Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2583146 A1 2006/05/04

(21) 2 583 146

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2005/09/12
(87) Date publication PCT/PCT Publication Date: 2006/05/04
(85) Entrée phase nationale/National Entry: 2007/04/03
(86) N° demande PCT/PCT Application No.: EP 2005/009784
(87) N° publication PCT/PCT Publication No.: 2006/045373
(30) Priorités/Priorities: 2004/10/20 (US60/620,401); 2004/12/07 (US11/006,492)

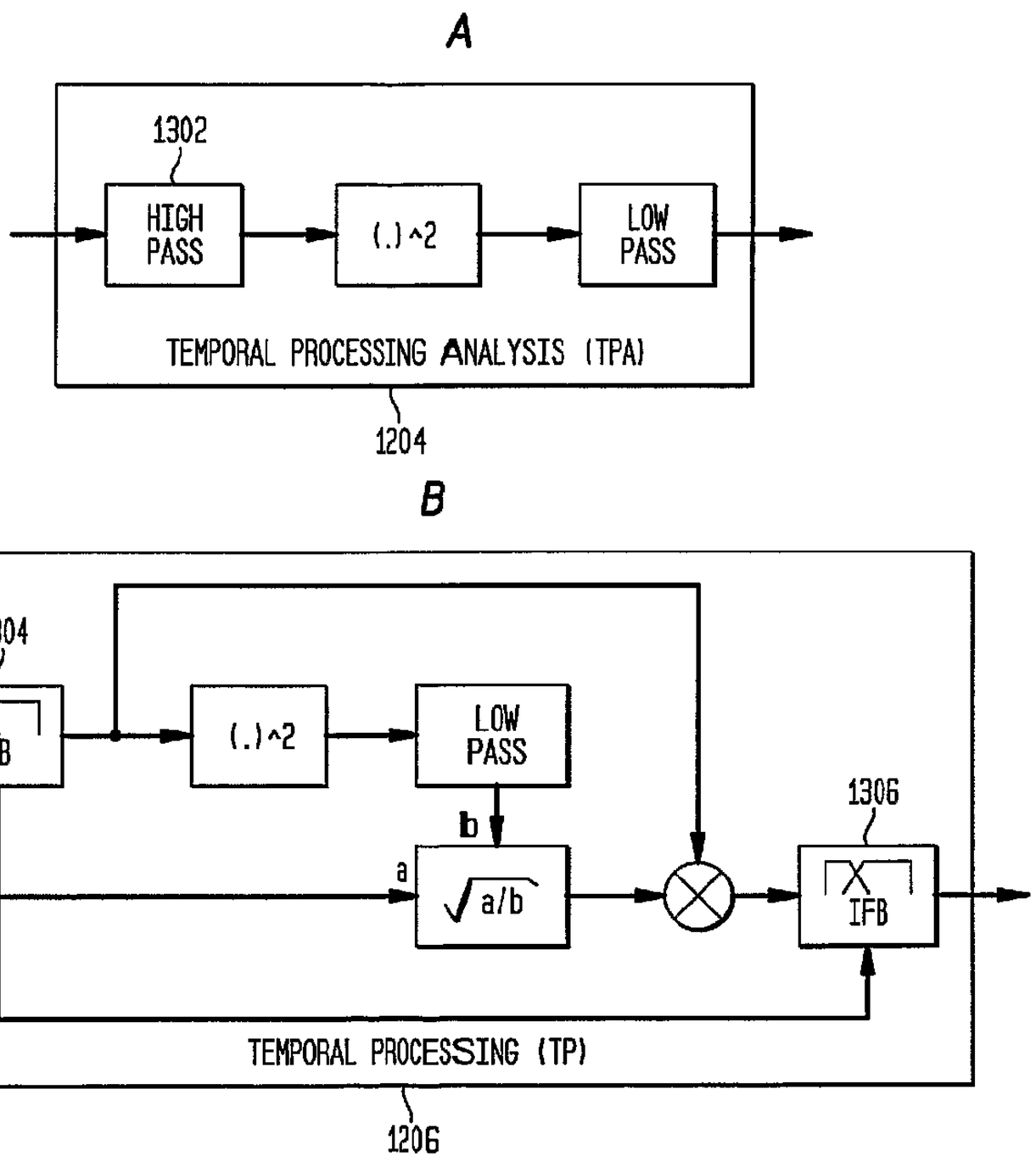
(51) Cl.Int./Int.Cl. *H04S 3/00* (2006.01), *G10L 19/00* (2006.01)

(71) Demandeurs/Applicants:
FRAUNHOFER-GELELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V., DE;
AGERE SYSTEMS INC., US

(72) Inventeurs/Inventors:
ALLAMANCHE, ERIC, US;
DISCH, SASCHA, DE;
FALLER, CHRISTOF, CH;
HERRE, JUERGEN, DE

(74) Agent: BCF LLP

(54) Titre : MISE EN FORME D'ENVELOPPE SONORE DIFFUSE POUR BCC ET ANALOGUE
(54) Title: DIFFUSE SOUND ENVELOPE SHAPING FOR BINAURAL CUE CODING SCHEMES AND THE LIKE



(57) Abrégé/Abstract:
An input audio signal having an input temporal envelope is converted into an output audio signal having an output temporal envelope. The input temporal envelope of the input audio signal is characterized. The input audio signal is processed to generate a processed audio signal, wherein the processing de-correlates the input audio signal. The processed audio signal is adjusted based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 May 2006 (04.05.2006)

PCT

(10) International Publication Number
WO 2006/045373 A1

(51) International Patent Classification:
H04S 3/00 (2006.01) *G10L 19/00* (2006.01)

(21) International Application Number:
PCT/EP2005/009784

(22) International Filing Date:
12 September 2005 (12.09.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/620,401	20 October 2004 (20.10.2004)	US
11/006,492	7 December 2004 (07.12.2004)	US

(71) Applicants *(for all designated States except US)*: **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastrasse 27c, 80686 Munich (DE). **AGERE SYSTEMS INC.** [US/US]; 1110 American Parkway NE, Allentown, Pennsylvania 18109 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ALLAMANCHE, Eric** [DE/US]; 6400 Christie Avenue, Apt. 3101, Emeryville, California 94608 (US). **DISCH, Sascha** [DE/DE]; Turnstrasse 7, 90763 Fürth (DE). **FALLER, Christof** [CH/CH]; Guetrain 1, CH-8274 Tägerwilen (CH). **HERRE, Jürgen** [DE/DE]; Hallerstrasse 24, 91054 Buckenhof (DE).

(74) Agents: **ZINKLER, Franz** et al.; SCHOPPE, ZIMMERMANN, STÖCKELER & ZINKLER, Postfach 246, 82043 Pullach bei München (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

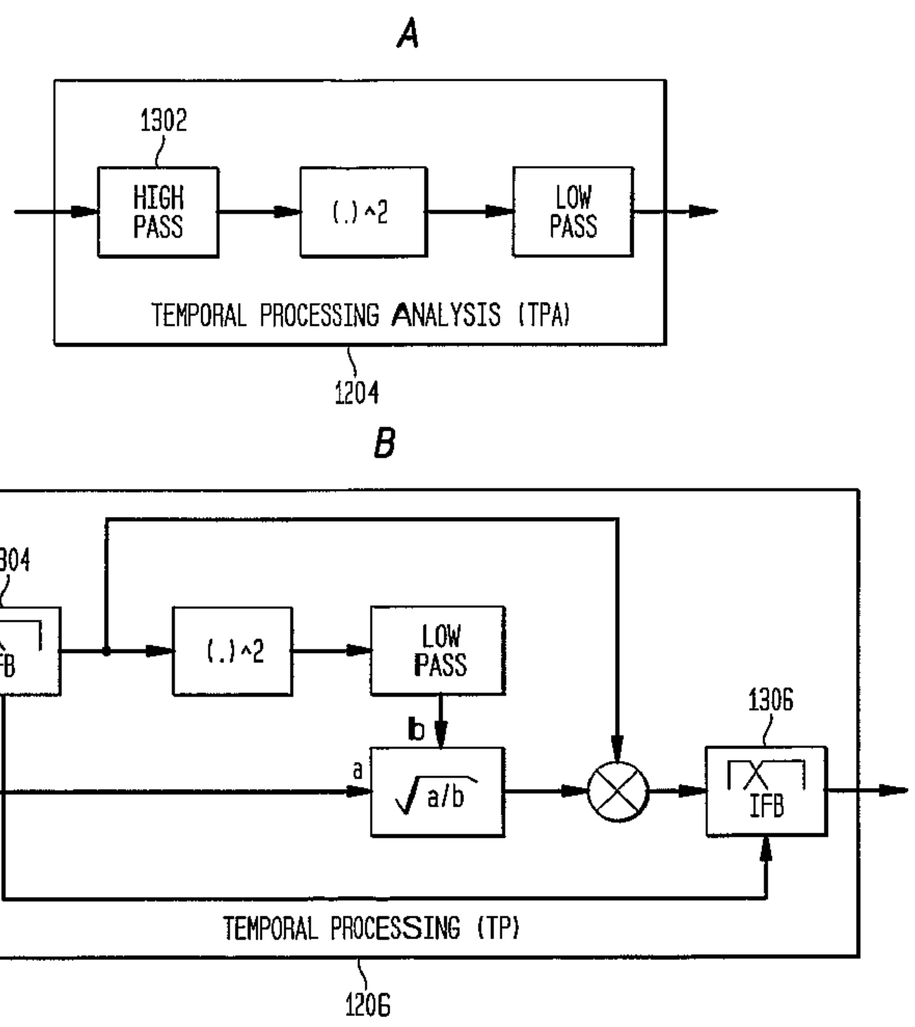
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

[Continued on next page]

(54) **Title:** DIFFUSE SOUND ENVELOPE SHAPING FOR BINAURAL CUE CODING SCHEMES AND THE LIKE



(57) **Abstract:** An input audio signal having an input temporal envelope is converted into an output audio signal having an output temporal envelope. The input temporal envelope of the input audio signal is characterized. The input audio signal is processed to generate a processed audio signal, wherein the processing de-correlates the input audio signal. The processed audio signal is adjusted based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

WO 2006/045373 A1

WO 2006/045373 A1

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DIFFUSE SOUND ENVELOPE SHAPING FOR BINAURAL CUE CODING SCHEMES AND THE LIKE

BACKGROUND OF THE INVENTION

Cross-Reference to Related Applications

This application claims the benefit of the filing date of U.S. provisional application no. 60/620,401, filed on 10/20/04 as attorney docket no. Allamanche 1-2-17-3, the teachings of which are incorporated herein by reference.

In addition, the subject matter of this application is related to the subject matter of the following U.S. applications , the teachings of all of which are incorporated herein by reference:

- U.S. application serial number 09/848,877, filed on 05/04/01 as attorney docket no. Faller 5;
- U.S. application serial number 10/045,458, filed on 11/07/01 as attorney docket no. Baumgarte 1-6-8, which itself claimed the benefit of the filing date of U.S. provisional application no. 60/311,565, filed on 08/10/01;
- U.S. application serial number 10/155,437, filed on 05/24/02 as attorney docket no. Baumgarte 2-10;
- U.S. application serial number 10/246,570, filed on 09/18/02 as attorney docket no. Baumgarte 3-11;
- U.S. application serial number 10/815,591, filed on 04/01/04 as attorney docket no. Baumgarte 7-12;
- U.S. application serial number 10/936,464, filed on 09/08/04 as attorney docket no. Baumgarte 8-7-15;
- U.S. application serial number 10/762,100, filed on 01/20/04 (Faller 13-1); and
- U.S. application serial number 10/xxx,xxx, filed on the same date as this application as attorney docket no. Allamanche 2-3-18-4.

The subject matter of this application is also related to subject matter described in the following papers, the teachings of all of which are incorporated herein by reference:

- F. Baumgarte and C. Faller, "Binaural Cue Coding - Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003;
- C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003; and
- C. Faller, "Coding of spatial audio compatible with different playback formats," *Preprint 117th Conv. Aud. Eng. Soc.*, October 2004.

Field of the Invention

The present invention relates to the encoding of audio signals and the subsequent synthesis of auditory scenes from the encoded audio data.

Description of the Related Art

When a person hears an audio signal (i.e., sounds) generated by a particular audio source, the audio signal will typically arrive at the person's left and right ears at two different times and with two different audio (e.g., decibel) levels, where those different times and levels are functions of the differences in the paths through which the audio signal travels to reach the left and right ears, respectively. The person's brain interprets these differences in time and level to give the person the perception that the received audio signal is being generated by an audio source located at a particular position (e.g., direction and distance) relative to the person. An auditory scene is the net effect of a person simultaneously hearing audio signals generated by one or more different audio sources located at one or more different positions relative to the person.

The existence of this processing by the brain can be used to synthesize auditory scenes, where audio signals from one or more different audio sources are purposefully modified to generate left and right audio signals that give the perception that the different audio sources are located at different positions relative to the listener.

Fig. 1 shows a high-level block diagram of conventional binaural signal synthesizer **100**, which converts a single audio source signal (e.g., a mono signal) into the left and right audio signals of a binaural signal, where a binaural signal is defined to be the two signals received at the eardrums of a listener. In addition to the audio source signal, synthesizer **100** receives a set of spatial cues corresponding to the desired position of the audio source relative to the listener. In typical implementations, the set of spatial cues comprises an inter-channel level difference (ICLD) value (which identifies the difference in audio level between the left and right audio signals as received at the left and right ears, respectively) and an inter-channel time difference (ICTD) value (which identifies the difference in time of arrival between the left and right audio signals as received at the left and right ears, respectively). In addition or as an alternative, some synthesis techniques involve the modeling of a direction-dependent transfer function for sound from the signal source to the eardrums, also referred to as the head-related transfer function (HRTF). See, e.g., J. Blauert, *The Psychophysics of Human Sound Localization*, MIT Press, 1983, the teachings of which are incorporated herein by reference.

Using binaural signal synthesizer **100** of Fig. 1, the mono audio signal generated by a single sound source can be processed such that, when listened to over headphones, the sound source is spatially placed by applying an appropriate set of spatial cues (e.g., ICLD, ICTD, and/or HRTF) to generate the

audio signal for each ear. See, e.g., D.R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA, 1994.

Binaural signal synthesizer **100** of Fig. 1 generates the simplest type of auditory scenes: those having a single audio source positioned relative to the listener. More complex auditory scenes comprising two or more audio sources located at different positions relative to the listener can be generated using an auditory scene synthesizer that is essentially implemented using multiple instances of binaural signal synthesizer, where each binaural signal synthesizer instance generates the binaural signal corresponding to a different audio source. Since each different audio source has a different location relative to the listener, a different set of spatial cues is used to generate the binaural audio signal for each different audio source.

SUMMARY OF THE INVENTION

According to one embodiment, the present invention is a method and apparatus for converting an input audio signal having an input temporal envelope into an output audio signal having an output temporal envelope. The input temporal envelope of the input audio signal is characterized. The input audio signal is processed to generate a processed audio signal, wherein the processing de-correlates the input audio signal. The processed audio signal is adjusted based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

According to another embodiment, the present invention is a method and apparatus for encoding C input audio channels to generate E transmitted audio channel(s). One or more cue codes are generated for two or more of the C input channels. The C input channels are downmixed to generate the E transmitted channel(s), where $C > E \geq 1$. One or more of the C input channels and the E transmitted channel(s) are analyzed to generate a flag indicating whether or not a decoder of the E transmitted channel(s) should perform envelope shaping during decoding of the E transmitted channel(s).

According to another embodiment, the present invention is an encoded audio bitstream generated by the method of the previous paragraph.

According to another embodiment, the present invention is an encoded audio bitstream comprising E transmitted channel(s), one or more cue codes, and a flag. The one or more cue codes are generated by generating one or more cue codes for two or more of the C input channels. The E transmitted channel(s) are generated by downmixing the C input channels, where $C > E \geq 1$. The flag is generated by analyzing one or more of the C input channels and the E transmitted channel(s), wherein the flag indicates whether or not a decoder of the E transmitted channel(s) should perform envelope shaping during decoding of the E transmitted channel(s).

BRIEF DESCRIPTION OF THE DRAWINGS

Other aspects, features, and advantages of the present invention will become more fully apparent from the following detailed description, the appended claims, and the accompanying drawings in which like reference numerals identify similar or identical elements.

Fig. 1 shows a high-level block diagram of conventional binaural signal synthesizer;

Fig. 2 is a block diagram of a generic binaural cue coding (BCC) audio processing system;

Fig. 3 shows a block diagram of a downmixer that can be used for the downmixer of Fig. 2;

Fig. 4 shows a block diagram of a BCC synthesizer that can be used for the decoder of Fig. 2;

Fig. 5 shows a block diagram of the BCC estimator of Fig. 2, according to one embodiment of the present invention;

Fig. 6 illustrates the generation of ICTD and ICLD data for five-channel audio;

Fig. 7 illustrates the generation of ICC data for five-channel audio;

Fig. 8 shows a block diagram of an implementation of the BCC synthesizer of Fig. 4 that can be used in a BCC decoder to generate a stereo or multi-channel audio signal given a single transmitted sum signal $s(n)$ plus the spatial cues;

Fig. 9 illustrates how ICTD and ICLD are varied within a subband as a function of frequency;

Fig. 10 shows a block diagram representing at least a portion of a BCC decoder, according to one embodiment of the present invention;

Fig. 11 illustrates an exemplary application of the envelope shaping scheme of Fig. 10 in the context of the BCC synthesizer of Fig. 4;

Fig. 12 illustrates an alternative exemplary application of the envelope shaping scheme of Fig. 10 in the context of the BCC synthesizer of Fig. 4, where envelope shaping is applied to in the time domain;

Figs. 13(a) and (b) show possible implementations of the TPA and the TP of Fig. 12, where envelope shaping is applied only at frequencies higher than the cut-off frequency f_{TP};

Fig. 14 illustrates an exemplary application of the envelope shaping scheme of Fig. 10 in the context of the late reverberation-based ICC synthesis scheme described in U.S. application serial number 10/815,591, filed on 04/01/04 as attorney docket no. Baumgarte 7-12;

Fig. 15 shows a block diagram representing at least a portion of a BCC decoder, according to an embodiment of the present invention that is an alternative to the scheme shown in Fig. 10;

Fig. 16 shows a block diagram representing at least a portion of a BCC decoder, according to an embodiment of the present invention that is an alternative to the schemes shown in Figs. 10 and 15;

Fig. 17 illustrates an exemplary application of the envelope shaping scheme of Fig. 15 in the context of the BCC synthesizer of Fig. 4; and

Figs. 18(a)-(c) show block diagrams of possible implementations of the TPA, ITP, and TP of Fig. 17.

DETAILED DESCRIPTION

In binaural cue coding (BCC), an encoder encodes C input audio channels to generate E transmitted audio channels, where $C>E\geq 1$. In particular, two or more of the C input channels are provided in a frequency domain, and one or more cue codes are generated for each of one or more different frequency bands in the two or more input channels in the frequency domain. In addition, the C input channels are downmixed to generate the E transmitted channels. In some downmixing implementations, at least one of the E transmitted channels is based on two or more of the C input channels, and at least one of the E transmitted channels is based on only a single one of the C input channels.

In one embodiment, a BCC coder has two or more filter banks, a code estimator, and a downmixer. The two or more filter banks convert two or more of the C input channels from a time domain into a frequency domain. The code estimator generates one or more cue codes for each of one or more different frequency bands in the two or more converted input channels. The downmixer downmixes the C input channels to generate the E transmitted channels, where $C>E\geq 1$.

In BCC decoding, E transmitted audio channels are decoded to generate C playback audio channels. In particular, for each of one or more different frequency bands, one or more of the E transmitted channels are upmixed in a frequency domain to generate two or more of the C playback channels in the frequency domain, where $C>E\geq 1$. One or more cue codes are applied to each of the one or more different frequency bands in the two or more playback channels in the frequency domain to generate two or more modified channels, and the two or more modified channels are converted from the frequency domain into a time domain. In some upmixing implementations, at least one of the C playback channels is based on at least one of the E transmitted channels and at least one cue code, and at least one of the C playback channels is based on only a single one of the E transmitted channels and independent of any cue codes.

