

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2014年5月22日(22.05.2014)

WIPO | PCT

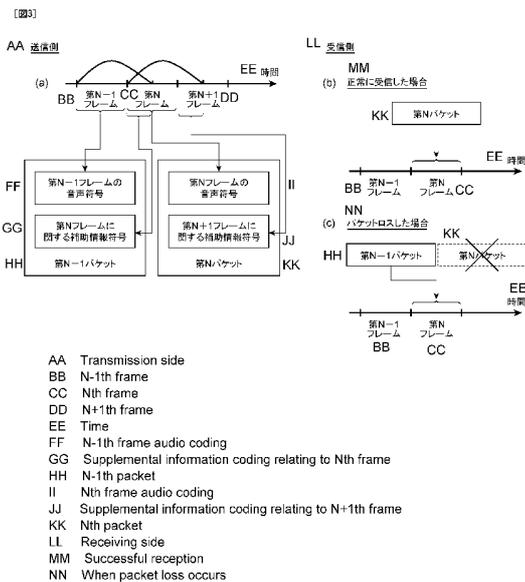
(10) 国際公開番号
WO 2014/077254 A1

- (51) 国際特許分類:
G10L 19/005 (2013.01) G10L 19/09 (2013.01)
G10L 19/00 (2013.01) H03M 7/30 (2006.01)
- (21) 国際出願番号: PCT/JP2013/080589
- (22) 国際出願日: 2013年11月12日(12.11.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2012-251646 2012年11月15日(15.11.2012) JP
- (71) 出願人: 株式会社NTTドコモ(NTT DOCOMO, INC.) [JP/JP]; 〒1006150 東京都千代田区永田町二丁目1番1号 Tokyo (JP).
- (72) 発明者: 堤 公孝(TSUTSUMI Kimitaka); 〒1006150 東京都千代田区永田町二丁目1番1号 山王パークタワー 株式会社NTTドコモ 知的財産部内 Tokyo (JP). 菊入 圭(KIKUIRI Kei); 〒1006150 東京都千代田区永田町二丁目1番1号 山王パークタワー 株式会社NTTドコモ 知的財産部内 Tokyo (JP). 山口 貴史(YAMAGUCHI Atsushi); 〒1006150 東京都千代田区永田町二丁目1番1号 山王パークタワー 株式会社NTTドコモ 知的財産部内 Tokyo (JP).
- (74) 代理人: 長谷川 芳樹, 外(HASEGAWA Yoshiki et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号丸の内 M Y P L A Z A (明治安田生命ビル) 9階 創英国際特許法律事務所 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[続葉有]

(54) Title: AUDIO CODING DEVICE, AUDIO CODING METHOD, AUDIO CODING PROGRAM, AUDIO DECODING DEVICE, AUDIO DECODING METHOD, AND AUDIO DECODING PROGRAM

(54) 発明の名称: 音声符号化装置、音声符号化方法、音声符号化プログラム、音声復号装置、音声復号方法及び音声復号プログラム



(57) Abstract: An objective of the present invention is, when packet loss occurs in audio coding, to recover audio quality without increasing algorithm delay. An audio signal transmission device which codes an audio signal comprises an audio coding unit which codes an audio signal, and a supplemental information coding unit which, from a prefetch, computes and codes supplemental information. Conversely, an audio signal receiving device which decodes the audio coding and outputs the audio signal comprises: an audio coding buffer which detects packet loss on the basis of a receiving state of an audio packet; an audio parameter decoding unit which decodes the audio coding upon successful receipt of the audio packet; a supplemental information decoding unit which decodes the supplemental information coding upon successful receipt of the audio packet; a supplemental information accumulation unit which accumulates supplemental information which is obtained by decoding supplemental information coding; an audio parameter loss processing unit which outputs an audio parameter when the audio packet loss is detected; and an audio compositing unit which composites the decoded audio from the audio parameter.

(57) 要約:

[続葉有]

WO 2014/077254 A1



添付公開書類:

- 国際調査報告 (条約第 21 条(3))

音声符号化におけるパケットロスにおいて、アルゴリズム遅延を増加させずに音声品質を回復することを目的とする。音声信号を符号化する音声信号送信装置は、音声信号を符号化する音声符号化部と、先読み信号から補助情報を算出して符号化する補助情報符号化部と、を備える。一方、音声信号を復号して音声信号を出力する音声信号受信装置は、音声パケットの受信状態からパケットロスを検出する音声符号バッファと、音声パケット正常受信時に音声信号を復号する音声パラメータ復号部と、音声パケット正常受信時に補助情報符号を復号する補助情報復号部と、補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積部と、音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理部と、音声パラメータから復号音声合成する音声合成部と、を備える。

明 細 書

発明の名称：

音声符号化装置、音声符号化方法、音声符号化プログラム、音声復号装置、音声復号方法及び音声復号プログラム

技術分野

[0001] 本発明は、音声パケットを、IP網や移動体通信網経由で伝送する際のエラー隠蔽に関するものであり、さらに詳しくは、エラー隠蔽を実現する、高精度パケットロス隠蔽信号生成のための音声符号化装置、音声符号化方法、音声符号化プログラム、音声復号装置、音声復号方法及び音声復号プログラムに関する。

背景技術

[0002] 音声・音響信号（以下「音声信号」と総称する）をIP網や移動体通信において伝送する際には、ある一定時間長ごとに音声信号を符号化して音声パケットを生成し、通信網経由で伝送する。受信側では、通信網を通じて音声パケットを受け取り、受信側のサーバ、MCU（Multipoint Control Unit）、端末等において復号し、復号音声信号とする。

[0003] 音声信号は、通常、デジタル形式で収録される。すなわち、1秒間あたりサンプリング周波数と同数の数列として計測・蓄積する。この数列の個々の要素をサンプルと呼ぶ。音声符号化においては、決まったサンプル数の音声信号が、内蔵したバッファに所定量蓄積するたびにバッファ内の音声信号を符号化する。上記所定のサンプル数をフレーム長といい、フレーム長と同数のサンプルの集合をフレームと呼ぶ。例えば、32kHzのサンプリング周波数において、フレーム長を20msとした場合、フレーム長は640サンプルである。なお、バッファの長さは1フレームより長くてもよい。

[0004] 通信網を通じて音声パケットを伝送する際には、通信網の輻輳状態等に起因して、一部の音声パケットが失われる、あるいは音声パケットに書き込まれた情報の一部に誤りが生じるといった現象（いわゆるパケットロス）が起

こりうる。そのような場合には、受信側において音声パケットを正しく復号することができないため、所望の復号音声信号を得ることが出来ない。また、パケットロスが生じた音声パケットに対応する復号音声信号は雑音として知覚されるため、受聴する人間に対して与える主観品質を著しく損なう。

[0005] 上記のような不都合を解消するため、パケットロスにより失われた部分の音声音響信号を補間する技術として、パケットロス隠蔽技術が用いられる。パケットロス隠蔽技術には、受信側のみでパケットロス隠蔽を行う「補助情報を用いないパケットロス隠蔽技術」と、送信側でパケットロス隠蔽に役立つパラメータを求めた上で、受信側に伝送し、受信側では受け取ったパラメータを利用してパケットロス隠蔽を行う「補助情報を用いるパケットロス隠蔽技術」がある。

[0006] このうち「補助情報を用いないパケットロス隠蔽技術」では、例えば、非特許文献1の技術のように過去に正常に受信したパケットに含まれていた復号音声信号をピッチ単位でコピーした上で、予め決めた減衰係数を乗算することにより、パケットロスした部分に対応する音声信号を生成する。「補助情報を用いないパケットロス隠蔽技術」は、パケットロスした部分の音声の性質がパケットロスする直前の音声と似ていることを前提としているため、パケットロスした部分がロスする直前の音声と異なる性質を持つ場合や、パワーが急激に変化する場合に、十分な隠蔽効果を発揮することができない。

[0007] 一方、「補助情報を用いるパケットロス隠蔽技術」には、特許文献1のように、送信側でパケットロス隠蔽に必要なパラメータを符号化して伝送し、受信側でのパケットロス隠蔽に利用する技術がある。特許文献1では、主要符号化・冗長符号化の2つの符号化方式により音声を符号化する。冗長符号化は、主要符号化よりも低いビットレートで、主要符号化が符号化するフレームより1つ前のフレームを符号化する（図1（a）参照）。例えば、第N番目のパケットには、第Nフレームを主要符号化により符号化して得た音声符号と、第N-1フレームを冗長符号化により符号化して得た補助情報符号を含めて伝送する。

[0008] 受信側では、時間的に連続する2つ以上のパケットの到着を待ってから、時間的に早い方のパケットを復号して、復号音声信号を得る。例えば、第Nフレームに対応する信号を得る場合、第N+1パケットが到着するのを待ってから復号を行う。第Nパケット、第N+1パケットを正常に受信した場合、第Nパケットに含まれる音声符号を復号することにより第Nフレームの音声信号を得る（図1（b）参照）。一方、パケットロスした場合（第Nパケットがロスした状況で、第N+1パケットを得た場合）、第N+1パケットに含まれる補助情報符号を復号することにより、第Nフレームの音声信号を得ることができる（図1（c）参照）。

[0009] 特許文献1の方法では、復号対象となるパケットが到着しても、さらに1パケット以上到着するまで、復号を待たなければならず、アルゴリズム遅延が1パケット分以上増加する。従って、特許文献1の方法では、パケットロス隠蔽による音質向上は期待できるものの、アルゴリズム遅延が増加し、音声通話品質が低下する。

[0010] さらに、上記のようなパケットロス隠蔽技術をCELP（Code Excited Linear Prediction、符号励振線形予測）符号化に対して適用する際には、CELPの動作の特徴に起因する別の課題も生じる。CELPは、線形予測に基づく音声モデルであり、音声信号を高精度にかつ高い圧縮率で符号化できるため、多くの国際標準で用いられている。

[0011] CELPでは、全極型合成フィルタによって励振信号 $e(n)$ をフィルタリングすることにより合成される。すなわち、次式に従い音声信号 $s(n)$ を合成する。

[数1]

$$s(n) = e(n) - \sum_{i=1}^P a(i) \cdot s(n-i)$$

$a(i)$ は線形予測係数（LP係数）であり、次数として例えば $P=16$ などの値を用いる。

[0012] 励振信号は、適応符号帳と呼ばれるバッファに蓄積される。あらたなフレ

ームの音声を合成するに当たっては、ピッチラグと呼ばれる位置情報を元に、適応符号帳から読み出された適応符号帳ベクトルと、励振信号の時間変化を表す固定符号帳ベクトルを加算することにより、新たに励振信号を生成する。新たに生成された励振信号は、適応符号帳に蓄積されると同時に、全極型合成フィルタによりフィルタリングされ、復号信号が合成される。

[0013] CELPでは、全てのフレームについてLP係数が算出される。LP係数を算出するに当たっては、10ms程度の先読み信号が必要である。すなわち、符号化対象フレームに加えて、先読み信号もバッファに蓄積した上で、LP係数算出およびその後の処理を実施する（図2参照）。各フレームは4つ程度のサブフレームに分割され、サブフレーム単位で、上記ピッチラグ算出、適応符号帳ベクトル算出、固定符号帳ベクトル算出、適応符号帳更新といった処理を行う。上記サブフレーム単位の処理にあたっては、LP係数も補間処理を行うことにより、サブフレーム毎に異なる係数に変化させる。また、LP係数は量子化・補間処理の都合上、LP係数の等価表現であるISP (Impedance Spectral Pair) パラメータ、ISF (Impedance Spectral Frequency) パラメータに変換した上で符号化される。LP係数とISPパラメータ・ISFパラメータの相互変換の処理手順については、非特許文献2に記載されている。

[0014] CELP符号化では、符号化側と復号側がそれぞれ適応符号帳をもち、それら適応符号帳が常に同期することを前提として符号化・復号を行う。パケットを正常に受信し、復号が正常に行われる状況では、符号化側の適応符号帳と復号側の適応符号帳は同期しているが、一度パケットロスが起これば、適応符号帳の同期を取ることができなくなる。

[0015] 例えば、ピッチラグとして用いる値が、符号化側と復号側で異なれば、適応符号帳ベクトルは時間的にずれたものとなる。この適応符号帳ベクトルで適応符号帳の更新を行うため、次のフレームを正常に受信したとしても、符号化側で求まる適応符号帳ベクトルと、復号側で求まる適応符号帳ベクトルは一致せず、適応符号帳の同期が回復することはない。このような適応符号

帳の不一致によって、パケットロスしたフレーム以降、数フレームに渡って音質劣化が生じる。

[0016] CELP符号化におけるパケットロス隠蔽において、より高度な技術として特許文献2の手法がある。特許文献2では、パケットロスによる影響が大きい特定のフレームにおいて、ピッチラグや適応符号帳ゲインの代わりに、遷移モードコードブックのインデクスを伝送する。特許文献2では、パケットロスによる影響が大きいフレームとして、遷移フレーム（無音の音声セグメントから有音の音声セグメントへの遷移、あるいは2つの母音間の遷移）に着目している。この遷移フレームにおいて、遷移モードコードブックを用いた励振信号生成を行うことにより、過去の適応符号帳に非依存の励振信号を生成し、過去のパケットロスによる適応符号帳不一致から回復することができる。

[0017] 特許文献2の方法は、例えば長めの母音が継続するフレームなどでは、遷移フレームコードブックを利用しないため、従来と同様、適応符号帳の不一致から回復することができない。さらに、遷移モードコードブックを含むパケットが失われた場合には、これまでと同様に、パケットロス以降のフレームにもロスの影響が残ってしまう。これは、遷移モードコードブックを含むパケットの次のパケットが失われた場合も同様である。

[0018] 遷移フレームコードブックのような過去のフレームに依存しないコードブックを、全てのフレームに適応することもできるが、符号化の効率が著しく落ちるため、低ビットレート・高音質を達成できない。

先行技術文献

特許文献

[0019] 特許文献1：特表2003-533916号公報

特許文献2：特表2010-507818号公報

非特許文献

[0020] 非特許文献1：ITU-T G. 711 Appendix I

非特許文献2：3GPP TS26-191

非特許文献3：3 G P P T S 2 6 - 1 9 0

非特許文献4：I T U - T G . 7 1 8

発明の概要

発明が解決しようとする課題

[0021] 特許文献1の方法を用いると、復号対象のパケットが到着した後、後続のパケット到着を待ってから復号を開始する。そのため、パケットロス隠蔽による音質向上はあるものの、アルゴリズム遅延が増加し、通話品質が低下する。

[0022] C E L P 符号化におけるパケットロス時に、符号化部・復号部間における適応符号帳の不一致が原因で、音声品質の劣化が生じる。特許文献2のような方法により、適応符号帳の不一致から回復することもできるが、遷移フレームの直前以外のフレームがロスした場合に十分な効果が得られない。

[0023] 本発明は、上記課題の解決のためになされたものであり、音声符号化におけるパケットロスにおいて、アルゴリズム遅延を増加させずに音声品質を回復する音声符号化装置、音声符号化方法、音声符号化プログラム、音声復号装置、音声復号方法及び音声復号プログラムを提供することを目的とする。

課題を解決するための手段

[0024] 上記目的を達成するために、本発明の一側面に係る音声符号化装置は、音声信号を符号化する音声符号化装置であって、音声信号を符号化する音声符号化部と、先読み信号から補助情報を算出して符号化する補助情報符号化部と、を備える。

[0025] 補助情報は、先読み信号におけるピッチラグに関すること、又は、補助情報は、先読み信号におけるピッチゲインに関すること、又は、先読み信号におけるピッチラグおよびピッチゲインに関することとしてもよい。また、補助情報は、補助情報の利用可否に関する情報を含むこととしてもよい。

[0026] 補助情報符号化部は、先読み信号部分について補助情報を算出して符号化し、さらに隠蔽信号を生成し、入力音声信号と補助情報符号化部が出力する隠蔽信号の誤差信号を符号化する誤差信号符号化部と、入力音声信号そのも

のを符号化する主要符号化部と、をさらに備えることとしてもよい。

[0027] また、本発明の一側面に係る音声復号装置は、音声符号を復号して音声信号を出力する音声復号装置であって、音声パケットの受信状態からパケットロスを検出する音声符号バッファと、音声パケット正常受信時に音声符号を復号する音声パラメータ復号部と、音声パケット正常受信時に補助情報符号を復号する補助情報復号部と、補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積部と、音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理部と、音声パラメータから復号音声を作成する音声合成部と、を備える。

[0028] 補助情報は、先読み信号におけるピッチラグに関すること、又は、先読み信号におけるピッチゲインに関すること、又は、先読み信号におけるピッチラグおよびピッチゲインに関することとしてもよい。また、補助情報は、補助情報の利用可否に関する情報を含むこととしてもよい。

[0029] 補助情報復号部は、補助情報符号を復号して補助情報を作成し、かつ補助情報を利用して先読み部分に関する隠蔽信号を作成し、音声信号と隠蔽信号との誤差信号に関する符号を復号する誤差復号部と、音声信号に関する符号を復号する主要復号部と、補助情報復号部が出力した隠蔽信号を蓄積する隠蔽信号蓄積部と、をさらに備えることとしてもよい。

[0030] 音声パケット正常受信時には、隠蔽信号蓄積部から読み出した隠蔽信号と、誤差復号部が出力した復号誤差信号を加算することにより復号信号の一部を生成し、補助情報復号部が出力した隠蔽信号により隠蔽信号蓄積部を更新する、こととしてもよい。

[0031] 音声パケットロス検出時には、隠蔽信号蓄積部から読み出した隠蔽信号を、復号信号の一部あるいは全部とする、こととしてもよい。

[0032] 音声パケットロス検出時には、音声パラメータ紛失処理部により予測した音声パラメータを用いて復号信号を生成し、その一部を用いて隠蔽信号蓄積部を更新する、こととしてもよい。

[0033] 音声パラメータ紛失処理部は、音声パケットロス検出時に、補助情報蓄積

部から読み出した補助情報を、音声パラメータの予測値の一部として利用する、こととしてもよい。

[0034] 音声合成部は、音声パケットロス検出時に、補助情報蓄積部から読み出した補助情報を用いて、音声パラメータの1つである適応符号帳ベクトルを修正する、こととしてもよい。

[0035] また、本発明の一側面に係る音声符号化方法は、音声信号を符号化する音声符号化装置による音声符号化方法であって、音声信号を符号化する音声符号化ステップと、先読み信号から補助情報を算出して符号化する補助情報符号化ステップと、を含む。

[0036] また、本発明の一側面に係る音声復号方法は、音声符号を復号して音声信号を出力する音声復号装置による音声復号方法であって、音声パケットの受信状態からパケットロスを検出する音声符号バッファステップと、音声パケット正常受信時に音声符号を復号する音声パラメータ復号ステップと、音声パケット正常受信時に補助情報符号を復号する補助情報復号ステップと、補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積ステップと、音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理ステップと、音声パラメータから復号音声を合成する音声合成ステップと、を含む。

[0037] また、本発明の一側面に係る音声符号化プログラムは、コンピュータを、音声信号を符号化する音声符号化部と、先読み信号から補助情報を算出して符号化する補助情報符号化部、として機能させる。

[0038] また、本発明の一側面に係る音声復号プログラムは、コンピュータを、音声パケットの受信状態からパケットロスを検出する音声符号バッファと、音声パケット正常受信時に音声符号を復号する音声パラメータ復号部と、音声パケット正常受信時に補助情報符号を復号する補助情報復号部と、補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積部と、音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理部と、音声パラメータから復号音声を合成する音声合成部、として機能させる。

発明の効果

[0039] 音声符号化におけるパケットロスにおいて、アルゴリズム遅延を増加させずに音声品質を回復することができる。特に、C E L P符号化において、パケットロス時におこる適応符号帳の劣化を低減し、パケットロス時の音声品質を改善することができる。

図面の簡単な説明

[0040] [図1]特許文献1に記載の先行発明におけるパケットと復号信号の時間的關係を示す図である。

[図2]C E L P符号化におけるL P分析対象信号と先読み信号の時間的關係を示す図である。

[図3]本発明の実施形態におけるパケットと復号信号の時間的關係を示す図である。

[図4]本発明の実施例1における音声信号送信装置の機能構成例を示す図である。

[図5]本発明の実施例1における音声信号受信装置の機能構成例を示す図である。

[図6]本発明の実施例1における音声信号送信装置の処理手順を示す図である。

[図7]本発明の実施例1における音声信号受信装置の処理手順を示す図である。

[図8]本発明の実施例1における補助情報符号化部の機能構成例を示す図である。

[図9]本発明の実施例1における補助情報符号化部の処理手順を示す図である。

[図10]本発明の実施例1におけるL P係数算出部の処理手順を示す図である。

[図11]本発明の実施例1におけるターゲット信号算出部の処理手順を示す図である。

[図12]本発明の実施例1における音声パラメータ紛失処理部の機能構成例を示す図である。

[図13]本発明の実施例1における音声パラメータ予測の処理手順を示す図である。

[図14]本発明の実施例1の変形例1-1における励振ベクトル合成部の処理手順を示す図である。

[図15]本発明の実施例1における音声合成部の機能構成図を示す図である。

[図16]本発明の実施例1における音声合成部の処理手順を示す図である。

[図17]本発明の実施例1の変形例1-2における補助情報符号化部（補助情報出力判定部を設けた場合）の機能構成例を示す図である。

[図18]本発明の実施例1の変形例1-2における補助情報符号化部（補助情報出力判定部を設けた場合）の処理手順を示す図である。

[図19]本発明の実施例1の変形例1-2における音声パラメータ予測の処理手順を示す図である。

[図20]本発明の実施例2における音声信号送信装置の機能構成例を示す図である。

[図21]本発明の実施例2における主要符号化部の機能構成例を示す図である。

[図22]本発明の実施例2における音声信号送信装置の処理手順を示す図である。

[図23]本発明の実施例2における音声信号受信装置の機能構成例を示す図である。

[図24]本発明の実施例2における音声信号受信装置の処理手順を示す図である。

[図25]本発明の実施例2における音声合成部の機能構成図を示す図である。

[図26]本発明の実施例2における音声パラメータ復号部の機能構成例を示す図である。

[図27]本発明の実施例3における補助情報符号化部の機能構成例を示す図で

ある。

[図28]本発明の実施例3における補助情報符号化部の処理手順を示す図である。

[図29]本発明の実施例3におけるピッチラグ選定部の処理手順を示す図である。

[図30]本発明の実施例3における補助情報復号部の処理手順を示す図である。

[図31]本発明の実施形態に係る音声符号化プログラムの構成を記憶媒体と共に示す図である。

[図32]本発明の実施形態に係る音声復号プログラムの構成を記憶媒体と共に示す図である。

[図33]本発明の実施例4における補助情報符号化部の機能構成例を示す図である。

[図34]本発明の実施例4における補助情報符号化部の処理手順を示す図である。

[図35]本発明の実施例4におけるピッチラグ予測部の処理手順を示す図（その1）である。

[図36]本発明の実施例4におけるピッチラグ予測部の処理手順を示す図（その2）である。

[図37]本発明の実施例4におけるピッチラグ予測部の処理手順を示す図（その3）である。

[図38]本発明の実施例4における適応符号帳算出部の処理手順を示す図である。

[図39]本発明の実施例5における補助情報符号化部の機能構成例を示す図である。

[図40]本発明の実施例5におけるピッチラグ符号化部の処理手順を示す図である。

[図41]本発明の実施例5における補助情報復号部の処理手順を示す図である。

。

[図42]本発明の実施例5におけるピッチラグ予測部の処理手順を示す図である。

[図43]本発明の実施例5における適応符号帳算出部の処理手順を示す図である。

発明を実施するための形態

- [0041] 添付図面を参照しながら本発明の実施形態を説明する。可能な場合には、同一の部分には同一の符号を付して、重複する説明を省略する。
- [0042] 本発明の実施形態は、符号化器側で算出した補助情報を符号化して伝送し、復号側でのパケットロス隠蔽に利用する「補助情報を用いるパケットロス隠蔽技術」を実現する符号化器、および復号器である。
- [0043] 本発明の実施形態では、パケットロス隠蔽に用いる補助情報は1つ前のパケットに含める。パケットに含まれる音声符号と補助情報符号の時間的關係を図3に示す。図3からも明らかなように、本発明の実施形態における補助情報は、CELP符号化における先読み信号について求めたパラメータである（ピッチラグ、適応符号帳ゲイン等）。
- [0044] 補助情報符号を1つ前のパケットに含めたことにより、復号対象のパケットよりも後続のパケットを待つことなく、復号を行うことができる。パケットロスを検出した際も、隠蔽対象となるフレームに関する補助情報が、直前のパケットにおいて得られているので、後続のパケットを待つことなく、高精度なパケットロス隠蔽を実現することができる。
- [0045] また、補助情報として先読み信号におけるCELP符号化のパラメータを伝送することにより、パケットがロスしても適応符号帳の不一致を軽減することができる。
- [0046] 本発明の実施形態は、音声信号送信装置（音声符号化装置）、音声信号受信装置（音声復号装置）からなる。音声信号送信装置の機能構成例を図4に示し、処理手順を図6に示す。また、音声信号受信装置の機能構成例を図5に示し、処理手順を図7に示す。

- [0047] 音声信号送信装置は、図4に示す通り、音声符号化部111、補助情報符号化部112からなる。音声信号受信装置は、図5に示す通り、音声符号バッファ121、音声パラメータ復号部122、音声パラメータ紛失処理部123、音声合成部124、補助情報復号部125、補助情報蓄積部126からなる。
- [0048] 音声信号送信装置は、図6に示す処理手順により、音声信号をフレーム毎に符号化して伝送する。
- [0049] 音声符号化部111は、符号化対象フレームに対して音声パラメータを算出して、音声符号を出力する（図6ステップS131）。
- [0050] 補助情報符号化部112は、先読み信号に対して音声パラメータを算出して、補助情報符号を出力する（図6ステップS132）。
- [0051] 音声信号が終了したか否かを判定し、音声信号が終了するまで上記を繰り返す（図6ステップS133）。
- [0052] 音声信号受信装置は、図7に示す処理手順により、到着する音声パケットを復号して音声信号を出力する。
- [0053] 音声符号バッファ121は、音声パケットの到着を待って、音声符号を蓄積する。音声パケットが正常に到着した場合は、処理を音声パラメータ復号部122に切替える。一方、音声パケットが正常に到着しない場合は、処理を音声パラメータ紛失処理部123に切り替える（図7ステップS141）。
- [0054] <音声パケットを正常に受信した場合>
音声パラメータ復号部122は、音声符号を復号して音声パラメータを出力する（図7ステップS142）。
- [0055] 補助情報復号部125は、補助情報符号を復号して、補助情報を出力する。出力した補助情報は、補助情報蓄積部126に送られる（図7ステップS143）。
- [0056] 音声合成部124は、音声パラメータ復号部122が出力した音声パラメータから音声信号を合成して出力する（図7ステップS144）。

[0057] 音声パラメータ紛失処理部 1 2 3 は、音声パラメータ復号部 1 2 2 が出力した音声パラメータを、パケットロス時に備えて蓄積する（図 7 ステップ S 1 4 5）。

[0058] 音声符号バッファ 1 2 1 は、音声パケットの送信が終了したかどうかを判断し、音声パケットの送信が終了した場合は処理を停止する。音声パケットの送信が継続する間は、上記ステップ S 1 4 1 - S 1 4 6 を繰り返す（図 7 ステップ S 1 4 7）。

[0059] <音声パケットがロスした場合>

音声パラメータ紛失処理部 1 2 3 は、補助情報蓄積部 1 2 6 から補助情報を読み出し、補助情報に含まれないパラメータについては予測を行うことで、音声パラメータを出力する（図 7 ステップ S 1 4 6）。

[0060] 音声合成部 1 2 4 は、音声パラメータ紛失処理部 1 2 3 が出力した音声パラメータから音声信号を合成して出力する（図 7 ステップ S 1 4 4）。

[0061] 音声パラメータ紛失処理部 1 2 3 は、音声パラメータ紛失処理部 1 2 3 が出力した音声パラメータを、パケットロス時に備えて蓄積する（図 7 ステップ S 1 4 5）。

[0062] 音声符号バッファ 1 2 1 は、音声パケットの送信が終了したかどうかを判断し、音声パケットの送信が終了した場合は処理を停止する。音声パケットの送信が継続する間は、上記ステップ S 1 4 1 - S 1 4 6 を繰り返す（図 7 ステップ S 1 4 7）。

[0063] [実施例 1]

本実施例では、補助情報としてピッチラグを伝送し、復号側ではパケットロス隠蔽信号の生成に用いる例について記載する。

[0064] 音声信号送信装置の機能構成例を図 4 に、音声信号受信装置の機能構成例を図 5 にそれぞれ示し、音声信号送信装置の処理手順を図 6 に、音声信号受信装置の処理手順を図 7 にそれぞれ示す。

[0065] <送信側>

音声信号送信装置において、入力音声信号は、音声符号化部 1 1 1 に送ら

れる。

[0066] 音声符号化部 111 は、符号化対象フレームを、CELP 符号化により符号化する（図 6 ステップ 131）。CELP 符号化の詳細については、例えば非特許文献 3 に記載の方法を用いる。CELP 符号化の処理手順の詳細については省略する。なお、CELP 符号化では符号化側でローカルデコードが行われる。ローカルデコードとは、符号化側でも音声符号を復号して、音声合成に必要なパラメータ（ISP パラメータおよび対応する ISF パラメータ、ピッチラグ、長期予測パラメータ、適応符号帳、適応符号帳ゲイン、固定符号帳ゲイン、固定符号帳ベクトル等）を得ることである。ローカルデコードにより得られたパラメータのうち、少なくとも ISP パラメータおよび ISF パラメータのいずれか又は両方、ピッチラグ、適応符号帳が、補助情報符号化部 112 に送られる。音声符号化部 111 として、非特許文献 4 のような音声符号化を用いる場合には、さらに符号化対象フレームの性質を表すインデクスを送ってもよい。また、音声符号化部 111 として CELP 符号化以外の符号化を用いることもできる。この場合、入力信号あるいはローカルデコードにより得られた復号信号から、少なくとも ISP パラメータおよび ISF パラメータのいずれか又は両方、ピッチラグ、適応符号帳を別途算出し、補助情報符号化部 112 に伝送する。

[0067] 補助情報符号化部 112 は、音声符号化部 111 で算出したパラメータと先読み信号を用いて補助情報符号を算出する（図 6 ステップ 132）。補助情報符号化部 112 は、図 8 に示す通り、LP 係数算出部 151、ターゲット信号算出部 152、ピッチラグ算出部 153、適応符号帳算出部 154、励振ベクトル合成部 155、適応符号帳バッファ 156、合成フィルタ 157、ピッチラグ符号化部 158 からなる。補助情報符号化部の処理手順を図 9 に示す。

[0068] LP 係数算出部 151 は、音声符号化部 111 で算出した ISF パラメータと、過去数フレームにおいて算出した ISF パラメータを用いて LP 係数を算出する（図 9 ステップ 161）。LP 係数算出部 151 の処理手順を図

