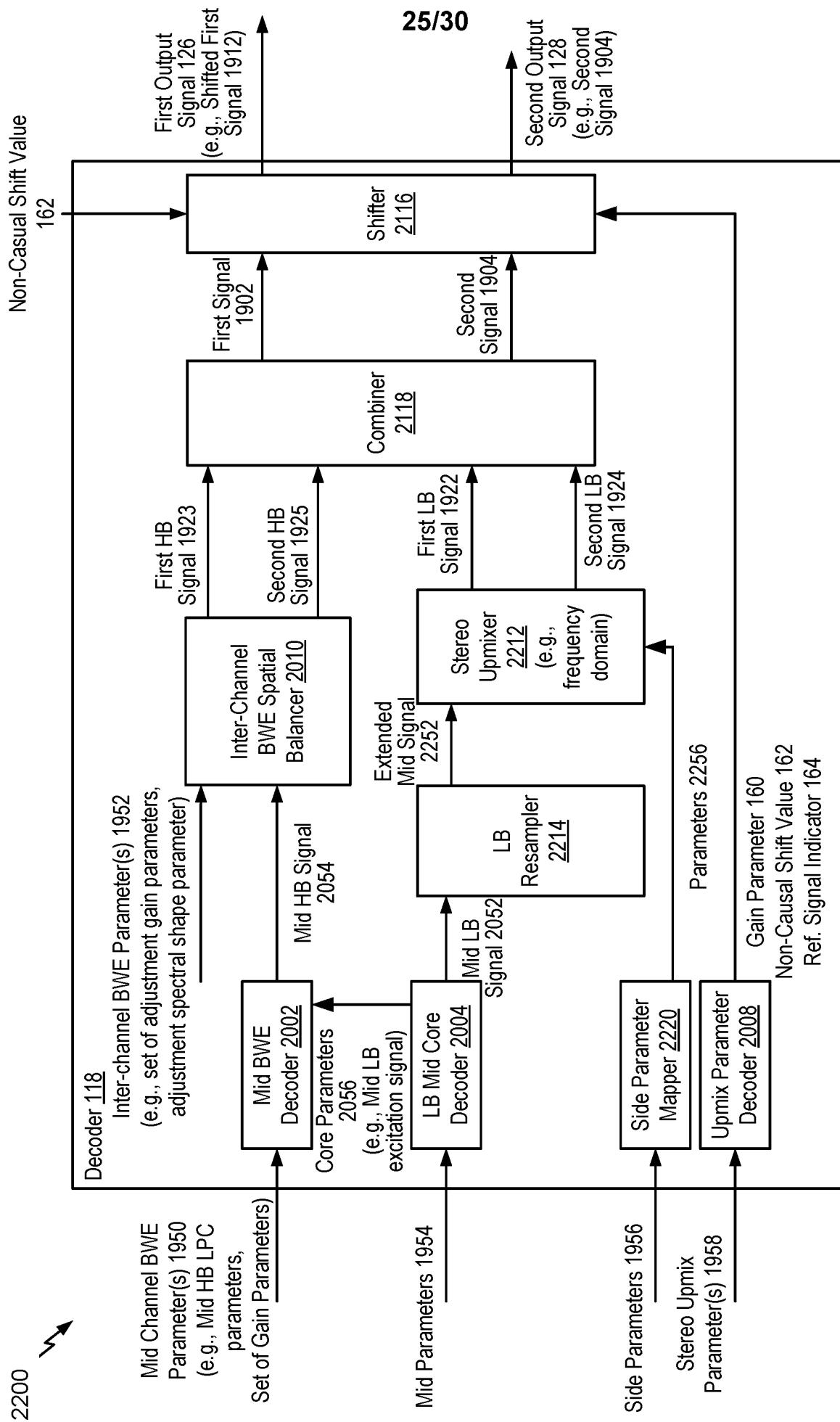