In one embodiment, a BCC decoder has an upmixer, a synthesizer, and one or more inverse filter banks. For each of one or more different frequency bands, the upmixer upmixes one or more of the E transmitted channels in a frequency domain to generate two or more of the C playback channels in the frequency domain, where $C>E\geq 1$. The synthesizer applies one or more cue codes to each of the one or more different frequency bands in the two or more playback channels in the frequency domain to generate two or more modified channels. The one or more inverse filter banks convert the two or more modified channels from the frequency domain into a time domain.

Depending on the particular implementation, a given playback channel may be based on a single transmitted channel, rather than a combination of two or more transmitted channels. For example, when there is only one transmitted channel, each of the *C* playback channels is based on that one transmitted channel. In these situations, upmixing corresponds to copying of the corresponding transmitted channel. As such, for applications in which there is only one transmitted channel, the upmixer may be implemented using a replicator that copies the transmitted channel for each playback channel.

BCC encoders and/or decoders may be incorporated into a number of systems or applications including, for example, digital video recorders/players, digital audio recorders/players, computers, satellite transmitters/receivers, cable transmitters/receivers, terrestrial broadcast transmitters/receivers, home entertainment systems, and movie theater systems.

Generic BCC Processing

Fig. 2 is a block diagram of a generic binaural cue coding (BCC) audio processing system **200** comprising an encoder **202** and a decoder **204**. Encoder **202** includes downmixer **206** and BCC estimator **208**.

Downmixer **206** converts *C* input audio channels $x_i(n)$ into *E* transmitted audio channels $y_i(n)$, where $C>E\geq 1$. In this specification, signals expressed using the variable *n* are time-domain signals, while signals expressed using the variable *k* are frequency-domain signals. Depending on the particular implementation, downmixing can be implemented in either the time domain or the frequency domain. BCC estimator **208** generates BCC codes from the *C* input audio channels and transmits those BCC codes as either in-band or out-of-band side information relative to the *E* transmitted audio channels. Typical BCC codes include one or more of inter-channel time difference (ICTD), inter-channel level difference (ICLD), and inter-channel correlation (ICC) data estimated between certain pairs of input channels as a function of frequency and time. The particular implementation will dictate between which particular pairs of input channels, BCC codes are estimated.

ICC data corresponds to the coherence of a binaural signal, which is related to the perceived width of the audio source. The wider the audio source, the lower the coherence between the left and right channels of the resulting binaural signal. For example, the coherence of the binaural signal corresponding to an orchestra spread out over an auditorium stage is typically lower than the coherence of the binaural signal corresponding to a single violin playing solo. In general, an audio signal with lower coherence is usually perceived as more spread out in auditory space. As such, ICC data is typically related to the apparent source width and degree of listener envelopment. See, e.g., J. Blauert, *The Psychophysics of Human Sound Localization*, MIT Press, 1983.

Depending on the particular application, the E transmitted audio channels and corresponding BCC codes may be transmitted directly to decoder **204** or stored in some suitable type of storage device for subsequent access by decoder **204**. Depending on the situation, the term "transmitting" may refer to either direct transmission to a decoder or storage for subsequent provision to a decoder. In either case, decoder **204** receives the transmitted audio channels and side information and performs upmixing and BCC synthesis using the BCC codes to convert the E transmitted audio channels into more than E (typically, but not necessarily, C) playback audio channels $\hat{x}_i(n)$ for audio playback. Depending on the particular implementation, upmixing can be performed in either the time domain or the frequency domain.

In addition to the BCC processing shown in Fig. 2, a generic BCC audio processing system may include additional encoding and decoding stages to further compress the audio signals at the encoder and then decompress the audio signals at the decoder, respectively. These audio codecs may be based on conventional audio compression/decompression techniques such as those based on pulse code modulation (PCM), differential PCM (DPCM), or adaptive DPCM (ADPCM).

When downmixer **206** generates a single sum signal (i.e., E=1), BCC coding is able to represent multi-channel audio signals at a bitrate only slightly higher than what is required to represent a mono audio signal. This is so, because the estimated ICTD, ICLD, and ICC data between a channel pair contain about two orders of magnitude less information than an audio waveform.

Not only the low bitrate of BCC coding, but also its backwards compatibility aspect is of interest. A single transmitted sum signal corresponds to a mono downmix of the original stereo or multi-channel signal. For receivers that do not support stereo or multi-channel sound reproduction, listening to the transmitted sum signal is a valid method of presenting the audio material on low-profile mono reproduction equipment. BCC coding can therefore also be used to enhance existing services involving the delivery of mono audio material towards multi-channel audio. For example, existing mono audio radio broadcasting systems can be enhanced for stereo or multi-channel playback if the BCC side information can be embedded into the existing transmission channel. Analogous capabilities exist when downmixing multi-channel audio to two sum signals that correspond to stereo audio.

BCC processes audio signals with a certain time and frequency resolution. The frequency resolution used is largely motivated by the frequency resolution of the human auditory system. Psychoacoustics suggests that spatial perception is most likely based on a critical band representation of the acoustic input signal. This frequency resolution is considered by using an invertible filterbank (e.g., based on a fast Fourier transform (FFT) or a quadrature mirror filter (QMF)) with subbands with bandwidths equal or proportional to the critical bandwidth of the human auditory system.

Generic Downmixing

In preferred implementations, the transmitted sum signal(s) contain all signal components of the input audio signal. The goal is that each signal component is fully maintained. Simply summation of the audio input channels often results in amplification or attenuation of signal components. In other words, the power of the signal components in a "simple" sum is often larger or smaller than the sum of the power of the corresponding signal component of each channel. A downmixing technique can be used that equalizes the sum signal such that the power of signal components in the sum signal is approximately the same as the corresponding power in all input channels.

Fig. 3 shows a block diagram of a downmixer **300** that can be used for downmixer **206** of Fig. 2 according to certain implementations of BCC system **200**. Downmixer **300** has a filter bank (FB) **302** for each input channel $x_i(n)$, a downmixing block **304**, an optional scaling/delay block **306**, and an inverse FB (IFB) **308** for each encoded channel $y_i(n)$.

Each filter bank **302** converts each frame (e.g., 20 msec) of a corresponding digital input channel $x_i(n)$ in the time domain into a set of input coefficients $\tilde{x}_i(k)$ in the frequency domain. Downmixing block **304** downmixes each sub-band of C corresponding input coefficients into a corresponding sub-band of E downmixed frequency-domain coefficients. Equation (1) represents the downmixing of the kth sub-band of input coefficients $\left(\tilde{x}_1(k), \tilde{x}_2(k), \ldots, \tilde{x}_C(k)\right)$ to generate the kth sub-band of downmixed coefficients $\left(\hat{y}_1(k), \hat{y}_2(k), \ldots, \hat{y}_E(k)\right)$ as follows:

$$\begin{bmatrix} \hat{y}_1(k) \\ \hat{y}_2(k) \\ \vdots \\ \hat{y}_E(k) \end{bmatrix} = \mathbf{D}_{CE} \begin{bmatrix} \tilde{x}_1(k) \\ \tilde{x}_2(k) \\ \vdots \\ \tilde{x}_C(k) \end{bmatrix}, \tag{1}$$

where $\mathbf{D}_{CE}$ is a real-valued C-by–E downmixing matrix.

Optional scaling/delay block **306** comprises a set of multipliers **310**, each of which multiplies a corresponding downmixed coefficient $\hat{y}_i(k)$ by a scaling factor $e_i(k)$ to generate a corresponding scaled coefficient $\tilde{y}_i(k)$. The motivation for the scaling operation is equivalent to equalization generalized for

downmixing with arbitrary weighting factors for each channel. If the input channels are independent, then the power $p_{\tilde{y}_i(k)}$ of the downmixed signal in each sub-band is given by Equation (2) as follows:

$$\begin{bmatrix} p_{\tilde{y}_1(k)} \\ p_{\tilde{y}_2(k)} \\ \vdots \\ p_{\tilde{y}_E(k)} \end{bmatrix} = \overline{\mathbf{D}}_{CE} \begin{bmatrix} p_{\tilde{x}_1(k)} \\ p_{\tilde{x}_2(k)} \\ \vdots \\ p_{\tilde{x}_C(k)} \end{bmatrix}, \tag{2}$$

where $\overline{\mathbf{D}}_{CE}$ is derived by squaring each matrix element in the *C*-by-*E* downmixing matrix $\mathbf{D}_{CE}$ and $p_{\tilde{x}_i(k)}$ is the power of sub-band k of input channel i.

If the sub-bands are not independent, then the power values $p_{\tilde{y}_i(k)}$ of the downmixed signal will be larger or smaller than that computed using Equation (2), due to signal amplifications or cancellations when signal components are in-phase or out-of-phase, respectively. To prevent this, the downmixing operation of Equation (1) is applied in sub-bands followed by the scaling operation of multipliers **310**. The scaling factors $e_i(k)$ $(1 \le i \le E)$ can be derived using Equation (3) as follows:

$$e_i(k) = \sqrt{\frac{p_{\tilde{y}_i(k)}}{p_{\hat{y}_i(k)}}}, \tag{3}$$

where $p_{\tilde{y}_i(k)}$ is the sub-band power as computed by Equation (2), and $p_{\hat{y}_i(k)}$ is power of the corresponding downmixed sub-band signal $\hat{y}_i(k)$.

In addition to or instead of providing optional scaling, scaling/delay block **306** may optionally apply delays to the signals.

Each inverse filter bank **308** converts a set of corresponding scaled coefficients $\tilde{y}_i(k)$ in the frequency domain into a frame of a corresponding digital, transmitted channel $y_i(n)$.

Although Fig. 3 shows all C of the input channels being converted into the frequency domain for subsequent downmixing, in alternative implementations, one or more (but less than C–1) of the C input channels might bypass some or all of the processing shown in Fig. 3 and be transmitted as an equivalent number of unmodified audio channels. Depending on the particular implementation, these unmodified

audio channels might or might not be used by BCC estimator **208** of Fig. 2 in generating the transmitted BCC codes.

In an implementation of downmixer **300** that generates a single sum signal $y(n)$, E=1 and the signals $\tilde{x}_c(k)$ of each subband of each input channel c are added and then multiplied with a factor $e(k)$, according to Equation (4) as follows:

$$\tilde{y}(k) = e(k)\sum_{c=1}^{C} \tilde{x}_c(k)\,. \tag{4}$$

the factor $e(k)$ is given by Equation (5) as follows:

$$e(k) = \sqrt{\frac{\sum_{c=1}^{C} p_{\tilde{x}_c}(k)}{p_{\tilde{x}}(k)}}\,, \tag{5}$$

where $p_{\tilde{x}_c}(k)$ is a short-time estimate of the power of $\tilde{x}_c(k)$ at time index k, and $p_{\tilde{x}}(k)$ is a short-time estimate of the power of $\sum_{c=1}^{C} \tilde{x}_c(k)$. The equalized subbands are transformed back to the time domain resulting in the sum signal $y(n)$ that is transmitted to the BCC decoder.

Generic BCC Synthesis

Fig. 4 shows a block diagram of a BCC synthesizer **400** that can be used for decoder **204** of Fig. 2 according to certain implementations of BCC system **200**. BCC synthesizer **400** has a filter bank **402** for each transmitted channel $y_i(n)$, an upmixing block **404**, delays **406**, multipliers **408**, correlation block **410**, and an inverse filter bank **412** for each playback channel $\hat{x}_i(n)$.

Each filter bank **402** converts each frame of a corresponding digital, transmitted channel $y_i(n)$ in the time domain into a set of input coefficients $\tilde{y}_i(k)$ in the frequency domain. Upmixing block **404** upmixes each sub-band of E corresponding transmitted-channel coefficients into a corresponding sub-band of C upmixed frequency-domain coefficients. Equation (4) represents the upmixing of the kth sub-

band of transmitted-channel coefficients $\left(\tilde{y}_1(k), \tilde{y}_2(k), \ldots, \tilde{y}_E(k)\right)$ to generate the kth sub-band of upmixed coefficients $\left(\tilde{s}_1(k), \tilde{s}_2(k), \ldots, \tilde{s}_C(k)\right)$ as follows:

$$\begin{bmatrix} \tilde{s}_1(k) \\ \tilde{s}_2(k) \\ \vdots \\ \tilde{s}_C(k) \end{bmatrix} = \mathbf{U}_{EC} \begin{bmatrix} \tilde{y}_1(k) \\ \tilde{y}_2(k) \\ \vdots \\ \tilde{y}_E(k) \end{bmatrix}, \tag{6}$$

where $\mathbf{U}_{EC}$ is a real-valued E-by-C upmixing matrix. Performing upmixing in the frequency-domain enables upmixing to be applied individually in each different sub-band.

Each delay 406 applies a delay value $d_i(k)$ based on a corresponding BCC code for ICTD data to ensure that the desired ICTD values appear between certain pairs of playback channels. Each multiplier 408 applies a scaling factor $a_i(k)$ based on a corresponding BCC code for ICLD data to ensure that the desired ICLD values appear between certain pairs of playback channels. Correlation block 410 performs a decorrelation operation A based on corresponding BCC codes for ICC data to ensure that the desired ICC values appear between certain pairs of playback channels. Further description of the operations of correlation block 410 can be found in U.S. Patent Application No. 10/155,437, filed on 05/24/02 as Baumgarte 2-10.

The synthesis of ICLD values may be less troublesome than the synthesis of ICTD and ICC values, since ICLD synthesis involves merely scaling of sub-band signals. Since ICLD cues are the most commonly used directional cues, it is usually more important that the ICLD values approximate those of the original audio signal. As such, ICLD data might be estimated between all channel pairs. The scaling factors $a_i(k)$ $(1 \le i \le C)$ for each sub-band are preferably chosen such that the sub-band power of each playback channel approximates the corresponding power of the original input audio channel.

One goal may be to apply relatively few signal modifications for synthesizing ICTD and ICC values. As such, the BCC data might not include ICTD and ICC values for all channel pairs. In that case, BCC synthesizer 400 would synthesize ICTD and ICC values only between certain channel pairs.

Each inverse filter bank 412 converts a set of corresponding synthesized coefficients $\tilde{\hat{x}}_i(k)$ in the frequency domain into a frame of a corresponding digital, playback channel $\hat{x}_i(n)$.

Although Fig. 4 shows all *E* of the transmitted channels being converted into the frequency domain for subsequent upmixing and BCC processing, in alternative implementations, one or more (but not all) of the *E* transmitted channels might bypass some or all of the processing shown in Fig. 4. For example, one or more of the transmitted channels may be unmodified channels that are not subjected to any upmixing. In addition to being one or more of the *C* playback channels, these unmodified channels, in turn, might be, but do not have to be, used as reference channels to which BCC processing is applied to synthesize one or more of the other playback channels. In either case, such unmodified channels may be subjected to delays to compensate for the processing time involved in the upmixing and/or BCC processing used to generate the rest of the playback channels.

Note that, although Fig. 4 shows *C* playback channels being synthesized from *E* transmitted channels, where *C* was also the number of original input channels, BCC synthesis is not limited to that number of playback channels. In general, the number of playback channels can be any number of channels, including numbers greater than or less than *C* and possibly even situations where the number of playback channels is equal to or less than the number of transmitted channels.

"Perceptually relevant differences" between audio channels

Assuming a single sum signal, BCC synthesizes a stereo or multi-channel audio signal such that ICTD, ICLD, and ICC approximate the corresponding cues of the original audio signal. In the following, the role of ICTD, ICLD, and ICC in relation to auditory spatial image attributes is discussed.

Knowledge about spatial hearing implies that for one auditory event, ICTD and ICLD are related to perceived direction. When considering binaural room impulse responses (BRIRs) of one source, there is a relationship between width of the auditory event and listener envelopment and ICC data estimated for the early and late parts of the BRIRs. However, the relationship between ICC and these properties for general signals (and not just the BRIRs) is not straightforward.

Stereo and multi-channel audio signals usually contain a complex mix of concurrently active source signals superimposed by reflected signal components resulting from recording in enclosed spaces or added by the recording engineer for artificially creating a spatial impression. Different source signals and their reflections occupy different regions in the time-frequency plane. This is reflected by ICTD, ICLD, and ICC, which vary as a function of time and frequency. In this case, the relation between instantaneous ICTD, ICLD, and ICC and auditory event directions and spatial impression is not obvious. The strategy of certain embodiments of BCC is to blindly synthesize these cues such that they approximate the corresponding cues of the original audio signal.

Filterbanks with subbands of bandwidths equal to two times the equivalent rectangular bandwidth (ERB) are used. Informal listening reveals that the audio quality of BCC does not notably

improve when choosing higher frequency resolution. A lower frequency resolution may be desired, since it results in less ICTD, ICLD, and ICC values that need to be transmitted to the decoder and thus in a lower bitrate.

Regarding time resolution, ICTD, ICLD, and ICC are typically considered at regular time intervals. High performance is obtained when ICTD, ICLD, and ICC are considered about every 4 to 16 ms. Note that, unless the cues are considered at very short time intervals, the precedence effect is not directly considered. Assuming a classical lead-lag pair of sound stimuli, if the lead and lag fall into a time interval where only one set of cues is synthesized, then localization dominance of the lead is not considered. Despite this, BCC achieves audio quality reflected in an average MUSHRA score of about 87 (i.e., "excellent" audio quality) on average and up to nearly 100 for certain audio signals.

The often-achieved perceptually small difference between reference signal and synthesized signal implies that cues related to a wide range of auditory spatial image attributes are implicitly considered by synthesizing ICTD, ICLD, and ICC at regular time intervals. In the following, some arguments are given on how ICTD, ICLD, and ICC may relate to a range of auditory spatial image attributes.

Estimation of spatial cues

In the following, it is described how ICTD, ICLD, and ICC are estimated. The bitrate for transmission of these (quantized and coded) spatial cues can be just a few kb/s and thus, with BCC, it is possible to transmit stereo and multi-channel audio signals at bitrates close to what is required for a single audio channel.

Fig. 5 shows a block diagram of BCC estimator **208** of Fig. 2, according to one embodiment of the present invention. BCC estimator **208** comprises filterbanks (FB) **502**, which may be the same as filterbanks **302** of Fig. 3, and estimation block **504**, which generates ICTD, ICLD, and ICC spatial cues for each different frequency subband generated by filterbanks **502**.

Estimation of ICTD, ICLD, and ICC for stereo signals

The following measures are used for ICTD, ICLD, and ICC for corresponding subband signals $\tilde{x}_1(k)$ and $\tilde{x}_2(k)$ of two (e.g., stereo) audio channels:

- ICTD [samples]:

$$\tau_{12}(k) = \arg\max_{d}\left\{\Phi_{12}(d,k)\right\}, \tag{7}$$

with a short-time estimate of the normalized cross-correlation function given by Equation (8) as follows:

$$\Phi_{12}(d,k) = \frac{p_{\tilde{x}_1\tilde{x}_2}(d,k)}{\sqrt{p_{\tilde{x}_1}(k-d_1)p_{\tilde{x}_2}(k-d_2)}}, \tag{8}$$

where

$$\begin{aligned} d_1 &= \max\{-d,0\} \\ d_2 &= \max\{d,0\} \end{aligned}, \tag{9}$$

and $p_{\tilde{x}_1\tilde{x}_2}(d,k)$ is a short–time estimate of the mean of $\tilde{x}_1(k-d_1)\tilde{x}_2(k-d_2)$.

- ICLD [dB]:

$$\Delta L_{12}(k) = 10\log_{10}\left(\frac{p_{\tilde{x}_2}(k)}{p_{\tilde{x}_1}(k)}\right). \tag{10}$$

- ICC:

$$c_{12}(k) = \max_d \left|\Phi_{12}(d,k)\right|. \tag{11}$$

Note that the absolute value of the normalized cross-correlation is considered and $c_{12}(k)$ has a range of [0,1].

Estimation of ICTD, ICLD, and ICC for multi-channel audio signals

When there are more than two input channels, it is typically sufficient to define ICTD and ICLD between a reference channel (e.g., channel number 1) and the other channels, as illustrated in Fig. 6 for the case of C=5 channels. where $\tau_{1c}(k)$ and $\Delta L_{12}(k)$ denote the ICTD and ICLD, respectively, between the reference channel 1 and channel c.