10に示す。

[0069] 最初に、音声符号化部111から得たISFパラメータを用いてバッファを更新する(図10ステップ171)。次に、先読み信号におけるISFパラメータ

[数2]

$$\dot{\omega}_i$$

を算出する。ISFパラメータ

[数3]

$$\dot{\omega}_i$$

は次式により算出する(図10ステップ172)。

[数4]

$$\dot{\omega}_i = \alpha \omega_i^{(-1)} + (1 - \alpha) \bar{\omega}_i$$

[数5]

$$\bar{\omega}_i = \beta \omega_i^C + (1 - \beta) \frac{\omega_i^{(-3)} + \omega_i^{(-2)} + \omega_i^{(-1)}}{3}$$

ここで、

[数6]

$$\omega_i^{(-j)}$$

はバッファに格納されたjフレーム前のISFパラメータである。

また、

[数7]

$$\omega_i^C$$

は事前に学習等で求めた発話区間でのISFパラメータである。 β は定数であり、例えば0.75のような値とすることができるが、これには限られない。また、 α も定数であり、0.9のような値とすることができるが、これに限定されない。

[数8]

$$\omega_i^c, \alpha, \beta$$

は、例えば非特許文献4記載のISFコンシールメントのように、符号化対象フレームの性質を表すインデクスにより変化させてもよい。

[0070] 次に、

[数9]

$$\dot{\omega}_i$$

が

[数10]

$$0 < \dot{\omega}_0 < \dot{\omega}_1 < \dots < \dot{\omega}_{14}$$

を満たすよう*i*の値を並び替え、隣り合う

[数11]

$$\dot{\omega}_i$$

同士が接近しすぎないように調整する。

[数12]

$$\dot{\omega}_i$$

の値を調整する手順には、例えば非特許文献4（式151）を用いることができる（図10ステップ173）。

[0071] 次に、ISFパラメータ

[数13]

$$\dot{\omega}_i$$

をISPパラメータに変換した上で、サブフレーム毎に補間する。ISFパラメータからISPパラメータを算出する方法として非特許文献4の6.4.4節記載の方法を、補間の方法としては、非特許文献4の6.8.3節に記載の処理手順を用いることができる（図10ステップ174）。

[0072] 次にサブフレーム毎のISPパラメータをLP係数

[数14]

$$\dot{a}_i^j \quad (0 < i \leq P, 0 \leq j < M_{la})$$

に変換する。ここで、先読み信号に含まれるサブフレームの数を M_{la} とした。ISPパラメータからLP係数への変換には、非特許文献4の6.4.5節に記載の処理手順を用いることができる(図10ステップ175)。

[0073] ターゲット信号算出部152は、LP係数

[数15]

$$\dot{a}_i^j$$

を用いて、ターゲット信号 $x(n)$ およびインパルス応答 $h(n)$ を算出する(図9ステップ162)。非特許文献4の6.8.4.1.3節に記載の通り、ターゲット信号は線形予測残差信号を聴覚重み付けフィルタに通すことにより得られる(図11)。

[0074] まず、LP係数を用いて先読み信号

[数16]

$$s_{pre}^l(n) \quad (0 \leq n < L')$$

の残差信号 $r(n)$ を次式に従い算出する(図11ステップ181)。

[数17]

$$r(n) = s_{pre}^l(n) + \sum_{i=1}^P \dot{a}_i^j \cdot s_{pre}^l(n-i)$$

[0075] ただし、 L' はサブフレームのサンプル数を表し、 L は符号化対象フレーム $s_{pre}(n)$ ($0 \leq n < L$) のサンプル数を表す。このとき、

[数18]

$$s_{pre}^l(n-p) = s_{pre}(n+L-p)$$

である。

[0076] 次に、ターゲット信号 $x(n)$ ($0 \leq n < L'$) を次式に従い算出する(図11ステップ182)。

[数19]

$$e(n) = r(n) - \sum_{i=1}^P \hat{a}_i^j \cdot e(n-i) \quad (0 \leq n < L')$$

[数20]

$$e(n) = s(n+L-1) - \hat{s}(n+L-1) \quad (-P \leq n < 0)$$

[数21]

$$\dot{e}(n) = r(n) + \sum_{i=1}^P \hat{a}_i^j \cdot \dot{e}(n-i)$$

[数22]

$$x(n) = e(n) + \gamma \cdot e(n-1)$$

ここで、聴覚重み付けフィルタ $\gamma = 0.68$ である。聴覚重み付けフィルタの値は、音声符号化の設計方針により別の値としてもよい。

[0077] 次に、次式に従いインパルス応答 $h(n)$ ($0 \leq n < L'$) を算出する (図11ステップ183)。

[数23]

$$\dot{h}(n) = \hat{a}_i^j - \sum_{i=1}^P \hat{a}_i^j \cdot \dot{h}(n-i)$$

[数24]

$$h(n) = \dot{h}(n) + \gamma \cdot \dot{h}(n-1)$$

[0078] ピッチラグ算出部153は、次式を最大化する k を求めることにより、サブフレーム毎にピッチラグを算出する (図9ステップ163)。ここで、演算量削減のために、上記ターゲット信号算出 (図11ステップ182) およびインパルス応答算出 (図11ステップ183) を省略して、残差信号そのものをターゲット信号として用いてもよい。

$$T_p = \text{argmax } T_k$$

[数25]

$$T_k = \frac{\sum_{n=0}^{L'-1} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{L'-1} y_k(n)y_k(n)}}$$

[数26]

$$y_k(n) = \sum_{i=0}^n v'(i) \cdot h(n-i)$$

[数27]

$$v'(n) = \sum_{i=-l}^l Int(i) \cdot u(n + N_{adapt} - T_p + i)$$

なお、 $y_k(n)$ は線形予測残差にインパルス応答を畳み込むことにより得られる。ここで、 $Int(i)$ は補間フィルタを表す。補間フィルタの詳細は非特許文献4の6.8.4.1.4.1節に記載の通りである。当然、補間にフィルタを用いず、 $v'(n) = u(n + N_{adapt} - T_p + i)$ としてもよい。

[0079] 上記、算出方法により、ピッチラグは整数値で求まるが、上記 T_k を補間することにより、ピッチラグの精度を小数点以下まで上げた場合についても求めることができる。補間により小数点以下のピッチラグを求める処理手順の詳細については、非特許文献4の6.8.4.1.4.1節に記載の処理方法を利用できる。

[0080] 適応符号帳算出部154はピッチラグ T_p 、適応符号帳バッファ156に格納された適応符号帳 $u(n)$ から、次式に従い適応符号帳ベクトル $v'(n)$ および長期予測パラメータを算出する(図9ステップ164)。

[数28]

$$v'(n) = \sum_{i=-l}^l Int(i) \cdot u(n + N_{adapt} - T_p + i)$$

長期パラメータ算出の詳細な処理手順として非特許文献3の5.7節に記載の方法を用いることができる。

[0081] 励振ベクトル合成部 155 は、適応符号帳ベクトル $v'(n)$ に事前に定めた適応符号帳ゲイン

[数29]

$$g_p^C$$

を乗算して、次式に従い励振信号ベクトルを出力する（図9ステップ165）。

[数30]

$$e(n) = g_p^C \cdot v'(n)$$

適応符号帳ゲイン

[数31]

$$g_p^C$$

の値には、例えば 1.0 などを用いるが、事前の学習により求めた値を用いてもよいし、符号化対象フレームの性質を表すインデクスにより変化させてもよい。

[0082] 次に、以下の式に従い、励振信号ベクトルにより、適応符号帳バッファ 156 に格納された適応符号帳 $u(n)$ の状態を更新する（図9ステップ166）。

$$u(n) = u(n+L) \quad (0 \leq n < N-L)$$

$$u(n+N-L) = e(n) \quad (0 \leq n < L)$$

[0083] 合成フィルタ 157 は、励振信号ベクトルを励振源とする、線形予測逆フィルタリングにより、次式に従い復号信号を合成する（図9ステップ167）。

[数32]

$$\hat{s}(n) = e(n) - \sum_{i=1}^P \hat{a}_i \cdot \hat{s}(n-i)$$

[0084] 上記図9ステップ162～ステップ167を先読み信号終了まで、サブフ

レーム毎に繰り返す（図9ステップ168）。

[0085] ピッチラグ符号化部158は、先読み信号において算出したピッチラグ
[数33]

$$T_p^{(j)} (0 \leq j < M_{la})$$

を符号化する（図9のステップ169）。ここで、先読み信号に含まれるサブフレームの数を M_{la} とした。

[0086] 符号化の方法としては、例えば以下が考えられるが、符号化の方法には何を用いても良い。

1. ピッチラグ

[数34]

$$T_p^{(j)} (0 \leq j < M_{la})$$

の一部あるいは全部をバイナリ符号化あるいはスカラ量子化あるいはベクトル量子化あるいは算術符号化して伝送する方法。

2. 直前サブフレームのピッチラグとの差分

[数35]

$$T_p^{(j)} - T_p^{(j-1)} (0 \leq j < M_{la})$$

の一部あるいは全部をバイナリ符号化あるいはスカラ量子化あるいはベクトル量子化あるいは算術符号化して伝送する方法。ただし、

[数36]

$$T_p^{(-1)}$$

は符号化対象フレームにおける最後のサブフレームのピッチラグである。

3. ピッチラグ

[数37]

$$T_p^{(j)} (0 \leq j < M_{la})$$

の一部あるいは全部と、符号化対象フレームで算出したピッチラグの一部ま

たは全部をまとめてベクトル量子化あるいは算術符号化して伝送する方法。

4. ピッチラグ

[数38]

$$T_p^{(j)} (0 \leq j < M_{la})$$

の一部あるいは全部を手がかりに、事前に定めた補間手法のうちの1つを選択し、その補間手法のインデクスを伝送する方法。この時、過去に音声の合成に用いた複数のサブフレームのピッチラグを、併せて補間手法の選択に用いてもよい。

[0087] スカラ量子化およびベクトル量子化には、経験的に定めたコードブックや、学習により事前に算出したコードブックを用いることができる。また、上記ピッチラグにオフセットの値を加算した上で符号化する方法も当然、本発明の実施形態の思想に含まれる。

[0088] <復号側>

図5に示す通り、音声信号受信装置は、音声符号バッファ121、音声パラメータ復号部122、音声パラメータ紛失処理部123、音声合成部124、補助情報復号部125、補助情報蓄積部126からなる。音声信号受信装置の処理手順は図7に示す通りである。

[0089] 音声符号バッファ121は、パケットを正常に受信したか否かを判断し、パケットを正常に受信したと判断した場合には、音声パラメータ復号部122および補助情報復号部125に処理を切替え、パケットを正常に受信できなかったと判断した場合には、音声パラメータ紛失処理部123に処理を切り替える(図7ステップ141)。

[0090] <正常にパケットを受信した場合>

音声パラメータ復号部122は、受信した音声符号を復号し、符号化対象フレームに関する音声を合成するのに必要な音声パラメータ(ISPパラメータおよび対応するISFパラメータ、ピッチラグ、長期予測パラメータ、適応符号帳、適応符号帳ゲイン、固定符号帳ゲイン、固定符号帳ベクトル等)を算出する(図7ステップ142)。

[0091] 補助情報復号部 1 2 5 は、補助情報符号を復号してピッチラグ

[数39]

$$\hat{T}_p^{(j)} \quad (0 \leq j < M_{la})$$

を算出し、補助情報蓄積部 1 2 6 に格納する。補助情報復号部 1 2 5 では、符号化側で用いた符号化方法に対応する復号方法を用いて、補助情報符号を復号する（図 7 ステップ 1 4 3）。

[0092] 音声合成部 1 2 4 は、音声パラメータ復号部 1 2 2 が出力したパラメータから符号化対象フレームに対応する音声信号を合成する（図 7 ステップ 1 4 4）。音声合成部 1 2 4 の機能構成例を図 1 5 に示し、処理手順を図 1 6 に示す。なお、信号の流れを示すために音声パラメータ紛失処理部 1 2 3 を記載しているが、音声パラメータ紛失処理部 1 2 3 は音声合成部 1 2 4 の機能構成には含まれない。

[0093] LP 係数算出部 1 1 2 1 は、ISF パラメータを ISP パラメータに変換した上で、補間処理を実施して、サブフレーム毎の ISP 係数を得る。次に ISP 係数を線形予測係数（LP 係数）に変換し、サブフレーム毎の LP 係数を得る（図 1 6 ステップ 1 1 3 0 1）。ISP 係数の補間処理、および ISP-LP 係数については、例えば非特許文献 4 の 6. 4. 5 節に記載の方法を用いることができる。これらパラメータ変換処理の手順は本発明の実施形態の本質ではないため、詳細については省略する。

[0094] 適応符号帳算出部 1 1 2 3 は、ピッチラグおよび長期予測パラメータと、適応符号帳 1 1 2 2 を用いて適応符号帳ベクトルを算出する（図 1 6 ステップ 1 1 3 0 2）。ピッチラグ

[数40]

$$\hat{T}_p^{(j)}$$

、適応符号帳 $u(n)$ から、次式に従い、適応符号帳ベクトル $v'(n)$ を算出する。

[数41]

$$v'(n) = \sum_{i=-l}^l \text{Int}(i) \cdot u(n + N_{\text{adapt}} - \hat{T}_p^{(j)} + i) \quad (0 \leq n < L')$$

適応符号帳ベクトルは、適応符号帳 $u(n)$ を FIR フィルタ $\text{Int}(i)$ により補間することで算出する。ここで、適応符号帳の長さを N_{adapt} とした。補間に用いるフィルタ $\text{Int}(i)$ は、(数27) の補間フィルタと同じである。事前に定めた長さ $2l+1$ の FIR フィルタである。 L' はサブフレームのサンプル数である。エンコーダ側と同様、補間にフィルタを用いなくてもよい。

[0095] 適応符号帳算出部 1123 は、長期予測パラメータの値に応じて、上記、適応符号帳ベクトルに対してフィルタリングを行う (図16ステップ11303)。長期予測パラメータがフィルタリングを指示する値をとる場合は、以下の式により適応符号帳ベクトルにフィルタリングを行う。

$$v'(n) = 0.18 v'(n-1) + 0.64 v'(n) + 0.18 v'(n+1)$$

[0096] 一方、長期予測パラメータがフィルタリングを指示しない値をとる場合は、フィルタリングを行わず、 $v(n) = v'(n)$ とする。

[0097] 励振ベクトル合成部 1124 は、適応符号帳ベクトルに適応符号帳ゲイン g_p を乗算する (図16ステップ11304)。さらに、励振ベクトル合成部 1124 は、固定符号帳ベクトル $c(n)$ に固定符号帳ゲイン g_c を乗算する (図16ステップ11305)。さらに、励振ベクトル合成部 1124 は、適応符号帳ベクトルと固定符号帳ベクトルを加算して、励振信号ベクトルを出力する (図16ステップ11306)。

$$e(n) = g_p \cdot v'(n) + g_c \cdot c(n)$$

[0098] ポストフィルタ 1125 は、励振信号ベクトルに対して、例えば、ピッチ強調、ノイズ強調、低域強調といった後処理を加える。ピッチ強調、ノイズ強調、低域強調と言った技術の詳細は、非特許文献3の6.1節に記載の通りである。ポストフィルタにおける処理は、本発明の実施形態の本質との関

係が薄いので、詳細を省略する（図16ステップ11307）。

[0099] 適応符号帳1122は、以下の式に従い、励振信号ベクトルにより状態を更新する（図16ステップ11308）。

$$u(n) = u(n+L) \quad (0 \leq n < N-L)$$

$$u(n+N-L) = e(n) \quad (0 \leq n < L)$$

[0100] 合成フィルタ1126は、励振信号ベクトルを励振源とする、線形予測逆フィルタリングにより、次式に従い、復号信号を合成する（図16ステップ11309）。

[数42]

$$\hat{s}(n) = e(n) - \sum_{i=1}^P \hat{a}(i) \cdot \hat{s}(n-i)$$

[0101] 聴覚重み逆フィルタ1127は、復号信号に対して、次式に従い聴覚重み逆フィルタを適用する（図16ステップ11310）。

[数43]

$$\hat{s}(n) = \hat{s}(n) + \beta \cdot \hat{s}(n-1)$$

β の値としては典型的には0.68等を用いるが、この値に限定されない。

[0102] 音声パラメータ紛失処理部123は、音声合成部124で用いた音声パラメータ（ISFパラメータ、ピッチラグ、適応符号帳ゲイン、固定符号帳ゲイン）をバッファに格納する（図7ステップ145）。

[0103] <パケットロスを検出した場合>

音声パラメータ紛失処理部123は、補助情報蓄積部126からピッチラグ

[数44]

$$\hat{T}_p^{(j)} \quad (0 \leq j < M_{la})$$

を読み出し、音声パラメータを予測する。音声パラメータ紛失処理部123の機能構成例を図12に示し、音声パラメータ予測の処理手順を図13に示

す。

[0104] ISF予測部191は、直前フレームに関するISFパラメータと、過去数フレームにおいて算出したISFパラメータを用いてISFパラメータを算出する(図13ステップ1101)。ISF予測部191の処理手順を図10に示す。

[0105] 最初に、直前フレームのISFパラメータを用いてバッファを更新する(図10ステップ171)。次に、以下の式に従いISFパラメータ

[数45]

$$\dot{\omega}_i$$

を算出する(図10ステップ172)。

[数46]

$$\dot{\omega}_i = \alpha \omega_i^{(-1)} + (1 - \alpha) \bar{\omega}_i$$

[数47]

$$\bar{\omega}_i = \beta \omega_i^C + (1 - \beta) \frac{\omega_i^{(-3)} + \omega_i^{(-2)} + \omega_i^{(-1)}}{3}$$

ここで、

[数48]

$$\omega_i^{(-j)}$$

はバッファに格納されたjフレーム前のISFパラメータである。また、

[数49]

$$\omega_i^C$$

、 α 、 β は、符号化側で用いた値と同様である。

[0106] 次に、

[数50]

$$\dot{\omega}_i$$

が

[数51]

$$0 < \dot{\omega}_0 < \dot{\omega}_1 < \dots < \dot{\omega}_{14}$$

を満たすよう i の値を並び替え、隣り合う

[数52]

$$\dot{\omega}_i$$

同士が接近しすぎないように調整する。

[数53]

$$\dot{\omega}_i$$

の値を調整手順には、例えば非特許文献4（式151）を用いることができる（図10ステップ173）。

[0107] ピッチラグ予測部192は、補助情報蓄積部126から補助情報符号を復号してピッチラグ

[数54]

$$\hat{T}_p^{(i)} (0 \leq i < M_{la})$$

を得る。さらに、過去に復号に用いたピッチラグ

[数55]

$$\hat{T}_p^{(-j)} (0 \leq j < J)$$

を用いて、ピッチラグ

[数56]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

を出力する。ここで、1フレームに含まれるサブフレームの数はM、補助情報に含まれるピッチラグの数を M_{la} である。ピッチラグ

[数57]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

の予測に当たっては、例えば非特許文献4の7.11.1.3節に記載の処

理手順を用いることができる（図13ステップ1102）。

[0108] 適応符号帳ゲイン予測部193は、事前に定めた適応符号帳ゲイン

[数58]

$$g_p^C$$

と、過去に復号に用いた適応符号帳ゲイン

[数59]

$$g_p^{(j)} (0 \leq j < J)$$

を用いて、適応符号帳ゲイン

[数60]

$$g_p^{(i)} (M_{la} \leq i < M)$$

を出力する。ここで、1フレームに含まれるサブフレームの数はM、補助情報に含まれるピッチラグの数を M_{la} である。適応符号帳ゲイン

[数61]

$$g_p^{(i)} (M_{la} \leq i < M)$$

の予測に当たっては、例えば、非特許文献4の7.11.2.5.3節記載の処理手順を用いることができる（図13ステップ1103）。

[0109] 固定符号帳ゲイン予測部194は、過去に復号に用いた固定符号帳ゲイン

[数62]

$$g_c^{(j)} (0 \leq j < J)$$

を用いて、固定符号帳ゲイン

[数63]

$$g_c^{(i)} (0 \leq i < M)$$

を出力する。ここで、1フレームに含まれるサブフレームの数はMである。

固定符号帳ゲイン

[数64]

$$g_c^{(i)} \quad (0 \leq i < M)$$

の予測に当たっては、例えば、非特許文献4の7. 11. 2. 6節に記載の処理手順を用いることができる（図13ステップ1104）。

[0110] 雑音信号生成部195は、長さLの白色雑音を出力する（図13ステップ1105）。ここで、1フレームの長さをLとした。

[0111] 音声合成部124は、音声パラメータ紛失処理部123が出力した音声パラメータから復号信号を合成する（図7ステップ144）。音声合成部124の動作は、＜音声パケットを正常に受信した場合＞の音声合成部の動作と同様であるので、詳細を省略する（図7ステップ144）。

[0112] 音声パラメータ紛失処理部123は、音声合成部124で用いた音声パラメータ（ISFパラメータ、ピッチラグ、適応符号帳ゲイン、固定符号帳ゲイン）をバッファに格納する（図7ステップ145）。

[0113] 上記実施例では、先読み信号に含まれる全てのサブフレームに関する補助情報を符号化して伝送する例を述べたが、特定のサブフレームに関する補助情報のみを伝送する構成としてもよい。

[0114] [変形例1-1]

実施例1の変形例として、ピッチゲインを補助情報に追加する例を示す。変形例1-1と実施例1との差分は、励振ベクトル合成部155の動作のみであるので、その他の部分について説明を省略する。

[0115] <符号化側>

励振ベクトル合成部155の処理手順を図14に示す。

[0116] 適応符号帳ベクトル v' (n) とターゲット信号 x (n) から、適応符号帳ゲイン

[数65]

$$g_p^C$$

を次式に従い算出する（図14ステップ1111）。

[数66]

$$g_p = \frac{\sum_{n=0}^{L'-1} x(n)y(n)}{\sum_{n=0}^{L'-1} y(n)y(n)}, \quad \text{bounded by } 0 \leq g_p \leq 1.2,$$

但し、 $y(n)$ は適応符号帳ベクトルにインパルス応答を畳み込んで得られる信号 $y(n) = v(n) * h(n)$ である。

[0117] 算出した適応符号帳ゲインを符号化し、補助情報符号に含める（図14ステップ1112）。符号化には、事前に学習により求めたコードブックを用いたスカラ量子化を用いることができるが、符号化の手法そのものには何を用いてもよい。

[0118] 適応符号帳ゲインの符号化において求めた符号を、復号して得られる適応符号帳ゲイン

[数67]

$$\hat{g}_p$$

を適応符号帳ベクトルに乗算することにより次式に従い、励振ベクトルを算出する（図14ステップ1113）。

[数68]

$$e(n) = \hat{g}_p \cdot v'(n)$$

[0119] <復号側>

励振ベクトル合成部155は、適応符号帳ベクトル $v'(n)$ に補助情報符号を復号することにより得られる適応符号帳ゲイン

[数69]

$$\hat{g}_p$$

を乗算して、次式により励振信号ベクトルを出力する（図9ステップ165）。

[数70]

$$e(n) = \hat{g}_p \cdot v'(n)$$

[0120] [変形例 1 – 2]

実施例 1 の変形例として、補助情報の利用判断のためのフラグを補助情報に追加する例を示す。

[0121] <符号化側>

補助情報符号化部の機能構成例を図 1 7 に、補助情報符号化部の処理手順を図 1 8 に示す。実施例 1 との差分は、補助情報出力判定部 1 1 2 8 (図 1 8 ステップ 1 1 3 1) のみであるので、その他の部分について説明を省略する。

[0122] 補助情報出力判定部 1 1 2 8 は、次式に従い復号信号と先読み信号の *segmental SNR* を算出し、*segmental SNR* が閾値を越えるときのみフラグの値をオンにセットして補助情報に含める。

[数71]

$$segSNR = \frac{\sum_{n=0}^{L'-1} \hat{s}^2(n)}{\sum_{n=0}^{L'-1} (s(n) - \hat{s}(n))^2}$$

一方、*segmental SNR* が閾値を越えないときは、フラグの値をオフにして補助情報に含める (図 1 8 ステップ 1 1 3 1)。なお、フラグの値がオンのときのみ、ピッチラグやピッチゲインなどの補助情報をフラグに付加して伝送し、フラグの値がオフのときはフラグの値のみ伝送することにより補助情報のビット量を削減してもよい。

[0123] <復号側>

補助情報復号部は、補助情報符号に含まれるフラグを復号する。音声パラメータ紛失処理部は、フラグの値がオンの場合は、実施例 1 と同様の処理手順により復号信号を算出する。一方、フラグの値がオフの場合は、補助情報を用いないパケットロス隠蔽手法により復号信号を算出する (図 1 9 ステップ 1 1 5 1)。

[0124] [実施例 2]

本実施例では、先読み信号部分の復号音声を正常受信時にも利用する例について述べる。説明を容易にするため、1フレームに含まれるサブフレームの数をMサブフレーム、先読み信号の長さをM' サブフレームとする。

[0125] <符号化側>

音声信号送信装置は、図20に示す通り、主要符号化部211、補助情報符号化部212、隠蔽信号蓄積部213、誤差信号符号化部214からなる。音声信号送信装置の処理手順を図22に示す。

[0126] 誤差信号符号化部214は、隠蔽信号蓄積部213から1サブフレーム分の隠蔽信号を読み出し、音声信号から減算して、誤差信号を算出する(図22ステップ221)。

[0127] 誤差信号符号化部214は誤差信号を符号化する。具体的な処理手順として非特許文献4の6.8.4.1.5節に記載のAVQ等を利用する。誤差信号の符号化にあたって、ローカルデコードを行い、復号誤差信号を出力する(図22ステップ222)。

[0128] 復号誤差信号を隠蔽信号に加算することにより、1サブフレーム分の復号信号を出力する(図22ステップ223)。

[0129] 上記、ステップ221~223を隠蔽信号終了までM' サブフレーム分繰り返す。

[0130] 主要符号化部211の機能構成を図21に示す。主要符号化部211は、ISF符号化部2011、ターゲット信号算出部2012、ピッチラグ算出部2013、適応符号帳算出部2014、固定符号帳算出部2015、ゲイン算出部2016、励振ベクトル算出部2017、合成フィルタ2018、適応符号帳バッファ2019からなる。

[0131] ISF符号化部2011は、符号化対象フレームおよび先読み信号に対してレビンソンダービン法を適用してLP係数を得る。次にLP係数をISFパラメータに変換して符号化する。次に符号を復号して復号ISFパラメータを得る。最後に復号ISFパラメータを補間したうえで、サブフレーム毎の復号LP係数を得る。レビンソンダービン法、LP係数からのISFパラ

メータ変換の処理手順は実施例1と同様である。また、ISFパラメータの符号化には、例えば非特許文献4の6.8.2節に記載の処理手順を用いる。ISF符号化部2011により、ISFパラメータを符号化したインデクス、復号ISFパラメータ、および復号ISFパラメータをLP係数に変換して得られる復号LP係数が得られる(図22ステップ224)。

[0132] ターゲット信号算出部2012の詳細な処理手順は実施例1の図9ステップ162と同様である(図22ステップ225)。

[0133] ピッチラグ算出部2013は、適応符号帳バッファを参照して、ターゲット信号を用いてピッチラグ、および長期予測パラメータを算出する。ピッチラグ、および長期予測パラメータ算出の詳細な処理手順は実施例1と同様である(図22ステップ226)。

[0134] 適応符号帳算出部2014は、ピッチラグ算出部2013で求めたピッチラグおよび長期予測パラメータを用いて適応符号帳ベクトルを算出する。適応符号帳算出部2014の詳細な処理手順は、実施例1と同様である(図22ステップ227)。

[0135] 固定符号帳算出部2015は、ターゲット信号および適応符号帳ベクトルを用いて、固定符号帳ベクトルおよび固定符号帳ベクトルを符号化して得られるインデクスを算出する。詳細な手順は誤差信号符号化部214で用いたAVQの処理手順と同様である(図22ステップ228)。

[0136] ゲイン算出部2016は、ターゲット信号、適応符号帳ベクトル、固定符号帳ベクトルを用いて、適応符号帳ゲイン、固定符号帳ゲイン、およびこれら2つのゲインを符号化して得られるインデクスを算出する。詳細な処理手順として非特許文献4の6.8.4.1.6節に記載の処理手順を利用できる(図22ステップ229)。

[0137] 励振ベクトル算出部2017は、ゲインを適用した適応符号帳ベクトルおよび固定符号帳ベクトルを加算して励振ベクトルを算出する。詳細な処理手順は実施例1と同様である。さらに励振ベクトル算出部2017は、励振ベクトルを用いて適応符号帳バッファ2019の状態を更新する。詳細な処理

手順は実施例 1 と同様である（図 2 2 ステップ 2 2 1 0）。

[0138] 合成フィルタ 2 0 1 8 は、復号 LP 係数および励振ベクトルを用いて復号信号を合成する（図 2 2 ステップ 2 2 1 1）。

[0139] 上記、ステップ 2 2 4 ~ 2 2 1 1 を符号化対象フレーム終了まで M - M' サブフレーム分繰り返す。

[0140] 補助情報符号化部 2 1 2 は先読み信号 M' サブフレームに対して、補助情報を算出する。具体的な処理手順は実施例 1 と同様である（図 2 2 ステップ 2 2 1 2）。

[0141] 実施例 1 の手順に加えて、実施例 2 では、補助情報符号化部 2 1 2 の合成フィルタ 1 5 7 が出力する復号信号を隠蔽信号蓄積部 2 1 3 に蓄積する（図 2 2 ステップ 2 2 1 3）。

[0142] <復号部>

図 2 3 に示す通り、音声信号受信装置は、音声符号バッファ 2 3 1、音声パラメータ復号部 2 3 2、音声パラメータ紛失処理部 2 3 3、音声合成部 2 3 4、補助情報復号部 2 3 5、補助情報蓄積部 2 3 6、誤差信号復号部 2 3 7、隠蔽信号蓄積部 2 3 8 からなる。音声信号受信装置の処理手順を図 2 4 に示す。音声合成部 2 3 4 の機能構成を図 2 5 に示す。

[0143] 音声符号バッファ 2 3 1 は、パケットを正常に受信したか否かを判断し、パケットを正常に受信したと判断した場合には、音声パラメータ復号部 2 3 2、補助情報復号部 2 3 5、誤差信号復号部 2 3 7 に処理を切替え、パケットを正常に受信できなかったと判断した場合には、音声パラメータ紛失処理部 2 3 3 に処理を切り替える（図 2 4 ステップ 2 4 1）。

[0144] <正常にパケットを受信した場合>

誤差信号復号部 2 3 7 は誤差信号符合を復号して復号誤差信号を得る。具体的な処理手順として非特許文献 4 の 7. 1. 2. 1. 2 節に記載の AVQ 等、符号化側で用いた方法に対応した復号方法を用いる（図 2 4 ステップ 2 4 2）。

[0145] 先読み励振ベクトル合成部 2 3 1 8 は、隠蔽信号蓄積部 2 3 8 から 1 サブ

フレーム分の隠蔽信号を読み出し、復号誤差信号に加算することにより、1サブフレーム分の復号信号を出力する（図24ステップ243）。

[0146] 上記、ステップ241～243を隠蔽信号終了までM'サブフレーム分繰り返す。

[0147] 音声パラメータ復号部232は、ISF復号部2211、ピッチラグ復号部2212、ゲイン復号部2213、固定符号帳復号部2214からなる。音声パラメータ復号部232の機能構成例を図26に示す。

[0148] ISF復号部2211は、ISF符号を復号し、LP係数に変換することにより復号LP係数を得る。例えば非特許文献4の7.1.1節に記載の処理手順を用いる（図24ステップ244）。