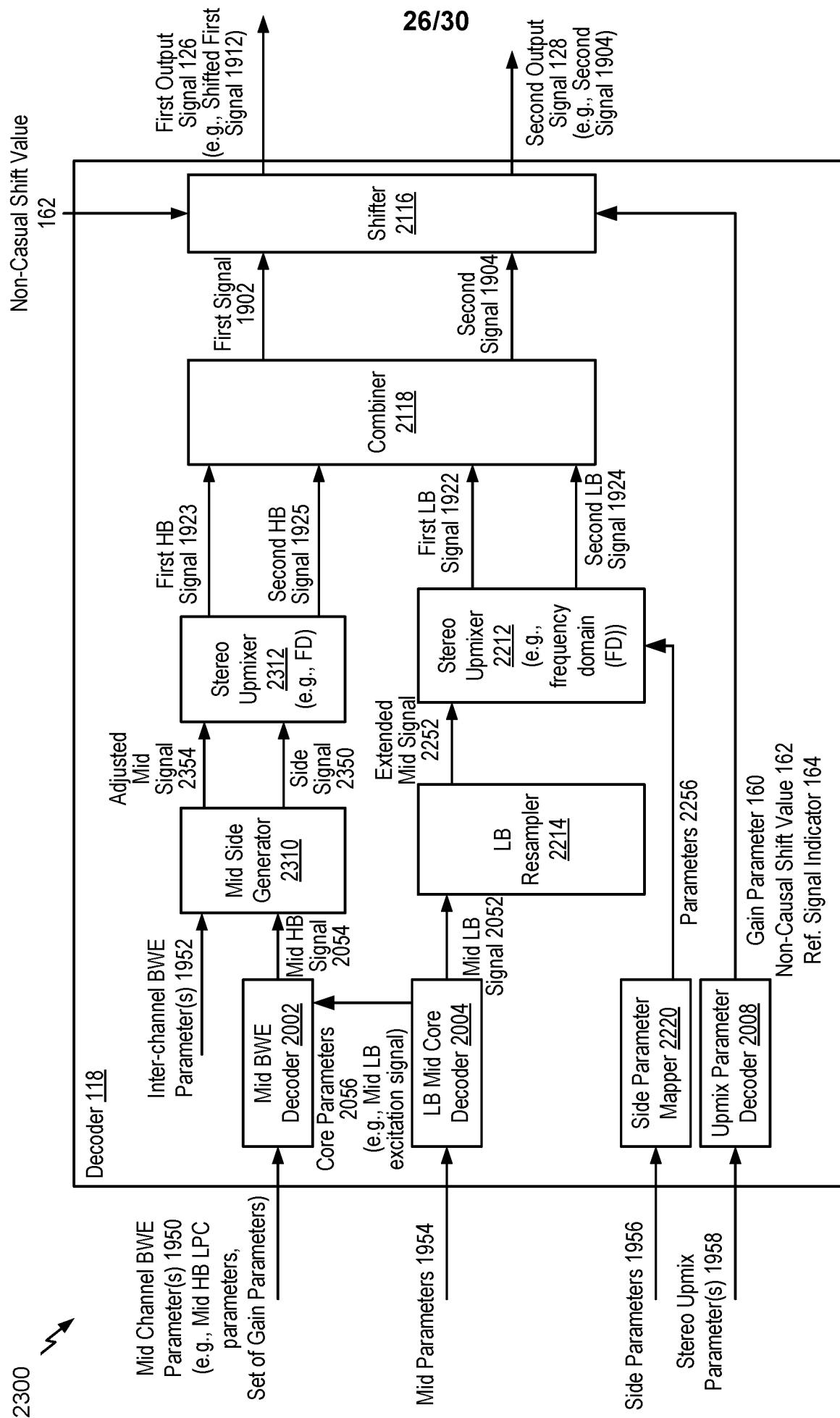