As opposed to ICTD and ICLD, ICC typically has more degrees of freedom. The ICC as defined can have different values between all possible input channel pairs. For C channels, there are $C(C\text{-}1)/2$ possible channel pairs; e.g., for 5 channels there are 10 channel pairs as illustrated in Fig. 7(a). However, such a scheme requires that, for each subband at each time index, $C(C\text{-}1)/2$ ICC values are estimated and transmitted, resulting in high computational complexity and high bitrate.

Alternatively, for each subband, ICTD and ICLD determine the direction at which the auditory event of the corresponding signal component in the subband is rendered. One single ICC parameter per

subband may then be used to describe the overall coherence between all audio channels. Good results can be obtained by estimating and transmitting ICC cues only between the two channels with most energy in each subband at each time index. This is illustrated in Fig. 7(b), where for time instants k-1 and k the channel pairs (3, 4) and (1, 2) are strongest, respectively. A heuristic rule may be used for determining ICC between the other channel pairs.

Synthesis of spatial cues

Fig. 8 shows a block diagram of an implementation of BCC synthesizer **400** of Fig. 4 that can be used in a BCC decoder to generate a stereo or multi-channel audio signal given a single transmitted sum signal $s(n)$ plus the spatial cues. The sum signal $s(n)$ is decomposed into subbands, where $\tilde{s}(k)$ denotes one such subband. For generating the corresponding subbands of each of the output channels, delays d_c, scale factors a_c, and filters h_c are applied to the corresponding subband of the sum signal. (For simplicity of notation, the time index k is ignored in the delays, scale factors, and filters.) ICTD are synthesized by imposing delays, ICLD by scaling, and ICC by applying de-correlation filters. The processing shown in Fig. 8 is applied independently to each subband.

ICTD synthesis

The delays d_c are determined from the ICTDs $\tau_{1c}(k)$, according to Equation (12) as follows:

$$d_c = \begin{cases} -\frac{1}{2}\left(\max_{2\le l\le C} \tau_{1l}(k) + \min_{2\le l\le C} \tau_{1l}(k)\right), & c = 1 \\ \tau_{1l}(k) + d_1 & 2 \le c \le C. \end{cases} \qquad (12)$$

The delay for the reference channel, d_1, is computed such that the maximum magnitude of the delays d_c is minimized. The less the subband signals are modified, the less there is a danger for artifacts to occur. If the subband sampling rate does not provide high enough time-resolution for ICTD synthesis, delays can be imposed more precisely by using suitable all-pass filters.

ICLD synthesis

In order that the output subband signals have desired ICLDs $\Delta L_{12}(k)$ between channel c and the reference channel 1, the gain factors a_c should satisfy Equation (13) as follows:

$$\frac{a_c}{a_1} = 10^{\frac{\Delta L_{1c}(k)}{20}} . \tag{13}$$

Additionally, the output subbands are preferably normalized such that the sum of the power of all output channels is equal to the power of the input sum signal. Since the total original signal power in each subband is preserved in the sum signal, this normalization results in the absolute subband power for each output channel approximating the corresponding power of the original encoder input audio signal. Given these constraints, the scale factors a_c are given by Equation (14) as follows:

$$a_c = \begin{cases} 1/\sqrt{1+\sum_{i=2}^{C} 10^{\Delta L_{1i}/10}}, & c = 1 \\ 10^{\Delta L_{1c}/20} a_1, & \text{otherwise.} \end{cases} \tag{14}$$

ICC synthesis

In certain embodiments, the aim of ICC synthesis is to reduce correlation between the subbands after delays and scaling have been applied, without affecting ICTD and ICLD. This can be achieved by designing the filters h_c in Fig. 8 such that ICTD and ICLD are effectively varied as a function of frequency such that the average variation is zero in each subband (auditory critical band).

Fig. 9 illustrates how ICTD and ICLD are varied within a subband as a function of frequency. The amplitude of ICTD and ICLD variation determines the degree of de–correlation and is controlled as a function of ICC. Note that ICTD are varied smoothly (as in Fig. 9(a)), while ICLD are varied randomly (as in Fig. 9(b)). One could vary ICLD as smoothly as ICTD, but this would result in more coloration of the resulting audio signals.

Another method for synthesizing ICC, particularly suitable for multi-channel ICC synthesis, is described in more detail in C. Faller, "Parametric multi-channel audio coding: Synthesis of coherence cues," *IEEE Trans. on Speech and Audio Proc.*, 2003, the teachings of which are incorporated herein by reference. As a function of time and frequency, specific amounts of artificial late reverberation are added to each of the output channels for achieving a desired ICC. Additionally, spectral modification can be applied such that the spectral envelope of the resulting signal approaches the spectral envelope of the original audio signal.

Other related and unrelated ICC synthesis techniques for stereo signals (or audio channel pairs) have been presented in E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *Preprint 114th Conv. Aud. Eng. Soc.*, Mar. 2003, and J. Engdegard, H. Purnhagen, J. Roden, and L. Liljeryd, "Synthetic ambience in parametric stereo coding,"

in *Preprint 117th Conv. Aud. Eng. Soc.*, May 2004, the teachings of both of which are incorporated here by reference.

C-to-*E* BCC

As described previously, BCC can be implemented with more than one transmission channel. A variation of BCC has been described which represents *C* audio channels not as one single (transmitted) channel, but as *E* channels, denoted *C*-to-*E* BCC. There are (at least) two motivations for *C*-to-*E* BCC:

- BCC with one transmission channel provides a backwards compatible path for upgrading existing mono systems for stereo or multi-channel audio playback. The upgraded systems transmit the BCC downmixed sum signal through the existing mono infrastructure, while additionally transmitting the BCC side information. *C*-to-*E* BCC is applicable to *E*-channel backwards compatible coding of *C*-channel audio.
- *C*-to-*E* BCC introduces scalability in terms of different degrees of reduction of the number of transmitted channels. It is expected that the more audio channels that are transmitted, the better the audio quality will be.

Signal processing details for *C*-to-*E* BCC, such as how to define the ICTD, ICLD, and ICC cues, are described in U.S. application serial number 10/762,100, filed on 01/20/04 (Faller 13-1).

Diffuse Sound Shaping

In certain implementations, BCC coding involves algorithms for ICTD, ICLD, and ICC synthesis. ICC cues can be synthesized by means of de-correlating the signal components in the corresponding subbands. This can be done by frequency-dependent variation of ICLD, frequency-dependent variation of ICTD and ICLD, all-pass filtering, or with ideas related to reverberation algorithms.

When these techniques are applied to audio signals, the temporal envelope characteristics of the signals are not preserved. Specifically, when applied to transients, the instantaneous signal energy is likely to be spread over a certain period of time. This results in artifacts such as "pre-echoes" or "washed-out transients."

A generic principle of certain embodiments of the present invention relates to the observation that the sound synthesized by a BCC decoder should not only have spectral characteristics that are similar to that of the original sound, but also resemble the temporal envelope of the original sound quite closely in order to have similar perceptual characteristics. Generally, this is achieved in BCC-like schemes by including a dynamic ICLD synthesis that applies a time-varying scaling operation to approximate each signal channel's temporal envelope. For the case of transient signals (attacks, percussive instruments, etc.), the temporal resolution of this process may, however, not be sufficient to

produce synthesized signals that approximate the original temporal envelope closely enough. This section describes a number of approaches to do this with a sufficiently fine time resolution.

Furthermore, for BCC decoders that do not have access to the temporal envelope of the original signals, the idea is to take the temporal envelope of the transmitted "sum signal(s)" as an approximation instead. As such, there is no side information necessary to be transmitted from the BCC encoder to the BCC decoder in order to convey such envelope information. In summary, the invention relies on the following principle:

- The transmitted audio channels (i.e., "sum channel(s)") - or linear combinations of these channels which BCC synthesis may be based on - are analyzed by a temporal envelope extractor for their temporal envelope with a high time resolution (e.g., significantly finer than the BCC block size).
- The subsequent synthesized sound for each output channel is shaped such that - even after ICC synthesis - it matches the temporal envelope determined by the extractor as closely as possible. This ensures that, even in the case of transient signals, the synthesized output sound is not significantly degraded by the ICC synthesis / signal de-correlation process.

Fig. 10 shows a block diagram representing at least a portion of a BCC decoder **1000**, according to one embodiment of the present invention. In Fig. 10, block **1002** represents BCC synthesis processing that includes, at least, ICC synthesis. BCC synthesis block **1002** receives base channels **1001** and generates synthesized channels **1003**. In certain implementations, block **1002** represents the processing of blocks **406**, **408**, and **410** of Fig. 4, where base channels **1001** are the signals generated by upmixing block **404** and synthesized channels **1003** are the signals generated by correlation block **410**. Fig. 10 represents the processing implemented for one base channel **1001'** and its corresponding synthesized channel. Similar processing is also applied to each other base channel and its corresponding synthesized channel.

Envelope extractor **1004** determines the fine temporal envelope *a* of base channel **1001'**, and envelope extractor **1006** determines the fine temporal envelope *b* of synthesized channel **1003'**. Inverse envelope adjuster **1008** uses temporal envelope *b* from envelope extractor **1006** to normalize the envelope (i.e., "flatten" the temporal fine structure) of synthesized channel **1003'** to produce a flattened signal **1005'** having a flat (e.g., uniform) time envelope. Depending on the particular implementation, the flattening can be applied either before or after upmixing. Envelope adjuster **1010** uses temporal envelope *a* from envelope extractor **1004** to re-impose the original signal envelope on the flattened signal **1005'** to generate output signal **1007'** having a temporal envelope substantially equal to the temporal envelope of base channel **1001**.

Depending on the implementation, this temporal envelope processing (also referred to herein as "envelope shaping") may be applied to the entire synthesized channel (as shown) or only to the orthogonalized part (e.g., late-reverberation part, de-correlated part) of the synthesized channel (as described subsequently). Moreover, depending on the implementation, envelope shaping may be applied either to time-domain signals or in a frequency-dependent fashion (e.g., where the temporal envelope is estimated and imposed individually at different frequencies).

Inverse envelope adjuster **1008** and envelope adjuster **1010** may be implemented in different ways. In one type of implementation, a signal's envelope is manipulated by multiplication of the signal's time-domain samples (or spectral / subband samples) with a time-varying amplitude modification function (e.g., $1/b$ for inverse envelope adjuster **1008** and a for envelope adjuster **1010**). Alternatively, a convolution / filtering of the signal's spectral representation over frequency can be used in a manner analogous to that used in the prior art for the purpose of shaping the quantization noise of a low bitrate audio coder. Similarly, the temporal envelope of signals may be extracted either directly by analysis the signal's time structure or by examining the autocorrelation of the signal spectrum over frequency.

Fig. 11 illustrates an exemplary application of the envelope shaping scheme of Fig. 10 in the context of BCC synthesizer **400** of Fig. 4. In this embodiment, there is a single transmitted sum signal $s(n)$, the C base signals are generated by replicating that sum signal, and envelope shaping is individually applied to different subbands. In alternative embodiments, the order of delays, scaling, and other processing may be different. Moreover, in alternative embodiments, envelope shaping is not restricted to processing each subband independently. This is especially true for convolution/filtering-based implementations that exploit covariance over frequency bands to derive information on the signal's temporal fine structure.

In Fig. 11(a), temporal process analyzer (TPA) **1104** is analogous to envelope extractor **1004** of Fig. 10, and each temporal processor (TP) **1106** is analogous to the combination of envelope extractor **1006**, inverse envelope adjuster **1008**, and envelope adjuster **1010** of Fig. 10.

Fig. 11(b) shows a block diagram of one possible time domain-based implementation of TPA **1104** in which the base signal samples are squared (**1110**) and then low-pass filtered (**1112**) to characterize the temporal envelope a of the base signal.

Fig. 11(c) shows a block diagram of one possible time domain-based implementation of TP **1106** in which the synthesized signal samples are squared (**1114**) and then low-pass filtered (**1116**) to characterize the temporal envelope b of the synthesized signal. A scale factor (e.g., sqrt (a/b)) is generated (**1118**) and then applied (**1120**) to the synthesized signal to generate an output signal having a temporal envelope substantially equal to that of the original base channel.

In alternative implementations of TPA **1104** and TP **1106**, the temporal envelopes are characterized using magnitude operations rather than by squaring the signal samples. In such implementations, the ratio *a/b* may be used as the scale factor without having to apply the square root operation.

Although the scaling operation of Fig. 11(c) corresponds to a time domain-based implementation of TP processing, TP processing (as well as TPA and inverse TP (ITP) processing) can also be implemented using frequency-domain signals, as in the embodiment of Figs. 17-18 (described below). As such, for purposes of this specification, the term "scaling function" should be interpreted to cover either time-domain or frequency-domain operations, such as the filtering operations of Figs. 18(b) and (c).

In general, TPA **1104** and TP **1106** are preferably designed such that they do not modify signal power (i.e., energy). Depending on the particular implementation, this signal power may be a short-time average signal power in each channel, e.g., based on the total signal power per channel in the time period defined by the synthesis window or some other suitable measure of power. As such, scaling for ICLD synthesis (e.g., using multipliers **408**) can be applied before or after envelope shaping.

Note that in Fig. 11(a), for each channel, there are two outputs, where TP processing is applied to only one of them. This reflects an ICC synthesis scheme that mixes two signal components: unmodified and orthogonalized signals, where the ratio of unmodified and orthogonalized signal components determines the ICC. In the embodiment shown in Fig. 11(a), TP is applied to only the orthogonalized signal component, where summation nodes **1108** recombine the unmodified signal components with the corresponding temporally shaped, orthogonalized signal components.

Fig. 12 illustrates an alternative exemplary application of the envelope shaping scheme of Fig. 10 in the context of BCC synthesizer **400** of Fig. 4, where envelope shaping is applied to in the time domain. Such an embodiment may be warranted when the time resolution of the spectral representation in which ICTD, ICLD, and ICC synthesis is carried out is not high enough for effectively preventing "pre-echoes" by imposing the desired temporal envelope. For example, this may be the case when BCC is implemented with a short-time Fourier transform (STFT).

As shown in Fig. 12(a), TPA **1204** and each TP **1206** are implemented in the time domain, where the full-band signal is scaled such that it has the desired temporal envelope (e.g., the envelope as estimated from the transmitted sum signal). Figs. 12(b) and (c) shows possible implementations of TPA **1204** and TP **1206** that are analogous to those shown in Figs. 11(b) and (c).

In this embodiment, TP processing is applied to the output signal, not only to the orthogonalized signal components. In alternative embodiments, time domain-based TP processing can be applied just to

the orthogonalized signal components if so desired, in which case unmodified and orthogonalized subbands would be converted to the time domain with separate inverse filterbanks.

Since full-band scaling of the BCC output signals may result in artifacts, envelope shaping might be applied only at specified frequencies, for example, frequencies larger than a certain cut-off frequency f_{TP} (e.g., 500 Hz). Note that the frequency range for analysis (TPA) may differ from the frequency range for synthesis (TP).

Figs. 13(a) and (b) show possible implementations of TPA **1204** and TP **1206** where envelope shaping is applied only at frequencies higher than the cut-off frequency f_{TP}. In particular, Fig. 13(a) shows the addition of high-pass filter **1302**, which filters out frequencies lower than f_{TP} prior to temporal envelope characterization. Fig. 13(b) shows the addition of two-band filterbank **1304** having with a cut-off frequency of f_{TP} between the two subbands, where only the high-frequency part is temporally shaped. Two-band inverse filterbank **1306** then recombines the low-frequency part with the temporally shaped, high-frequency part to generate the output signal.

Figs. 14 illustrates an exemplary application of the envelope shaping scheme of Fig. 10 in the context of the late reverberation-based ICC synthesis scheme described in U.S. application serial number 10/815,591, filed on 04/01/04 as attorney docket no. Baumgarte 7-12. In this embodiment, TPA **1404** and each TP **1406** are applied in the time domain, as in Fig. 12 or Fig. 13, but where each TP **1406** is applied to the output from a different late reverberation (LR) block **1402**.

Fig. 15 shows a block diagram representing at least a portion of a BCC decoder **1500**, according to an embodiment of the present invention that is an alternative to the scheme shown in Fig. 10. In Fig. 15, BCC synthesis block **1502**, envelope extractor **1504**, and envelope adjuster **1510** are analogous to BCC synthesis block **1002**, envelope extractor **1004**, and envelope adjuster **1010** of Fig. 10. In Fig. 15, however, inverse envelope adjuster **1508** is applied prior to BCC synthesis, rather than after BCC synthesis, as in Fig. 10. In this way, inverse envelope adjuster **1508** flattens the base channel before BCC synthesis is applied.

Fig. 16 shows a block diagram representing at least a portion of a BCC decoder **1600**, according to an embodiment of the present invention that is an alternative to the schemes shown in Figs. 10 and 15. In Fig. 16, envelope extractor **1604** and envelope adjuster **1610** are analogous to envelope extractor **1504** and envelope adjuster **1510** of Fig. 15. In the embodiment of Fig. 15, however, synthesis block **1602** represents late reverberation-based ICC synthesis similar to that shown in Fig. 16. In this case, envelope shaping is applied only to the uncorrelated late-reverberation signal, and summation node **1612** adds the temporally shaped, late-reverberation signal to the original base channel (which already has the desired temporal envelope). Note that, in this case, an inverse envelope adjuster does not need to be applied,

because the late-reverberation signal has an approximately flat temporal envelope due to its generation process in block **1602**.

Fig. 17 illustrates an exemplary application of the envelope shaping scheme of Fig. 15 in the context of BCC synthesizer **400** of Fig. 4. In Fig. 17, TPA **1704**, inverse TP (ITP) **1708**, and TP **1710** are analogous to envelope extractor **1504**, inverse envelope adjuster **1508**, and envelope adjuster **1510** of Fig. 15.

In this frequency-based embodiment, envelope shaping of diffuse sound is implemented by applying a convolution to the frequency bins of (e.g., STFT) filterbank **402** along the frequency axis. Reference is made to U.S. patent 5,781,888 (Herre) and U.S. patent 5,812,971 (Herre), the teachings of which are incorporated herein by reference, for subject matter related to this technique.

Fig. 18(a) shows a block diagram of one possible implementation of TPA **1704** of Fig. 17. In this implementation, TPA **1704** is implemented as a linear predictive coding (LPC) analysis operation that determines the optimum prediction coefficients for the series of spectral coefficients over frequency. Such LPC analysis techniques are well-known e.g., from speech coding and many algorithms for efficient calculation of LPC coefficients are known, such as the autocorrelation method (involving the calculation of the signal's autocorrelation function and a subsequent Levinson-Durbin recursion). As a result of this computation, a set of LPC coefficients are available at the output that represent the signal's temporal envelope.

Figs. 18(b) and (c) show block diagrams of possible implementations of ITP **1708** and TP **1710** of Fig. 17. In both implementations, the spectral coefficients of the signal to be processed are processed in order of (increasing or decreasing) frequency, which is symbolized here by rotating switch circuitry, converting these coefficients into a serial order for processing by a predictive filtering process (and back again after this processing). In the case of ITP **1708**, the predictive filtering calculates the prediction residual and in this way "flattens" the temporal signal envelope. In the case of TP **1710**, the inverse filter re-introduces the temporal envelope represented by the LPC coefficients from TPA **1704**.

For the calculation of the signal's temporal envelope by TPA **1704**, it is important to eliminate the influence of the analysis window of filterbank **402**, if such a window is used. This can be achieved by either normalizing the resulting envelope by the (known) analysis window shape or by using a separate analysis filterbank which does not employ an analysis window.

The convolution/filtering-based technique of Fig. 17 can also be applied in the context of the envelope shaping scheme of Fig. 16, where envelope extractor **1604** and envelope adjuster **1610** are based on the TPA of Fig. 18(a) and the TP of Fig. 18(c), respectively.