[0149] ピッチラグ復号部2212は、ピッチラグ符号を復号してピッチラグおよび長期予測パラメータを得る（図24ステップ245）。

[0150] ゲイン復号部2213は、ゲイン符号を復号して適応符号帳ゲイン、固定符号帳ゲインを得る。詳細な処理手順は非特許文献4の7.1.2.1.3節に記載の通りである（図24ステップ246）。

[0151] 適応符号帳算出部2313は、ピッチラグおよび長期予測パラメータを用いて適応符号帳ベクトルを算出する。適応符号帳算出部2313の詳細な処理手順は、実施例1に記載の通りである（図24ステップ247）。

[0152] 固定符号帳復号部2214は、固定符号帳符号を復号して、固定符号帳ベクトルを算出する。詳細な手順は非特許文献4の7.1.2.1.2節に記載の通りである（図24ステップ248）。

[0153] 励振ベクトル合成部2314は、ゲインを適用した適応符号帳ベクトルおよび固定符号帳ベクトルを加算して励振ベクトルを算出する。さらに励振ベクトル算出部は、励振ベクトルを用いて適応符号帳バッファを更新する（図24ステップ249）。詳細な処理手順は実施例1と同様である。

[0154] 合成フィルタ2316は、復号LP係数および励振ベクトルを用いて復号信号を合成する（図24ステップ2410）。詳細な処理手順は実施例1と同様である。

- [0155] 上記、ステップ244～2410を符号化対象フレーム終了までM-M'サブフレーム分繰り返す。
- [0156] 補助情報復号部235の機能構成は実施例1と同様である。補助情報復号部235は、補助情報符号を復号してピッチラグを算出する（図24ステップ2411）。
- [0157] 音声パラメータ紛失処理部233の機能構成は実施例1と同様である。
ISF予測部191は、直前フレームのISFパラメータを用いてISFパラメータを予測し、LP係数に変換する。処理手順は実施例1の図10のステップ172、173、174と同様である（図24ステップ2412）。
- [0158] 適応符号帳算出部2313は、補助情報復号部235が出力したピッチラグと、適応符号帳2312を用いて適応符号帳ベクトルを算出する（図24ステップ2413）。処理手順は図16ステップ11301、11302と同様である。
- [0159] 適応符号帳ゲイン予測部193は、適応符号帳ゲインを出力する。具体的な処理手順は図13ステップ1103と同様である（図24ステップ2414）。
- [0160] 固定符号帳ゲイン予測部194は、固定符号帳ゲインを出力する。具体的な処理手順は図13ステップ1104と同様である（図24ステップ2415）。
- [0161] 雑音信号生成部195は、白色雑音を出力し、固定符号帳ベクトルとする。処理手順は図13ステップ1105と同様である（図24ステップ2416）。
- [0162] 励振ベクトル合成部2314は、適応符号帳ベクトルおよび固定符号帳ベクトルに各々ゲインを適用した上で加算し、励振ベクトルを算出する。また励振ベクトルにより適応符号帳バッファを更新する（図24ステップ2417）。
- [0163] 合成フィルタ2316は、上記LP係数と励振ベクトルを用いて復号信号

を算出する。算出した復号信号で隠蔽信号蓄積部 238 を更新する（図 24 ステップ 2418）。

[0164] 上記ステップを M' サブフレーム分繰り返す、復号信号を音声信号として出力する。

[0165] <パケットをロスした場合>

隠蔽信号蓄積部から 1 サブフレーム分の隠蔽信号を読み出し、復号信号とする（図 24 ステップ 2419）。

[0166] 上記を M' サブフレーム分繰り返す。

[0167] ISF 予測部 191 は、ISF パラメータを予測する（図 24 ステップ 2420）。処理手順として、図 13 ステップ 1101 を用いる。

[0168] ピッチラグ予測部 192 は、過去に復号に用いたピッチラグを用いて予測ピッチラグを出力する（図 24 ステップ 2421）。予測に用いる処理手順は、実施例 1 の図 13 ステップ 1102 と同様である。

[0169] 適応符号帳ゲイン予測部 193、固定符号帳ゲイン予測部 194、雑音信号生成部 195、音声合成部 234 の動作は実施例 1 と同様である（図 24 ステップ 2422）。

[0170] 上記ステップを M サブフレーム分繰り返す、M-M' サブフレーム分の復号信号は音声信号として出力し、残りの M' サブフレーム分の復号信号で隠蔽信号蓄積部 238 を更新する。

[0171] [実施例 3]

適応符号帳ベクトルの算出に当たって、声門パルス同期を用いる場合について述べる。

[0172] <符号化側>

音声信号送信装置の機能構成は実施例 1 と同じである。機能構成および処理手順が異なるのは補助情報符号化部のみであるので、ここでは補助情報符号化部の動作についてのみ述べる。

[0173] 補助情報符号化部は、LP 係数算出部 311、ピッチラグ予測部 312、ピッチラグ選定部 313、ピッチラグ符号化部 314、適応符号帳バッファ

315からなる。補助情報符号化部の機能構成図を図27に、処理手順を図28に示す。

[0174] LP係数算出部311は、実施例1のLP係数算出部と同じであるので説明を省略する(図28ステップ321)。

[0175] ピッチラグ予測部312は、音声符号化部から得たピッチラグを用いてピッチラグ予測値

[数72]

$$\hat{T}_p$$

を算出する(図28ステップ322)。予測の具体的な処理は実施例1におけるピッチラグ予測部192における、ピッチラグ

[数73]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

の予測と同様である(図13ステップ1102と同じである)。

[0176] 次にピッチラグ選定部313は、補助情報として伝送するピッチラグを決定する(図28ステップ323)。ピッチラグ選定部313のより詳細な処理手順を図29に示す。

[0177] 最初に、ピッチラグ予測値

[数74]

$$\hat{T}_p$$

および過去のピッチラグの値

[数75]

$$\hat{T}_p^{(-j)} (0 \leq j < J)$$

から、次式に従いピッチラグコードブックを生成する(図29ステップ331)。

[数76]

< $\hat{T}_p - \hat{T}_p^{(-1)} \geq 0$ の場合>

$$\hat{T}_C^j = \begin{cases} \hat{T}_p & (j=0) \\ \hat{T}_p^{(-1)} - j \cdot \delta_j + \rho & (0 < j < I) \end{cases}$$

[数77]

< $\hat{T}_p - \hat{T}_p^{(-1)} < 0$ の場合>

$$\hat{T}_C^j = \begin{cases} \hat{T}_p & (j=0) \\ \hat{T}_p^{(-1)} + j \cdot \delta_j + \rho & (0 < j < I) \end{cases}$$

ここで、1サブフレーム前のピッチラグの値は

[数78]

$$\hat{T}_p^{(-1)}$$

である。またコードブックのインデクス数を I とする。また δ_j は事前に定めたステップ幅であり、 ρ は事前に定めた定数である。

[0178] 次に適応符号帳、ピッチラグ予測値

[数79]

$$\hat{T}_p$$

を用いて、次式に従い初期励振ベクトル $u_0(n)$ を生成する（図29ステップ332）。

[数80]

$$u_0(n) = \begin{cases} 0.18u_0(n - \hat{T}_p - 1) + 0.64u_0(n - \hat{T}_p) + 0.18u_0(n - \hat{T}_p + 1) & (0 \leq n < \hat{T}_p) \\ u_0(n - \hat{T}_p) & (\hat{T}_p \leq n < L) \end{cases}$$

初期励振ベクトル算出の処理手順は、非特許文献4の式(607)および式(608)と同様である。

[0179] 次に初期励振ベクトルに対して、ピッチラグコードブック中の全ての候補ピッチラグ

[数81]

$$\hat{T}_c^j (0 \leq j < J)$$

を用いて声門パルス同期を適用し、候補適応符号帳ベクトル $u^j(n)$ ($0 \leq j < J$) を生成する (図 29 ステップ 333)。声門パルス同期は非特許文献 4 の 7. 1 1. 2. 5 節におけるパルス位置が利用できない場合と同様の処理手順を用いる。ただし、非特許文献 4 における $u(n)$ は本発明の実施形態の $u_0(n)$ に対応し、`extrapolated pitch` は本発明の実施形態の

[数82]

$$\hat{T}_c^j$$

に対応し、`the last reliable pitch (T_c)` は本発明の実施形態の

[数83]

$$\hat{T}_p^{(-1)}$$

に対応する。

[0180] 候補適応符号帳ベクトル $u^j(n)$ ($0 \leq j < J$) について、評価尺度を計算する (図 29 ステップ 334)。評価尺度として `segmental SNR` を用いる場合には、LP 係数を用いた逆フィルタリングにより信号を合成し、入力信号との間で次式に従い `segmental SNR` を算出する。

[数84]

$$\hat{s}_j(n) = u^j(n) - \sum_{i=1}^P \hat{a}(i) \cdot \hat{s}_j(n-i)$$

[数85]

$$segSNR_j = \frac{\sum_{n=0}^{L-1} \hat{s}_j^2(n)}{\sum_{n=0}^{L-1} (s(n) - \hat{s}_j(n))^2}$$

[0181] 逆フィルタリングを行う代わりに、次式に従い、残差信号を用いて適応符号帳ベクトルの領域で $segmental$ SNR を算出してもよい。

[数86]

$$r(n) = s(n) + \sum_{i=1}^P \hat{a}(i) \cdot s(n-i)$$

[数87]

$$segSNR_j = \frac{\sum_{n=0}^{L'-1} u^j(n)}{\sum_{n=0}^{L'-1} (r(n) - u^j(n))^2}$$

この場合、LP係数を用いて先読み信号 $s(n)$ ($0 \leq n < L'$) の残差信号 $r(n)$ を算出する (図11ステップ181)。

[0182] ステップ334で算出した評価尺度のうち最大のものに対応するインデクスを選択し、当該インデクスに対応するピッチラグを求める。(図29ステップ335)。

[数88]

$$\arg \max_j [segSNR_j]$$

[0183] <復号側>

音声信号受信装置の機能構成は実施例1と同様である。実施例1との差分は音声パラメータ紛失処理部123、補助情報復号部125、補助情報蓄積部126の機能構成と処理手順であるので、これらについてのみ述べる。

[0184] <正常にパケットを受信した場合>

補助情報復号部125は、補助情報符号を復号してピッチラグ

[数89]

$$\hat{T}_C^{idx}$$

を算出し、補助情報蓄積部126に格納する。補助情報復号部125の処理手順を図30に示す。

[0185] ピッチラグ算出に当たって、最初にピッチラグ予測部312は、音声復号部から得たピッチラグを用いてピッチラグ予測値

[数90]

$$\hat{T}_p$$

を算出する（図30ステップ341）。予測の具体的な処理は実施例3における図28ステップ322と同様である。

[0186] 次に、ピッチラグ予測値

[数91]

$$\hat{T}_p$$

および過去のピッチラグの値

[数92]

$$\hat{T}_p^{(-j)} \quad (0 \leq j < J)$$

から、次式に従いピッチラグコードブックを生成する（図30ステップ342）。

[数93]

< $\hat{T}_p - \hat{T}_p^{(-1)} \geq 0$ の場合>

$$\hat{T}_C^j = \begin{cases} \hat{T}_p & (j=0) \\ \hat{T}_p^{(-1)} - j \cdot \delta_j + \rho & (0 < j < I) \end{cases}$$

[数94]

< $\hat{T}_p - \hat{T}_p^{(-1)} < 0$ の場合>

$$\hat{T}_C^j = \begin{cases} \hat{T}_p & (j=0) \\ \hat{T}_p^{(-1)} + j \cdot \delta_j + \rho & (0 < j < I) \end{cases}$$

処理手順は図29ステップ331と同様である。ここで、1サブフレーム前のピッチラグの値は

[数95]

$$\hat{T}_p^{(-1)}$$

である。またコードブックのインデクス数を l とする。また δ_j は事前に定めたステップ幅であり、 ρ は事前に定めた定数である。

[0187] 次に、ピッチラグコードブックを参照し、補助情報として伝送されたインデクス idx に対応するピッチラグ

[数96]

$$\hat{T}_C^{idx}$$

を求め、補助情報蓄積部 126 に格納する（図30ステップ343）。

[0188] <パケットロスを検出した場合>

音声合成部の機能構成も実施例1と同様であるが（図15と同様）、実施例1と動作が異なる適応符号帳算出部 1123 についてのみ以下に述べる。

[0189] 音声パラメータ紛失処理部 123 は、補助情報蓄積部 126 からピッチラグを読み出した上で次式に従ってピッチラグ予測値を算出し、ピッチラグ予測部 192 の出力の代わりに用いる。

[数97]

$$\hat{T}_p = \hat{T}_p^{(-1)} + \kappa \cdot (\hat{T}_C^{idx} - \hat{T}_p^{(-1)})$$

ここで、

[数98]

$$\kappa$$

は予め定めた定数である。

[0190] 次に、適応符号帳、ピッチラグ予測値

[数99]

$$\hat{T}_p$$

を用いて、次式に従い初期励振ベクトル $u_0(n)$ を生成する（図29ステップ332）。

[数100]

$$u_0(n) = \begin{cases} 0.18u_0(n - \hat{T}_p^{(-1)} - 1) + 0.64u_0(n - \hat{T}_p^{(-1)}) + 0.18u_0(n - \hat{T}_p^{(-1)} + 1) & (0 \leq n < \hat{T}_p^{(-1)}) \\ u_0(n - \hat{T}_p^{(-1)}) & (\hat{T}_p^{(-1)} \leq n < L) \end{cases}$$

[0191] 次に初期励振ベクトルに対して、ピッチラグ

[数101]

$$\hat{T}_c^{idx}$$

を用いて声門パルス同期を適用し、適応符号帳ベクトル $u(n)$ を生成する。声門パルス同期は図29ステップ333と同様の処理手順を用いる。

[0192] 続いて、上述した一連の音声信号送信装置による処理をコンピュータに実行させるための音声符号化プログラム70を説明する。図31に示すように、音声符号化プログラム70は、コンピュータに挿入されてアクセスされる、あるいはコンピュータが備える記録媒体60に形成されたプログラム格納領域61内に格納される。

[0193] 音声符号化プログラム70は、音声符号化モジュール700と、補助情報符号化モジュール701とを備えて構成される。音声符号化モジュール700と、補助情報符号化モジュール701とを実行させることにより実現される機能は、上述した音声信号送信装置の音声符号化部111と、補助情報符号化部112との機能とそれぞれ同様である。

[0194] なお、音声符号化プログラム70は、その一部若しくは全部が、通信回線等の伝送媒体を介して伝送され、他の機器により受信されて記録（インストールを含む）される構成としてもよい。また、音声符号化プログラム70の各モジュールは、1つのコンピュータでなく、複数のコンピュータのいずれかにインストールされてもよい。その場合、当該複数のコンピュータからなるコンピュータシステムによって上述した一連の音声符号化プログラム70の処理が行われる。

[0195] 続いて、上述した一連の音声信号受信装置による処理をコンピュータに実行させるための音声復号プログラム90を説明する。図32に示すように、

音声復号プログラム90は、コンピュータに挿入されてアクセスされる、あるいはコンピュータが備える記録媒体80に形成されたプログラム格納領域81内に格納される。

[0196] 音声復号プログラム90は、音声符号バッファモジュール900と、音声パラメータ復号モジュール901と、補助情報復号モジュール902と、補助情報蓄積モジュール903と、音声パラメータ紛失処理モジュール904と、音声合成モジュール905とを備えて構成される。音声符号バッファモジュール900と、音声パラメータ復号モジュール901と、補助情報復号モジュール902と、補助情報蓄積モジュール903と、音声パラメータ紛失処理モジュール904と、音声合成モジュール905とを実行させることにより実現される機能は、上述した音声信号受信装置の音声符号バッファ231と、音声パラメータ復号部232と、補助情報復号部235と、補助情報蓄積部236と、音声パラメータ紛失処理部233と、音声合成部234との機能とそれぞれ同様である。

[0197] なお、音声復号プログラム90は、その一部若しくは全部が、通信回線等の伝送媒体を介して伝送され、他の機器により受信されて記録（インストールを含む）される構成としてもよい。また、音声復号プログラム90の各モジュールは、1つのコンピュータでなく、複数のコンピュータのいずれかにインストールされてもよい。その場合、当該複数のコンピュータからなるコンピュータシステムによって上述した一連の音声復号プログラム90の処理が行われる。

[0198] [実施例4]

補助情報を復号側でのピッチラグ予測に用いる例について述べる。

[0199] <符号化側>

音声信号送信装置の機能構成は実施例1と同様である。機能構成および処理手順が異なるのは補助情報符号化部112のみであるので、ここでは補助情報符号化部112の動作についてのみ述べる。

[0200] 補助情報符号化部112の機能構成図を図33に、処理手順を図34に示

す。補助情報符号化部 1 1 2 は、LP 係数算出部 5 1 1、残差信号算出部 5 1 2、ピッチラグ算出部 5 1 3、適応符号帳算出部 5 1 4、適応符号帳バッファ 5 1 5、ピッチラグ符号化部 5 1 6 からなる。

[0201] LP 係数算出部 5 1 1 は、実施例 1 の図 8 の LP 係数算出部 1 5 1 と同様であるので説明を省略する。

[0202] 残差信号算出部 5 1 2 は、実施例 1 の図 1 1 のステップ 1 8 1 と同様の処理により、残差信号を算出する。

[0203] ピッチラグ算出部 5 1 3 は、次式を最大化する k を求めることにより、サブフレーム毎にピッチラグを算出する（図 3 4 のステップ 1 6 3）。ここで、 $u(n)$ は適応符号帳、 L' は 1 サブフレームに含まれるサンプル数を表す。

$$T_p = \arg \max_k T_k$$

[数 102]

$$T_k = \frac{\sum_{n=0}^{L'-1} r(n)u(n-k)}{\sqrt{\sum_{n=0}^{L'-1} u(n-k)u(n-k)}}$$

[0204] 適応符号帳算出部 5 1 4 はピッチラグ T_p 、適応符号帳 $u(n)$ から、適応符号帳ベクトル $v'(n)$ を算出する。ここで適応符号帳の長さを N_{adapt} とした。（図 3 4 のステップ 1 6 4）

$$v'(n) = u(n + N_{\text{adapt}} - T_p)$$

[0205] 適応符号帳バッファ 5 1 5 は、適応符号帳ベクトル $v'(n)$ により状態を更新する（図 3 4 のステップ 1 6 6）。

$$u(n) = u(n + L') \quad (0 \leq n < N - L')$$

$$u(n + N - L') = v'(n) \quad (0 \leq n < L')$$

[0206] ピッチラグ符号化部 5 1 6 は、実施例 1 と同様なので省略する（図 3 4 のステップ 1 6 9）。

[0207] <復号側>

音声信号受信装置は、実施例 1 と同様に、音声符号バッファ 1 2 1、音声パラメータ復号部 1 2 2、音声パラメータ紛失処理部 1 2 3、音声合成部 1

24、補助情報復号部125、補助情報蓄積部126からなる。音声信号受信装置の処理手順は図7に示す通りである。

[0208] 音声符号バッファ121の動作は実施例1と同様である。

[0209] <正常にパケットを受信した場合>

音声パラメータ復号部122の動作は実施例1と同様である。

[0210] 補助情報復号部125は、補助情報符号を復号してピッチラグ

[数103]

$$\hat{T}_p^{(j)} (0 \leq j < M_{ia})$$

を算出し、補助情報蓄積部126に格納する。補助情報復号部125では、符号化側で用いた符号化方法に対応する復号方法を用いて、補助情報符号を復号する。

[0211] 音声合成部124は、実施例1と同様である。

[0212] <パケットロスを検出した場合>

音声パラメータ紛失処理部123（図12参照）のISF予測部191は、実施例1と同様にしてISFパラメータを算出する。

[0213] ピッチラグ予測部192の処理手順を図35に示す。ピッチラグ予測部192は、実施例1と同様にして、補助情報蓄積部126から補助情報符号を読み出してピッチラグ

[数104]

$$\hat{T}_p^{(i)} (0 \leq i < M_{ia})$$

を得る（図35のステップ4051）。さらに、過去に復号に用いたピッチラグ

[数105]

$$\hat{T}_p^{(-j)} (0 \leq j < J)$$

を用いて、ピッチラグ

[数106]

$$\hat{T}_p^{(i)} (M_{ia} \leq i < M)$$

を出力する（図35のステップ4052）。ここで、1フレームに含まれるサブフレームの数はM、補助情報に含まれるピッチラグの数を M_{la} とする。

ピッチラグ

[数107]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

の予測に当たっては、非特許文献4のような処理手順を用いることができる（図13のステップ1102）。

[0214] ここで、ピッチラグ予測部192は、ピッチラグ

[数108]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

の予測に当たって、過去に復号に用いたピッチラグ

[数109]

$$\hat{T}_p^{(-j)} (1 \leq j \leq J)$$

とピッチラグ

[数110]

$$\hat{T}_p^{(i)} (0 \leq i < M_{la})$$

を用いてピッチラグ

[数111]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

を予測してもよい。また、

[数112]

$$\hat{T}_p^{(i)} = \hat{T}_p^{(M_{la})}$$

としてもよい。この場合のピッチラグ予測部の処理手順は図36のようになる。

[0215] さらに、ピッチラグ予測部192は、ピッチラグの予測値について信頼性が低い場合にのみ

[数113]

$$\hat{T}_p^{(i)} = \hat{T}_p^{(M_{ia})}$$

としてもよい。この場合のピッチラグ予測部 1 9 2 の処理手順を図 3 7 に示す。予測値を用いたか、補助情報により得られたピッチラグ

[数114]

$$\hat{T}_p^{(M_{ia})}$$

を用いたかに関する指示情報を適応符号帳算出部 1 5 4 に入力するようにしてもよい。

[0216] 適応符号帳ゲイン予測部 1 9 3、固定符号帳ゲイン予測部 1 9 4 は実施例 1 と同様である。

[0217] 雑音信号生成部 1 9 5 は、実施例 1 と同様である。

[0218] 音声合成部 1 2 4 は、音声パラメータ紛失処理部 1 2 3 が出力したパラメータから符号化対象フレームに対応する音声信号を合成する。

[0219] 音声合成部 1 2 4 (図 1 5 参照) の LP 係数算出部 1 1 2 1 は、実施例 1 と同様にして LP 係数を得る (図 1 6 のステップ 1 1 3 0 1)。

[0220] 適応符号帳算出部 1 1 2 3 は、実施例 1 と同様にして適応符号帳ベクトルを算出する。適応符号帳算出部 1 1 2 3 は、常に適応符号帳ベクトルにフィルタリングを行うようにしてもよいし、常にフィルタリングを行わないようにしてもよい。すなわち以下の式を用いて適応符号帳ベクトルを算出する。ここで、フィルタ係数を f_i とした。

$$v(n) = f_{-1} v'(n-1) + f_0 v'(n) + f_1 v'(n+1)$$

フィルタリングを指示しない値をとる場合は、 $v(n) = v'(n)$ とする (適応符号帳算出ステップ A)。

[0221] 適応符号帳算出部 1 1 2 3 は、次の手順で適応符号帳ベクトルを算出してもよい (適応符号帳算出ステップ B)。

[0222] ピッチラグおよび適応符号帳 1 1 2 2 を用いて初期適応符号帳ベクトルを算出する。

$$v(n) = f_{-1} v'(n-1) + f_0 v'(n) + f_1 v'(n+1)$$

設計方針によって、 $v(n) = v'(n)$ としてもよい。

- [0223] 次に初期適応符号帳ベクトルに対して、声門パルス同期を適用する。声門パルス同期は非特許文献4の7.11.2.5節におけるパルス位置が利用できない場合と同様の処理手順を用いる。ただし、非特許文献4における $u(n)$ は本発明の実施形態の $v(n)$ に対応し、`extrapolated pitch` は本発明の実施形態の

[数115]

$$\hat{T}_p^{(M-1)}$$

に対応し、`the last reliable pitch` (T_c) は本発明の実施形態の

[数116]

$$\hat{T}_p^{(M_p-1)}$$

に対応する。

- [0224] さらに、適応符号帳算出部1123は、ピッチラグ予測部192が、上記予測値の指示情報を出力している場合に、上記指示情報が、補助情報として送られてきたピッチラグを予測値として用いないことを示す場合（図38のステップ4082：NO）に上記適応符号帳算出ステップAを用い、それ以外の場合（図38のステップ4082：YES）には上記適応符号帳算出ステップBを用いるようにしてもよい。この場合の適応符号帳算出部1123の処理手順を図38に示す。

- [0225] 励振ベクトル合成部1124は、実施例1と同様にして、励振信号ベクトルを出力する（図16のステップ11306）。

- [0226] ポストフィルタ1125は、実施例1と同様にして、合成信号に後処理を加える。

- [0227] 適応符号帳1122は、実施例1と同様にして、励振信号ベクトルにより状態を更新する（図16のステップ11308）。

[0228] 合成フィルタ 1 1 2 6 は、実施例 1 と同様にして、復号信号を合成する（図 1 6 のステップ 1 1 3 0 9）。

[0229] 聴覚重み逆フィルタ 1 1 2 7 は、実施例 1 と同様にして、聴覚重み逆フィルタを適用する。

[0230] 音声パラメータ紛失処理部 1 2 3 は、実施例 1 と同様にして、音声合成部 1 2 4 で用いた音声パラメータ（ISFパラメータ、ピッチラグ、適応符号帳ゲイン、固定符号帳ゲイン）をバッファに格納する（図 7 のステップ 1 4 5）。

[0231] [実施例 5]

本実施例では、特定のフレームクラスにおいてのみ補助情報としてピッチラグを伝送し、それ以外では、ピッチラグを伝送しない構成について述べる。

[0232] <送信側>

音声信号送信装置において、入力音声信号は、音声符号化部 1 1 1 に送られる。

[0233] 本実施例における音声符号化部 1 1 1 は、符号化対象フレームの性質を表すインデックスを必ず算出し、補助情報符号化部 1 1 2 に伝送する。それ以外の動作については、実施例 1 と同様である。

[0234] 補助情報符号化部 1 1 2 において、実施例 1 - 4 との違いは、ピッチラグ符号化部 1 5 8 のみであるので、ピッチラグ符号化部 1 5 8 の動作について、以下に述べる。実施例 5 における補助情報符号化部 1 1 2 の構成図を図 3 9 に示す。

[0235] ピッチラグ符号化部 1 5 8 の処理手順を図 4 0 に示す。ピッチラグ符号化部 1 5 8 は、符号化対象フレームの性質を現すインデックスを読み出し（図 4 0 のステップ 5 0 2 1）、符号化対象フレームの性質を現すインデックスが、事前に定めた値と等しい場合は、補助情報に割り当てるビット数を B ビットとする（ $B > 1$ ）。一方、符号化対象フレームの性質を現すインデックスが事前に定めた値と異なる場合は、補助情報に割り当てるビット数を 1 ビットと

する（図40のステップ5022）。

[0236] 補助情報に割り当てるビット数が1ビットの場合（図40のステップ5022：NO）、補助情報インデクスに補助情報を伝送していないことを示す値をセットし、補助情報符号とする（図40のステップ5023）。

[0237] 一方、補助情報に割り当てるビット数がBビットの場合（図40のステップ5022：YES）、補助情報インデクスに補助情報を伝送することを示す値をセットし（図40のステップ5024）、さらにピッチラグを実施例1の方法により符号化して得られるB-1ビットの符号を含めて、補助情報符号とする（図40のステップ5025）。

[0238] <復号側>

音声信号受信装置は、実施例1と同様に、音声符号バッファ121、音声パラメータ復号部122、音声パラメータ紛失処理部123、音声合成部124、補助情報復号部125、補助情報蓄積部126からなる。音声信号受信装置の処理手順は図7に示す通りである。

[0239] 音声符号バッファ121の動作は実施例1と同様である。

[0240] <正常にパケットを受信した場合>

音声パラメータ復号部122の動作は実施例1と同様である。

[0241] 補助情報復号部125の処理手順を図41に示す。補助情報復号部125は、最初に補助情報符号に含まれる補助情報インデクスを復号する（図41のステップ5031）。補助情報インデクスが補助情報を伝送していないことを表す場合、それ以上の復号動作を行わない。また、補助情報インデクスの値を補助情報蓄積部126に格納する（図41のステップ5032）。

[0242] 一方、補助情報インデクスが、補助情報を伝送することを表す場合、さらにB-1ビットの復号を行い、ピッチラグ

[数117]

$$\hat{T}_p^{(j)} (0 \leq j < M_{la})$$

を算出し、補助情報蓄積部126に格納する（図41のステップ5033）。また、補助情報インデクスの値を補助情報蓄積部126に格納する。なお

、B-1ビットの補助情報の復号は、実施例1の補助情報復号部125と同様の動作である。

[0243] 音声合成部124は、実施例1と同様である。

[0244] <パケットロスを検出した場合>

音声パラメータ紛失処理部123（図12参照）のISF予測部191は、実施例1と同様にしてISFパラメータを算出する。

[0245] ピッチラグ予測部192の処理手順を図42に示す。ピッチラグ予測部192は、補助情報蓄積部126から補助情報インデクスを読み出して（図42のステップ5041）、補助情報を伝送することを表す値であるかどうかを調べる（図42のステップ5042）。

[0246] <補助情報インデクスが補助情報を伝送することを表す値である場合>

実施例1と同様にして、補助情報蓄積部126から補助情報符号を読み出してピッチラグ

[数118]

$$\hat{T}_p^{(i)} (0 \leq i < M_{la})$$

を得る（図42の5043）。さらに、過去に復号に用いたピッチラグ

[数119]

$$\hat{T}_p^{(-j)} (0 \leq j < J)$$

および補助情報として得られた

[数120]

$$\hat{T}_p^{(i)} (0 \leq i < M_{la})$$

を用いて、ピッチラグ

[数121]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

を出力する（図42のステップ5044）。ここで、1フレームに含まれるサブフレームの数はM、補助情報に含まれるピッチラグの数をM_{la}とする。ピッチラグ

[数122]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

の予測に当たっては、非特許文献4のような処理手順を用いることができる
(図13のステップ1102)。また、

[数123]

$$\hat{T}_p^{(i)} = \hat{T}_p^{(M_{la})}$$

としてもよい。

[0247] さらに、ピッチラグ予測部192は、ピッチラグの予測値について信頼性が低い場合にのみ

[数124]

$$\hat{T}_p^{(i)} = \hat{T}_p^{(M_{la})}$$

とし、それ以外の場合には予測値を

[数125]

$$\hat{T}_p^{(i)}$$

とするようにしてもよい(図42のステップ5046)。また、予測値を用いたか、補助情報により得られたピッチラグ

[数126]

$$\hat{T}_p^{(M_{la})}$$

を用いたかに関するピッチラグ指示情報を適応符号帳算出部1123に入力するようにしてもよい。

[0248] <補助情報インデクスが補助情報を伝送しないことを表す値である場合>
ピッチラグ予測部192は、ピッチラグ

[数127]

$$\hat{T}_p^{(i)} (M_{la} \leq i < M)$$

の予測に当たって、過去に復号に用いたピッチラグ

[数128]

$$\hat{T}_p^{(-j)} (1 \leq j < J)$$

を用いてピッチラグ

[数129]

$$\hat{T}_p^{(i)} (0 \leq i < M)$$

を予測する（図42のステップ5048）。

[0249] さらに、ピッチラグ予測部192は、ピッチラグの予測値について信頼性が低い場合にのみ

[数130]

$$\hat{T}_p^{(i)} = \hat{T}_p^{(-1)}$$

とし（図42のステップ5049）、それ以外の場合には予測値を

[数131]

$$\hat{T}_p^{(i)}$$

とするようにしてもよい。また、予測値を用いたか、過去復号に用いたピッチラグ

[数132]

$$\hat{T}_p^{(-1)}$$

を用いたかに関するピッチラグ指示情報を適応符号帳算出部1123に入力する（図42のステップ5050）。

[0250] 適応符号帳ゲイン予測部193、固定符号帳ゲイン予測部194は実施例1と同様である。

[0251] 雑音信号生成部195は、実施例1と同様である。

[0252] 音声合成部124は、音声パラメータ紛失処理部123が出力したパラメータから符号化対象フレームに対応する音声信号を合成する。

[0253] 音声合成部124（図15参照）のLP係数算出部1121は、実施例1と同様にしてLP係数を得る（図16のステップ11301）。

[0254] 適応符号帳算出部 1 1 2 3 の処理手順を図 4 3 に示す。適応符号帳算出部 1 1 2 3 は、実施例 1 と同様にして適応符号帳ベクトルを算出する。まず、ピッチラグ指示情報を参照し（図 4 3 のステップ 5 0 5 1）、予測値の信頼性が低い場合（図 4 3 のステップ 5 0 5 2 : YES）は、以下の式を用いて適応符号帳ベクトルを算出する（図 4 3 のステップ 5 0 5 5）。ここで、フィルタ係数を f_i とした。

$$v(n) = f_{-1} v'(n-1) + f_0 v'(n) + f_1 v'(n+1)$$

なお設計方針により、 $v(n) = v'(n)$ としてもよい。

[0255] ピッチラグ指示情報を参照し、予測値の信頼性が高い場合（図 4 3 のステップ 5 0 5 2 : NO）、適応符号帳算出部 1 1 2 3 は、次の手順で適応符号帳ベクトルを算出する。