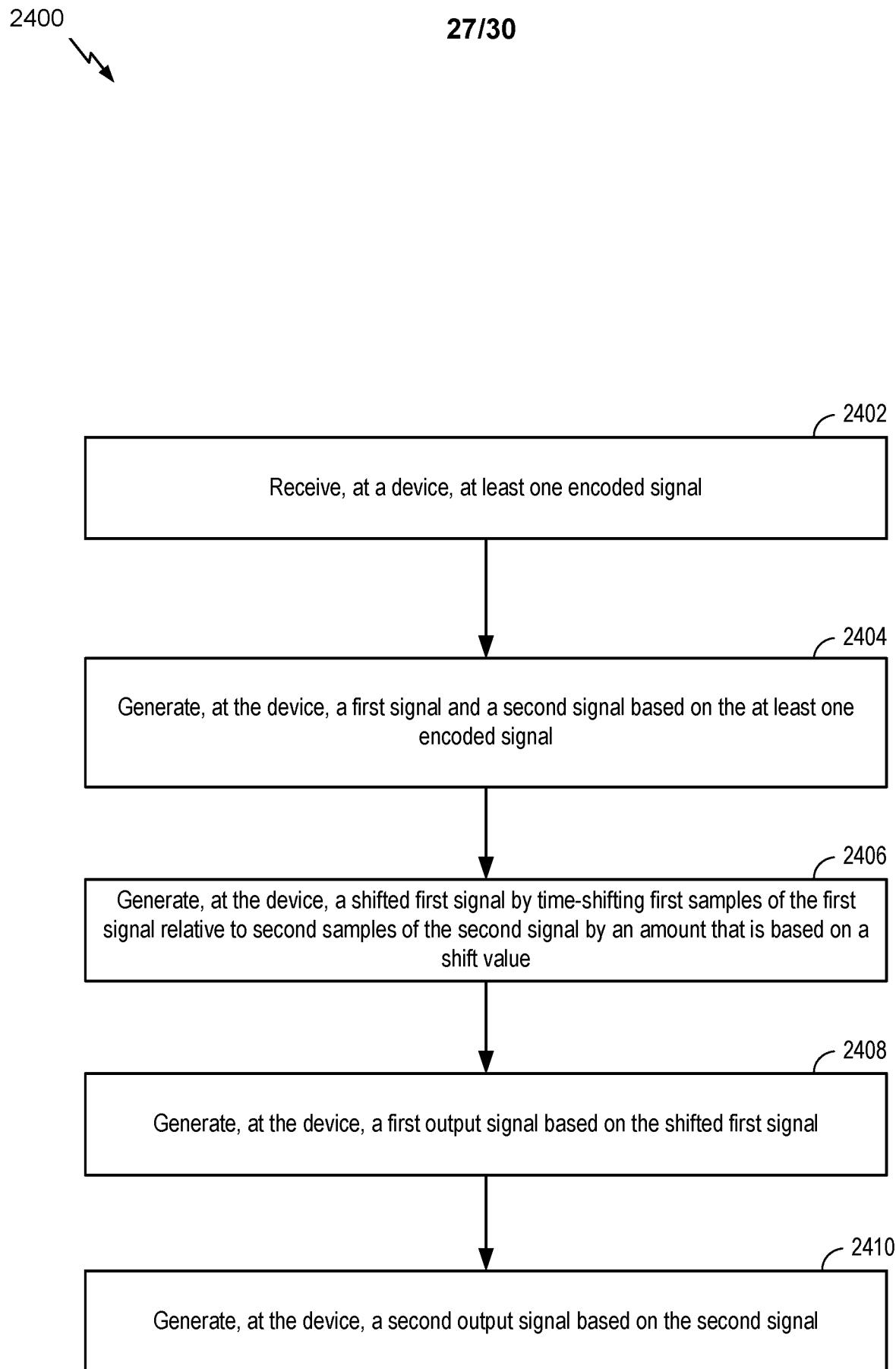