Further Alternative Embodiments

BCC decoders can be designed to selectively enable/disable envelope shaping. For example, a BCC decoder could apply a conventional BCC synthesis scheme and enable the envelope shaping when the temporal envelope of the synthesized signal fluctuates sufficiently such that the benefits of envelope shaping dominate over any artifacts that envelope shaping may generate. This enabling/disabling control can be achieved by:

(1) Transient detection: If a transient is detected, then TP processing is enabled. Transient detection can be implemented with in a look-ahead manner to effectively shape not only the transient but also the signal shortly before and after the transient. Possible ways of detecting transients include:

   o Observing the temporal envelope of the transmitted BCC sum signal(s) to determine when there is a sudden increase in power indicating the occurrence of a transient; and

   o Examining the gain of the predictive (LPC) filter. If the LPC prediction gain exceeds a specified threshold, it can be assumed that the signal is transient or highly fluctuating. The LPC analysis is computed on the spectrum's autocorrelation.

(2) Randomness detection: There are scenarios when the temporal envelope is fluctuating pseudo-randomly. In such a scenario, no transient might be detected but TP processing could still be applied (e.g., a dense applause signal corresponds to such a scenario).

Additionally, in certain implementations, in order to prevent possible artifacts in tonal signals, TP processing is not applied when the tonality of the transmitted sum signal(s) is high.

Furthermore, similar measures can be used in the BCC encoder to detect when TP processing should be active. Since the encoder has access to all original input signals, it may employ more sophisticated algorithms (e.g., a part of estimation block 208) to make a decision of when TP processing should be enabled. The result of this decision (a flag signaling when TP should be active) can be transmitted to the BCC decoder (e.g., as part of the side information of Fig. 2).

Although the present invention has been described in the context of BCC coding schemes in which there is a single sum signal, the present invention can also be implemented in the context of BCC coding schemes having two or more sum signals. In this case, the temporal envelope for each different "base" sum signal can be estimated before applying BCC synthesis, and different BCC output channels may be generated based on different temporal envelopes, depending on which sum signals were used to synthesize the different output channels. An output channel that is synthesized from two or more different sum channels could be generated based on an effective temporal envelope that takes into account (e.g., via weighted averaging) the relative effects of the constituent sum channels.

Although the present invention has been described in the context of BCC coding schemes involving ICTD, ICLD, and ICC codes, the present invention can also be implemented in the context of other BCC coding schemes involving only one or two of these three types of codes (e.g., ICLD and ICC, but not ICTD) and/or one or more additional types of codes. Moreover, the sequence of BCC synthesis processing and envelope shaping may vary in different implementations. For example, when envelope shaping is applied to frequency-domain signals, as in Figs. 14 and 16, envelope shaping could alternatively be implemented after ICTD synthesis (in those embodiments that employ ICTD synthesis), but prior to ICLD synthesis. In other embodiments, envelope shaping could be applied to upmixed signals before any other BCC synthesis is applied.

Although the present invention has been described in the context of BCC coding schemes, the present invention can also be implemented in the context of other audio processing systems in which audio signals are de-correlated or other audio processing that needs to de-correlate signals.

Although the present invention has been described in the context of implementations in which the encoder receives input audio signal in the time domain and generates transmitted audio signals in the time domain and the decoder receives the transmitted audio signals in the time domain and generates playback audio signals in the time domain, the present invention is not so limited. For example, in other implementations, any one or more of the input, transmitted, and playback audio signals could be represented in a frequency domain.

BCC encoders and/or decoders may be used in conjunction with or incorporated into a variety of different applications or systems, including systems for television or electronic music distribution, movie theaters, broadcasting, streaming, and/or reception. These include systems for encoding/decoding transmissions via, for example, terrestrial, satellite, cable, internet, intranets, or physical media (e.g., compact discs, digital versatile discs, semiconductor chips, hard drives, memory cards, and the like). BCC encoders and/or decoders may also be employed in games and game systems, including, for example, interactive software products intended to interact with a user for entertainment (action, role play, strategy, adventure, simulations, racing, sports, arcade, card, and board games) and/or education that may be published for multiple machines, platforms, or media. Further, BCC encoders and/or decoders may be incorporated in audio recorders/players or CD-ROM/DVD systems. BCC encoders and/or decoders may also be incorporated into PC software applications that incorporate digital decoding (e.g., player, decoder) and software applications incorporating digital encoding capabilities (e.g., encoder, ripper, recoder, and jukebox).

The present invention may be implemented as circuit-based processes, including possible implementation as a single integrated circuit (such as an ASIC or an FPGA), a multi-chip module, a single card, or a multi-card circuit pack. As would be apparent to one skilled in the art, various functions

of circuit elements may also be implemented as processing steps in a software program. Such software may be employed in, for example, a digital signal processor, micro-controller, or general-purpose computer.

The present invention can be embodied in the form of methods and apparatuses for practicing those methods. The present invention can also be embodied in the form of program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of program code, for example, whether stored in a storage medium, loaded into and/or executed by a machine, or transmitted over some transmission medium or carrier, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code segments combine with the processor to provide a unique device that operates analogously to specific logic circuits.

It will be further understood that various changes in the details, materials, and arrangements of the parts which have been described and illustrated in order to explain the nature of this invention may be made by those skilled in the art without departing from the scope of the invention as expressed in the following claims.

Although the steps in the following method claims, if any, are recited in a particular sequence with corresponding labeling, unless the claim recitations otherwise imply a particular sequence for implementing some or all of those steps, those steps are not necessarily intended to be limited to being implemented in that particular sequence.

CLAIMS

We claim:

1. A method for converting an input audio signal having an input temporal envelope into an output audio signal having an output temporal envelope, the method comprising:

characterizing the input temporal envelope of the input audio signal;

processing the input audio signal to generate a processed audio signal, wherein the processing de-correlates the input audio signal; and

adjusting the processed audio signal based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

2. The invention of claim 1, wherein the processing comprises inter-channel correlation (ICC) synthesis.

3. The invention of claim 2, wherein the ICC synthesis is part of binaural cue coding (BCC) synthesis.

4. The invention of claim 3, wherein the BCC synthesis further comprises at least one of inter-channel level difference (ICLD) synthesis and inter-channel time difference (ICTD) synthesis.

5. The invention of claim 2, wherein the ICC synthesis comprises late-reverberation ICC synthesis.

6. The invention of claim 1, wherein the adjusting comprises:

characterizing a processed temporal envelope of the processed audio signal; and

adjusting the processed audio signal based on both the characterized input and processed temporal envelopes to generate the output audio signal.

7. The invention of claim 6, wherein the adjusting comprises:

generating a scaling function based on the characterized input and processed temporal envelopes; and

applying the scaling function to the processed audio signal to generate the output audio signal.

8. The invention of claim 1, further comprising adjusting the input audio signal based on the characterized input temporal envelope to generate a flattened audio signal, wherein the processing is applied to the flattened audio signal to generate the processed audio signal.

9. The invention of claim 1, wherein:

the processing generates an uncorrelated processed signal and a correlated processed signal; and

the adjusting is applied to the uncorrelated processed signal to generate an adjusted processed signal, wherein the output signal is generated by summing the adjusted processed signal and the correlated processed signal.

10. The invention of claim 1, wherein:

the characterizing is applied only to specified frequencies of the input audio signal; and

the adjusting is applied only to the specified frequencies of the processed audio signal.

11 . The invention of claim 10, wherein:

the characterizing is applied only to frequencies of the input audio signal above a specified cutoff frequency; and

the adjusting is applied only to frequencies of the processed audio signal above the specified cutoff frequency.

12. The invention of claim 1, wherein each of the characterizing, the processing, and the adjusting is applied to a frequency-domain signal.

13. The invention of claim 12, wherein each of the characterizing, the processing, and the adjusting is individually applied to different signal subbands.

14. The invention of claim 12, wherein the frequency domain corresponds to a fast Fourier transform (FFT).

15. The invention of claim 12, wherein the frequency domain corresponds to a quadrature mirror filter (QMF).

16. The invention of claim 1, wherein each of the characterizing and the adjusting is applied to a time-domain signal.

17. The invention of claim 16, wherein the processing is applied to a frequency-domain signal.

18. The invention of claim 17, wherein the frequency domain corresponds to an FFT.

19. The invention of claim 17, wherein the frequency domain corresponds to a QMF.

20. The invention of claim 1, further comprising determining whether to enable or disable the characterizing and the adjusting.

21. The invention of claim 20, wherein the determining is based on an enable/disable flag generated by an audio encoder that generated the input audio signal.

22. The invention of claim 20, wherein the determining is based on analyzing the input audio signal to detect transients in the input audio signal such that the characterizing and the adjusting are enabled if occurrence of a transient is detected.

23. An apparatus for converting an input audio signal having an input temporal envelope into an output audio signal having an output temporal envelope, the apparatus comprising:

means for characterizing the input temporal envelope of the input audio signal;

means for processing the input audio signal to generate a processed audio signal, wherein the means for processing is adapted to de-correlate the input audio signal; and

means for adjusting the processed audio signal based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

24. Apparatus for converting an input audio signal having an input temporal envelope into an output audio signal having an output temporal envelope, the apparatus comprising:

an envelope extractor adapted to characterize the input temporal envelope of the input audio signal;

a synthesizer adapted to process the input audio signal to generate a processed audio signal, wherein the synthesizer is adapted to de-correlate the input audio signal; and

an envelope adjuster adapted to adjust the processed audio signal based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

25, The invention of claim 24, wherein:

the apparatus is a system selected from the group consisting of a digital video player, a digital audio player, a computer, a satellite receiver, a cable receiver, a terrestrial broadcast receiver, a home entertainment system, and a movie theater system; and

the system comprises the envelope extractor, the synthesizer, and the envelope adjuster.

26. A method for encoding *C* input audio channels to generate *E* transmitted audio channel(s), the method comprising:

generating one or more cue codes for two or more of the *C* input channels;

downmixing the *C* input channels to generate the *E* transmitted channel(s), where $C>E\geq 1$; and

analyzing one or more of the *C* input channels and the *E* transmitted channel(s) to generate a flag indicating whether or not a decoder of the *E* transmitted channel(s) should perform envelope shaping during decoding of the *E* transmitted channel(s).

27. The invention of claim 26, wherein the envelope shaping adjusts a temporal envelope of a decoded channel generated by the decoder to substantially match a temporal envelope of a corresponding transmitted channel.

28.~~26.~~ An apparatus for encoding *C* input audio channels to generate *E* transmitted audio, channel(s), the apparatus comprising:

means for generating one or more cue codes for two or more of the *C* input channels;

means for downmixing the *C* input channels to generate the *E* transmitted channel(s), where $C>E\geq 1$; and

means for analyzing one or more of the *C* input channels and the *E* transmitted channel(s) to generate a flag indicating whether or not a decoder of the *E* transmitted channel(s) should perform envelope shaping during decoding of the *E* transmitted channel(s).

29.~~27.~~ Apparatus for encoding *C* input audio channels to generate *E* transmitted audio channel(s), the apparatus comprising:

a code estimator adapted to generate one or more cue codes for two or more of the *C* input channels; and

a downmixer adapted to downmix the *C* input channels to generate the *E* transmitted channel(s), where $C>E\geq 1$, wherein the code estimator is further adapted to analyze one or more of the *C* input channels and the *E* transmitted channel(s) to generate a flag indicating whether or not a decoder of the *E* transmitted channel(s) should perform envelope shaping during decoding of the *E* transmitted channel(s).

30.~~28.~~ The invention of claim 29~~27~~, wherein:

the apparatus is a system selected from the group consisting of a digital video recorder, a digital audio recorder, a computer, a satellite transmitter, a cable transmitter, a terrestrial broadcast transmitter, a home entertainment system, and a movie theater system; and

the system comprises the code estimator and the downmixer.

31.~~29.~~ An encoded audio bitstream generated by encoding *C* input audio channels to generate *E* transmitted audio channel(s), wherein:

one or more cue codes are generated for two or more of the *C* input channels;

the *C* input channels are downmixed to generate *E* transmitted channel(s), where $C>E\geq 1$;

a flag is generated by analyzing one or more of the *C* input channels and the *E* transmitted channel(s), wherein the flag indicates whether or not a decoder of the *E* transmitted channel(s) should perform envelope shaping during decoding of the *E* transmitted channel(s); and

the *E* transmitted channel(s), the one or more cue codes, and the flag are encoded into the encoded audio bitstream.

32.~~30.~~ An encoded audio bitstream comprising *E* transmitted channel(s), one or more cue codes, and a flag, wherein:

the one or more cue codes are generated by generating one or more cue codes for two or more of the *C* input channels;

the *E* transmitted channel(s) are generated by downmixing the *C* input channels, where $C>E\geq 1$; and

the flag is generated by analyzing one or more of the *C* input channels and the *E* transmitted channel(s), wherein the flag indicates whether or not a decoder of the *E* transmitted channel(s) should perform envelope shaping during decoding of the *E* transmitted channel(s).

33.~~31.~~ A machine-readable medium, having encoded thereon program code, wherein, when the program code is executed by a machine, the machine implements a method for converting an input audio signal having an input temporal envelope into an output audio signal having an output temporal envelope, the method comprising:

characterizing the input temporal envelope of the input audio signal;

processing the input audio signal to generate a processed audio signal, wherein the processing de-correlates the input audio signal; and

adjusting the processed audio signal based on the characterized input temporal envelope to generate the output audio signal, wherein the output temporal envelope substantially matches the input temporal envelope.

34.~~32.~~ A machine-readable medium, having encoded thereon program code, wherein, when the program code is executed by a machine, the machine implements a method for encoding C input audio channels to generate E transmitted audio channel(s), the method comprising:

generating one or more cue codes for two or more of the C input channels;

downmixing the C input channels to generate the E transmitted channel(s), where $C>E\geq 1$; and

analyzing one or more of the C input channels and the E transmitted channel(s) to generate a flag indicating whether or not a decoder of the E transmitted channel(s) should perform envelope shaping during decoding of the E transmitted channel(s).

FIG. 1
(PRIOR ART)

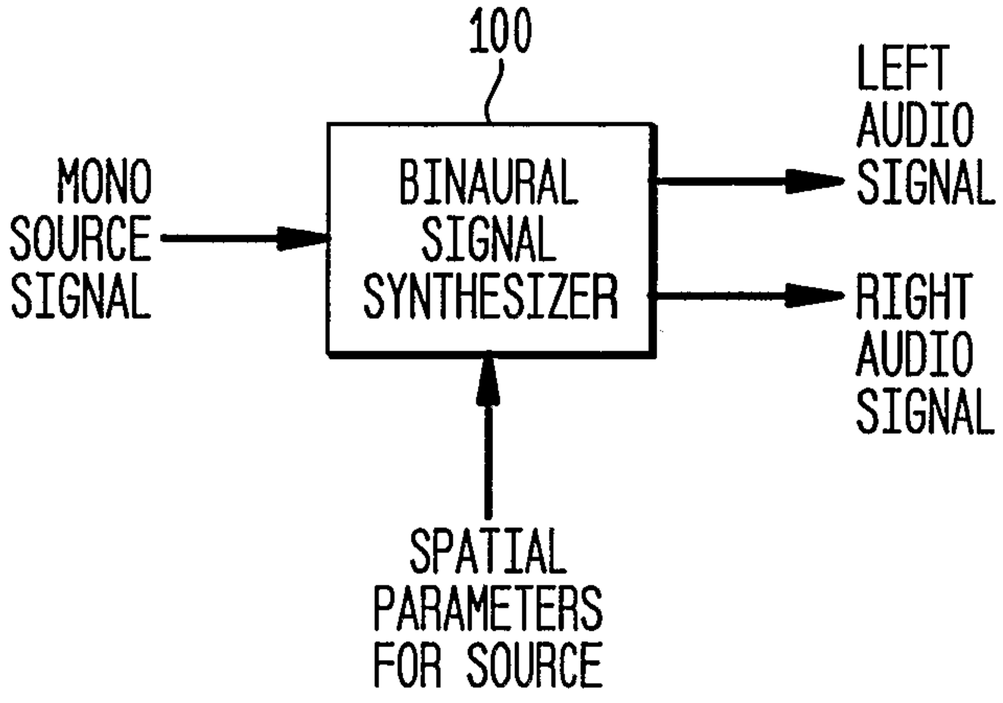
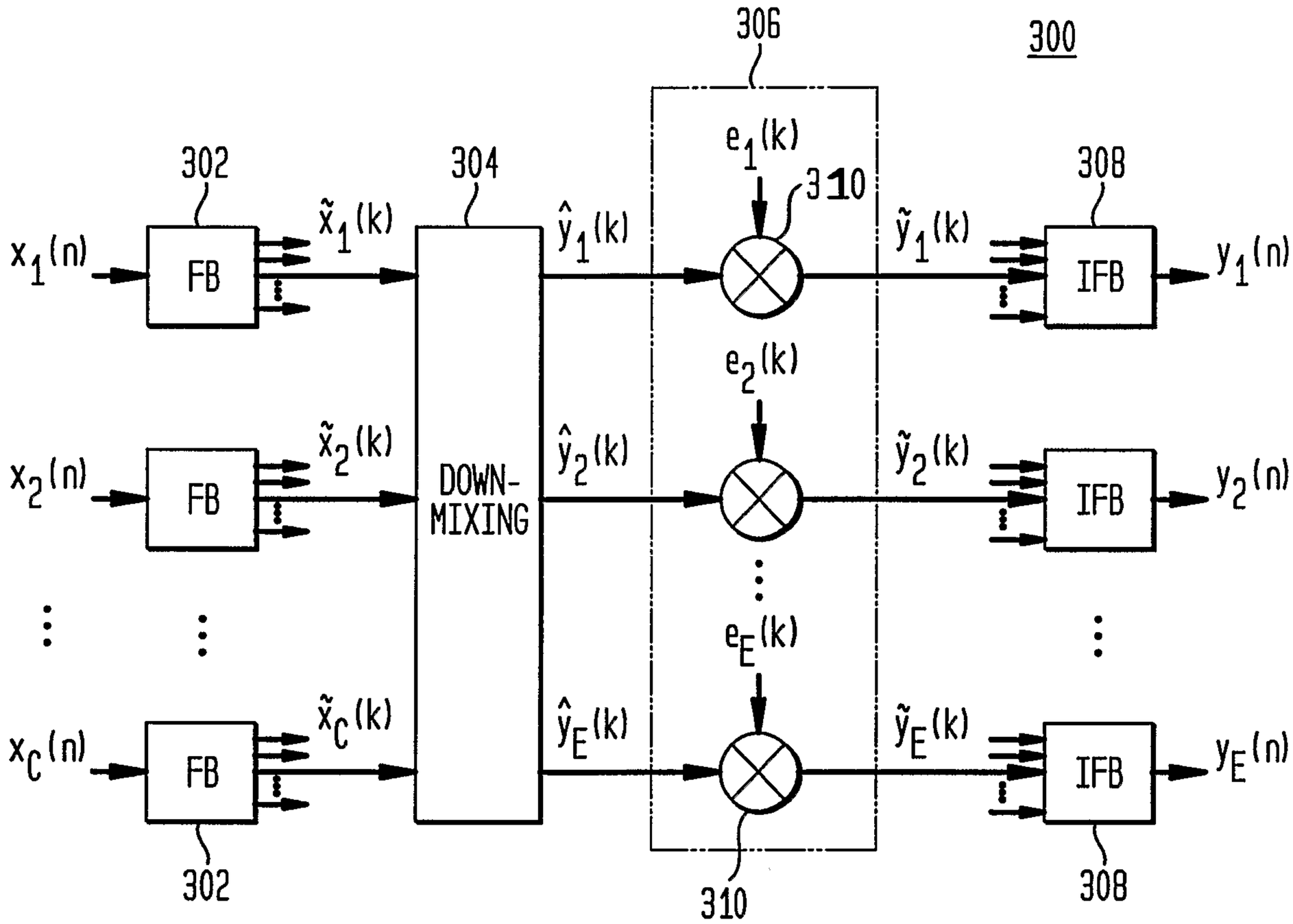