[0256] まず、ピッチラグおよび適応符号帳 1 1 2 2 を用いて初期適応符号帳ベクトルを算出する（図 4 3 のステップ 5 0 5 3）。

$$v(n) = f_{-1} v'(n-1) + f_0 v'(n) + f_1 v'(n+1)$$

設計方針によって、 $v(n) = v'(n)$ としてもよい。

[0257] 次に初期適応符号帳ベクトルに対して、声門パルス同期を適用する。声門パルス同期は非特許文献 4 の 7. 1 1. 2. 5 節におけるパルス位置が利用できない場合と同様の処理手順を用いる（図 4 3 のステップ 5 0 5 4）。ただし、非特許文献 4 における $u(n)$ は本発明の実施形態の $v(n)$ に対応し、`extrapolated pitch` は本発明の実施形態の

[数133]

$$\hat{T}_p^{(M-1)}$$

に対応し、`the last reliable pitch` (T_c) は本発明の実施形態の

[数134]

$$\hat{T}_p^{(-1)}$$

に対応する。

- [0258] 励振ベクトル合成部 1 1 2 4 は、実施例 1 と同様にして、励振信号ベクトルを出力する（図 1 6 のステップ 1 1 3 0 6）。
- [0259] ポストフィルタ 1 1 2 5 は、実施例 1 と同様にして、合成信号に後処理を加える。
- [0260] 適応符号帳 1 1 2 2 は、実施例 1 と同様にして、励振信号ベクトルにより状態を更新する（図 1 6 のステップ 1 1 3 0 8）。
- [0261] 合成フィルタ 1 1 2 6 は、実施例 1 と同様にして、復号信号を合成する（図 1 6 のステップ 1 1 3 0 9）。
- [0262] 聴覚重み逆フィルタ 1 1 2 7 は、実施例 1 と同様にして、聴覚重み逆フィルタを適用する。
- [0263] 音声パラメータ紛失処理部 1 2 3 は、実施例 1 と同様にして、音声合成部 1 2 4 で用いた音声パラメータ（ISFパラメータ、ピッチラグ、適応符号帳ゲイン、固定符号帳ゲイン）をバッファに格納する（図 7 のステップ 1 4 5）。

符号の説明

- [0264] 6 0、8 0…記憶媒体、6 1、8 1…プログラム格納領域、7 0…音声符号化プログラム、9 0…音声復号プログラム、1 1 1…音声符号化部、1 1 2…補助情報符号化部、1 2 1、2 3 1…音声符号バッファ、1 2 2、2 3 2…音声パラメータ復号部、1 2 3、2 3 3…音声パラメータ紛失処理部、1 2 4、2 3 4…音声合成部、1 2 5、2 3 5…補助情報復号部、1 2 6、2 3 6…補助情報蓄積部、1 5 1、5 1 1、1 1 2 1…LP係数算出部、1 5 2、2 0 1 2…ターゲット信号算出部、1 5 3、5 1 3、2 0 1 3…ピッチラグ算出部、1 5 4、1 1 2 3、5 1 4、2 0 1 4、2 3 1 3…適応符号帳算出部、1 5 5、1 1 2 4、2 3 1 4…励振ベクトル合成部、1 5 6、3 1 5、5 1 5、2 0 1 9…適応符号帳バッファ、1 5 7、1 1 2 6、2 0 1 8、2 3 1 6…合成フィルタ、1 5 8、5 1 6…ピッチラグ符号化部、1 9 1…ISF予測部、1 9 2…ピッチラグ予測部、1 9 3…適応符号帳ゲイン予測部、1 9 4…固定符号帳ゲイン予測部、1 9 5…雑音信号生成部、2 1

1…主要符号化部、212…補助情報符号化部、213、238…隠蔽信号蓄積部、214…誤差信号符号化部、237…誤差信号復号部、311…LP係数算出部、312…ピッチラグ予測部、313…ピッチラグ選定部、314…ピッチラグ符号化部、512…残差信号算出部、700…音声符号化モジュール、701…補助情報符号化モジュール、900…音声パラメータ復号モジュール、901…音声パラメータ紛失処理モジュール、902…音声合成モジュール、903…補助情報復号モジュール、1128…補助情報出力判定部、1122、2312…適応符号帳、1125…ポストフィルタ、1127…聴覚重み逆フィルタ、2011…ISF符号化部、2015…固定符号帳算出部、2016…ゲイン算出部、2017…励振ベクトル算出部、2211…ISF復号部、2212…ピッチラグ復号部、2213…ゲイン復号部、2214…固定符号帳復号部、2318…先読み励振ベクトル合成部。

請求の範囲

- [請求項1] 音声信号を符号化する音声符号化装置であって、
音声信号を符号化する音声符号化部と、
先読み信号から補助情報を算出して符号化する補助情報符号化部と、
、
を備える音声符号化装置。
- [請求項2] 前記補助情報は、先読み信号におけるピッチラグに関する、請求項1に記載の音声符号化装置。
- [請求項3] 前記補助情報は、先読み信号におけるピッチゲインに関する、請求項1に記載の音声符号化装置。
- [請求項4] 前記補助情報は、先読み信号におけるピッチラグおよびピッチゲインに関する、請求項1に記載の音声符号化装置。
- [請求項5] 前記補助情報は、補助情報の利用可否に関する情報を含む、請求項1～4の何れか一項に記載の音声符号化装置。
- [請求項6] 前記補助情報符号化部は、先読み信号部分について補助情報を算出して符号化し、さらに隠蔽信号を生成し、
入力音声信号と前記補助情報符号化部が出力する隠蔽信号の誤差信号を符号化する誤差信号符号化部と、
入力音声信号そのものを符号化する主要符号化部と、
をさらに備える、請求項1～5の何れか一項に記載の音声符号化装置。
- [請求項7] 音声符号を復号して音声信号を出力する音声復号装置であって、
音声パケットの受信状態からパケットロスを検出する音声符号バッファと、
音声パケット正常受信時に音声符号を復号する音声パラメータ復号部と、
音声パケット正常受信時に補助情報符号を復号する補助情報復号部と、

補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積部と、

音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理部と、

音声パラメータから復号音声を作成する音声合成部と、

を備える音声復号装置。

[請求項8] 前記補助情報は、先読み信号におけるピッチラグに関する、請求項7に記載の音声復号装置。

[請求項9] 前記補助情報は、先読み信号におけるピッチゲインに関する、請求項7に記載の音声復号装置。

[請求項10] 前記補助情報は、先読み信号におけるピッチラグおよびピッチゲインに関する、請求項7に記載の音声復号装置。

[請求項11] 前記補助情報は、補助情報の利用可否に関する情報を含む、請求項7～10の何れか一項に記載の音声復号装置。

[請求項12] 前記補助情報復号部は、補助情報符号を復号して補助情報を出力し、かつ補助情報を利用して先読み部分に関する隠蔽信号を出力し、音声信号と隠蔽信号との誤差信号に関する符号を復号する誤差復号部と、

音声信号に関する符号を復号する主要復号部と、

前記補助情報復号部が出力した隠蔽信号を蓄積する隠蔽信号蓄積部と、

をさらに備える、請求項7～11の何れか一項に記載の音声復号装置。

[請求項13] 音声パケット正常受信時には、前記隠蔽信号蓄積部から読み出した隠蔽信号と、前記誤差復号部が出力した復号誤差信号を加算することにより復号信号の一部を生成し、前記補助情報復号部が出力した隠蔽信号により前記隠蔽信号蓄積部を更新する、請求項12に記載の音声復号装置。

- [請求項14] 音声パケットロス検出時には、前記隠蔽信号蓄積部から読み出した隠蔽信号を、復号信号の一部あるいは全部とする、請求項12又は13に記載の音声復号装置。
- [請求項15] 音声パケットロス検出時には、前記音声パラメータ紛失処理部により予測した音声パラメータを用いて復号信号を生成し、その一部を用いて前記隠蔽信号蓄積部を更新する、請求項12～14の何れか一項に記載の音声復号装置。
- [請求項16] 前記音声パラメータ紛失処理部は、音声パケットロス検出時に、前記補助情報蓄積部から読み出した補助情報を、音声パラメータの予測値の一部として利用する、請求項7～11の何れか一項に記載の音声復号装置。
- [請求項17] 前記音声合成部は、音声パケットロス検出時に、前記補助情報蓄積部から読み出した補助情報を用いて、音声パラメータの1つである適応符号帳ベクトルを修正する、請求項7～16の何れか一項に記載の音声復号装置。
- [請求項18] 音声信号を符号化する音声符号化装置による音声符号化方法であって、
音声信号を符号化する音声符号化ステップと、
先読み信号から補助情報を算出して符号化する補助情報符号化ステップと、
を含む音声符号化方法。
- [請求項19] 音声符号を復号して音声信号を出力する音声復号装置による音声復号方法であって、
音声パケットの受信状態からパケットロスを検出する音声符号バッファステップと、
音声パケット正常受信時に音声符号を復号する音声パラメータ復号ステップと、
音声パケット正常受信時に補助情報符号を復号する補助情報復号ス

テップと、

補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積ステップと、

音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理ステップと、

音声パラメータから復号音声を合成する音声合成ステップと、
を含む音声復号方法。

[請求項20]

コンピュータを、

音声信号を符号化する音声符号化部と、

先読み信号から補助情報を算出して符号化する補助情報符号化部、
として機能させる音声符号化プログラム。

[請求項21]

コンピュータを、

音声パケットの受信状態からパケットロスを検出する音声符号バッファと、

音声パケット正常受信時に音声符号を復号する音声パラメータ復号部と、

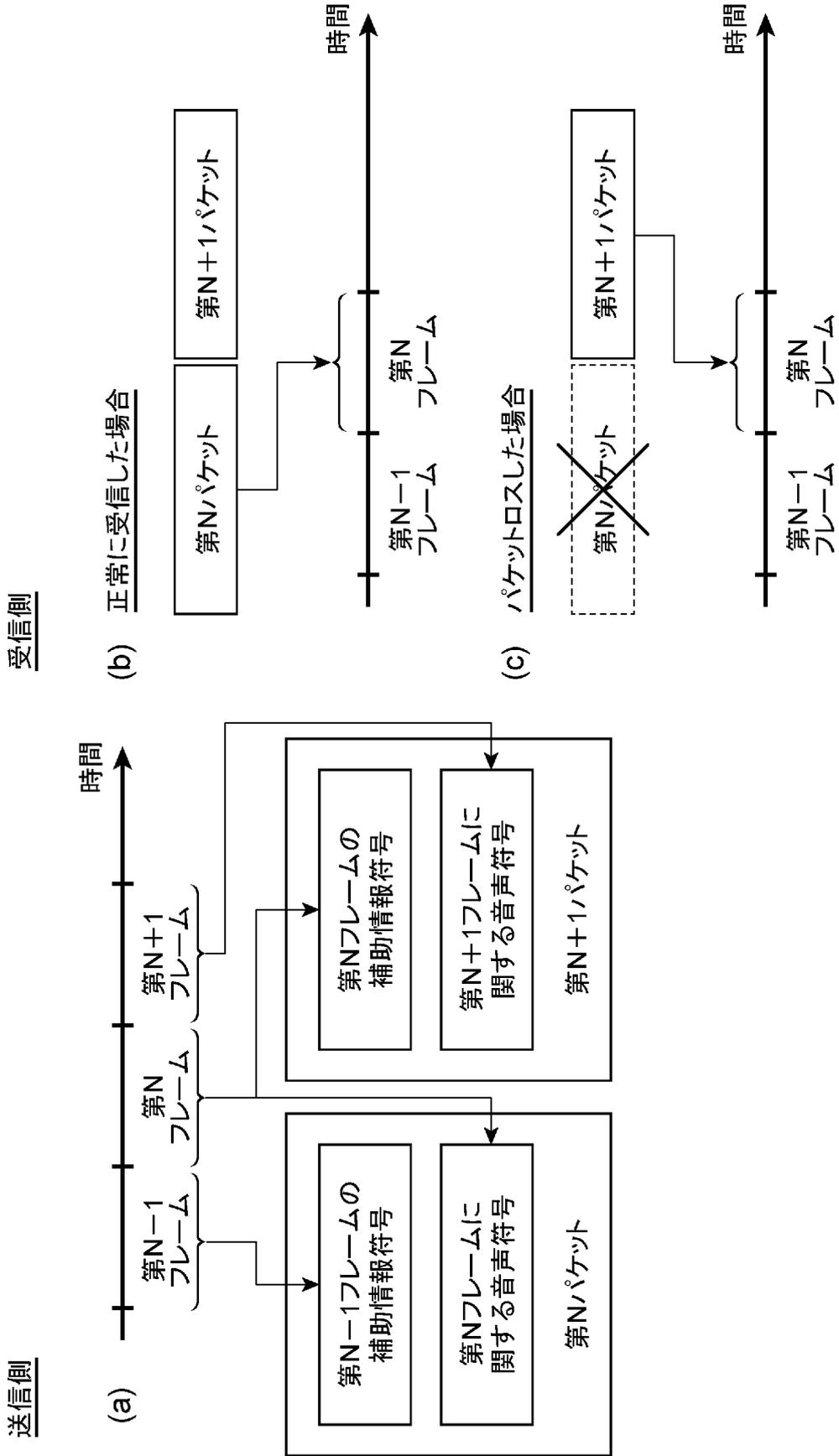
音声パケット正常受信時に補助情報符号を復号する補助情報復号部と、

補助情報符号を復号して得られる補助情報を蓄積する補助情報蓄積部と、

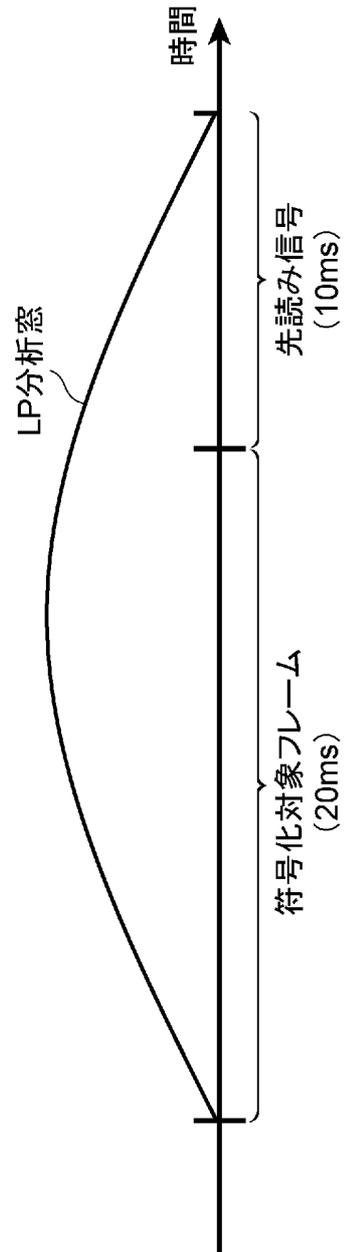
音声パケットロス検出時に音声パラメータを出力する音声パラメータ紛失処理部と、

音声パラメータから復号音声を合成する音声合成部、
として機能させる音声復号プログラム。

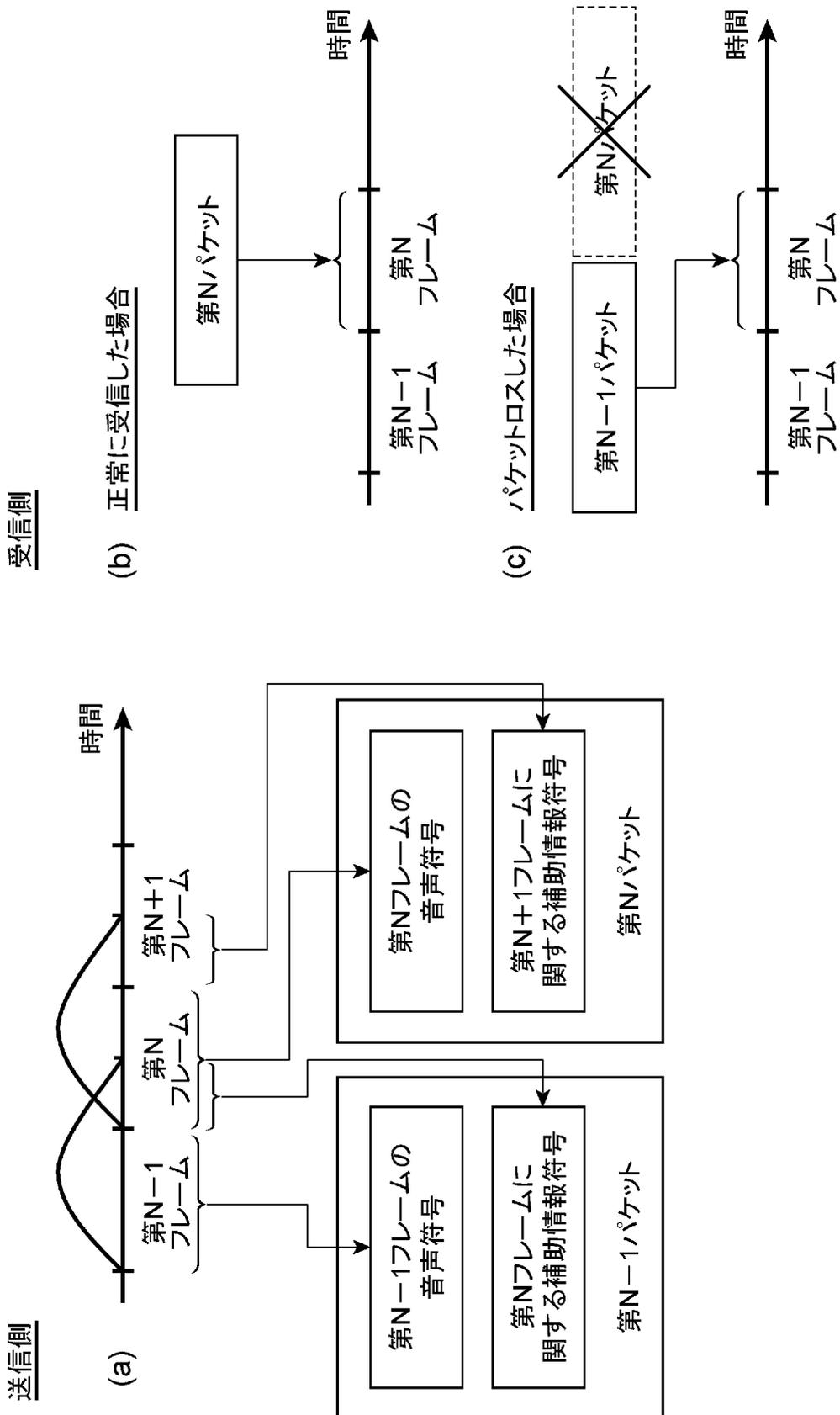
[図1]



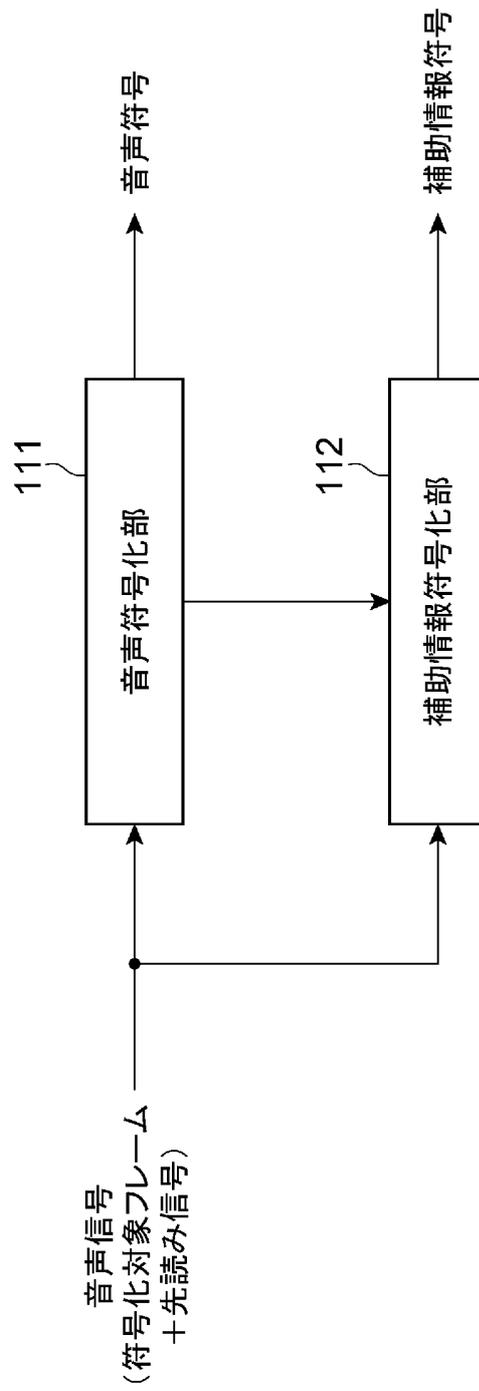
[図2]



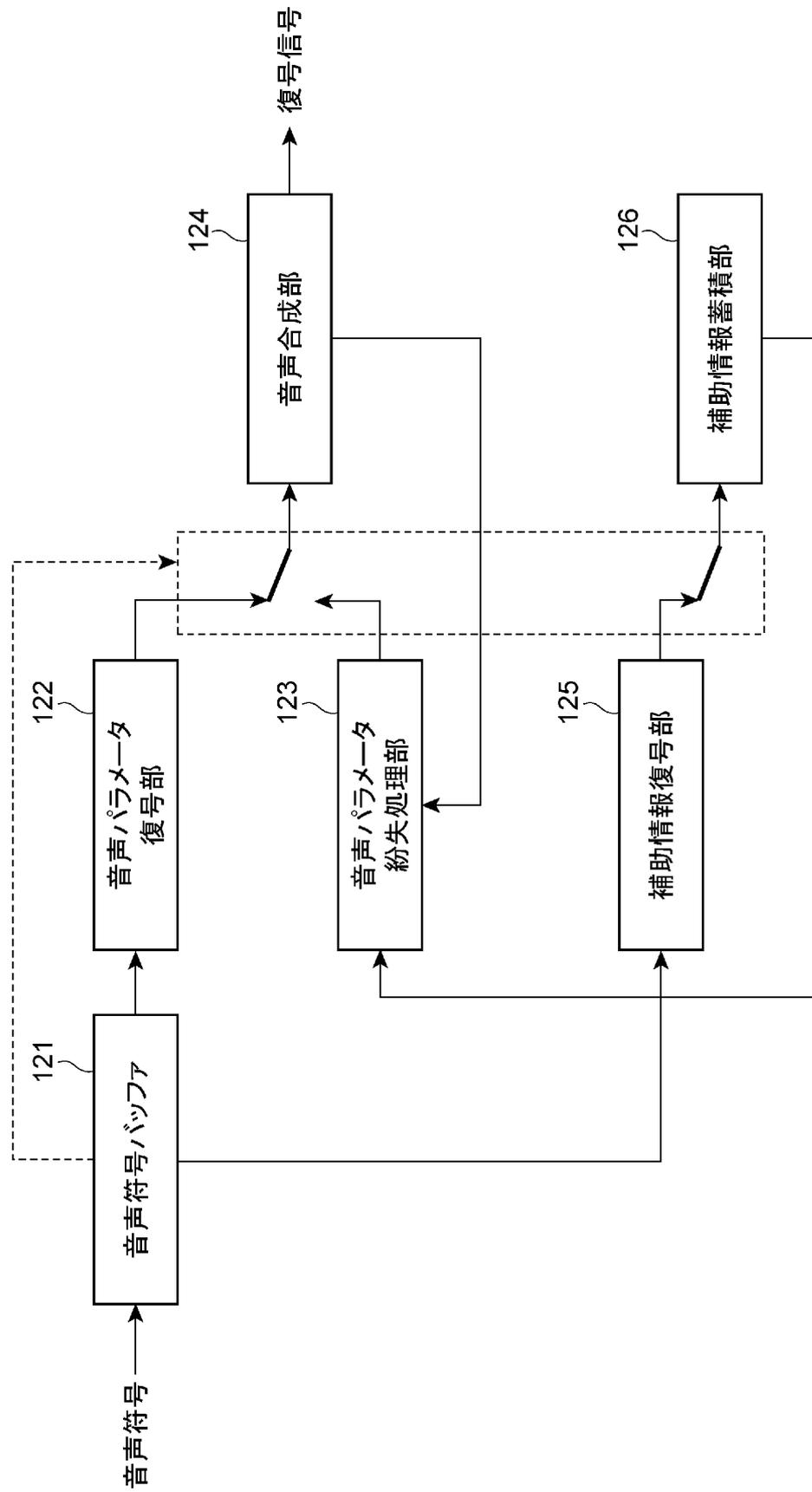
[図3]



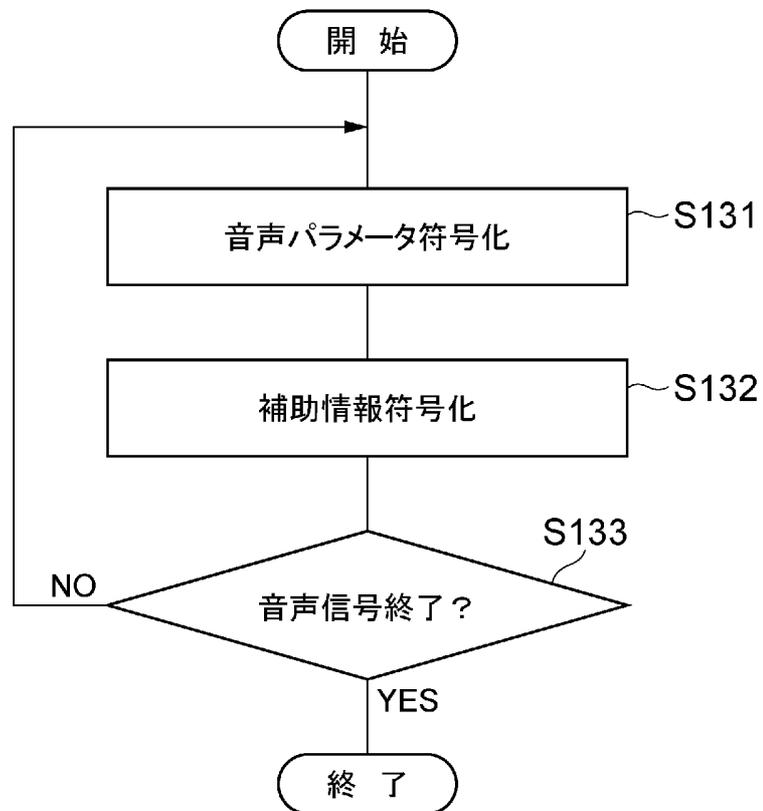
[図4]



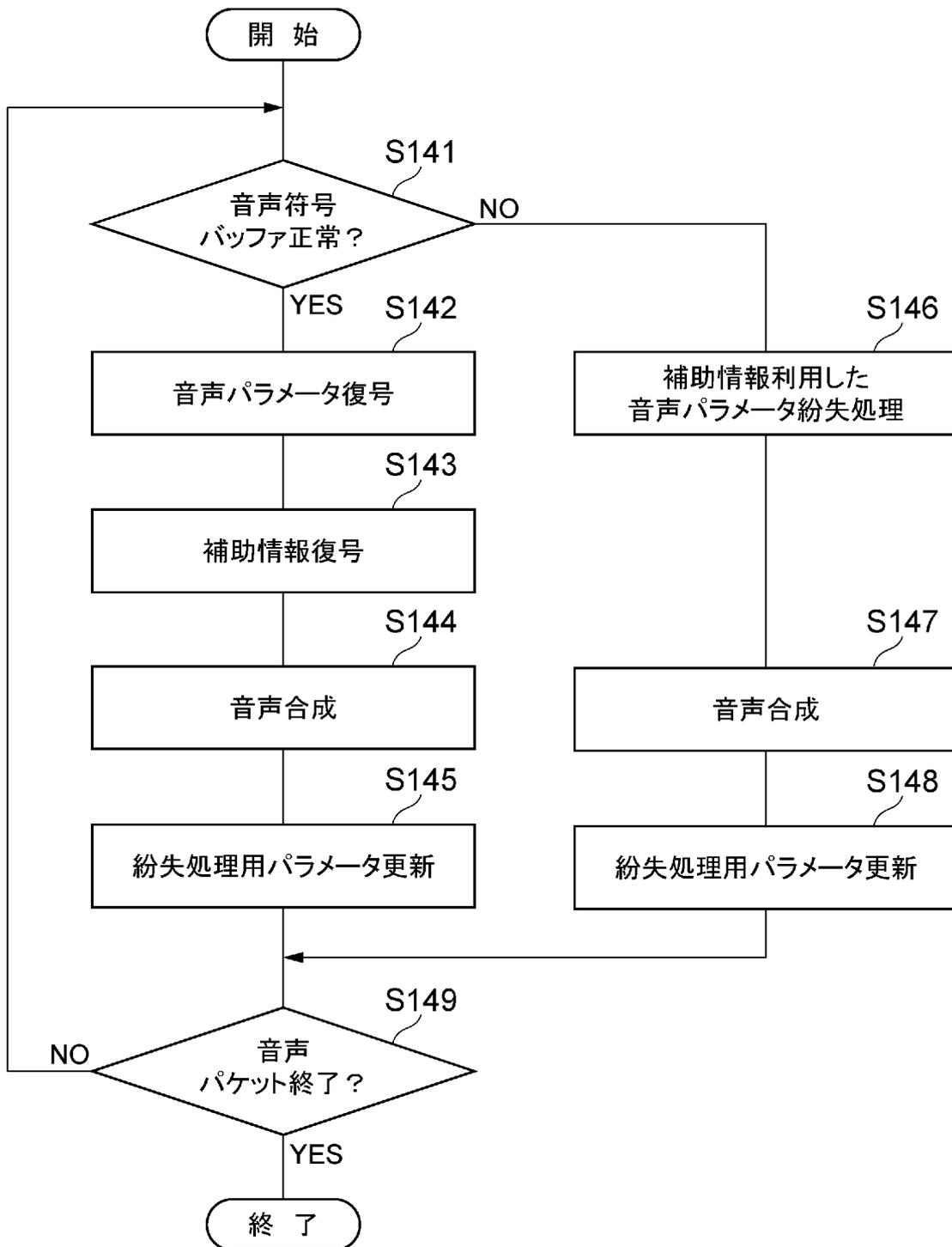
[図5]