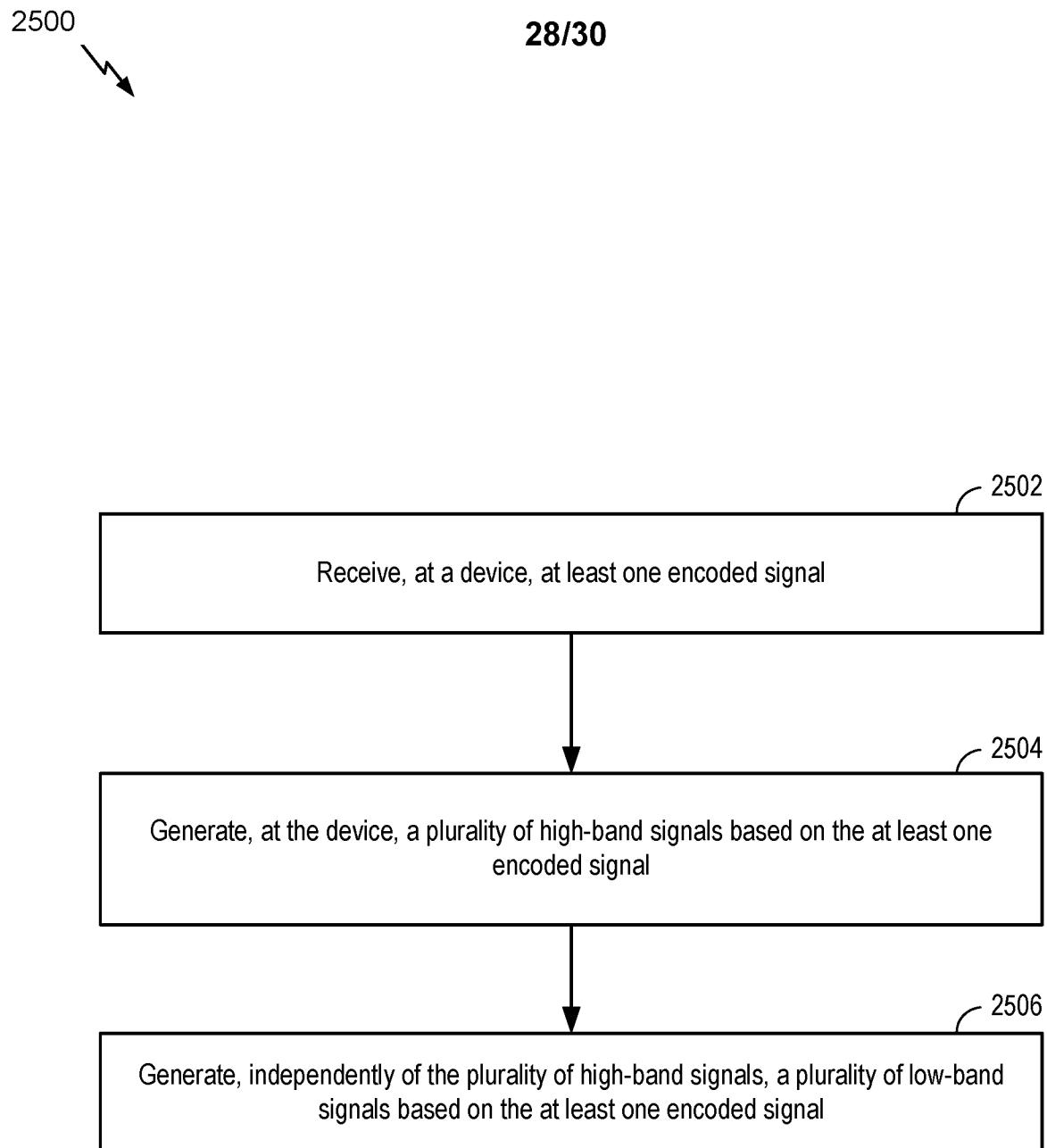