FIG. 2

FIG. 3


300
306
302
304
308
310
$x_1(n)$
FB
$\tilde{x}_1(k)$
$\hat{y}_1(k)$
$e_1(k)$
$\tilde{y}_1(k)$
IFB
$y_1(n)$
$x_2(n)$
FB
$\tilde{x}_2(k)$
DOWN-MIXING
$\hat{y}_2(k)$
$e_2(k)$
$\tilde{y}_2(k)$
IFB
$y_2(n)$
$x_C(n)$
FB
$\tilde{x}_C(k)$
$\hat{y}_E(k)$
$e_E(k)$
$\tilde{y}_E(k)$
IFB
$y_E(n)$
302
310
308

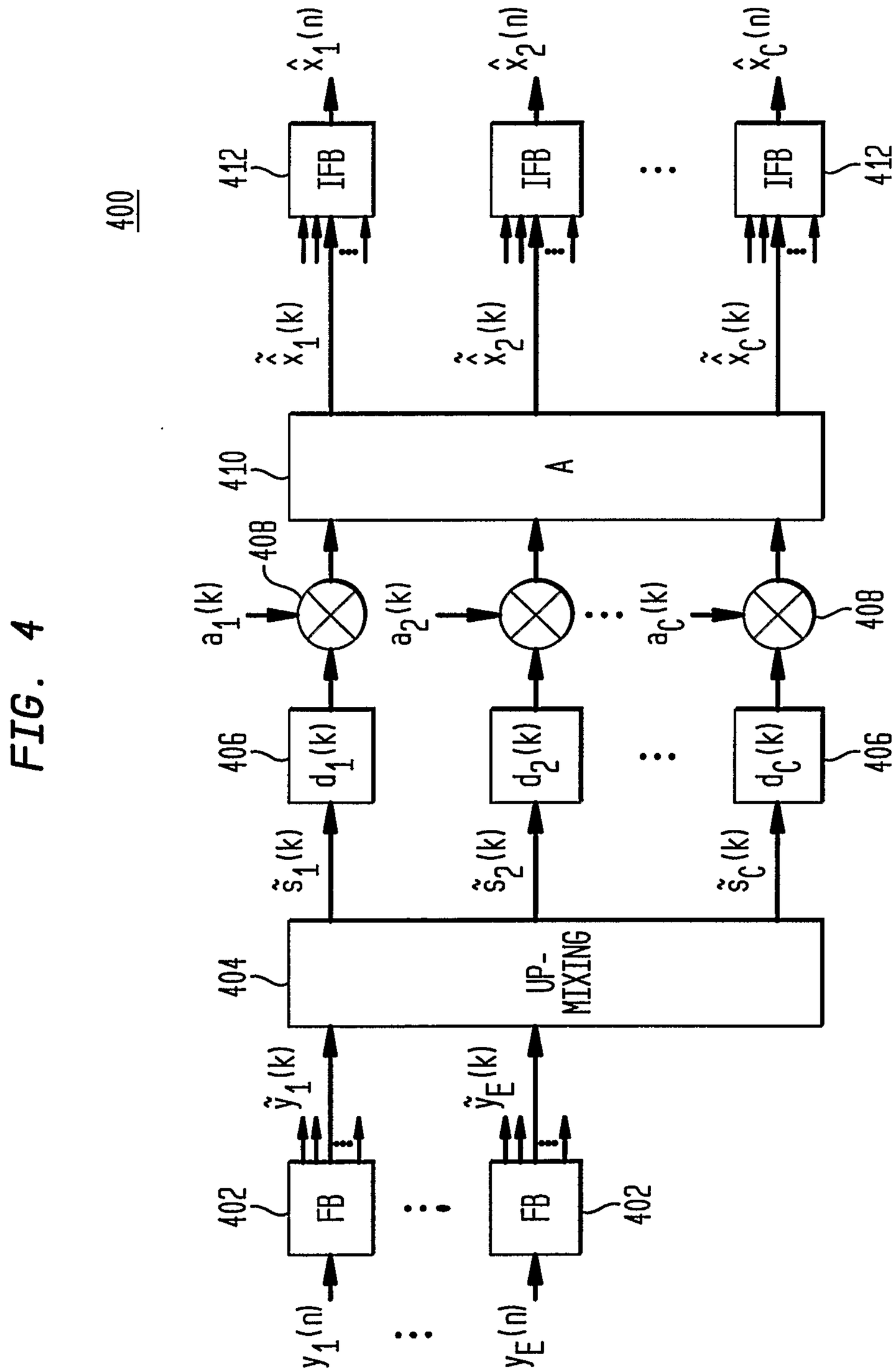

400
412
IFB
$\tilde{\tilde{x}}_1(k)$
$\tilde{\tilde{x}}_2(k)$
$\tilde{\tilde{x}}_C(k)$
410
A
408
$a_1(k)$
$a_2(k)$
$a_C(k)$
406
$d_1(k)$
$d_2(k)$
$d_C(k)$
$\tilde{s}_1(k)$
$\tilde{s}_2(k)$
$\tilde{s}_C(k)$
404
UP-MIXING
$\tilde{y}_1(k)$
$\tilde{y}_E(k)$
402
FB
$y_1(n)$
$y_E(n)$
$x_1(n)$
$x_2(n)$
$x_C(n)$

FIG. 4

FIG. 5

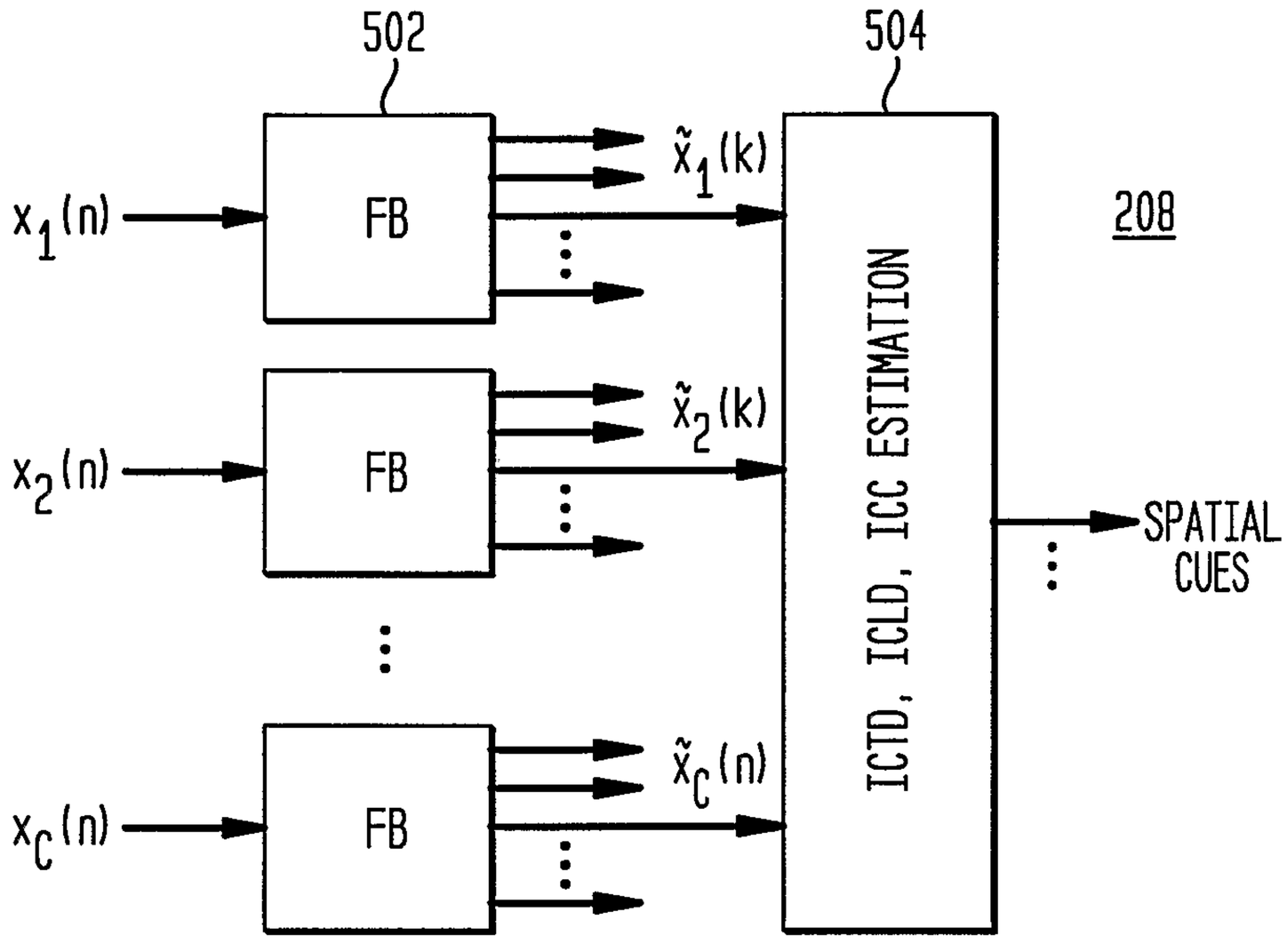

502
504
$x_1(n)$
FB
$\tilde{x}_1(k)$
208
$x_2(n)$
FB
$\tilde{x}_2(k)$
ICTD, ICLD, ICC ESTIMATION
SPATIAL CUES
$x_C(n)$
FB
$\tilde{x}_C(n)$

FIG. 6

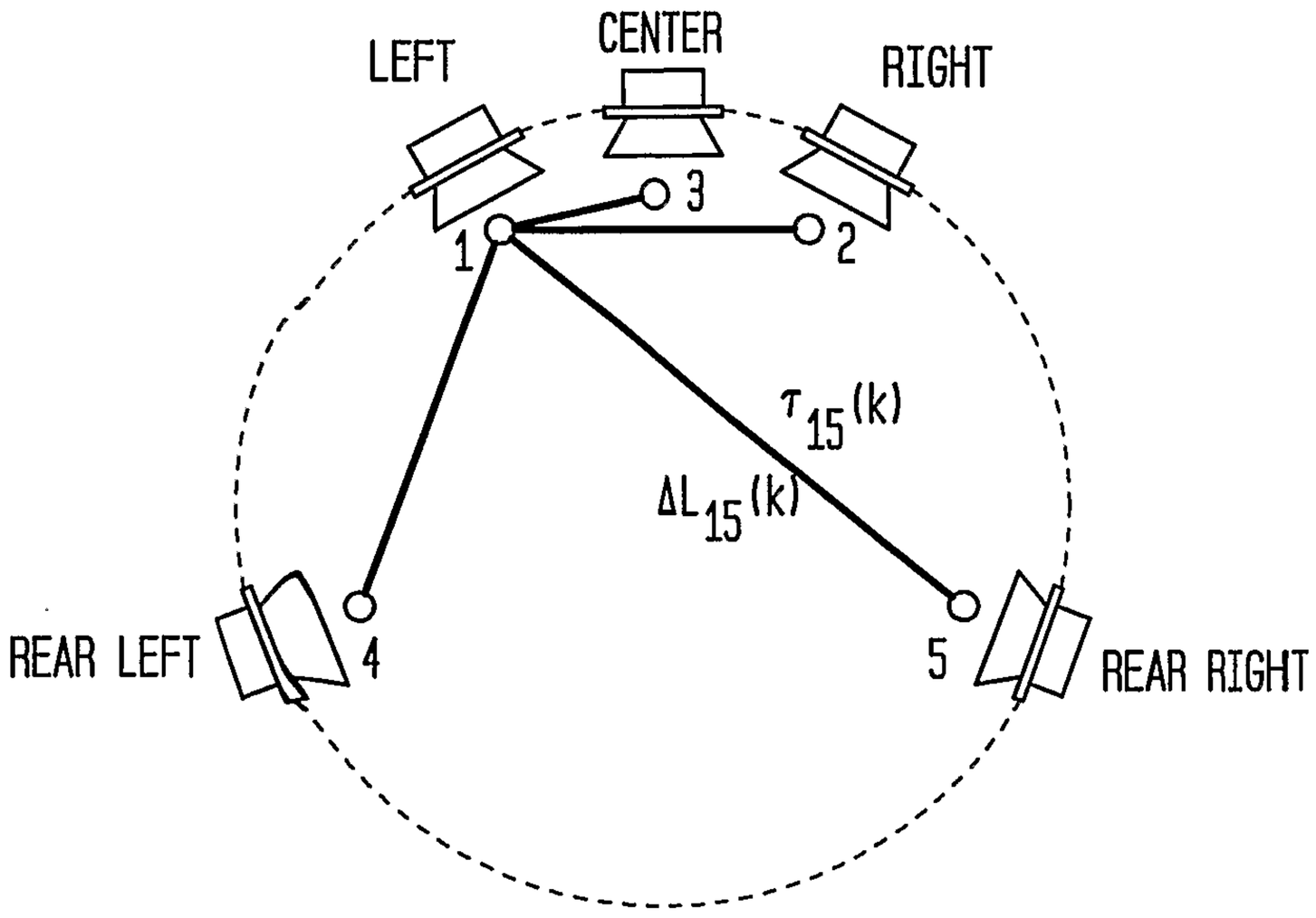

LEFT
CENTER
RIGHT
1
2
3
$\tau_{15}(k)$
$\Delta L_{15}(k)$
REAR LEFT
4
5
REAR RIGHT

FIG. 7A

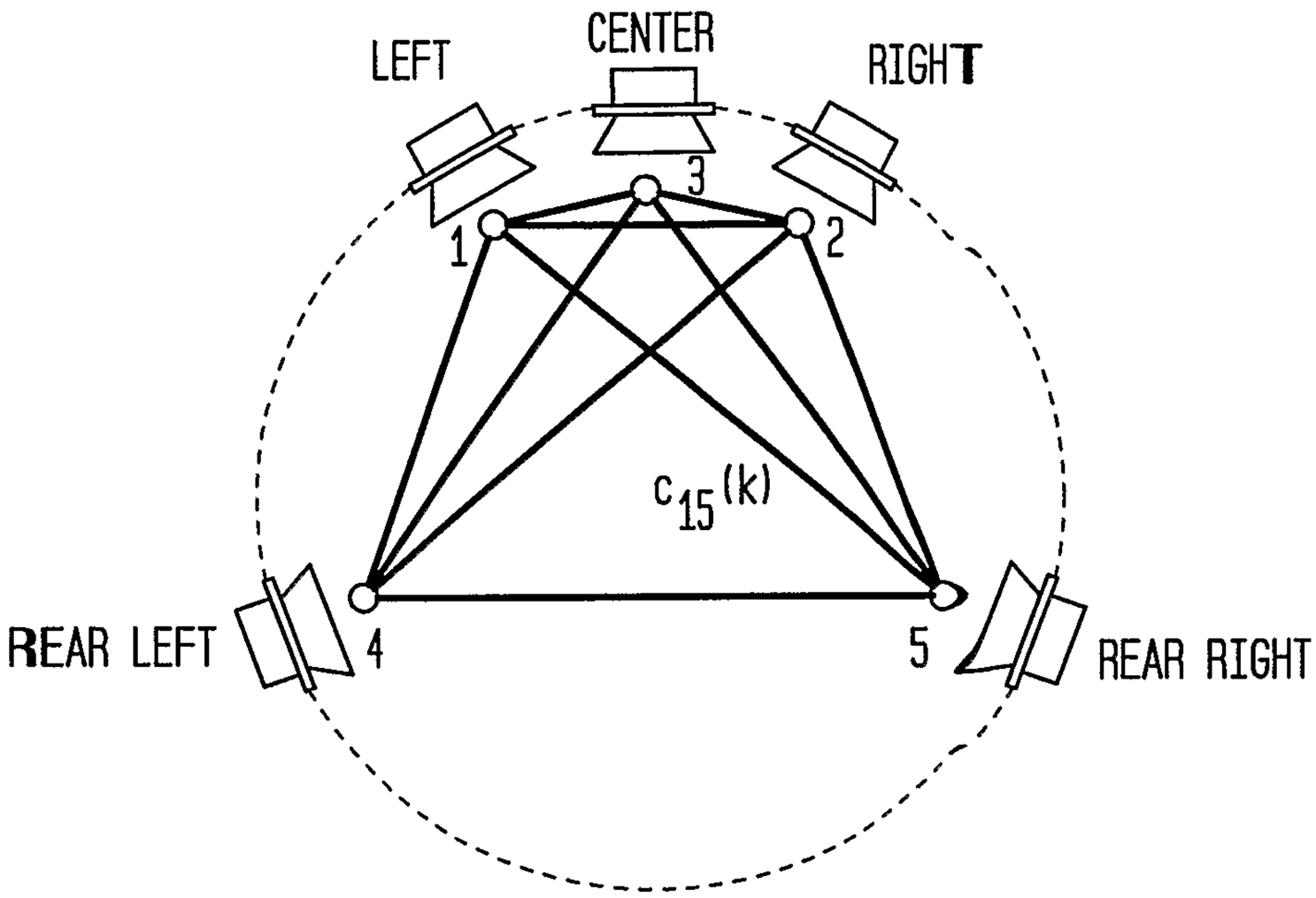
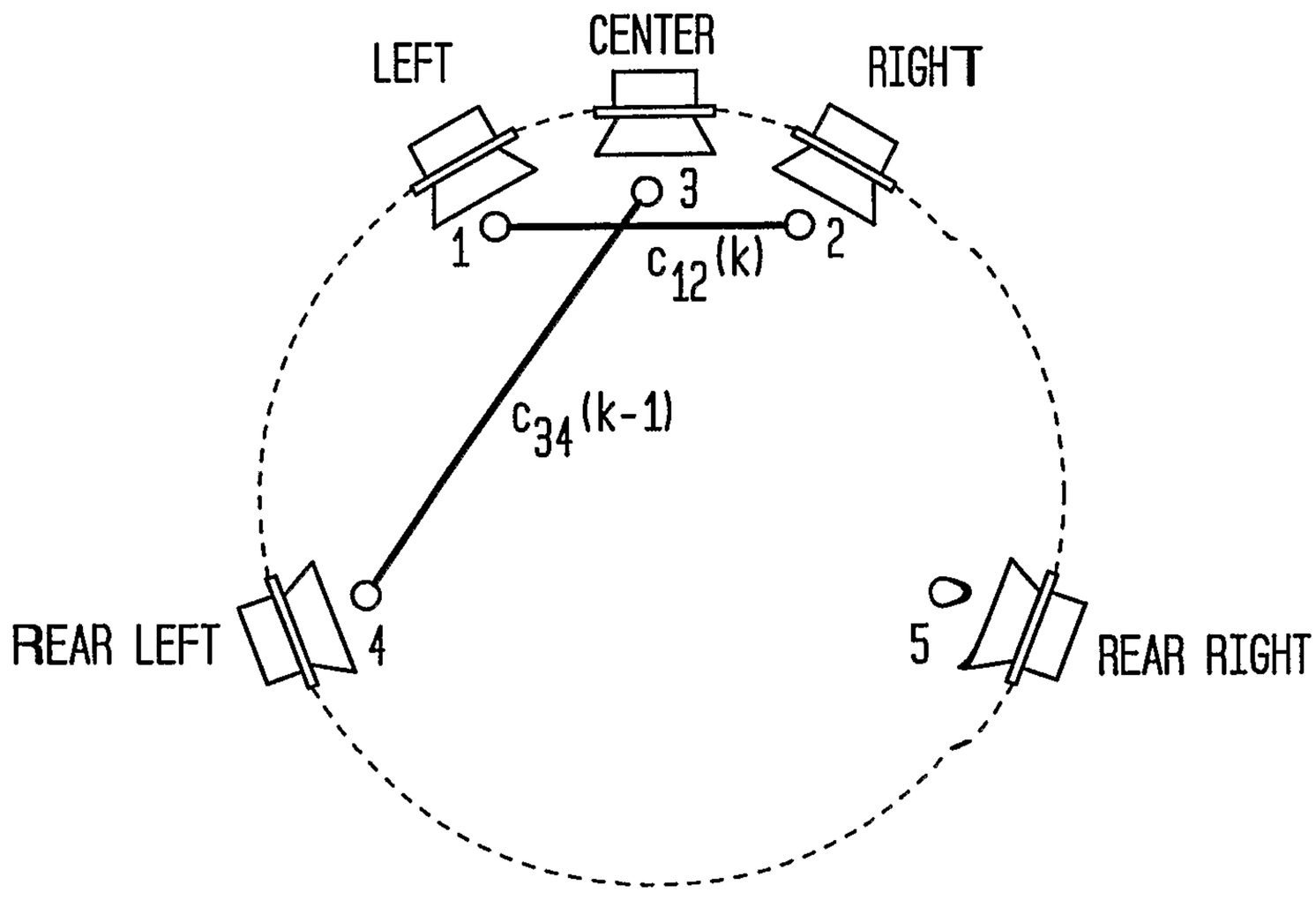