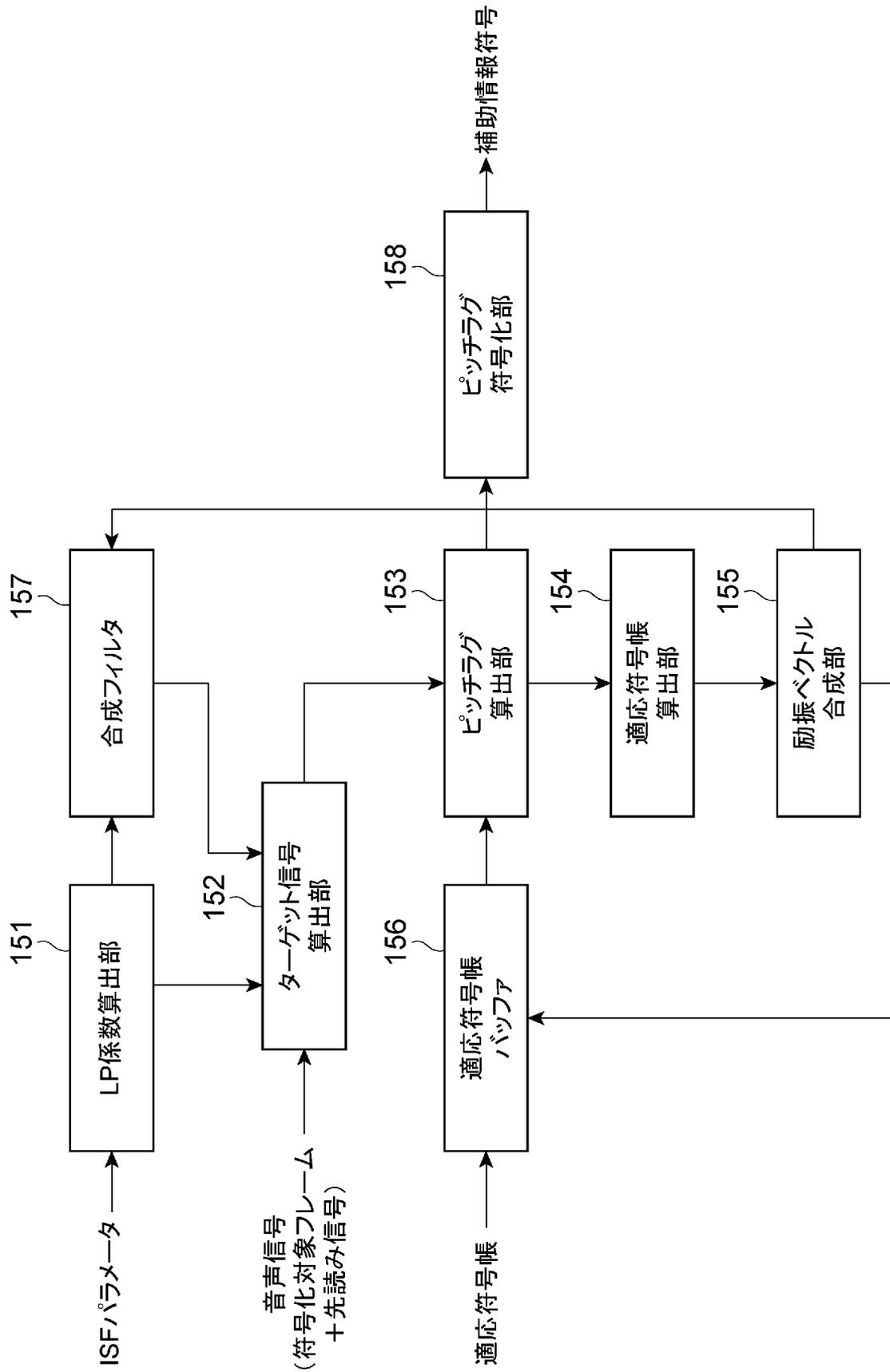
[図6]



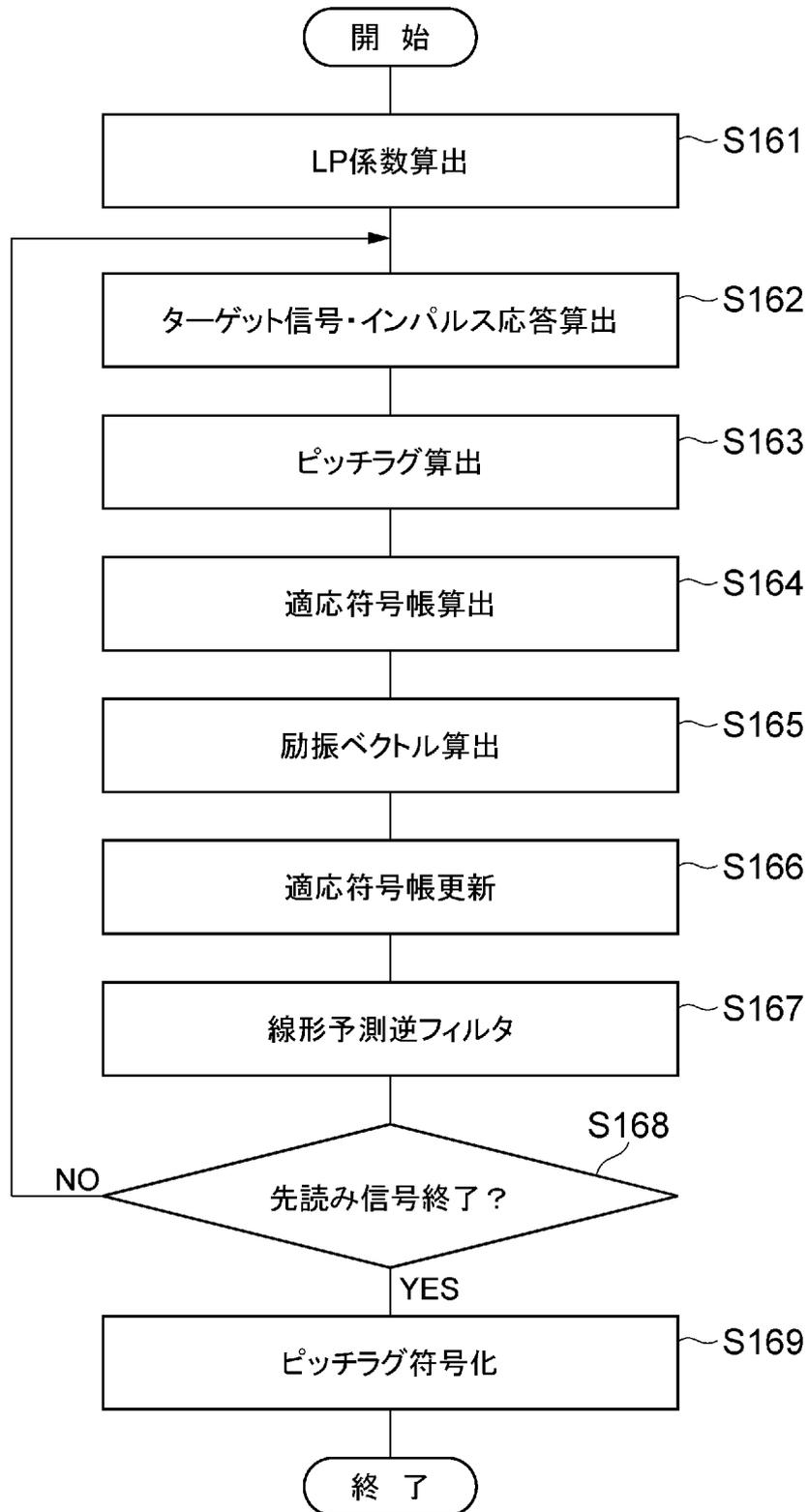
[図7]



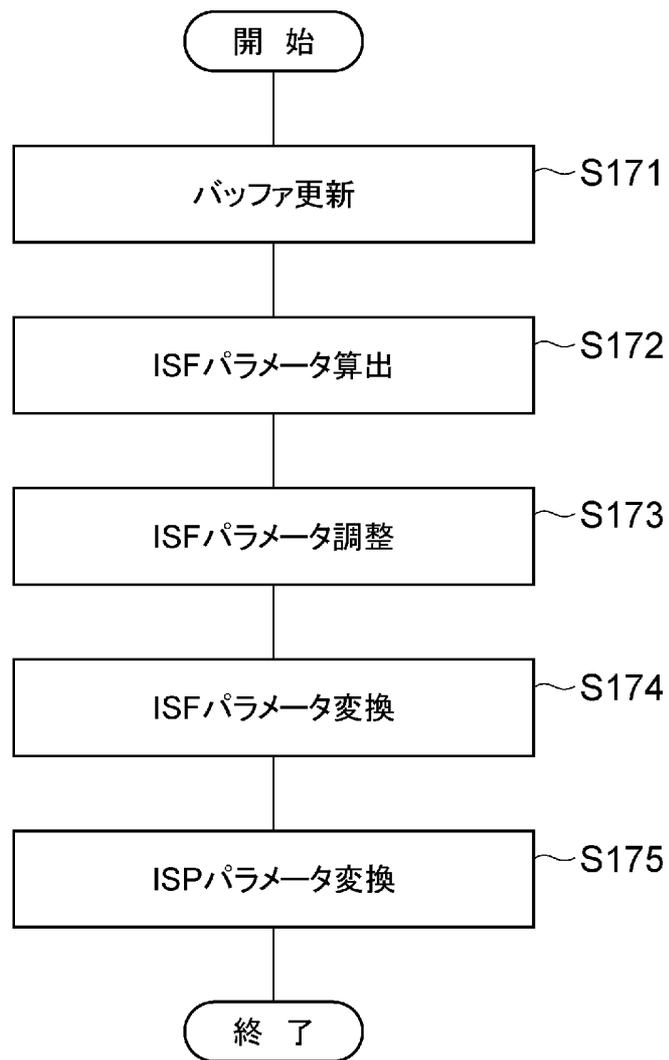
[図8]



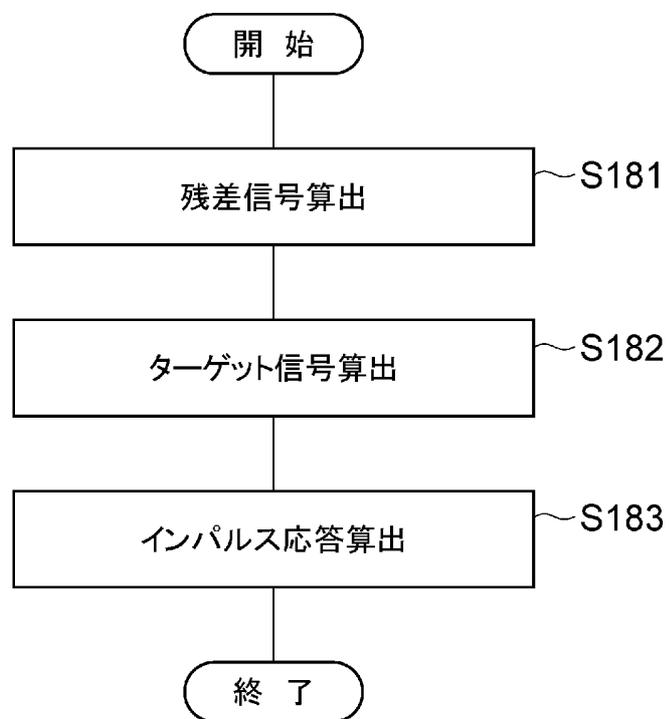
[図9]



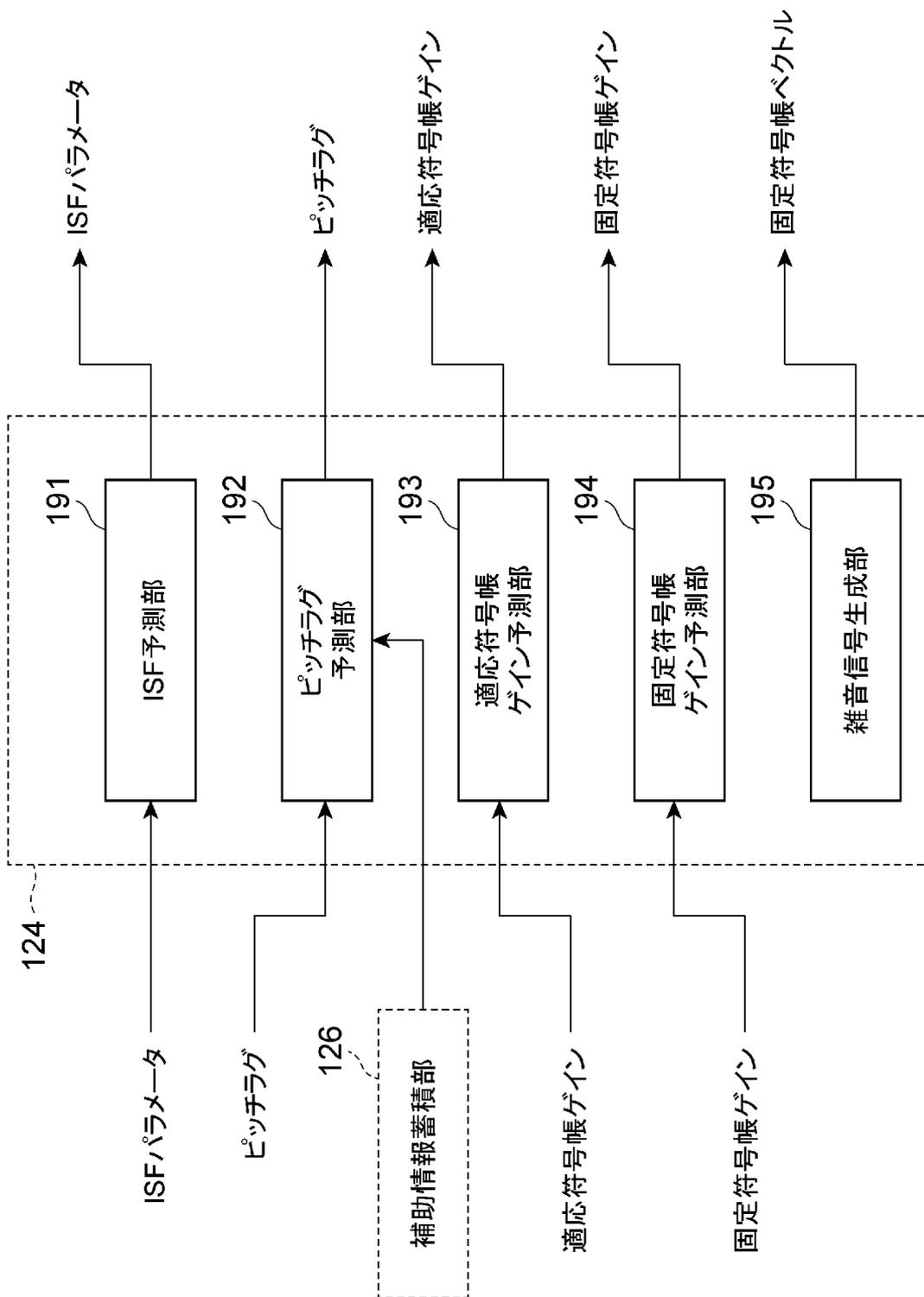
[図10]



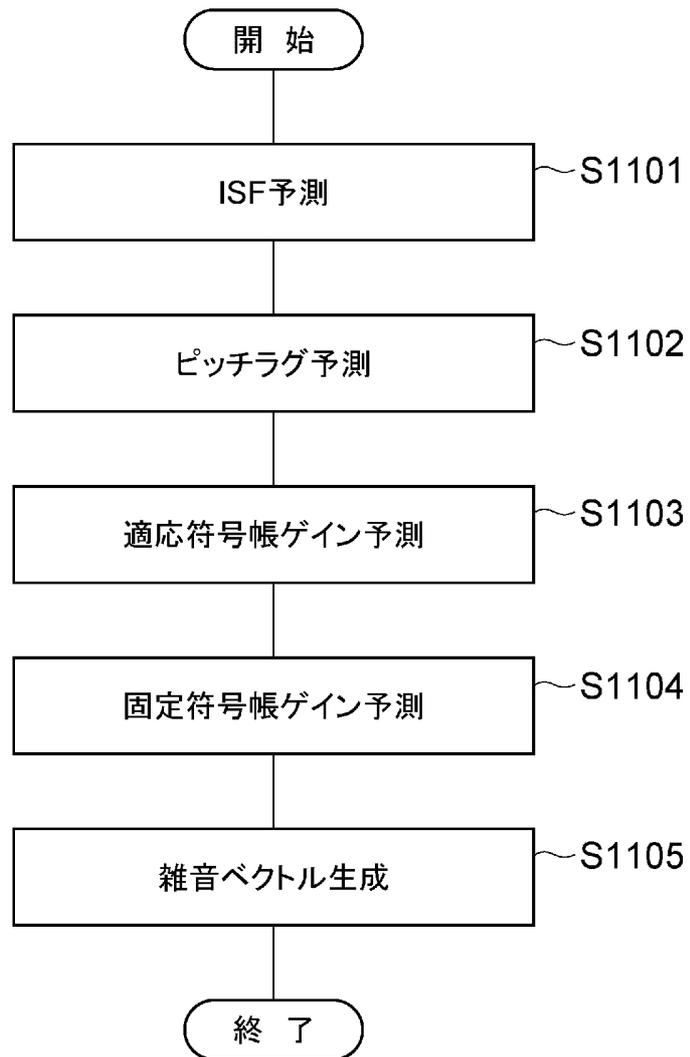
[図11]



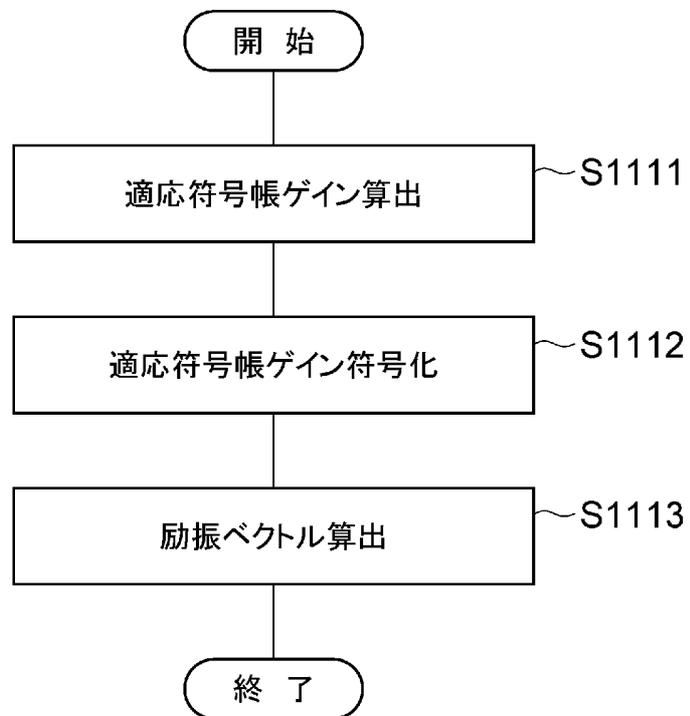
[図12]



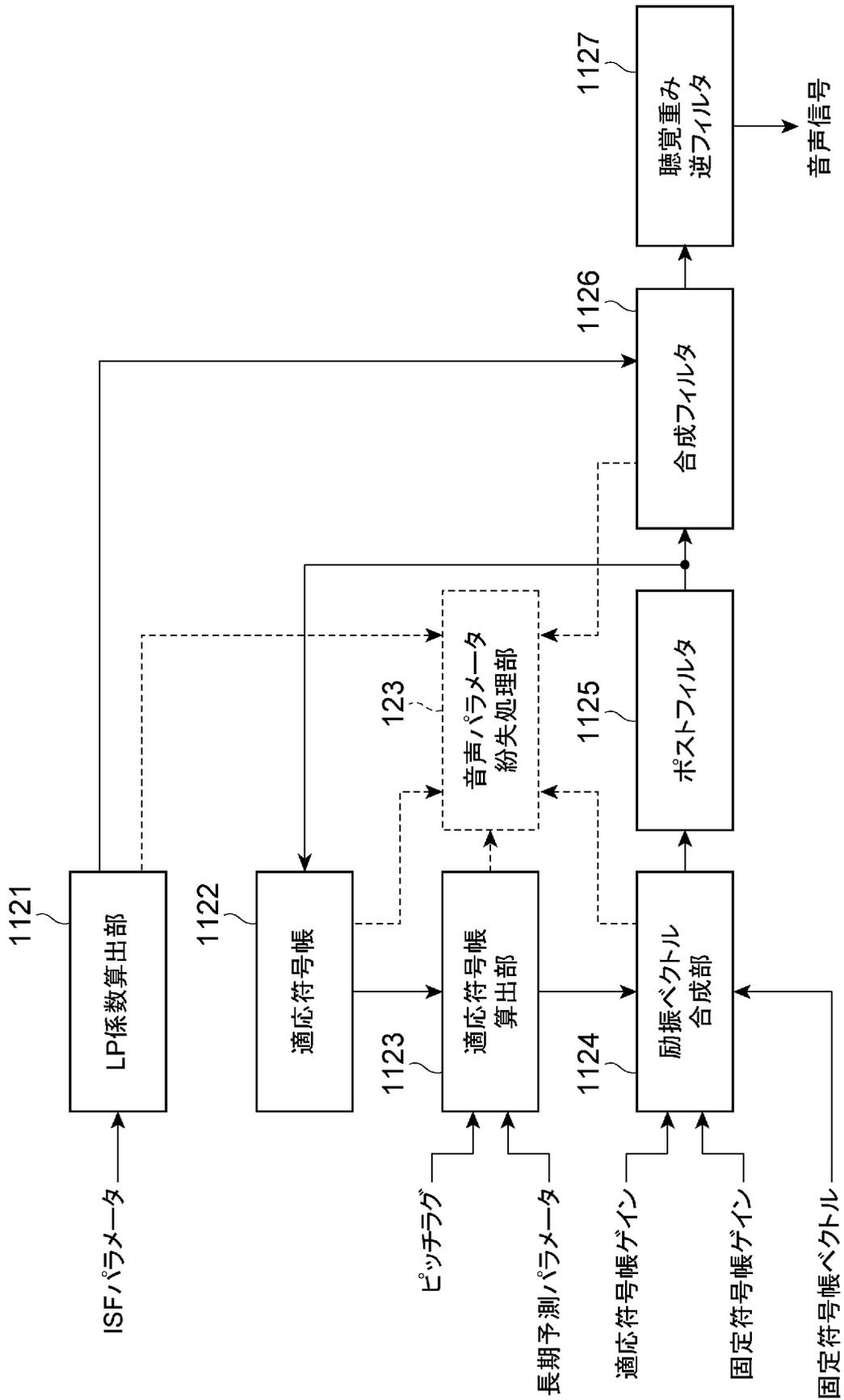
[図13]



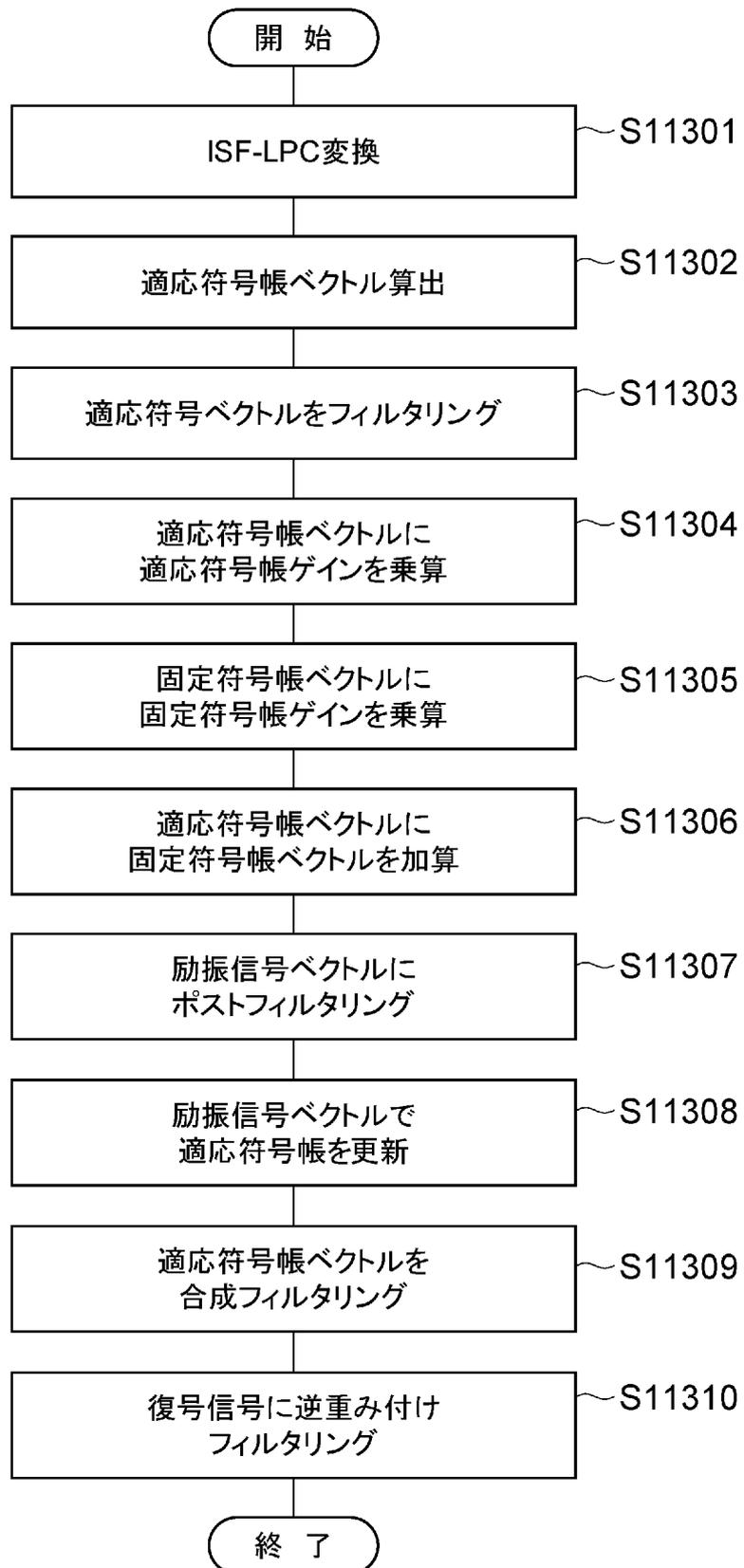
[図14]



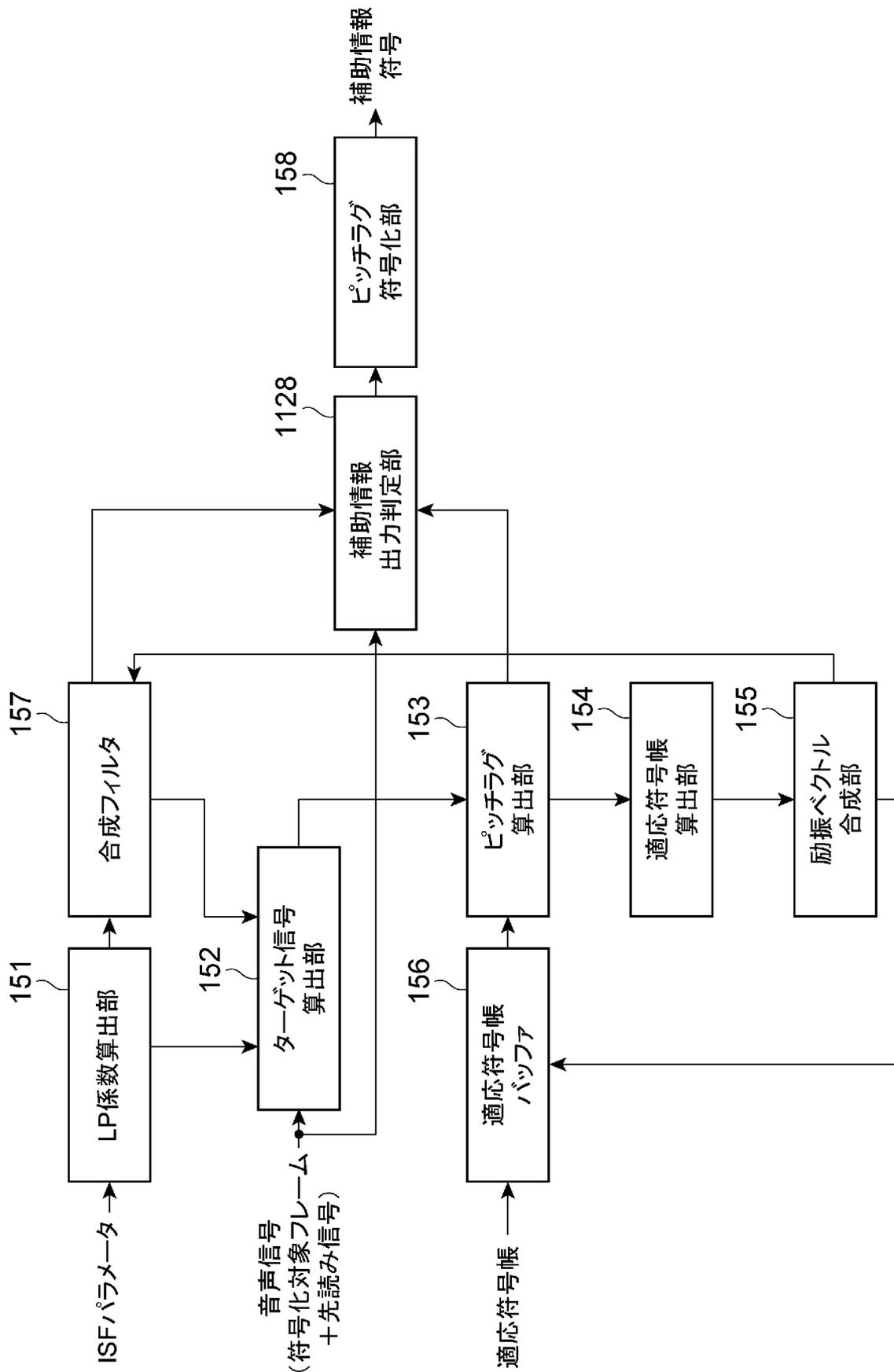
[図15]