FIG. 21

**FIG. 22**

**FIG. 23**

**FIG. 24**

**FIG. 25**

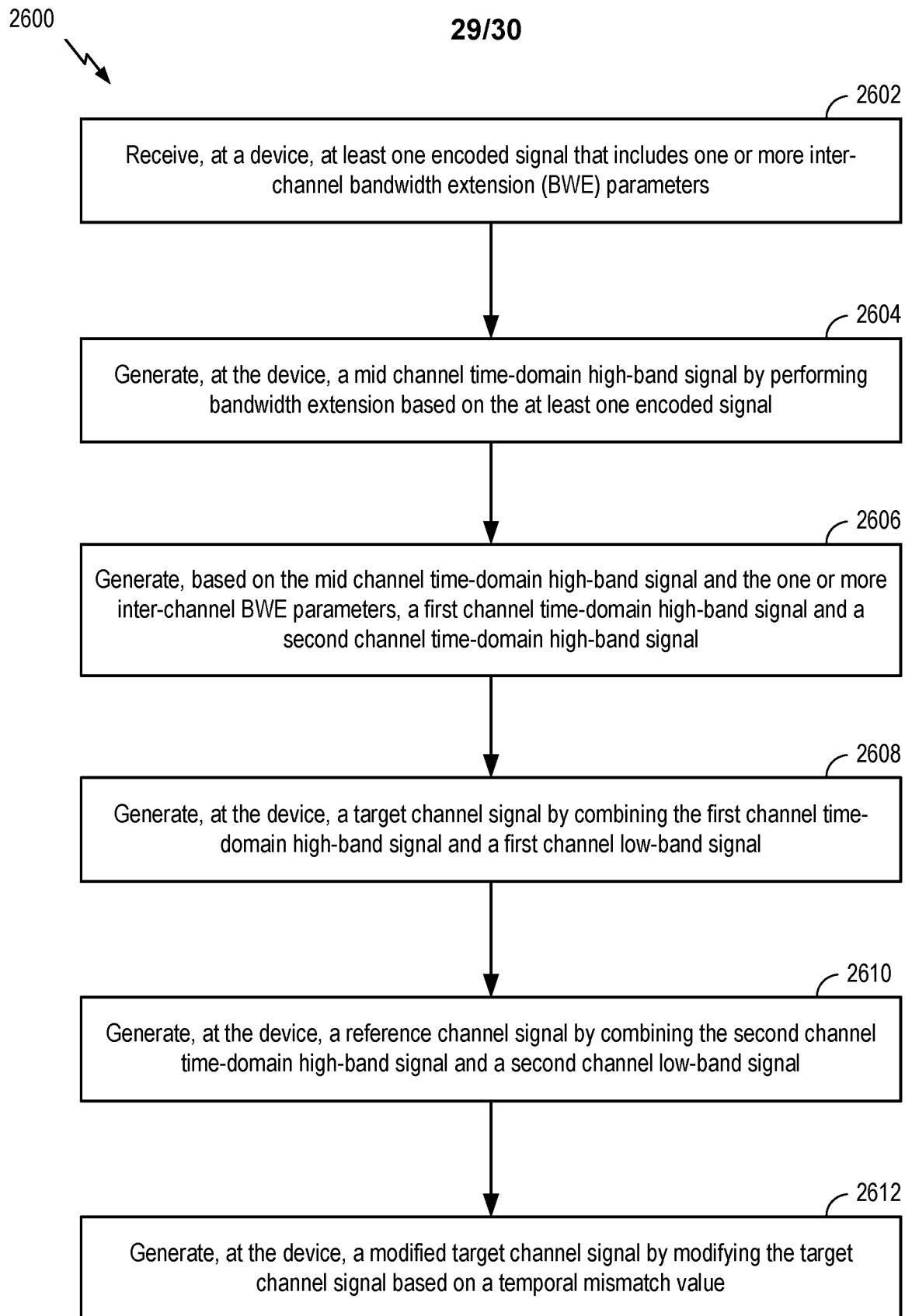


FIG. 26

30/30

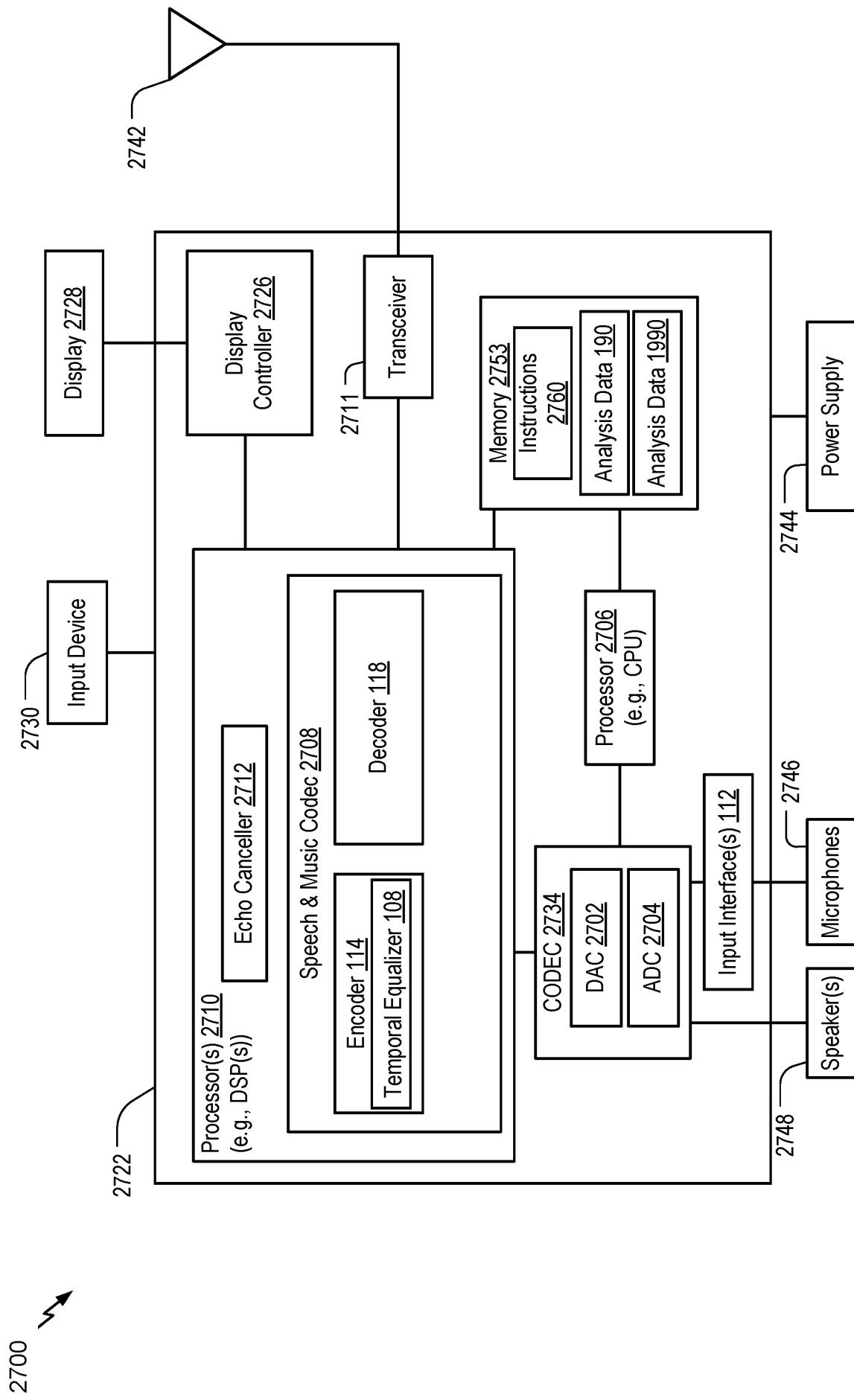


FIG. 27

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2017/023032

**A. CLASSIFICATION OF SUBJECT MATTER**  
 INV. G10L19/008 G10L19/04 G10L19/24 G10L19/02  
 ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009/325524 A1 (OH HYEN O [KR] ET AL) 31 December 2009 (2009-12-31) figures 1,4,7,8B paragraphs [0035], [0036], [0044] paragraphs [0047] - [0049], [0051] paragraphs [0069], [0079], [0080] paragraphs [0082], [0083], [0085] paragraphs [0086], [0100], [0108] paragraphs [0110], [0115] ----- A DVB ORGANIZATION: "29n154462.doc", DVB, DIGITAL VIDEO BROADCASTING, C/O EBU - 17A ANCIENNE ROUTE - CH-1218 GRAND SACONNEX, GENEVA - SWITZERLAND, 3 December 2015 (2015-12-03), XP017850098, pages 10-13 pages 128,129 pages 149-154 -----	1-33 1-33



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 May 2017	09/06/2017
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Tilp, Jan

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No  
PCT/US2017/023032

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009325524	A1 31-12-2009	EP 2124224 A1 US 2009325524 A1 WO 2009142465 A2	25-11-2009 31-12-2009 26-11-2009