LEFT
CENTER
RIGHT
3
1
2
$c_{15}(k)$
REAR LEFT
4
5
REAR RIGHT

FIG. 7B


LEFT
CENTER
RIGHT
3
1
2
$c_{12}(k)$
$c_{34}(k-1)$
REAR LEFT
4
5
REAR RIGHT

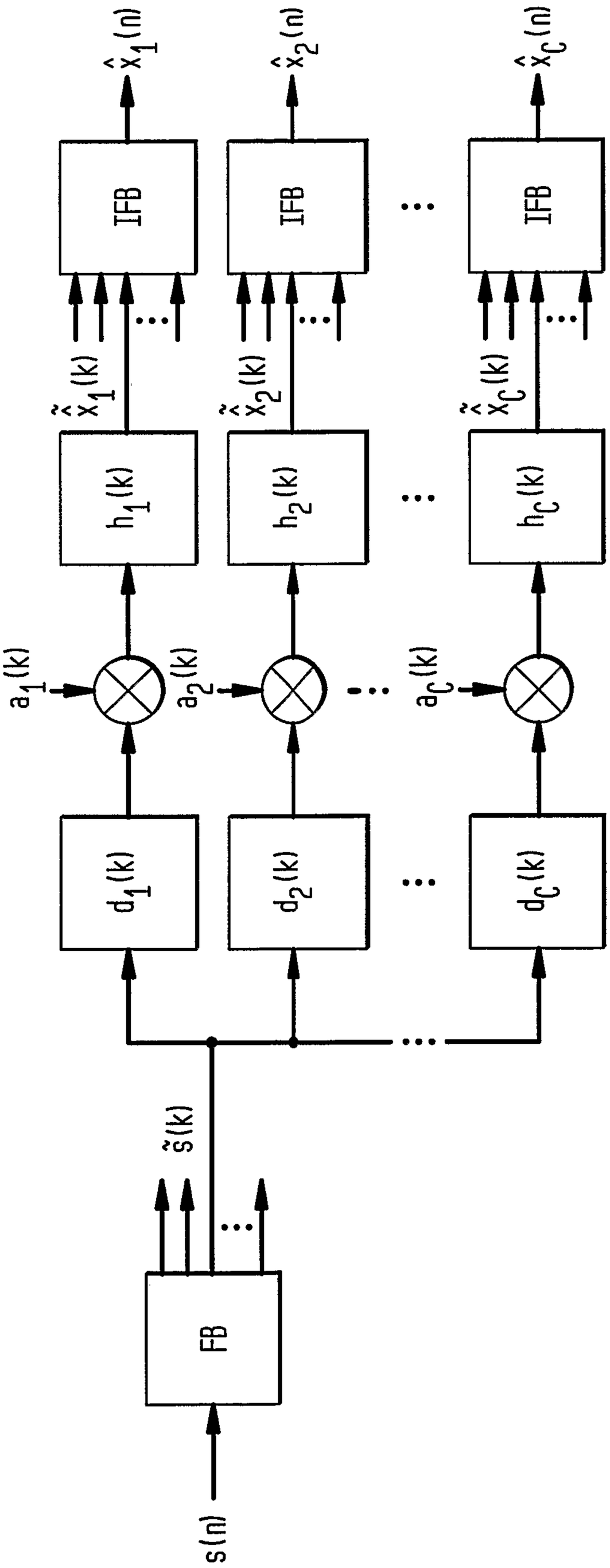

$\hat{x}_1(n)$
$\hat{x}_2(n)$
$\hat{x}_C(n)$
IFB
IFB
IFB
$\tilde{\hat{x}}_1(k)$
$\tilde{\hat{x}}_2(k)$
$\tilde{\hat{x}}_C(k)$
$h_1(k)$
$h_2(k)$
$h_C(k)$
$a_1(k)$
$a_2(k)$
$a_C(k)$
$d_1(k)$
$d_2(k)$
$d_C(k)$
$\tilde{s}(k)$
FB
s(n)

FIG. 8

FIG. 9

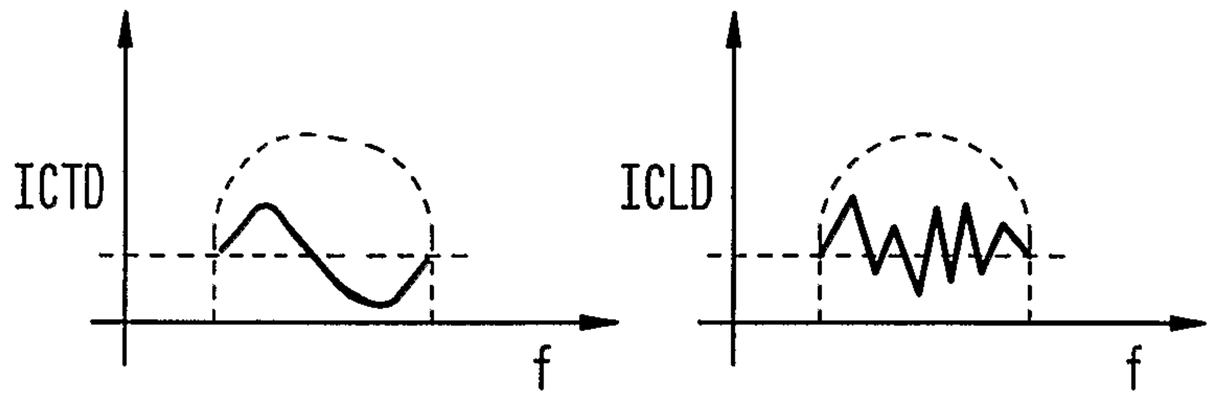
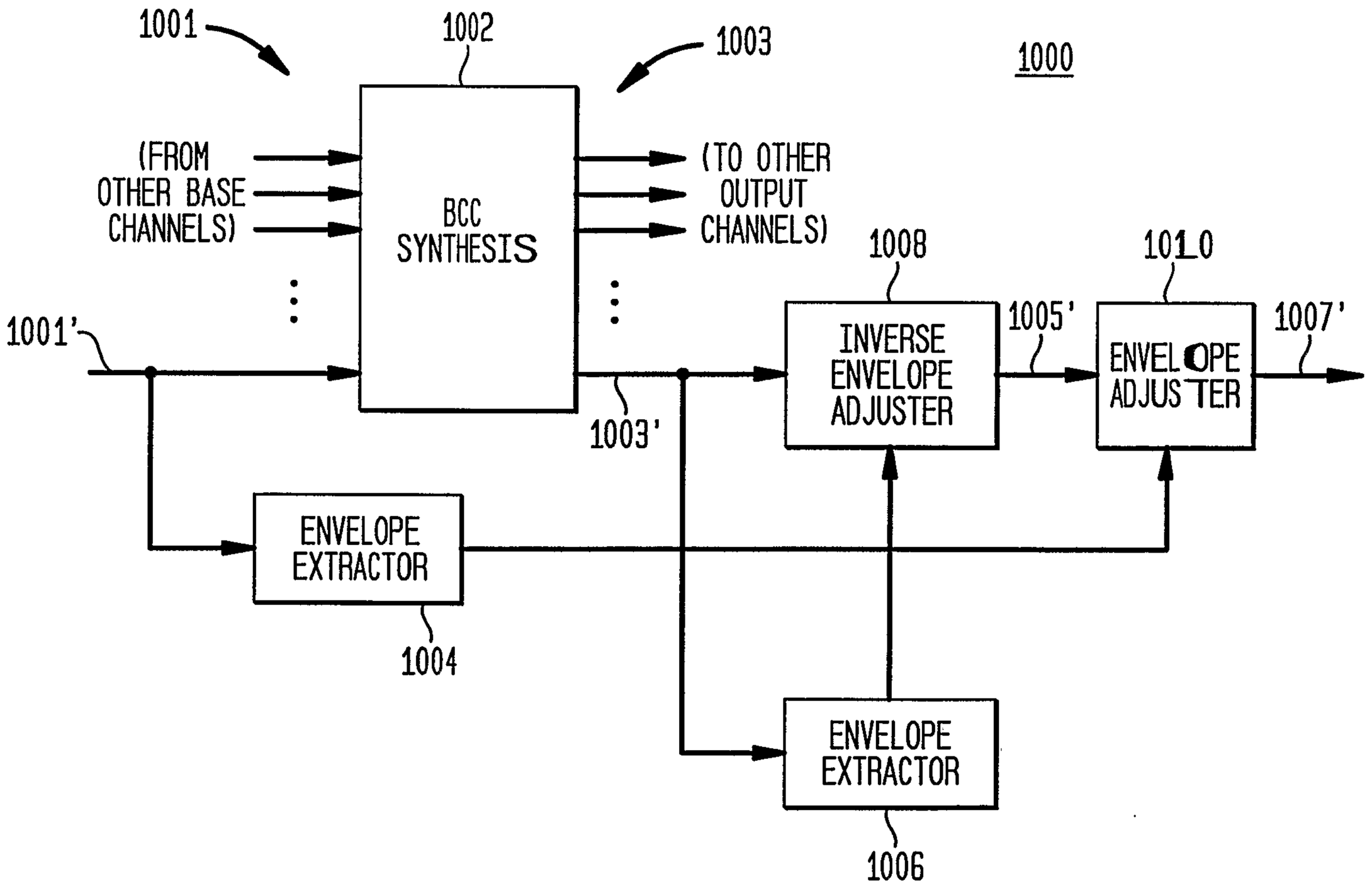

ICTD
f
ICLD
f

FIG. 10


1001
1002
1003
1000
(FROM OTHER BASE CHANNELS)
BCC SYNTHESIS
(TO OTHER OUTPUT CHANNELS)
1008
1010
1001'
1005'
1007'
INVERSE ENVELOPE ADJUSTER
ENVELOPE ADJUSTER
1003'
ENVELOPE EXTRACTOR
1004
ENVELOPE EXTRACTOR
1006