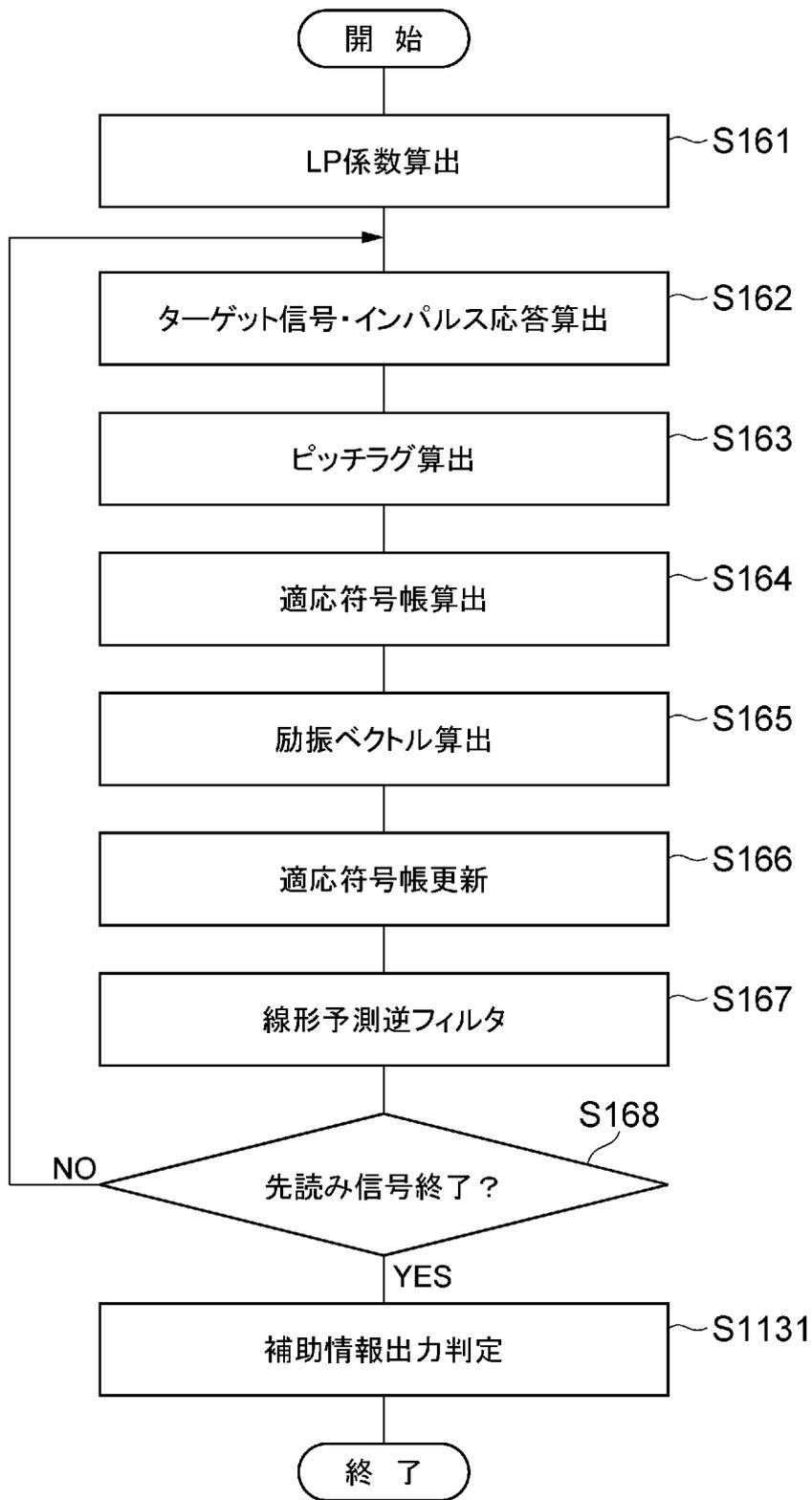
[図16]



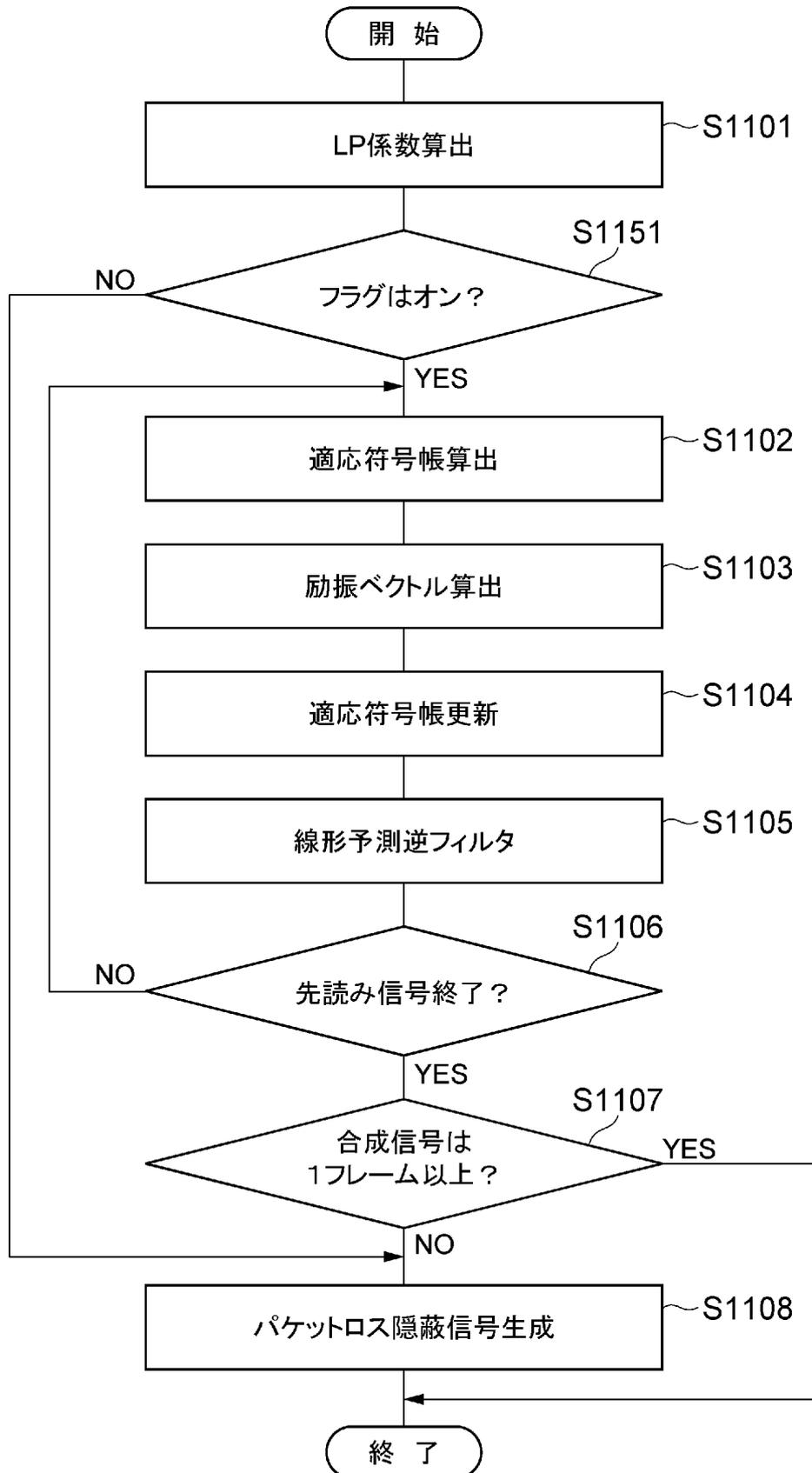
[図17]



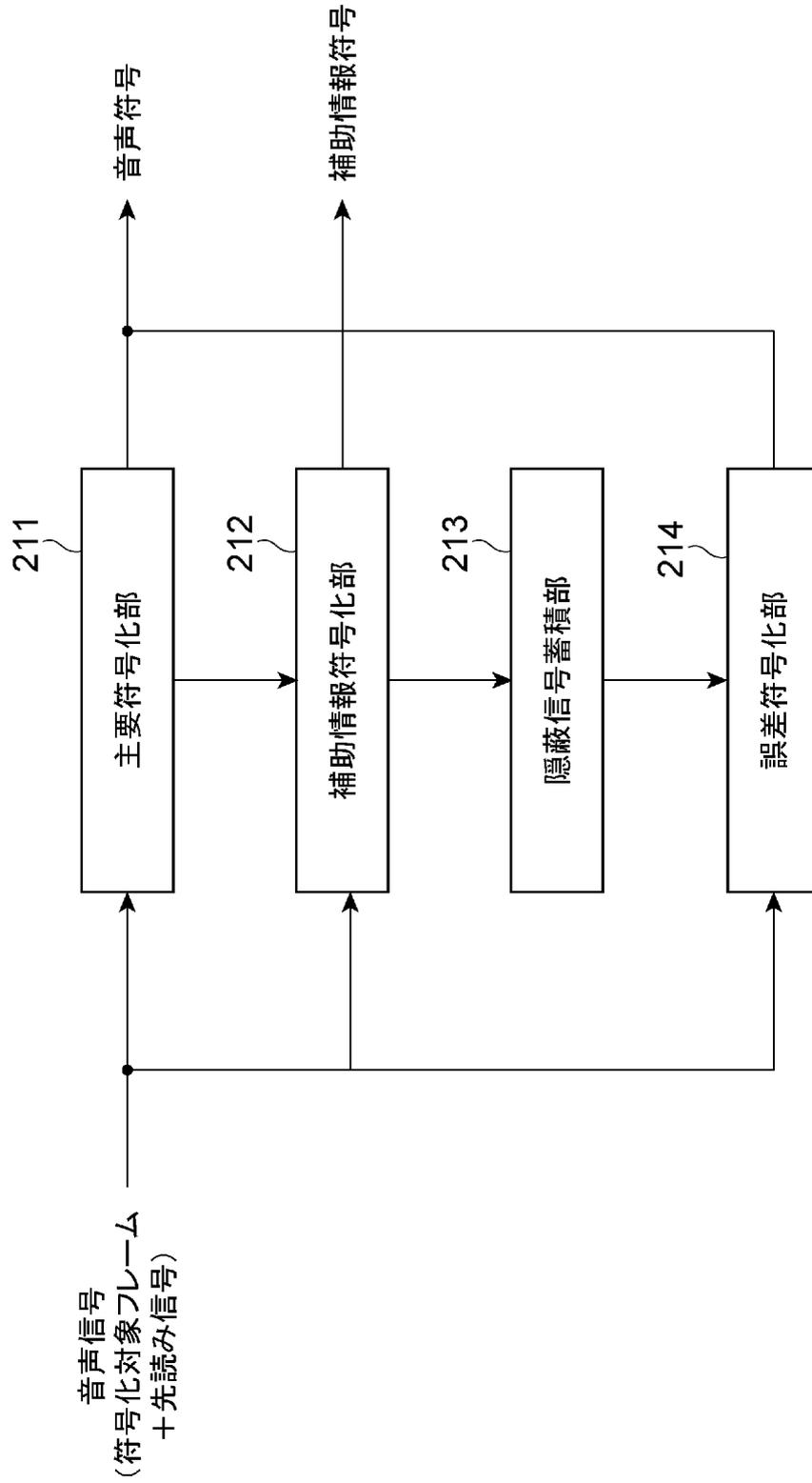
[図18]



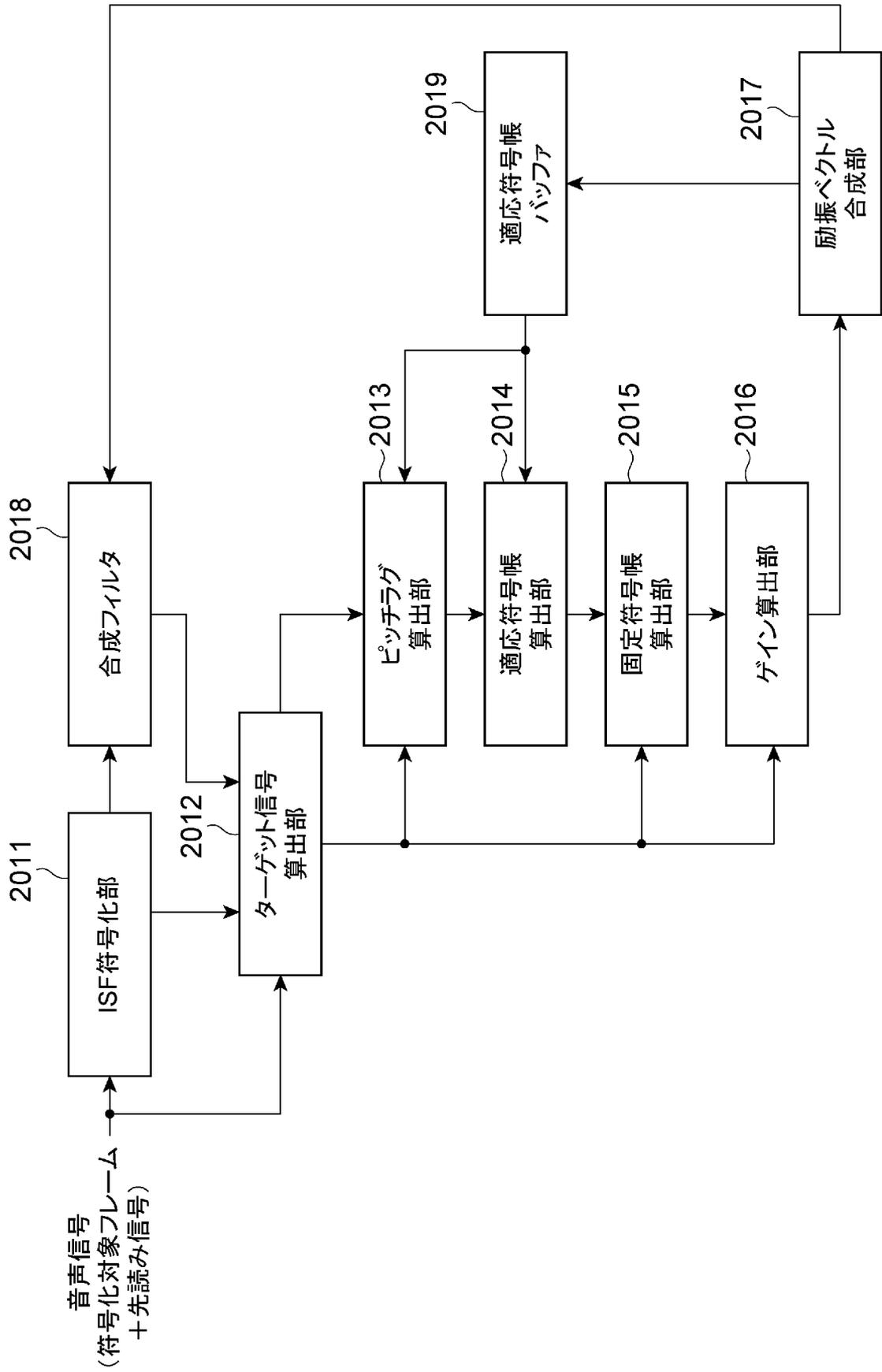
[図19]



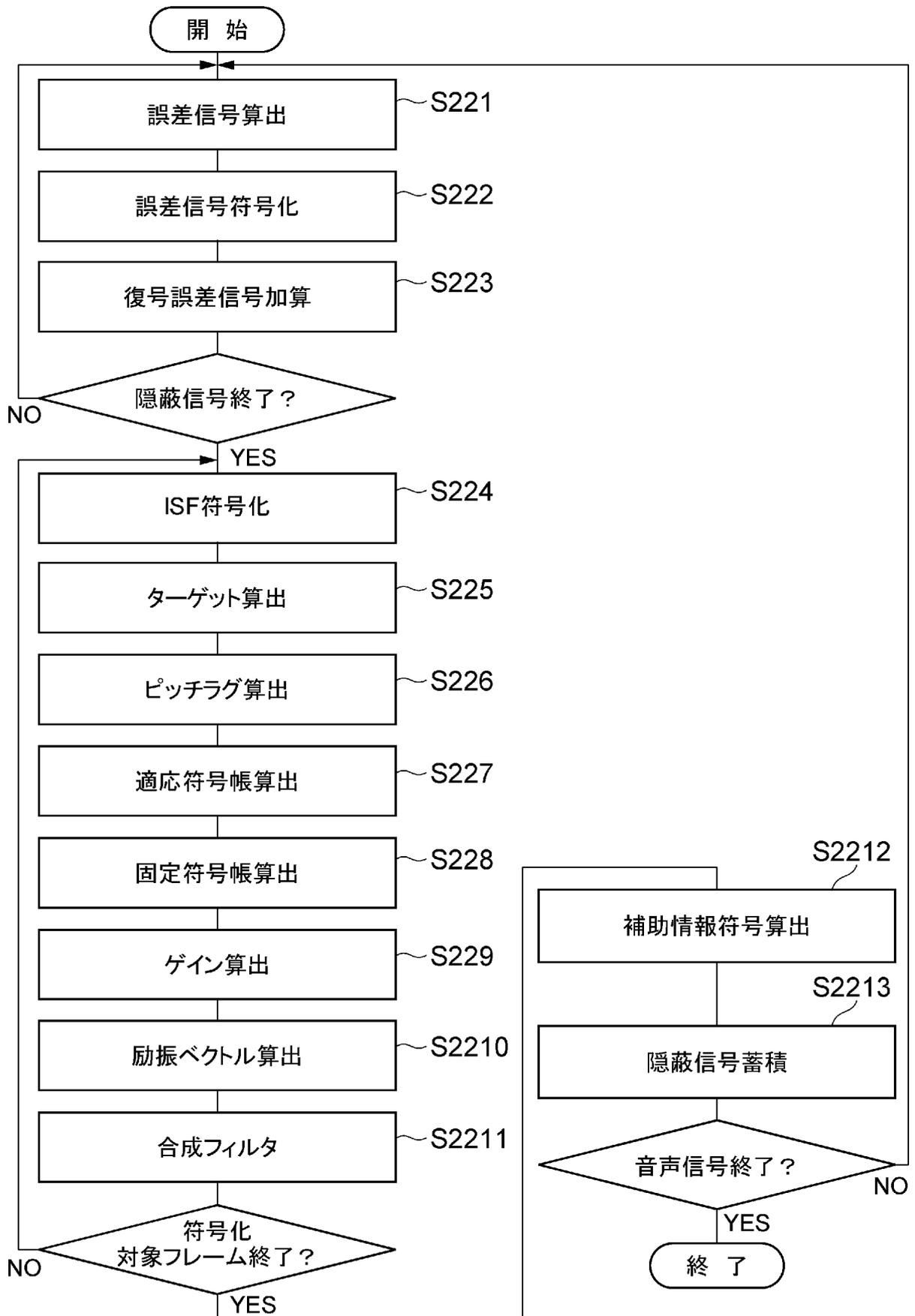
[図20]



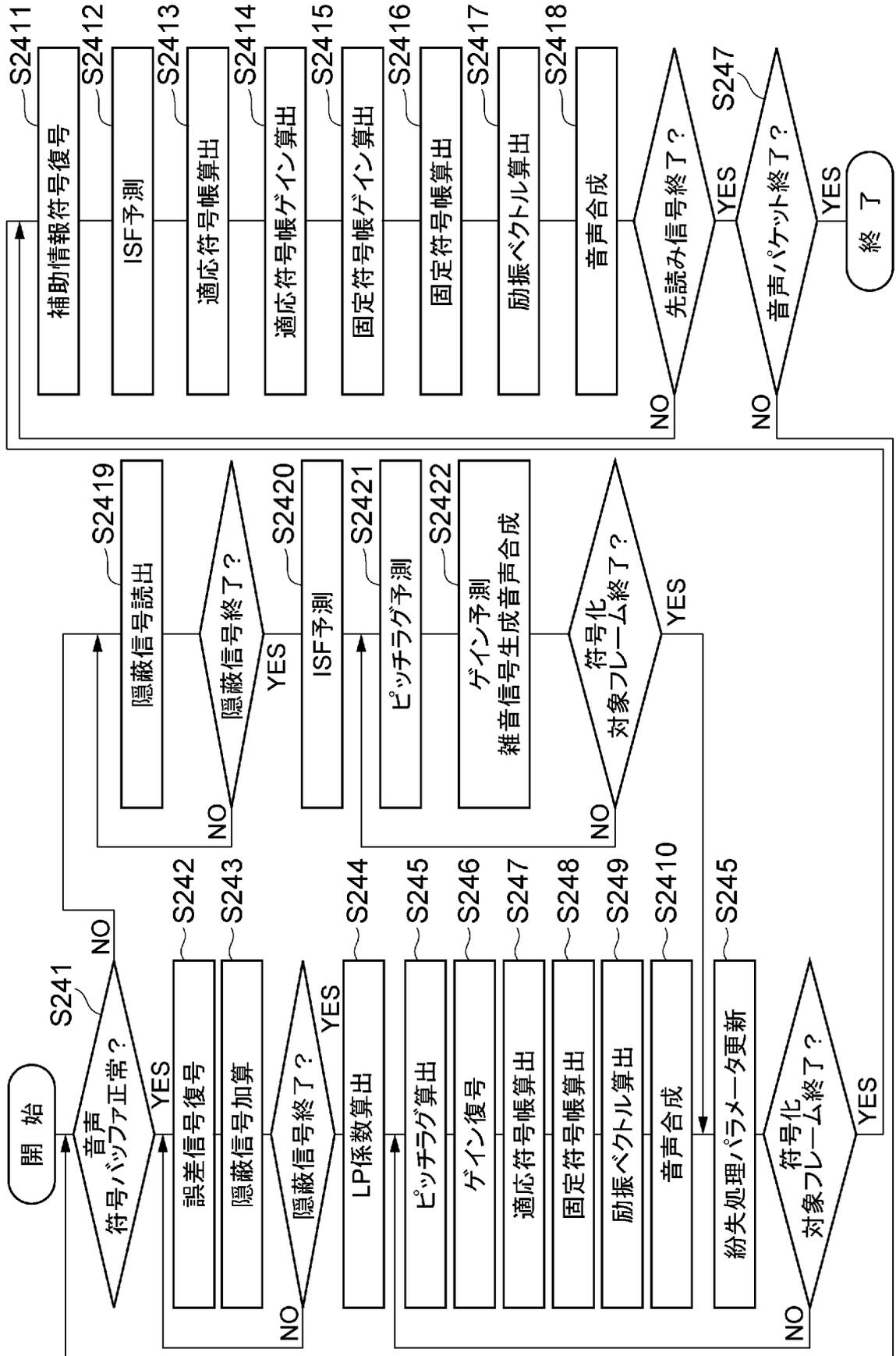
[図21]



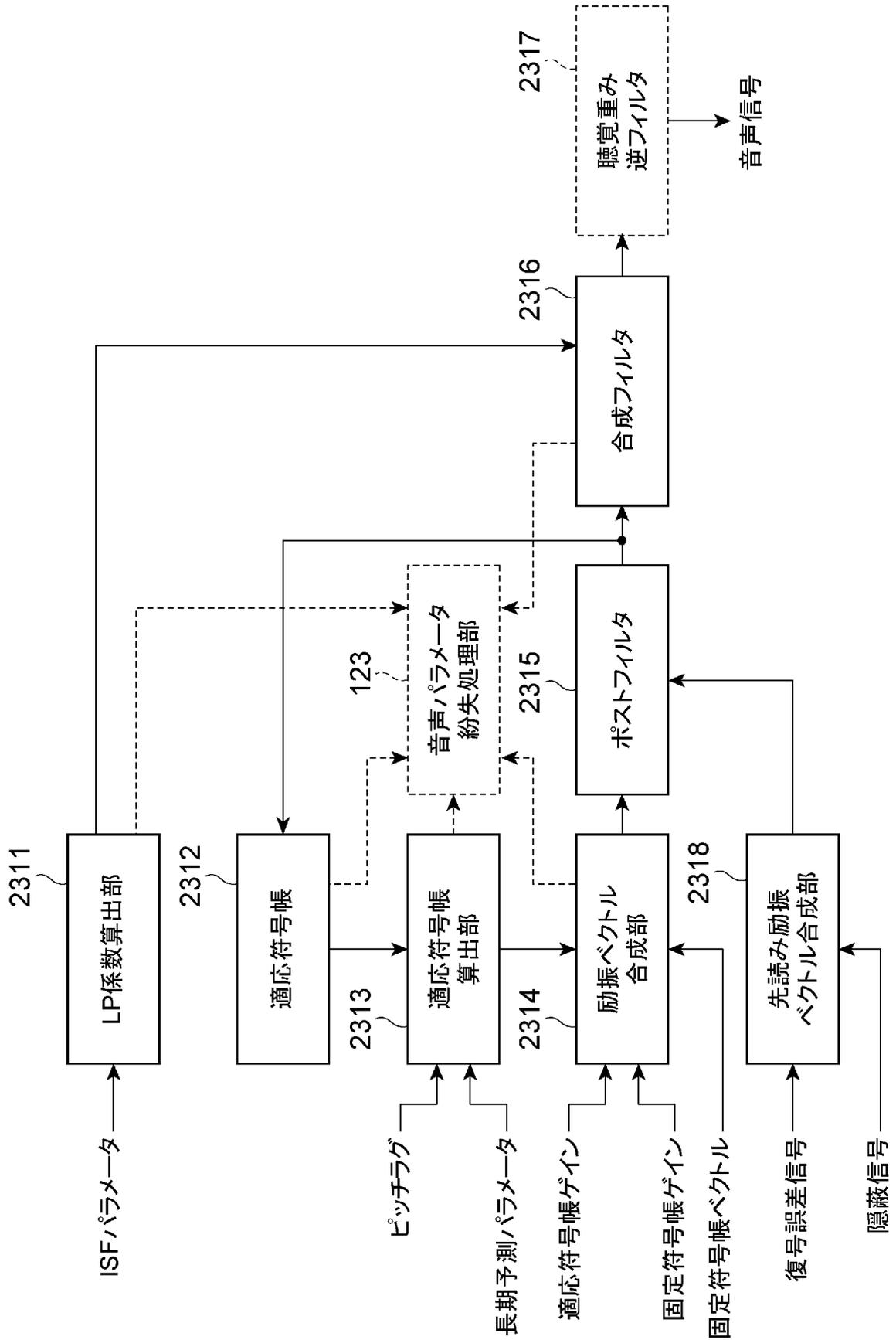
[図22]



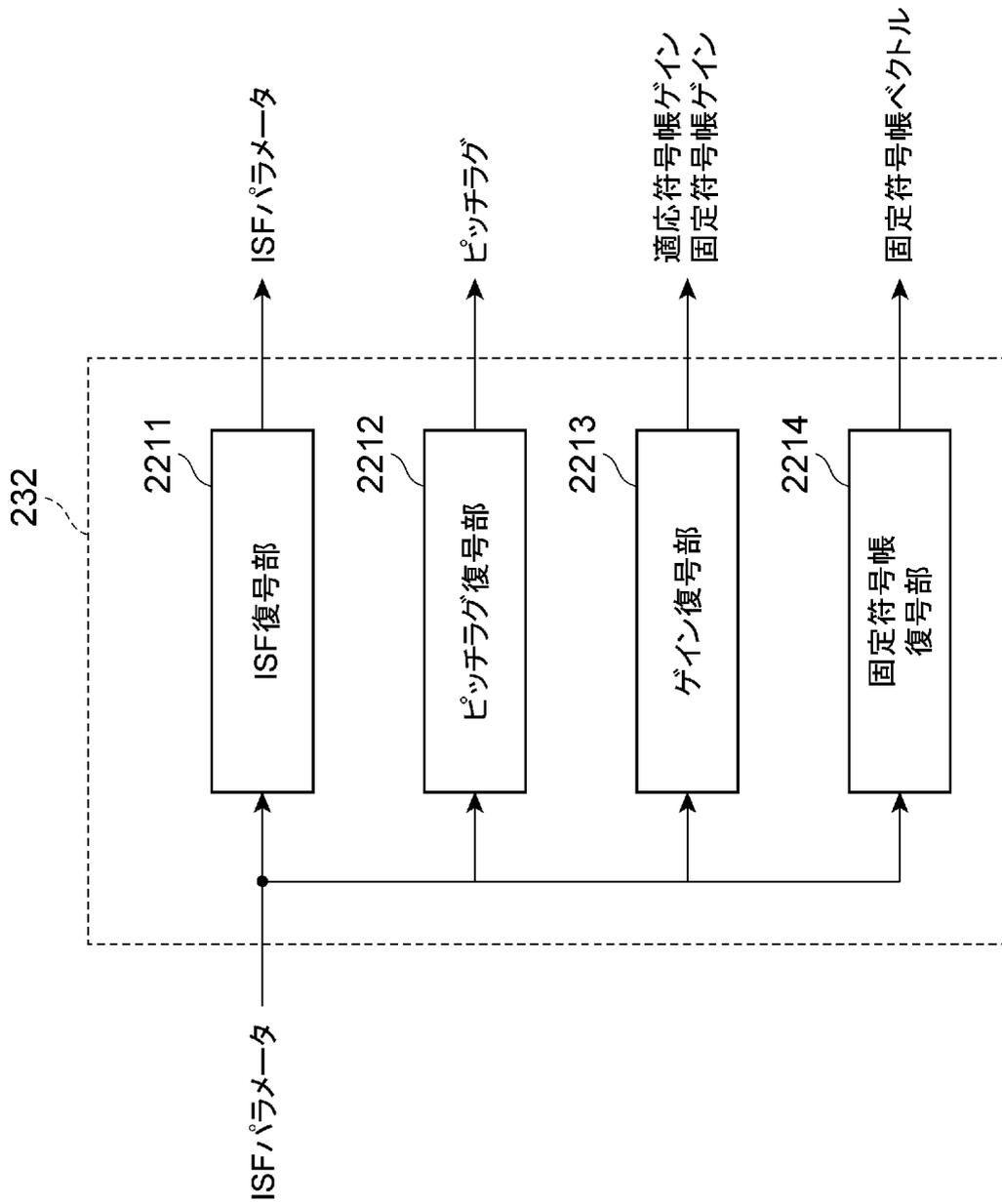
[図24]



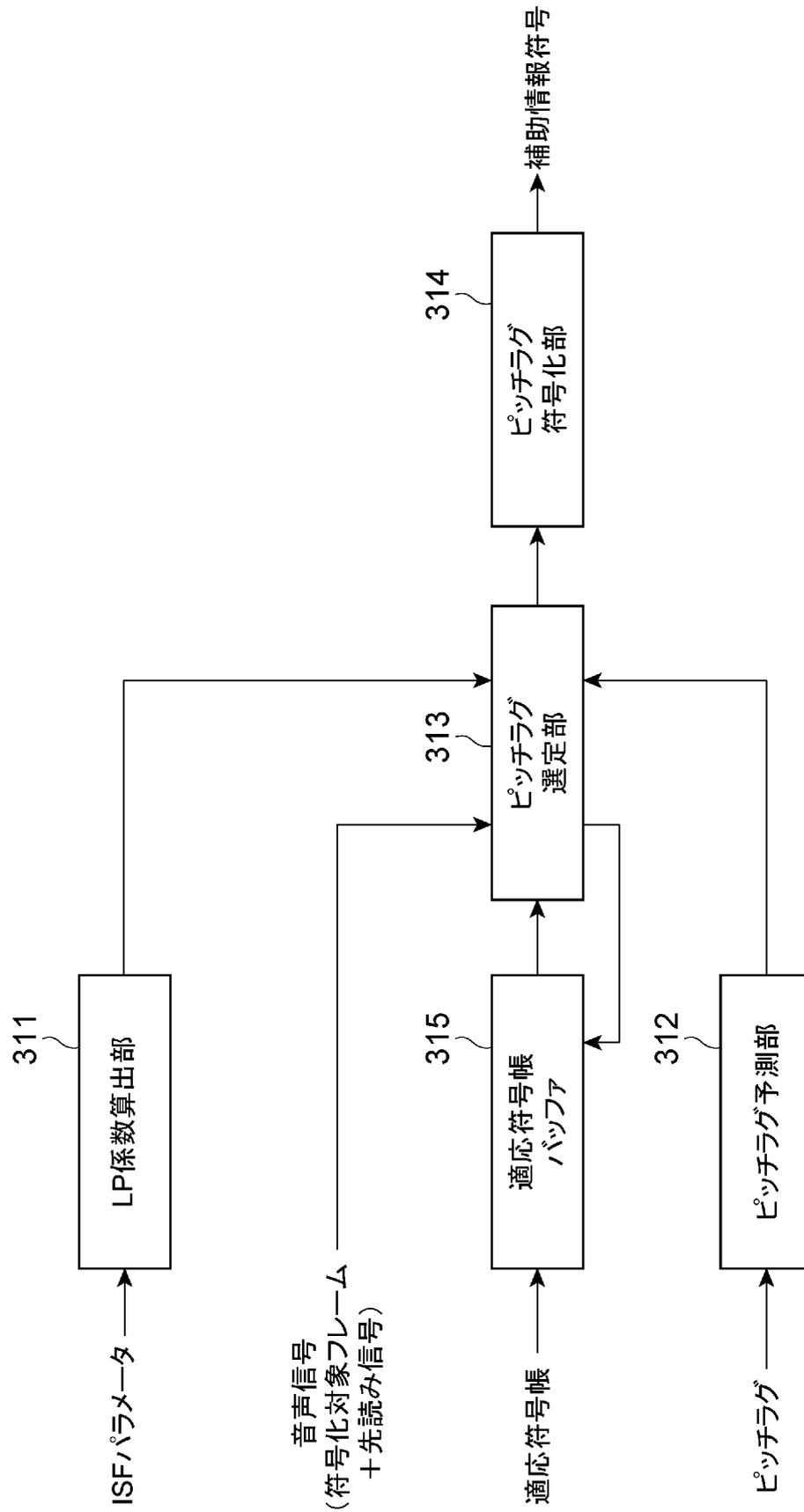
[図25]



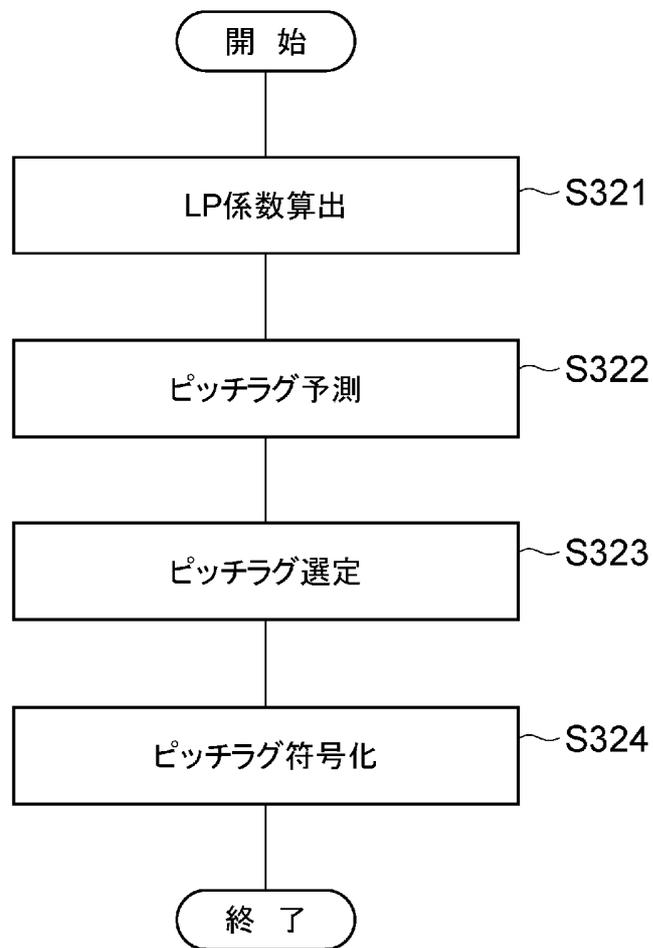
[図26]



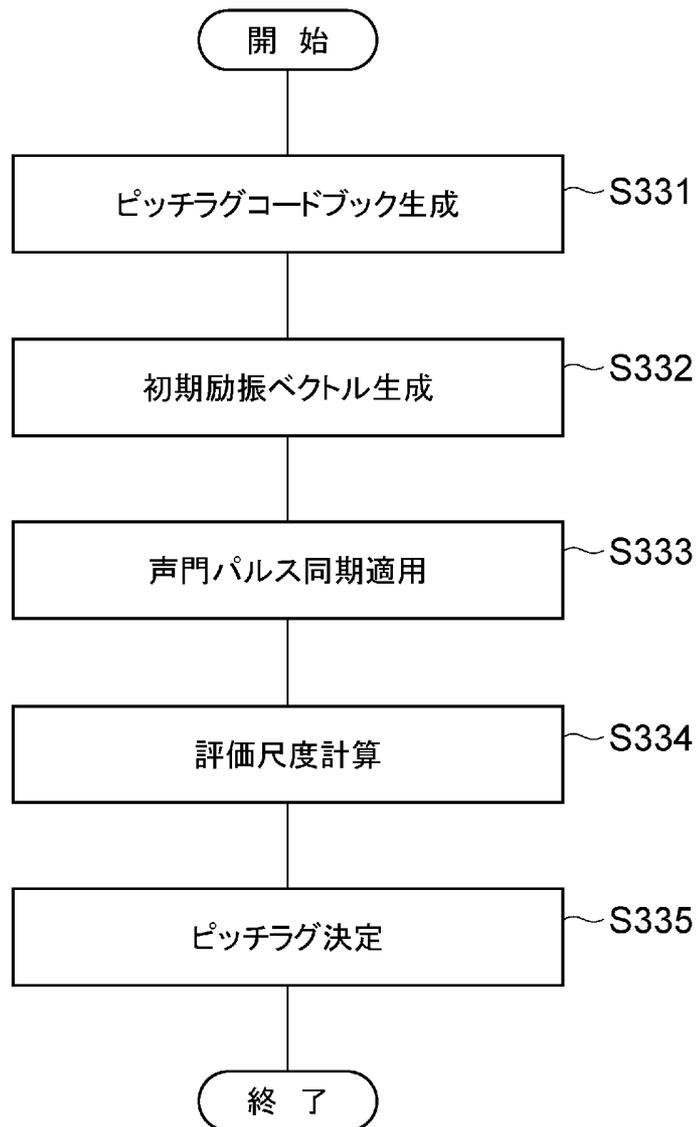
[図27]



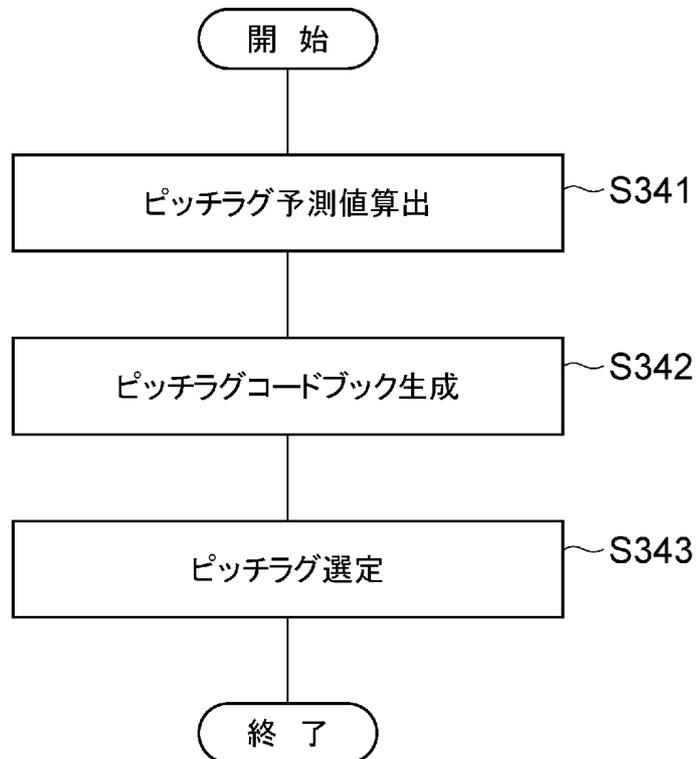
[図28]



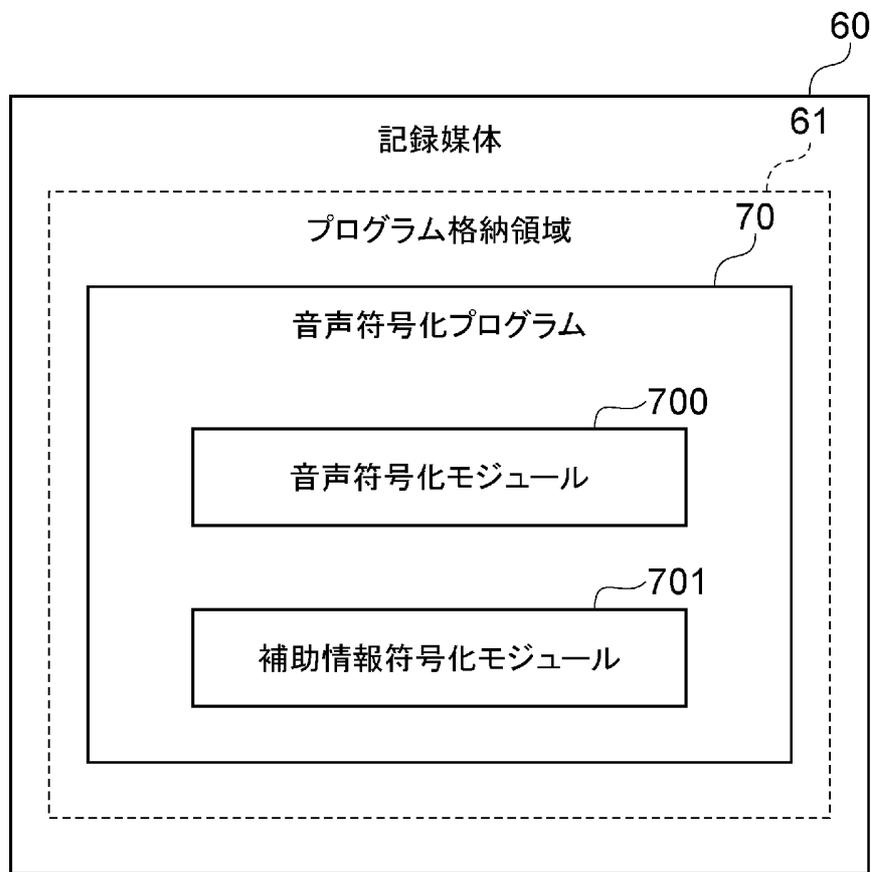
[図29]



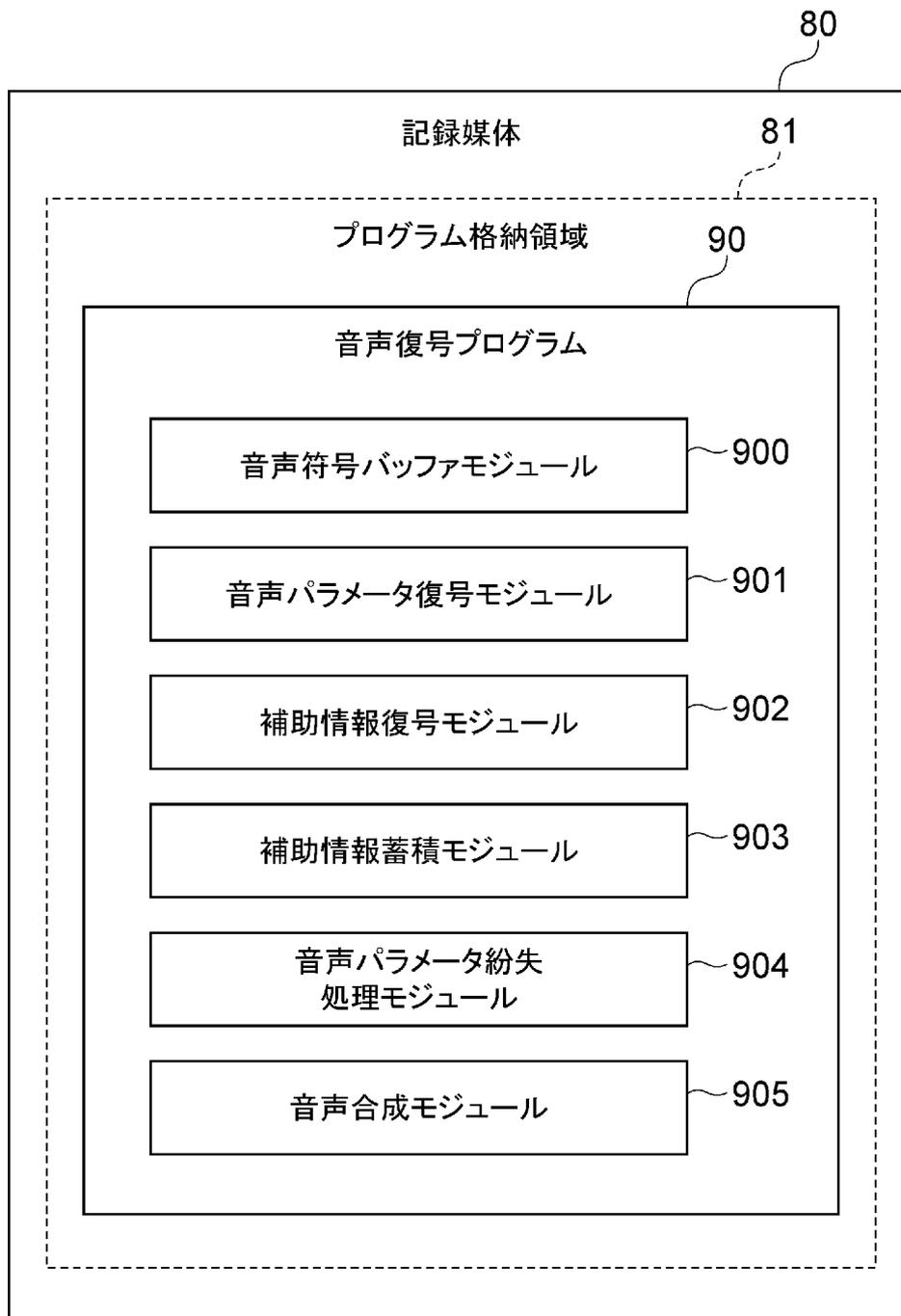
[図30]



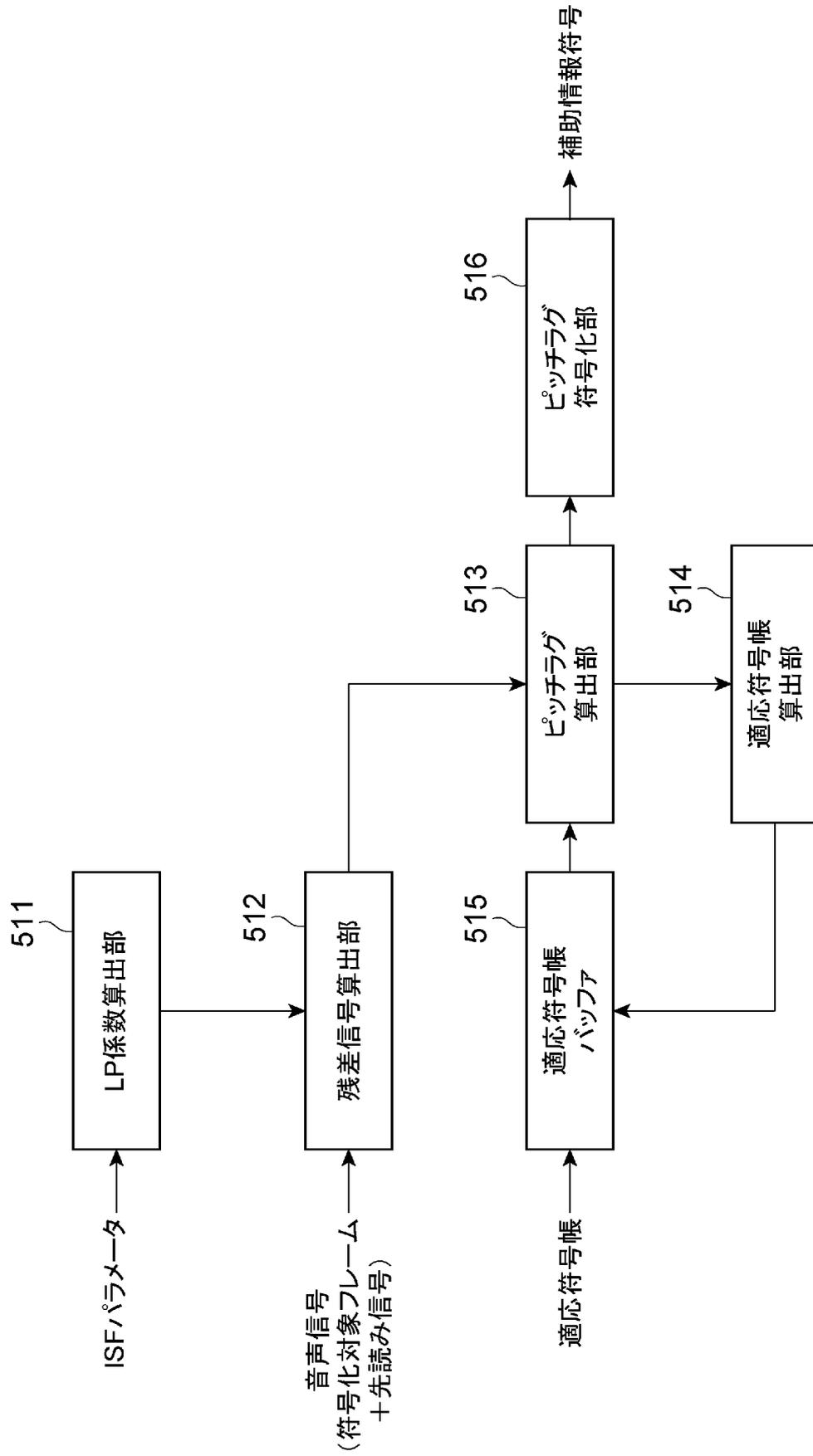
[図31]



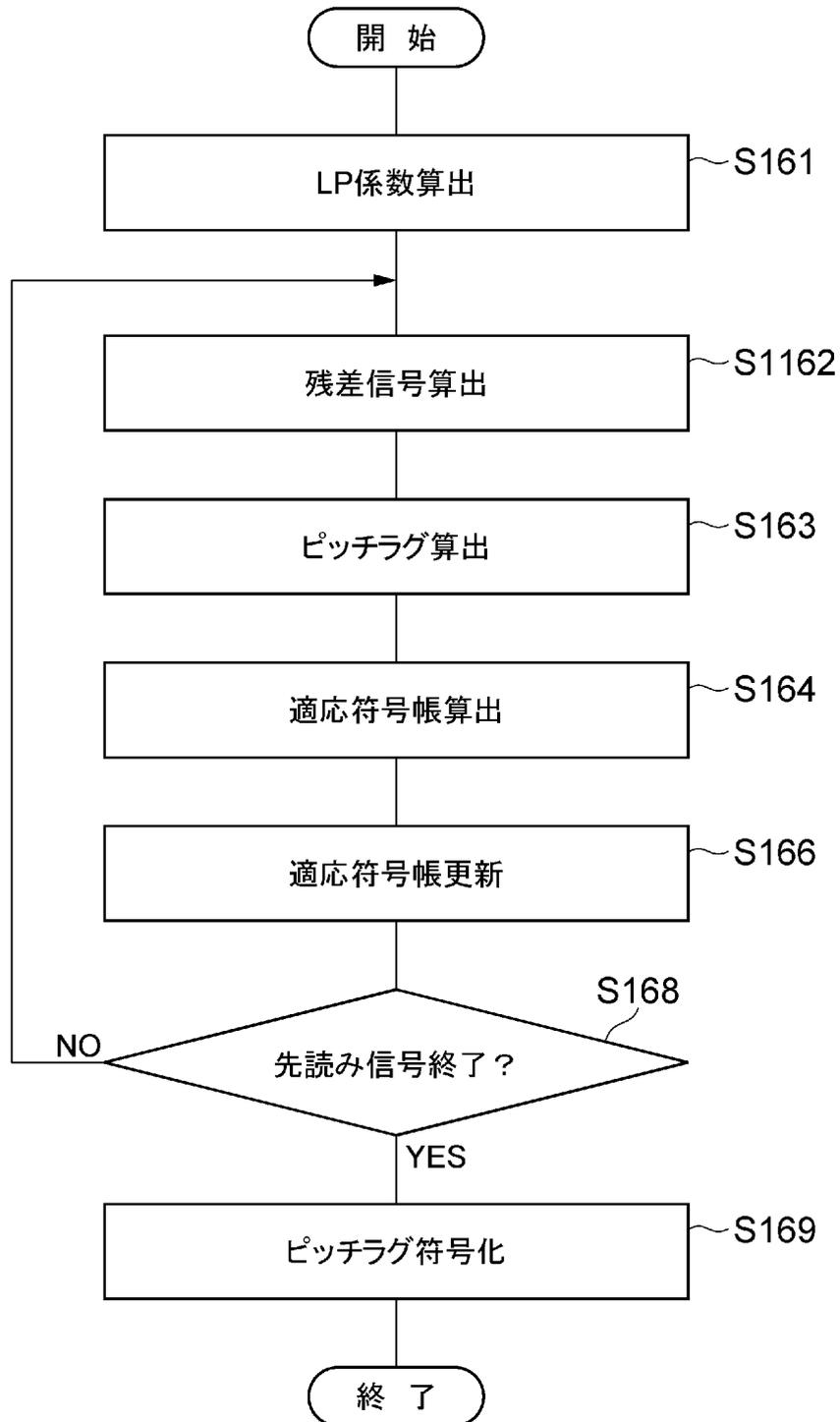
[図32]



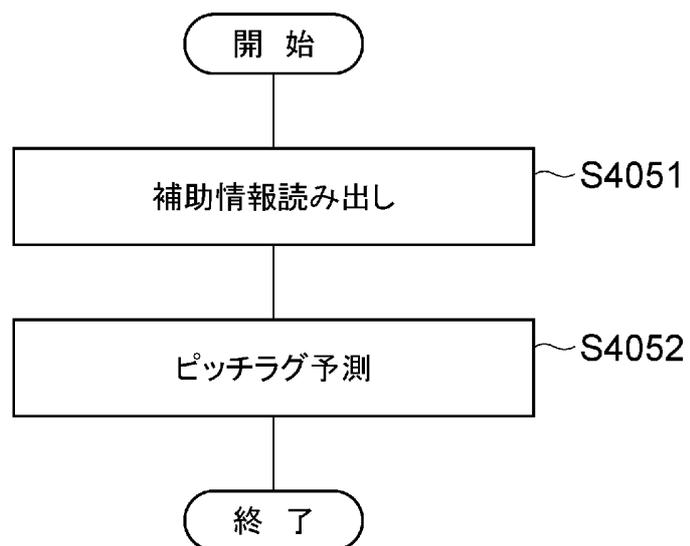
[図33]



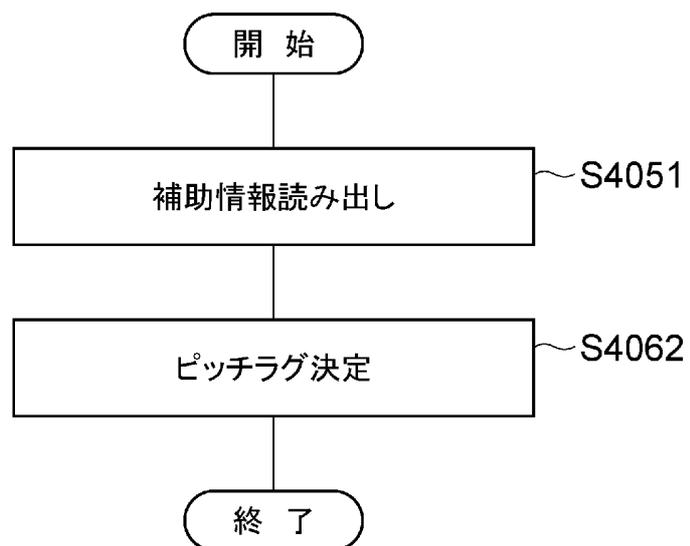
[図34]



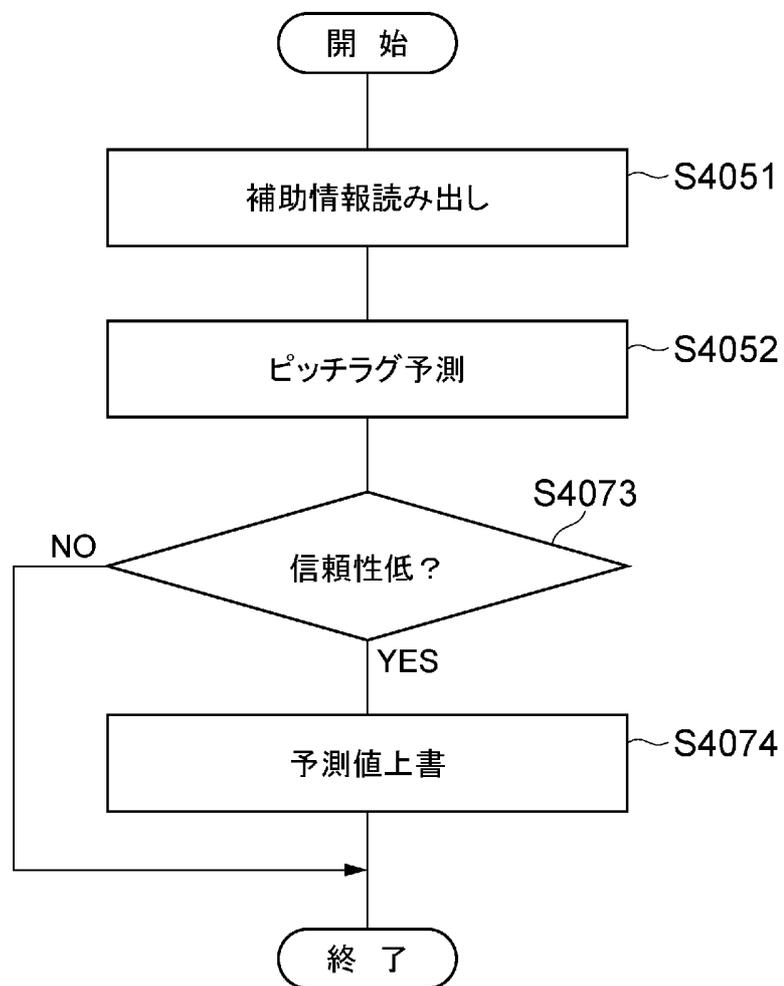
[図35]



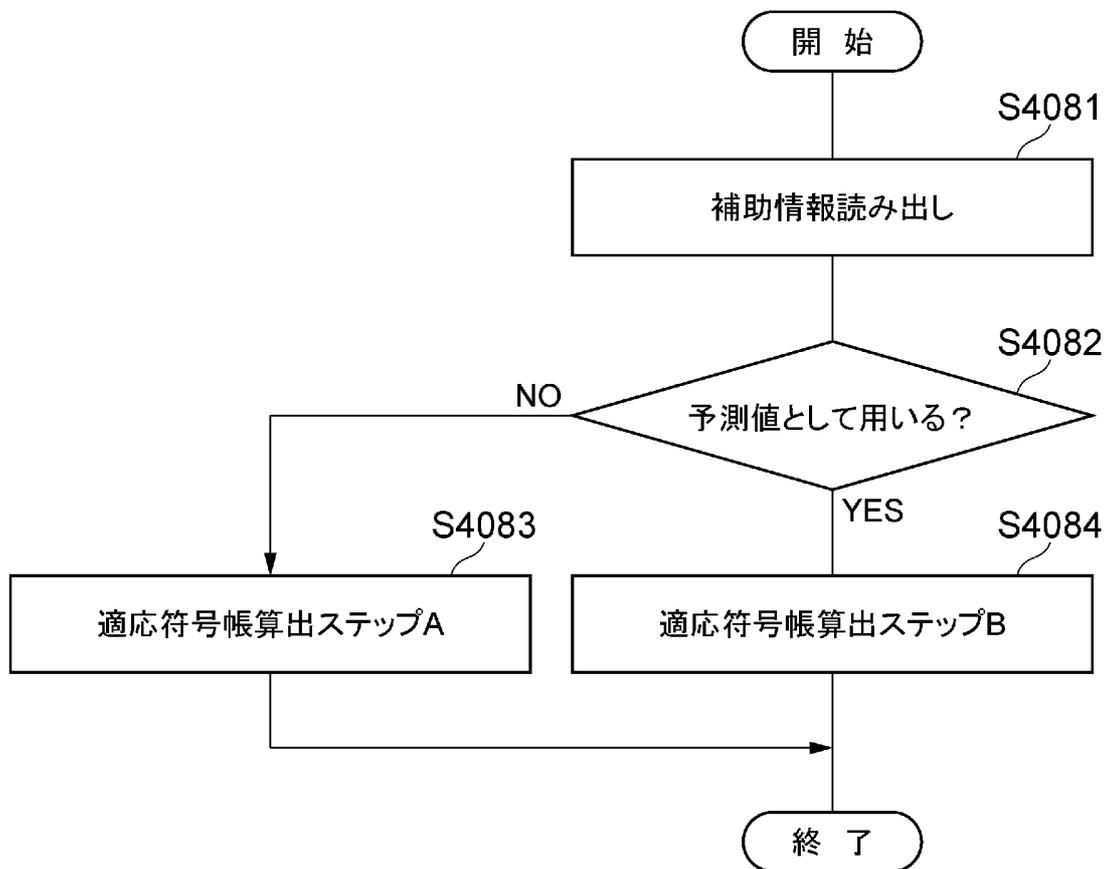
[図36]