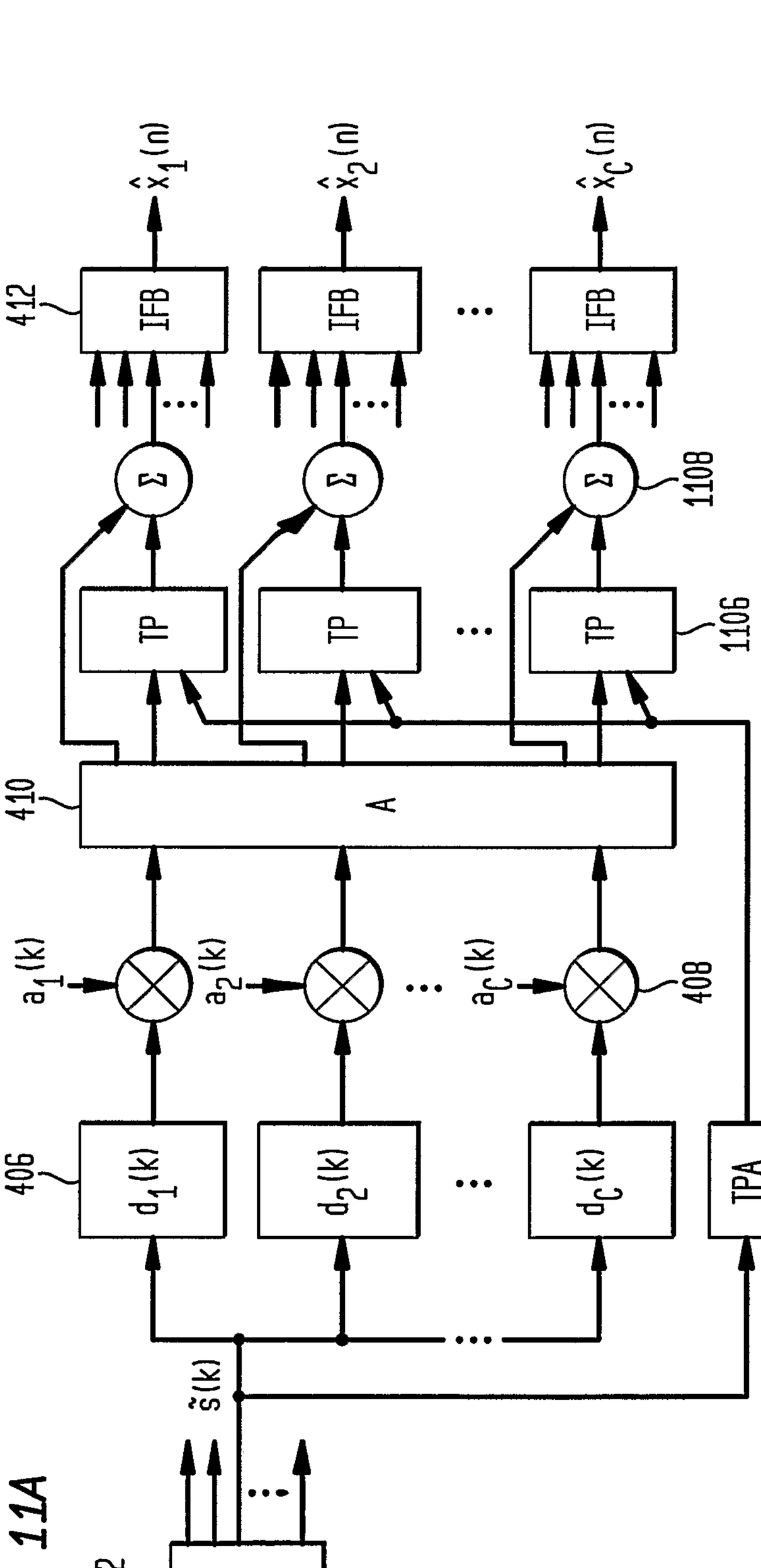


FIG. 11A

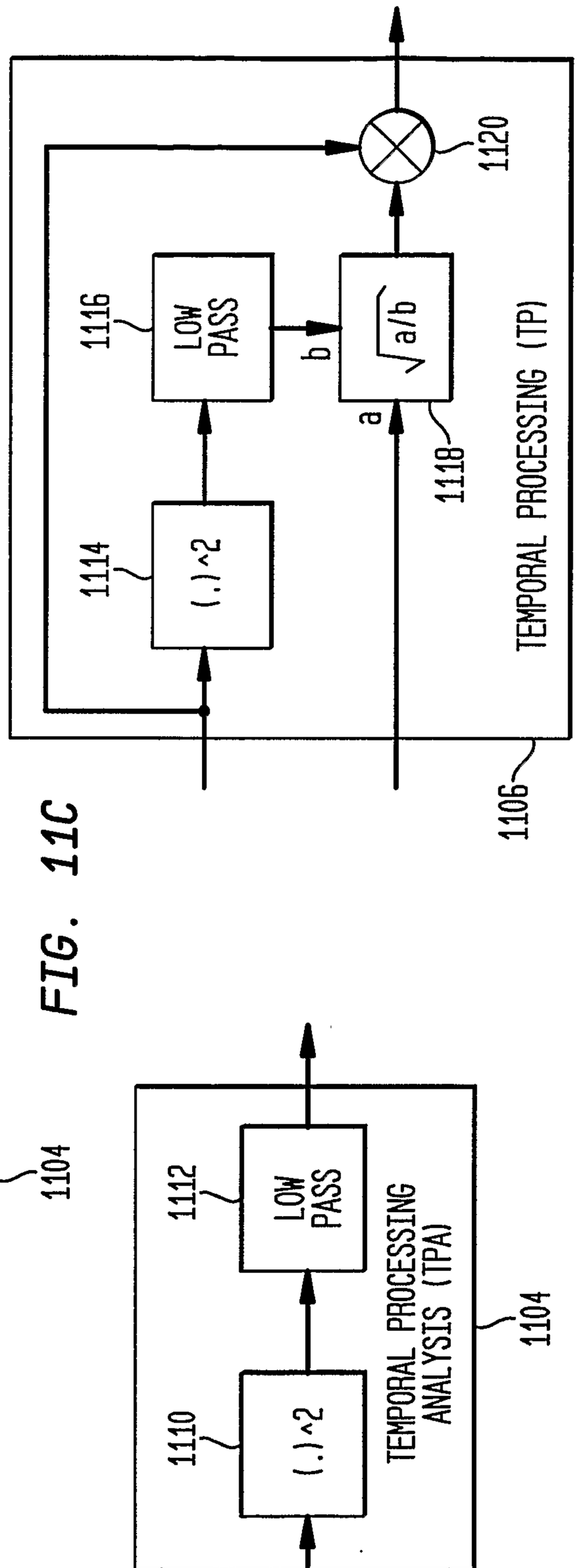


FIG. 11C

FIG. 11B

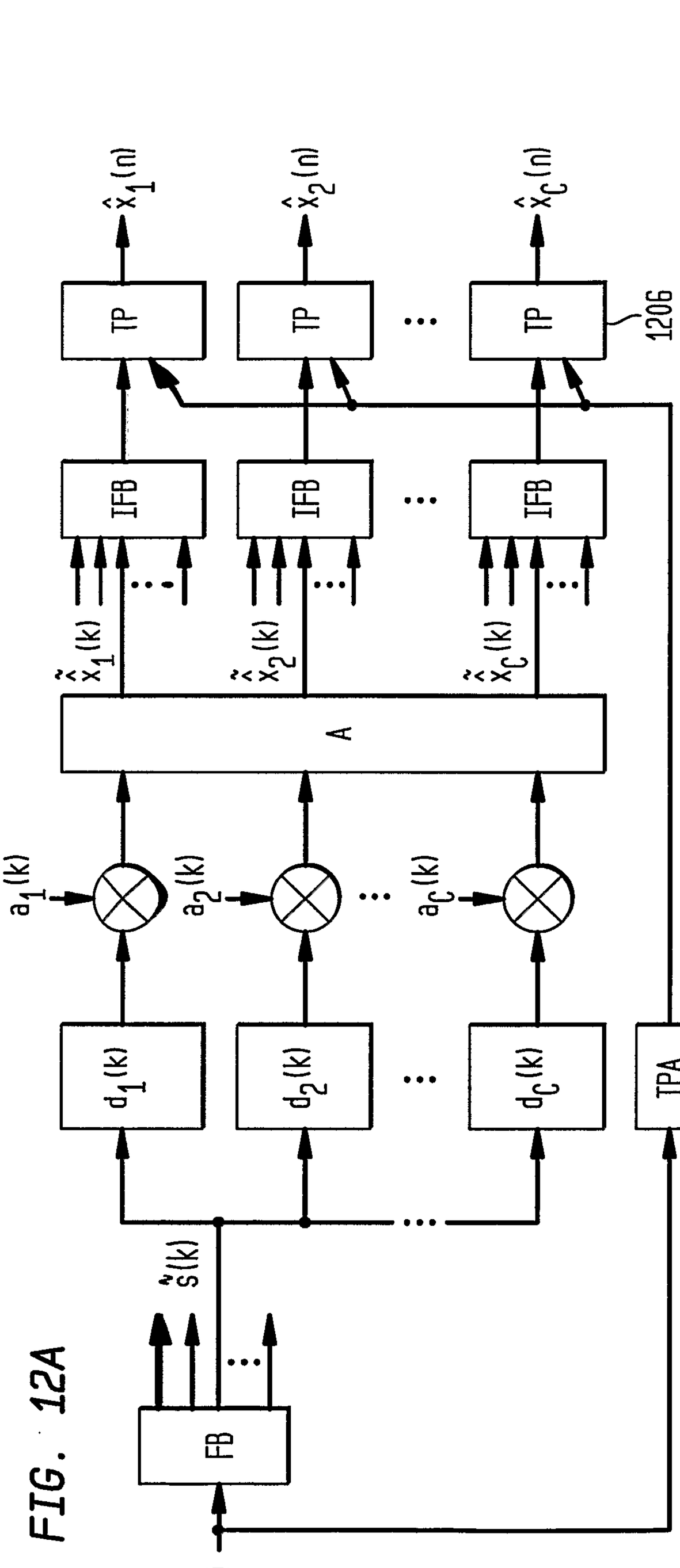


FIG. 12A

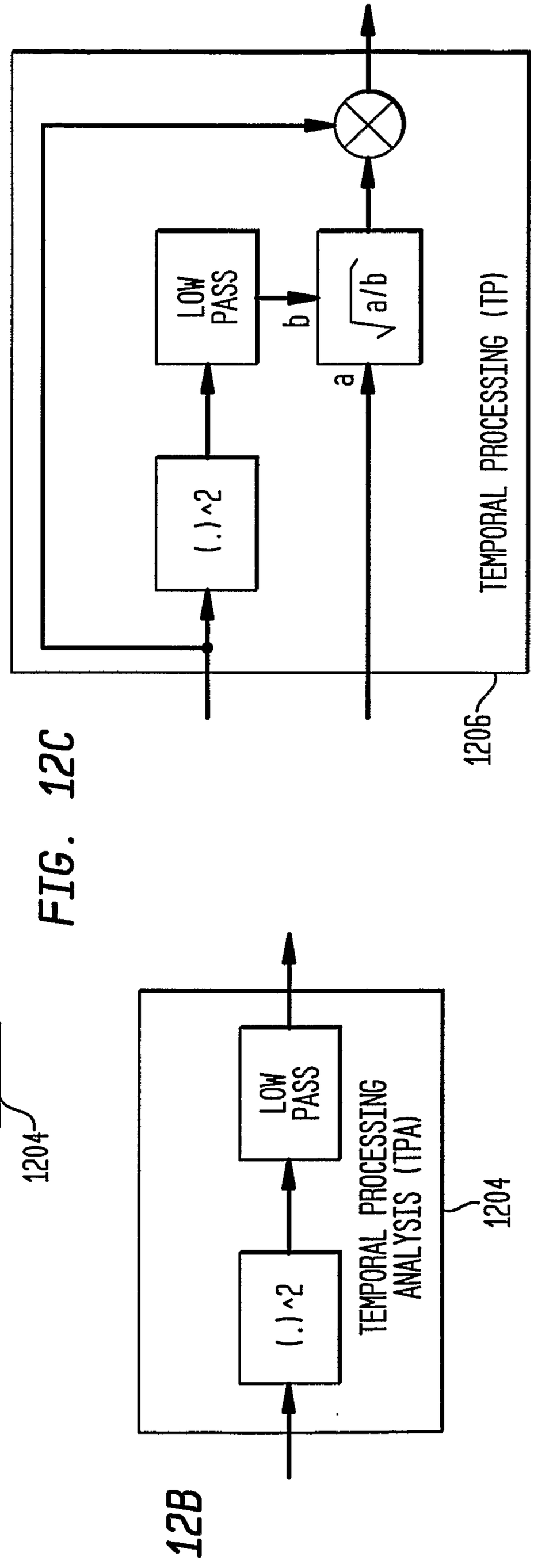


FIG. 12C

FIG. 12B

FIG. 13A

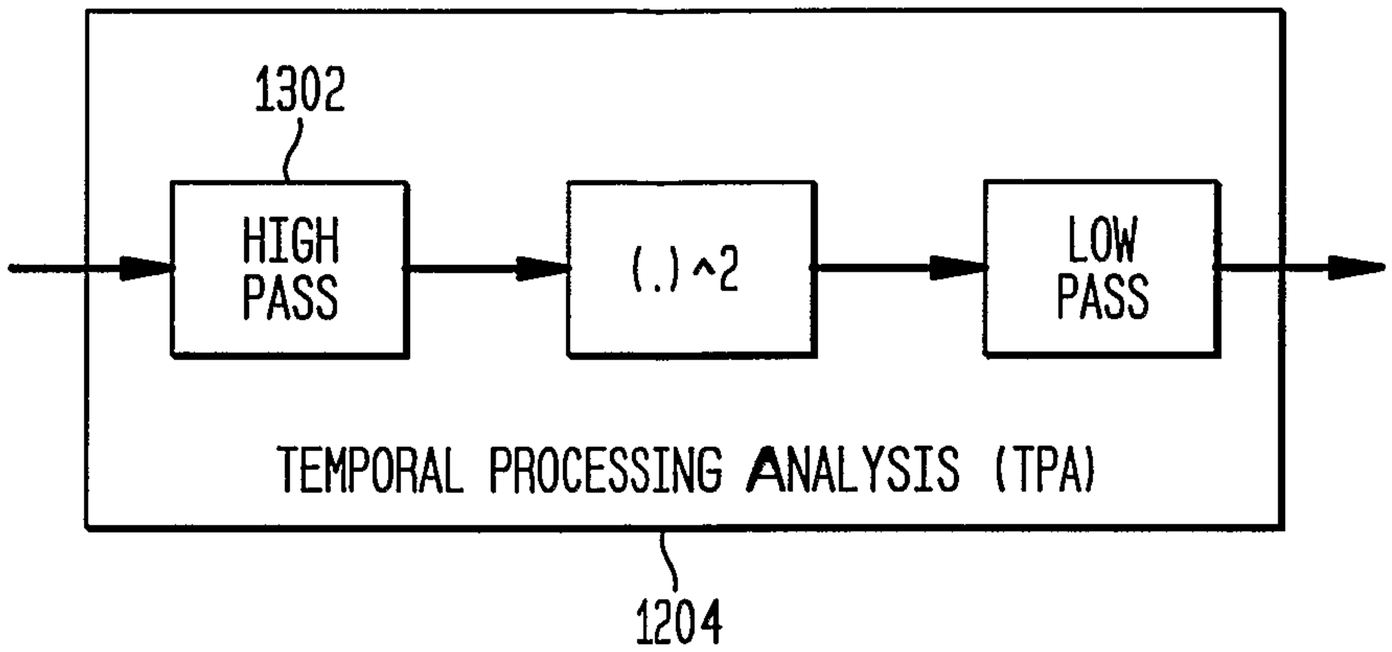


FIG. 13B

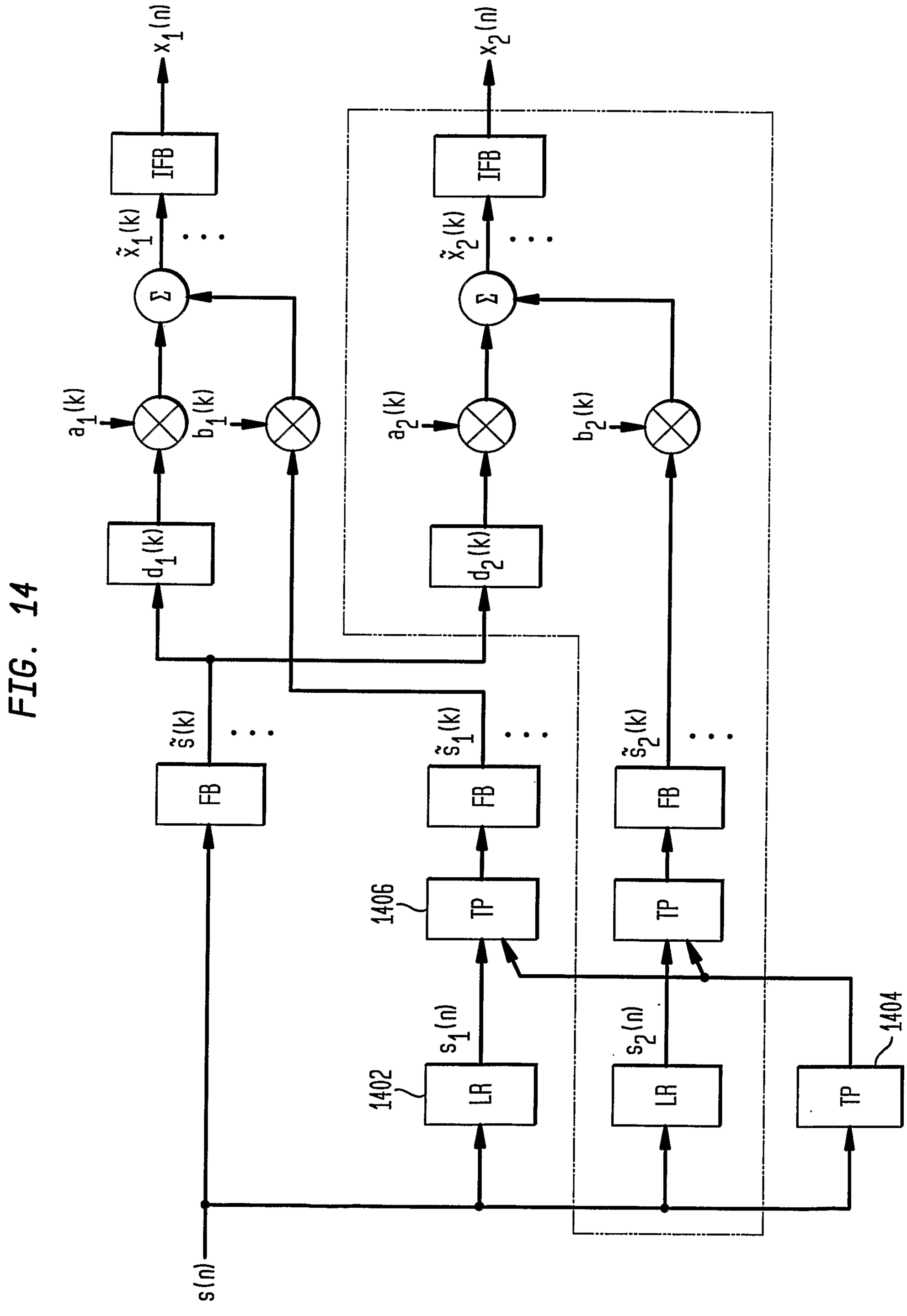

$x_1(n)$
$x_2(n)$
IFB
IFB
$\tilde{x}_1(k)$
$\tilde{x}_2(k)$
Σ
Σ
$a_1(k)$
$b_1(k)$
$a_2(k)$
$b_2(k)$
$d_1(k)$
$d_2(k)$
$\tilde{s}(k)$
$\tilde{s}_1(k)$
$\tilde{s}_2(k)$
FB
FB
FB
1406
TP
TP
$s_1(n)$
$s_2(n)$
1404
1402
LR
LR
TP
s(n)

FIG. 14

FIG. 15

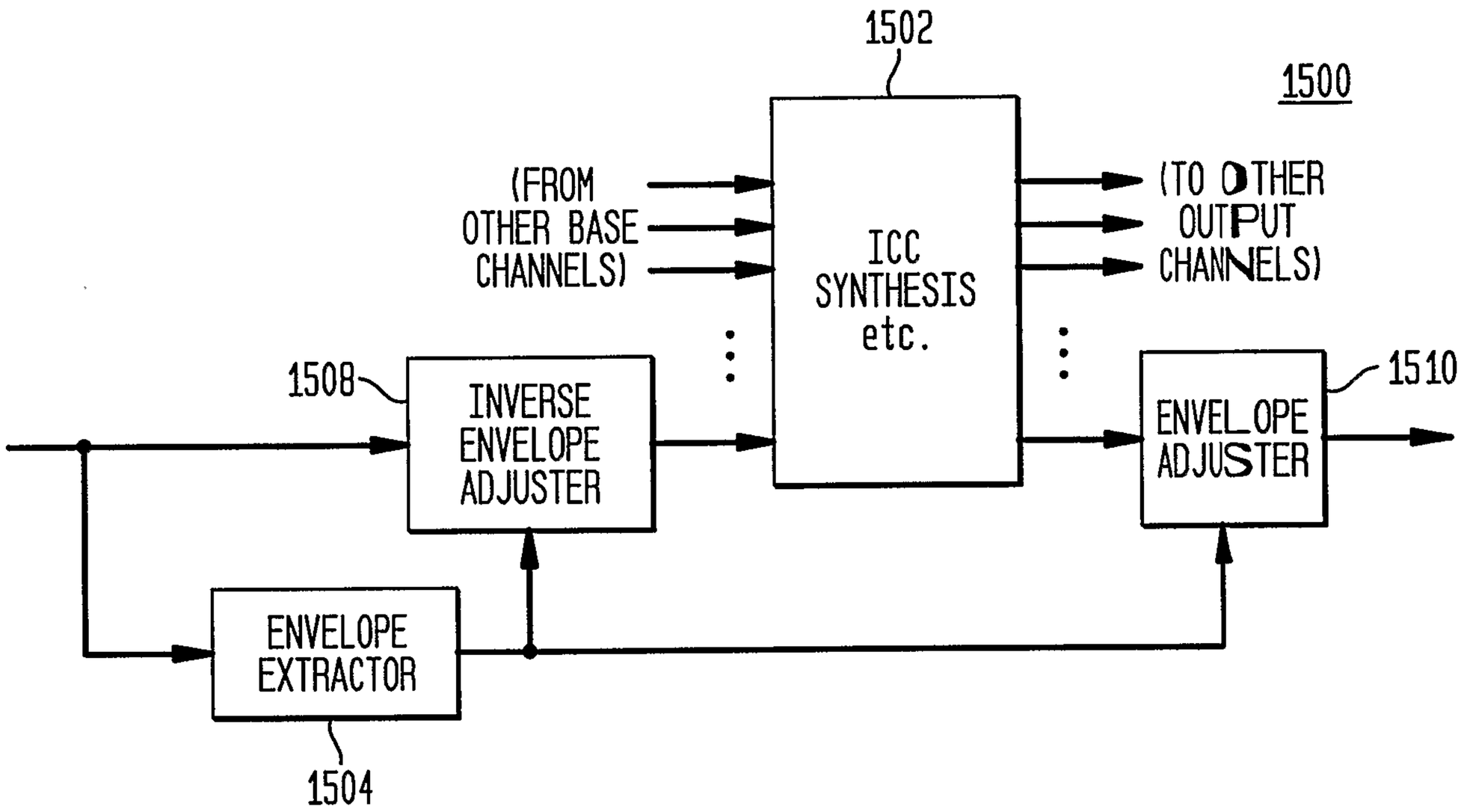
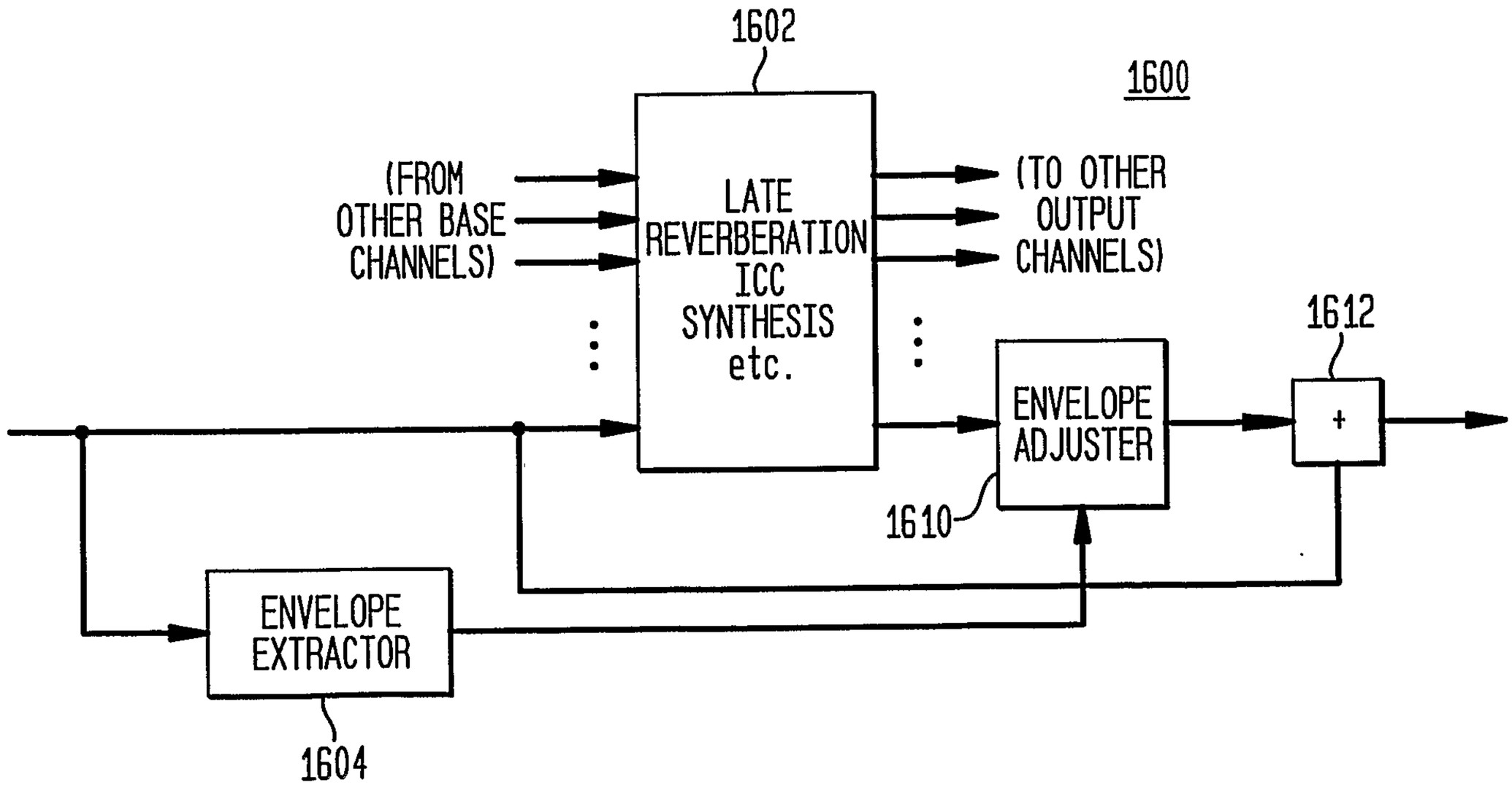

1502
1500
(FROM OTHER BASE CHANNELS)
ICC SYNTHESIS etc.
(TO OTHER OUTPUT CHANNELS)
1508
INVERSE ENVELOPE ADJUSTER
1510
ENVELOPE ADJUSTER
ENVELOPE EXTRACTOR
1504

FIG. 16


1602
1600
(FROM OTHER BASE CHANNELS)
LATE REVERBERATION ICC SYNTHESIS etc.
(TO OTHER OUTPUT CHANNELS)
1612
ENVELOPE ADJUSTER
+
1610
ENVELOPE EXTRACTOR
1604

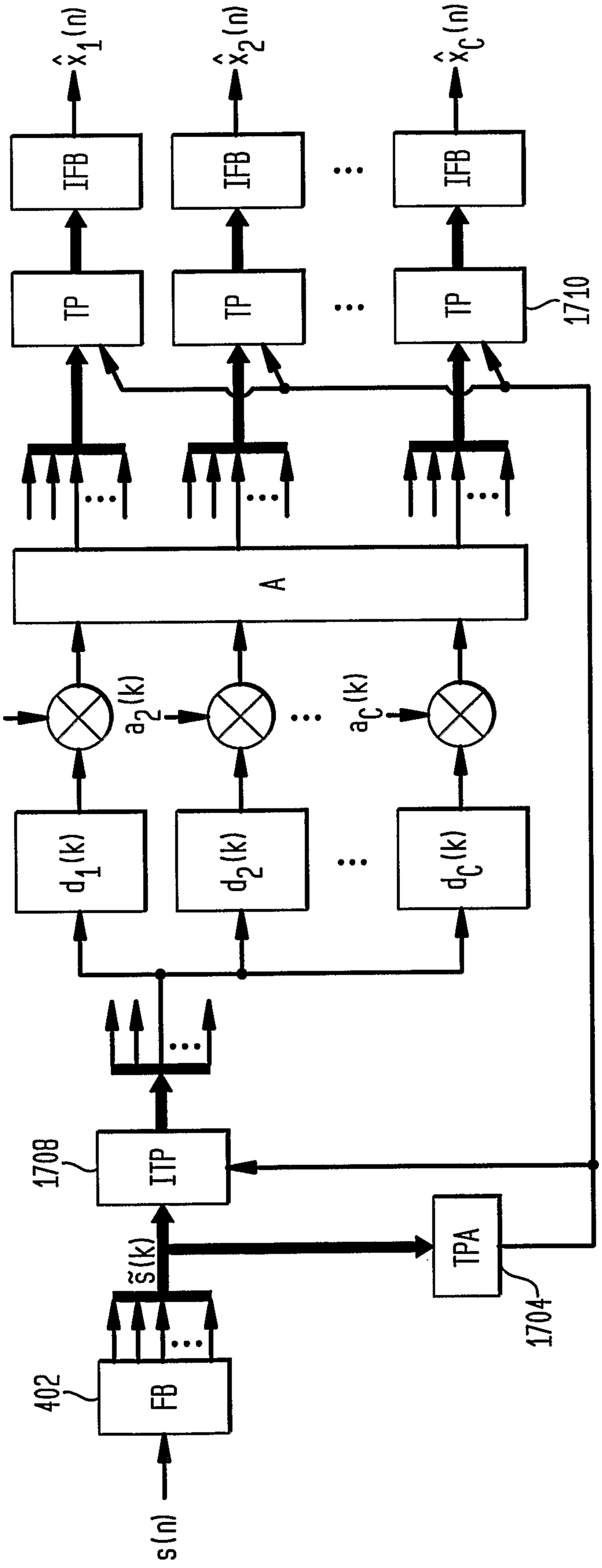

$\hat{x}_1(n)$
$\hat{x}_2(n)$
$\hat{x}_C(n)$
IFB
IFB
IFB
TP
TP
TP
1710
A
$a_1(k)$
$a_2(k)$
$a_C(k)$
$d_1(k)$
$d_2(k)$
$d_C(k)$
1708
ITP
$\tilde{s}(k)$
TPA
1704
402
FB
s(n)

FIG. 17

FIG. 18A

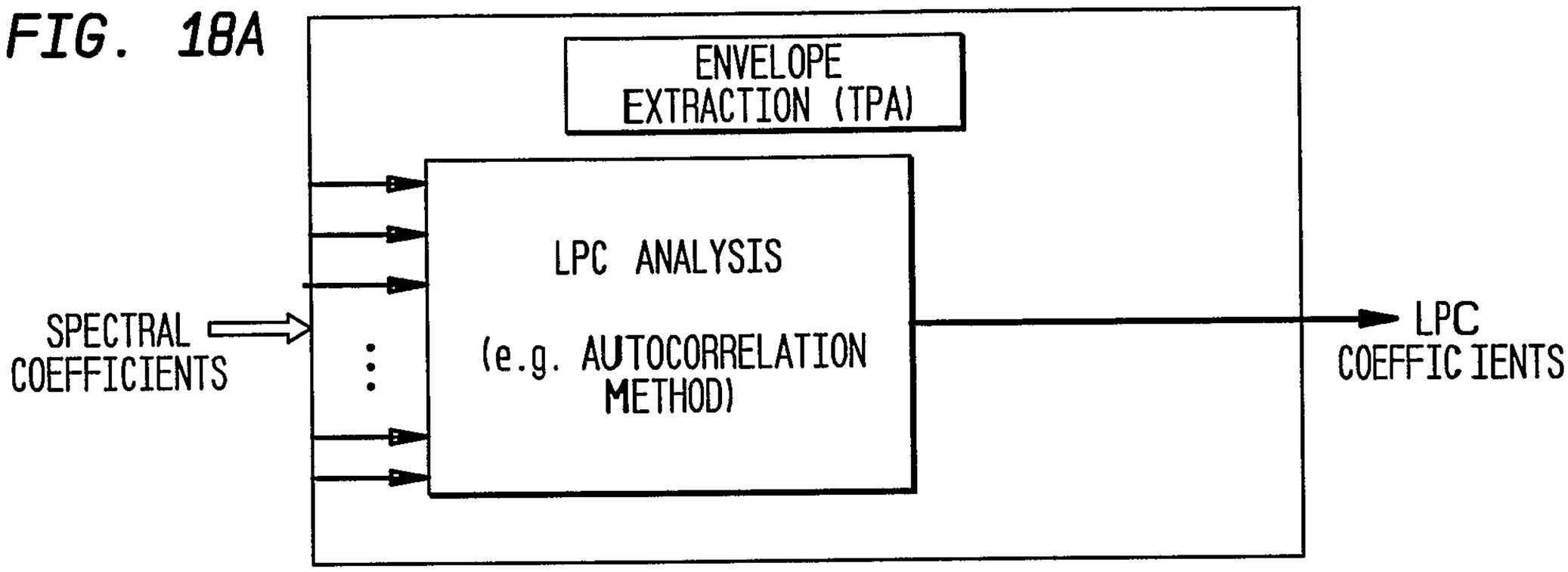


FIG. 18B

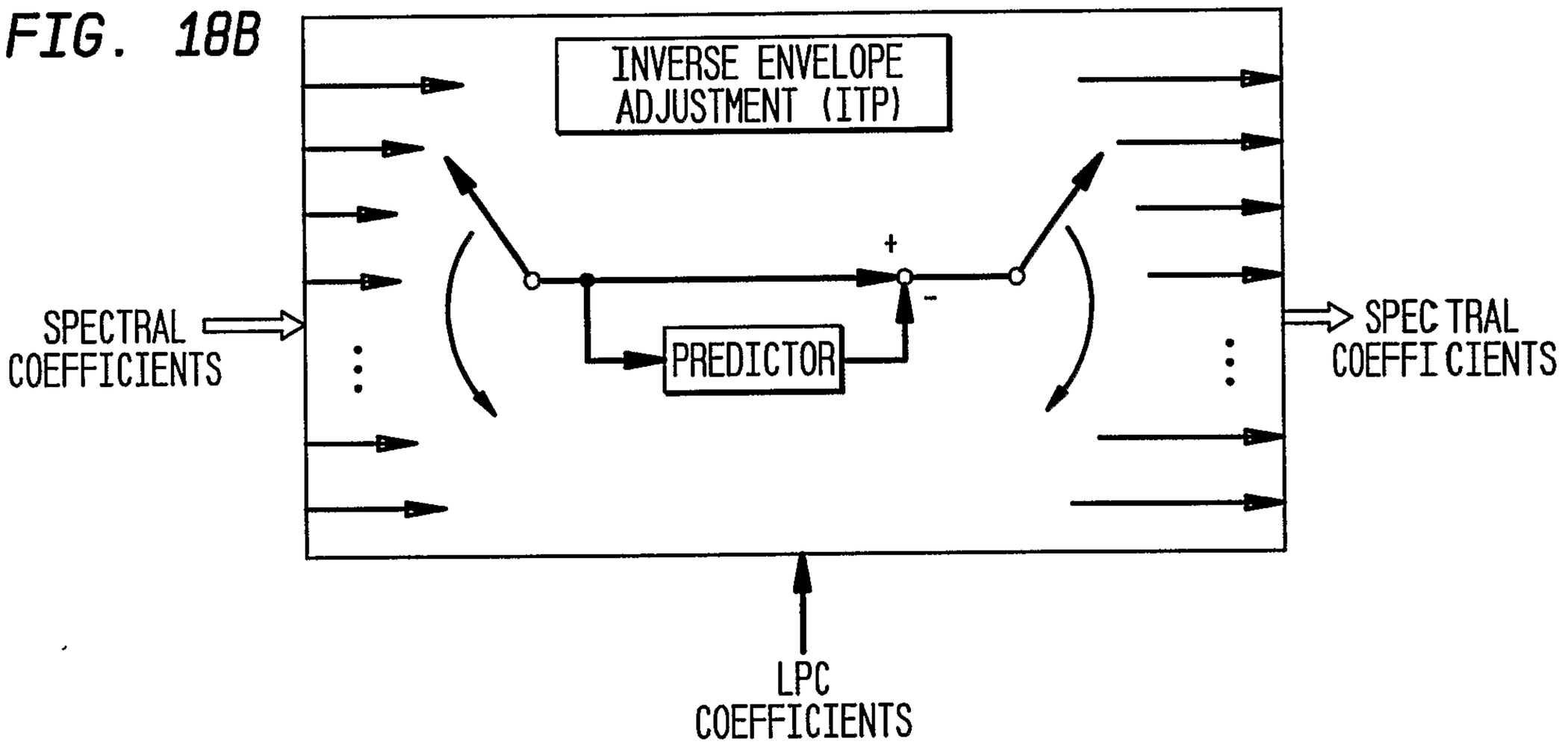
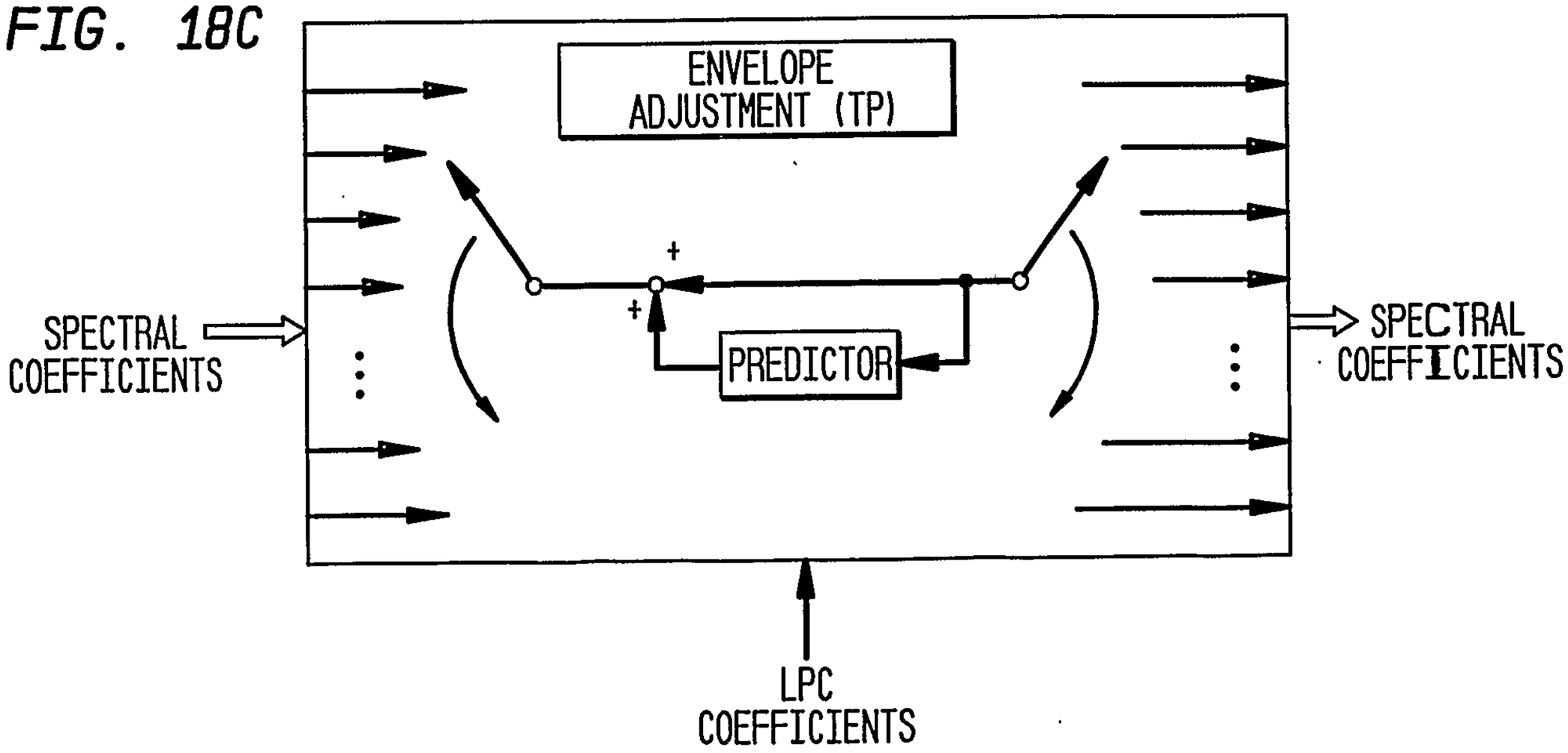
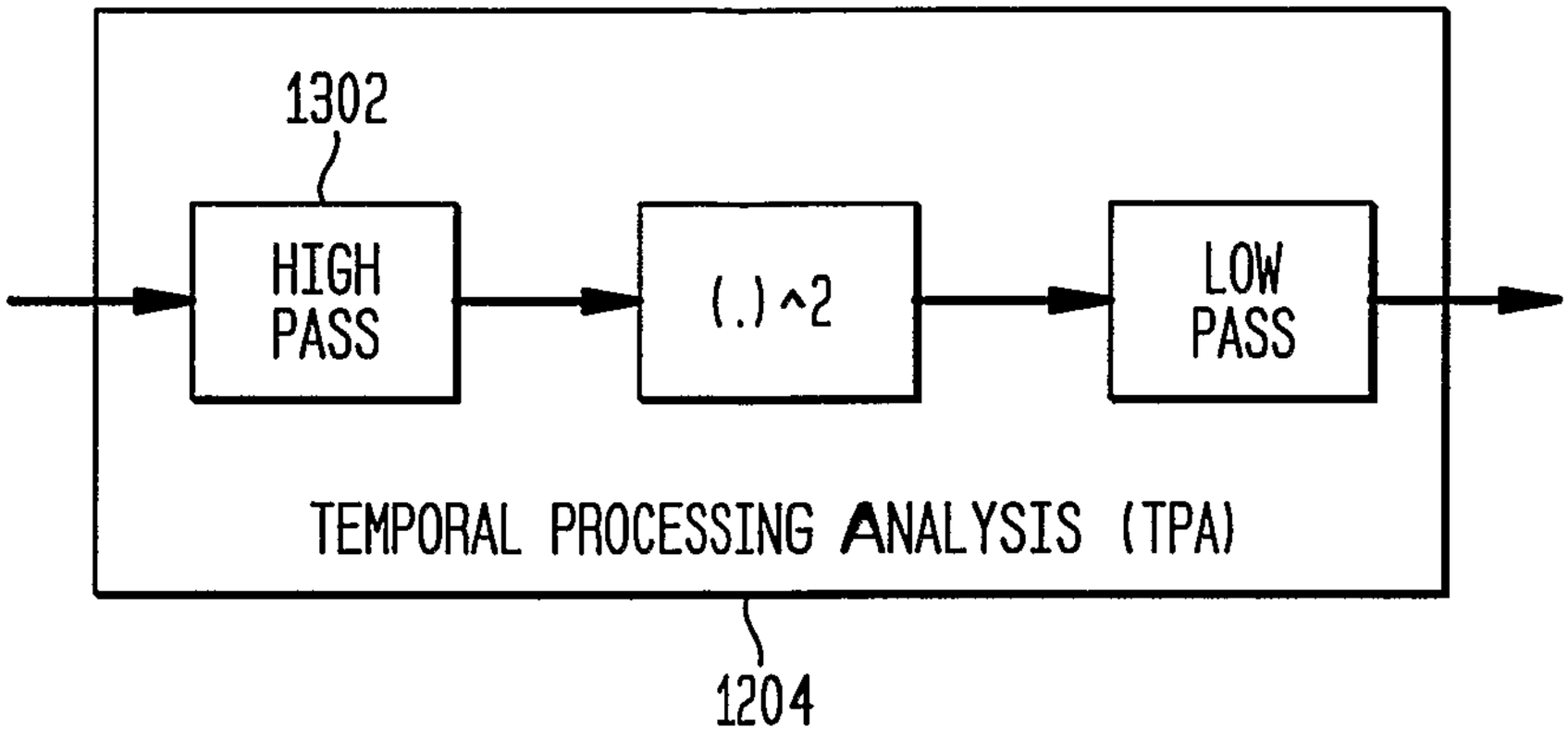


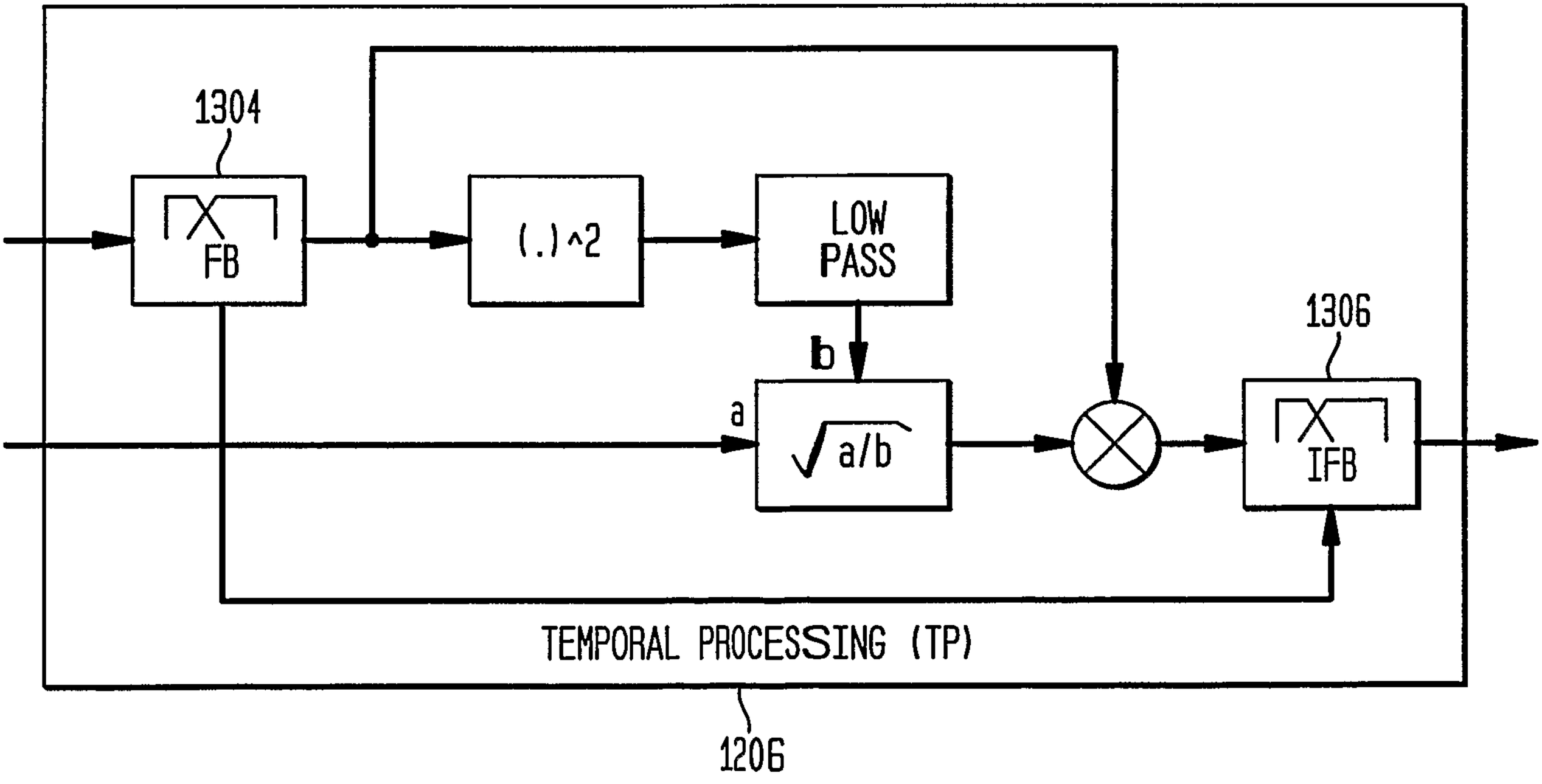
FIG. 18C



A


1302
HIGH PASS
(.)^2
LOW PASS
TEMPORAL PROCESSING ANALYSIS (TPA)
1204

B


1304
FB
(.)^2
LOW PASS
b
a
√a/b
1306
IFB
TEMPORAL PROCESSING (TP)
1206