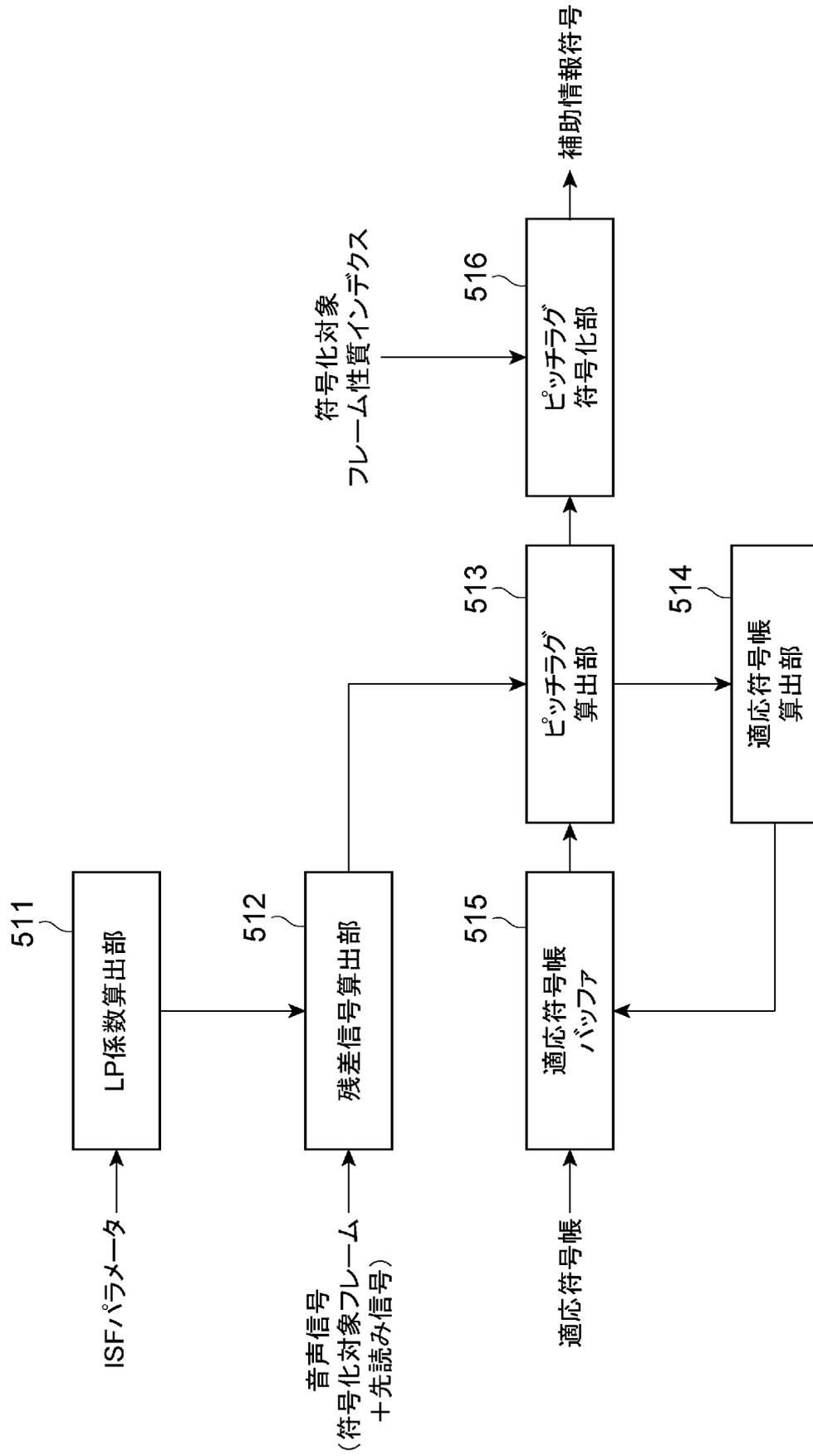
[図37]



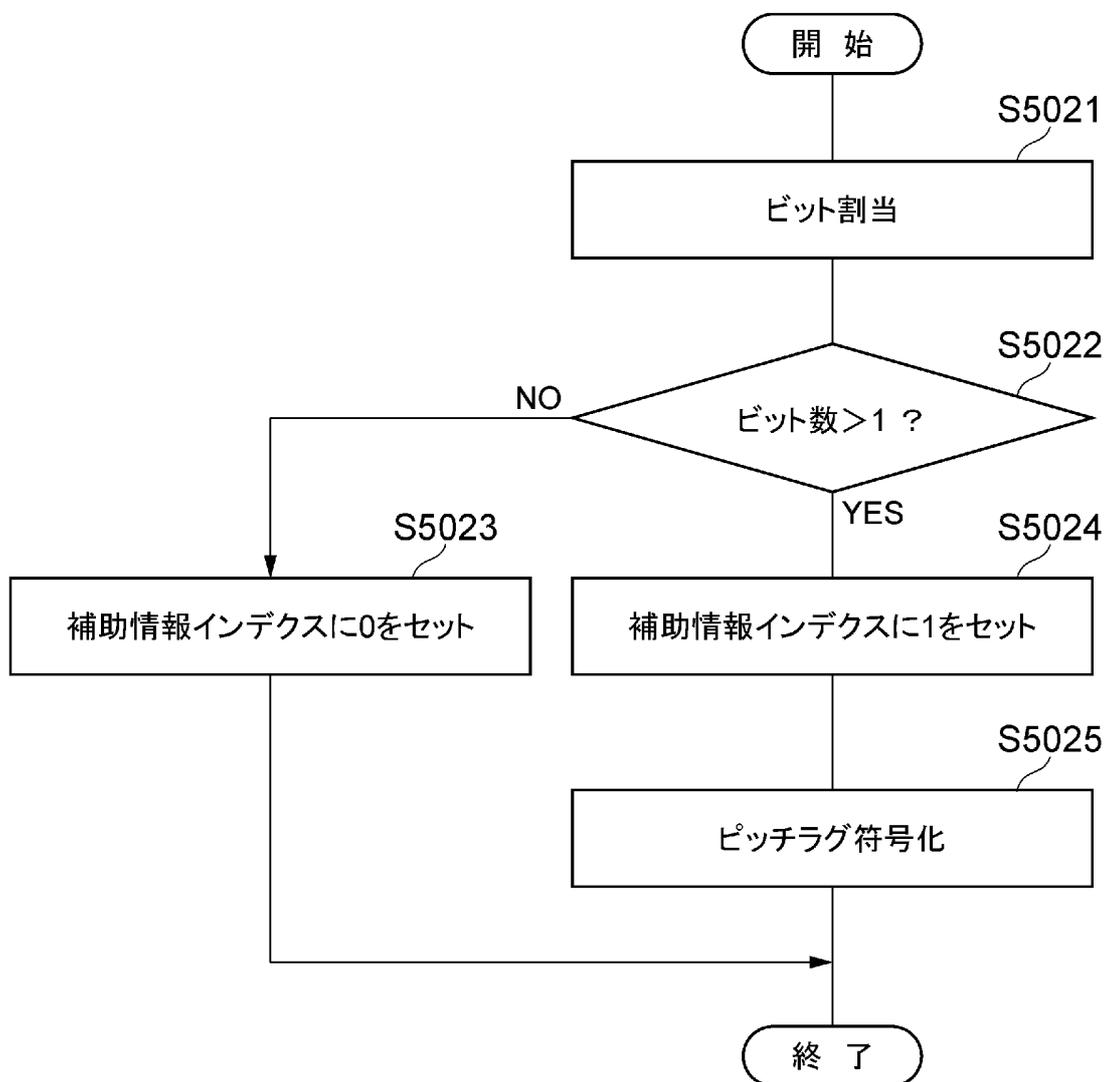
[図38]



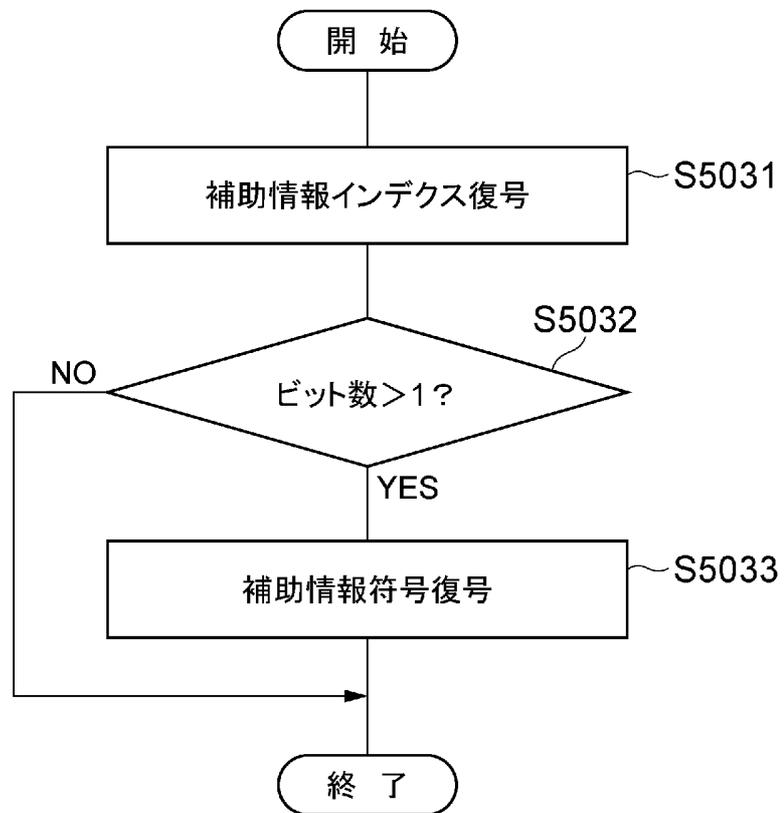
[図39]



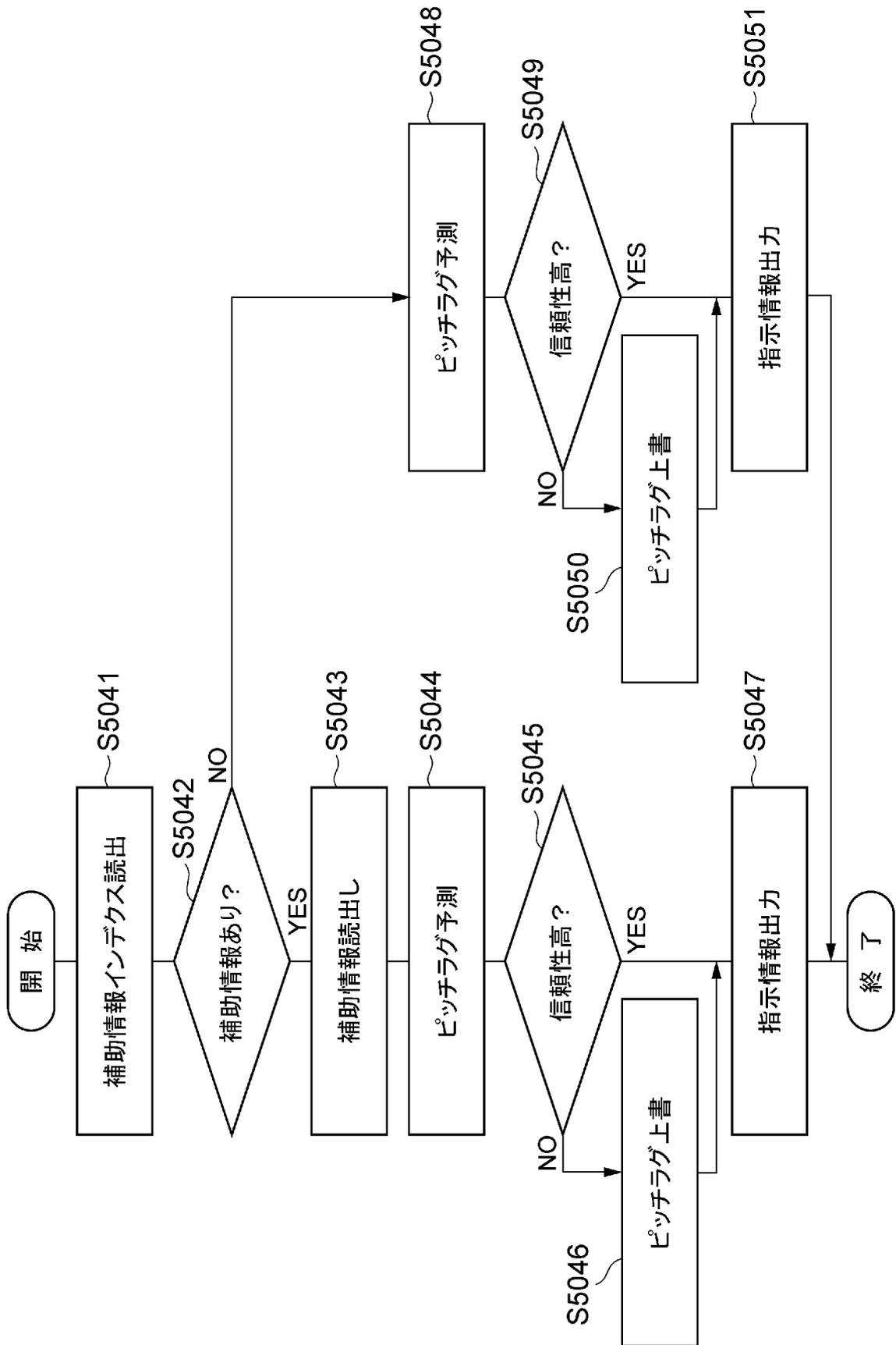
[図40]



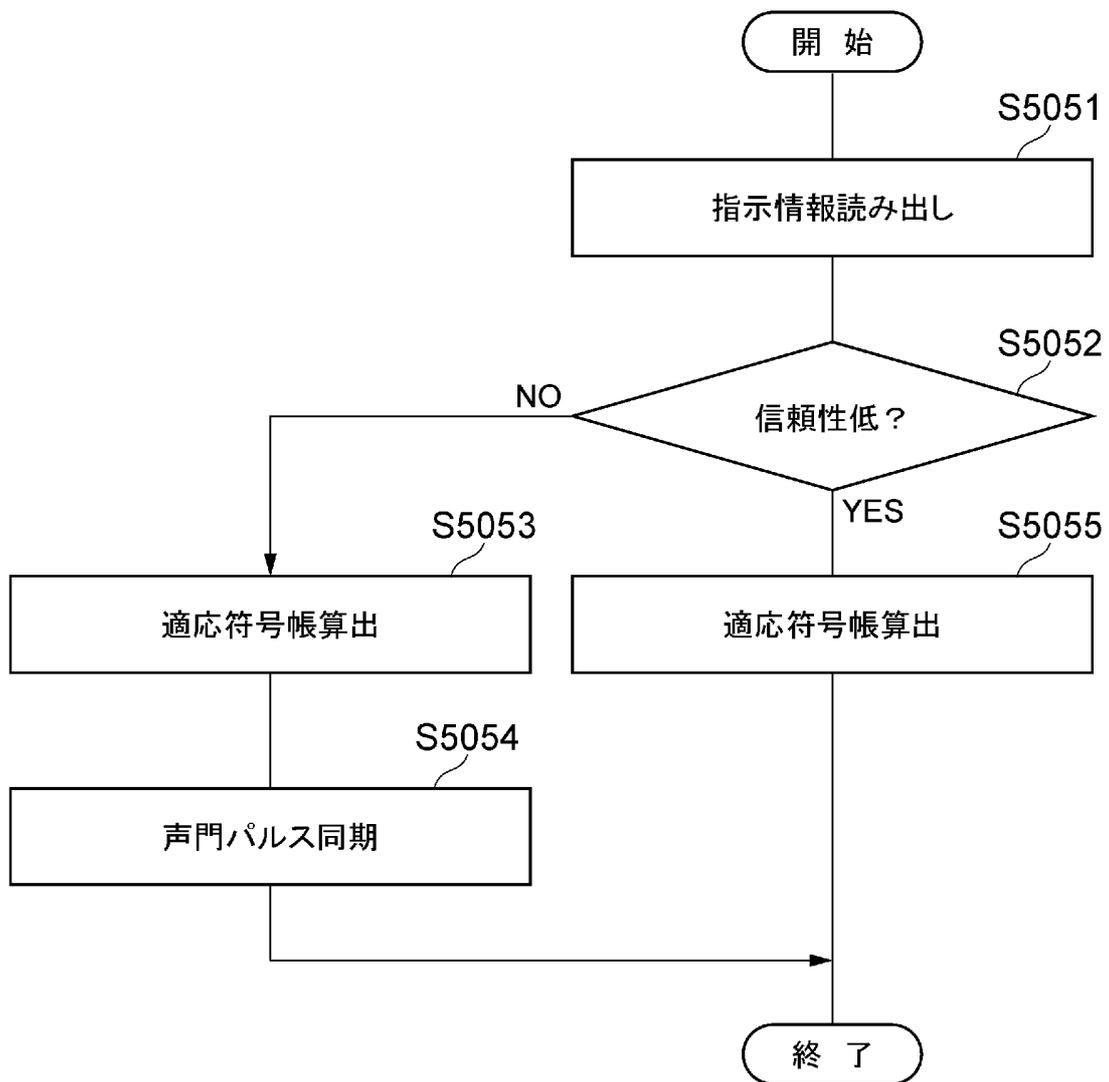
[図41]



[図42]



[図43]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/080589

A. CLASSIFICATION OF SUBJECT MATTER

G10L19/005(2013.01)i, G10L19/00(2013.01)i, G10L19/09(2013.01)i, H03M7/30(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L19/005, G10L19/00, G10L19/09, H03M7/30

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

Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2014
Kokai Jitsuyo Shinan Koho 1971-2014 Toroku Jitsuyo Shinan Koho 1994-2014

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	JP 2004-138756 A (Matsushita Electric Industrial Co., Ltd.), 13 May 2004 (13.05.2004), paragraphs [0060] to [0063]; fig. 1, 3 (Family: none)	1-5, 7-11, 16-21 6, 12-15
A	WO 2008/007698 A1 (Matsushita Electric Industrial Co., Ltd.), 17 January 2008 (17.01.2008), paragraph [0023] & US 2009/0248404 A1	3-6, 9-17

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
07 January, 2014 (07.01.14)

Date of mailing of the international search report
21 January, 2014 (21.01.14)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/080589

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-526173 A (Nokia Corp.), 26 August 2004 (26.08.2004), paragraph [0004] & US 6968309 B1 & EP 1330818 A & WO 2002/037475 A1 & DE 60121201 D & DE 60121201 T & AU 1513802 A & CA 2424202 A & BR 115057 A & AT 332002 T & CN 1489762 A & ES 2266281 T & PT 1330818 E & ZA 200302556 A	3-6, 9-17
X A	JP 2002-268696 A (Nippon Telegraph and Telephone Corp.), 20 September 2002 (20.09.2002), paragraph [0010] (Family: none)	1-4, 7-10, 18-21 5, 6, 11-17
X A	WO 2012/070370 A1 (NTT Docomo Inc.), 31 May 2012 (31.05.2012), fig. 5 & CN 103229234 A & TW 201243825 A	1, 7, 18-21 2-6, 8-17
X A	JP 2003-249957 A (Nippon Telegraph and Telephone Corp.), 05 September 2003 (05.09.2003), fig. 3 (Family: none)	1, 7, 18-21 2-6, 8-17
A	JP 2008-111991 A (NTT Docomo Inc.), 15 May 2008 (15.05.2008), entire text; all drawings (Family: none)	1-21

A. 発明の属する分野の分類 (国際特許分類 (IPC))
 Int.Cl. G10L19/005(2013.01)i, G10L19/00(2013.01)i, G10L19/09(2013.01)i, H03M7/30(2006.01)i

B. 調査を行った分野
 調査を行った最小限資料 (国際特許分類 (IPC))
 Int.Cl. G10L19/005, G10L19/00, G10L19/09, H03M7/30

最小限資料以外の資料で調査を行った分野に含まれるもの
 日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2014年
 日本国実用新案登録公報 1996-2014年
 日本国登録実用新案公報 1994-2014年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X	JP 2004-138756 A (松下電器産業株式会社) 2004.05.13, 段落【0060】 - 【0063】, 第1, 3図 (ファミリーなし)	1-5, 7-11, 16-21
A		6, 12-15
A	WO 2008/007698 A1 (松下電器産業株式会社) 2008.01.17, 段落 [0023] & US 2009/0248404 A1	3-6, 9-17

C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー	の日の後に公表された文献
「A」 特に関連のある文献ではなく、一般的技術水準を示すもの	「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)	「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「O」 口頭による開示、使用、展示等に言及する文献	「&」 同一パテントファミリー文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日 07.01.2014	国際調査報告の発送日 21.01.2014
--------------------------	--------------------------

国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 間宮 嘉誉 電話番号 03-3581-1101 内線 3591	5 Z	3 4 5 4
---	--	-----	---------

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2004-526173 A (ノキア コーポレーション) 2004.08.26, 段落【0004】 & US 6968309 B1 & EP 1330818 A & WO 2002/037475 A1 & DE 60121201 D & DE 60121201 T & AU 1513802 A & CA 2424202 A & BR 115057 A & AT 332002 T & CN 1489762 A & ES 2266281 T & PT 1330818 E & ZA 200302556 A	3-6, 9-17
X	JP 2002-268696 A (日本電信電話株式会社) 2002.09.20, 段落【0010】 (ファミリーなし)	1-4, 7-10, 18-21
A		5, 6, 11-17
X	WO 2012/070370 A1 (株式会社エヌ・ティ・ティ・ドコモ) 2012.05.31, 第5図 & CN 103229234 A & TW 201243825 A	1, 7, 18-21 2-6, 8-17
A		
X	JP 2003-249957 A (日本電信電話株式会社) 2003.09.05, 第3図 (フ ファミリーなし)	1, 7, 18-21 2-6, 8-17
A		
A	JP 2008-111991 A (株式会社エヌ・ティ・ティ・ドコモ) 2008.05.15, 全文, 全図 (ファミリーなし)	1-21



(11) **EP 2 922 053 A1**

(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 153(4) EPC

(43) Date of publication:
23.09.2015 Bulletin 2015/39

(51) Int Cl.:
G10L 19/005 (2013.01) G10L 19/00 (2013.01)
G10L 19/09 (2013.01) H03M 7/30 (2006.01)

(21) Application number: **13854879.7**

(86) International application number:
PCT/JP2013/080589

(22) Date of filing: **12.11.2013**

(87) International publication number:
WO 2014/077254 (22.05.2014 Gazette 2014/21)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

(72) Inventors:
• **TSUTSUMI Kimitaka**
Tokyo 100-6150 (JP)
• **KIKUIRI Kei**
Tokyo 100-6150 (JP)
• **YAMAGUCHI Atsushi**
Tokyo 100-6150 (JP)

(30) Priority: **15.11.2012 JP 2012251646**

(74) Representative: **Viering, Jentschura & Partner**
Patent- und Rechtsanwälte
Am Brauhaus 8
01099 Dresden (DE)

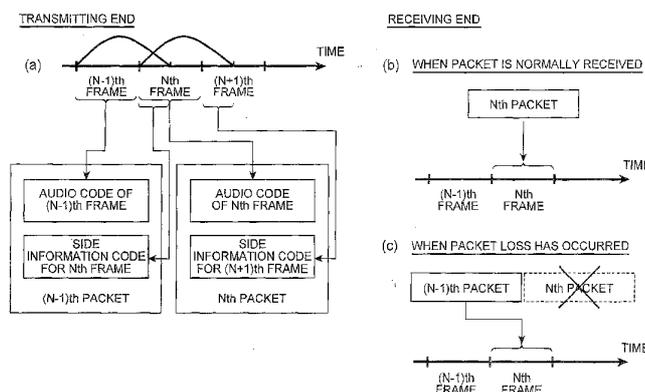
(71) Applicant: **NTT Docomo, Inc.**
Tokyo 100-6150 (JP)

(54) **AUDIO CODING DEVICE, AUDIO CODING METHOD, AUDIO CODING PROGRAM, AUDIO DECODING DEVICE, AUDIO DECODING METHOD, AND AUDIO DECODING PROGRAM**

(57) An objective of the present invention is, when packet loss occurs in audio encoding, to recover audio quality without increasing algorithmic delay. An audio signal transmission device for encoding an audio signal includes an audio encoding unit that encodes an audio signal and a side information encoding unit that calculates and encodes side information from a look-ahead signal. On the other hand, an audio signal receiving device for decoding an audio code and outputting an audio signal includes an audio code buffer that detects packet loss based on a received state of an audio packet, an

audio parameter decoding unit that decodes an audio code when an audio packet is correctly received, a side information decoding unit that decodes a side information code when an audio packet is correctly received, a side information accumulation unit that accumulates side information obtained by decoding a side information code, an audio parameter missing processing unit that outputs an audio parameter when audio packet loss is detected, and an audio synthesis unit that synthesizes a decoded audio from an audio parameter.

Fig.3



EP 2 922 053 A1

Description**Technical Field**

5 **[0001]** The present invention relates to error concealment for transmission of audio packets through an IP network or a mobile communication network and, more specifically, relates to an audio encoding device, an audio encoding method, an audio encoding program, an audio decoding device, an audio decoding method, and an audio decoding program for highly accurate packet loss concealment signal generation to implement error concealment.

10 **Background Art**

[0002] In the transmission of audio and acoustic signals (which are collectively referred to hereinafter as "audio signal") through an IP network or a mobile communication network, the audio signal is encoded into audio packets at regular time intervals and transmitted through a communication network. At the receiving end, the audio packets are received through the communication network and decoded into a decoded audio signal by server, a MCU (Multipoint Control Unit), a terminal or the like.

15 **[0003]** The audio signal is generally collected in digital format. Specifically, it is measured and accumulated as a sequence of numerals whose number is the same as a sampling frequency per second. Each element of the sequence is called a "sample". In audio encoding, each time a predetermined number of samples of an audio signal is accumulated in a built-in buffer, the audio signal in the buffer is encoded. The above-described specified number of samples is called a "frame length", and a set of the same number of samples as the frame length is called "frame". For example, at the sampling frequency of 32 kHz, when the frame length is 20 ms, the frame length is 640 samples. Note that the length of the buffer may be more than one frame.

20 **[0004]** When transmitting audio packets through a communication network, a phenomenon (so-called "packet loss") can occur where some of the audio packets are lost, or an error can occur in part of information written in the audio packets due to congestion in the communication network or the like. In such a case, the audio packets cannot be correctly decoded at the receiving end, and therefore a desired decoded audio signal cannot be obtained. Further, the decoded audio signal corresponding to the audio packet where packet loss has occurred is detected as noise, which significantly degrades the subjective quality to a person who listens to the audio.

30 **[0005]** In order to overcome the above-described inconveniences, packet loss concealment technology is used as a way to interpolate a part of the audio/acoustic signal that is lost by packet loss. There are two types of packet loss concealment technology: "packet loss concealment technology without using side information" where packet loss concealment is performed only at the receiving end and "packet loss concealment technology using side information" where parameters that help packet loss concealment are obtained at the transmitting end and transmitted to the receiving end, where packet loss concealment is performed using the received parameters at the receiving end.

35 **[0006]** The "packet loss concealment technology without using side information" generates an audio signal corresponding to a part where packet loss has occurred by copying a decoded audio signal contained in a packet that has been correctly received in the past on a pitch-by-pitch basis and then multiplying it by a predetermined attenuation coefficient as described in Non Patent Literature 1, for example. Because the "packet loss concealment technology without using side information" is based on the assumption that the properties of the part of the audio where packet loss has occurred are similar to those of the audio immediately before the occurrence of loss, the concealment effect cannot be sufficiently obtained when the part of the audio where packet loss has occurred has different properties from the audio immediately before the occurrence of loss or when there is a sudden change in power.

45 **[0007]** On the other hand, the "packet loss concealment technology using side information" includes a technique that encodes parameters required for packet loss concealment at the transmitting end and transmits them for use in packet loss concealment at the receiving end as described in Patent Literature 1. In Patent Literature 1, the audio is encoded by two encoding methods: main encoding and redundant encoding. The redundant encoding encodes the frame immediately before the frame to be encoded by the main encoding at a lower bit rate than the main encoding (see Fig. 1 (a)). For example, the Nth packet contains an audio code obtained by encoding the Nth frame by major encoding and a side information code obtained by encoding the (N-1)th frame by redundant encoding.

50 **[0008]** The receiving end waits for the arrival of two or more temporally successive packets and then decodes the temporally earlier packet and obtains a decoded audio signal. For example, to obtain a signal corresponding to the Nth frame, the receiving end waits for the arrival of the (N+1)th packet and then performs decoding. In the case where the Nth packet and the (N+1)th packet are correctly received, the audio signal of the Nth frame is obtained by decoding the audio code contained in the Nth packet (see Fig. 1(b)). On the other hand, in the case where packet loss has occurred (when the (N+1)th packet is obtained in the condition where the Nth packet is lost), the audio signal of the Nth frame can be obtained by decoding the side information code contained in the (N+1)th packet (see Fig. 1(c)).

55 **[0009]** According to the method of Patent Literature 1, after a packet to be decoded arrives, it is necessary to wait to

perform decoding until one or more packet arrives, and algorithmic delay increases by one packet or more. Accordingly, in the method of Patent Literature 1, although the audio quality can be improved by packet loss concealment, the algorithmic delay increases to cause the degradation of the voice communication quality.

5 **[0010]** Further, in the case of applying the above-described packet loss concealment technology to CELP (Code Excited Linear Prediction) encoding, another problem caused by the characteristics of the operation of CELP arises. Because CELP is an audio model based on linear prediction and is able to encode an audio signal with high accuracy and with a high compression ratio, it is used in many international standards.

[0011] In CELP, an audio signal is synthesized by filtering an excitation signal $e(n)$ using an all-pole synthesis filter. Specifically, an audio signal $s(n)$ is synthesized according to the following equation:

10

$$s(n) = e(n) - \sum_{i=1}^P a(i) \cdot s(n-1) \quad \text{Equation 1}$$

15 where $a(i)$ is a linear prediction coefficient (LP coefficient), and a value such as $P=16$, for example, is used as a degree.

[0012] The excitation signal is accumulated in a buffer called an adaptive codebook. When synthesizing the audio for a new frame, an excitation signal is newly generated by adding an adaptive codebook vector read from the adaptive codebook and a fixed codebook vector representing a change in excitation signal over time based on position information called a pitch lag. The newly generated excitation signal is accumulated in the adaptive codebook and is also filtered

20 by the all-pole synthesis filter, and thereby a decoded signal is synthesized.

[0013] In CELP, an LP coefficient is calculated for all frames. In the calculation of the LP coefficient, a look-ahead signal of about 10 ms is required. Specifically, in addition to a frame to be encoded, a look-ahead signal is accumulated in the buffer, and then the LP coefficient calculation and the subsequent processing are performed (see Fig. 2). Each frame is divided into about four sub-frames, and processing such as the above-described pitch lag calculation, adaptive

25 codebook vector calculation, fixed codebook vector calculation and adaptive codebook update are performed in each sub-frame. In the processing of each sub-frame, the LP coefficient is also interpolated so that the coefficient varies from sub-frame to sub-frame. Further, for quantization and interpolation, the LP coefficient is encoded after being converted into an ISP (Immittance Spectral Pair) parameter and an ISF (Immittance Spectral Frequency) parameter, which are equivalent representation(s) of the LP coefficient(s). A procedure for the interconversion of the LP coefficient(s) and the ISP parameter and the ISF parameter is described in Non Patent Literature 2.

[0014] In CELP encoding, encoding and decoding are performed based on the assumption that both of the encoding end and the decoding end have adaptive codebooks, and those adaptive codebooks are always synchronized with each other. Although the adaptive codebook at the encoding end and the adaptive codebook at the decoding end are synchronized under conditions where packets are correctly received and decoding correctly, once packet loss has occurred,

35 the synchronization of the adaptive codebooks cannot be achieved.

[0015] For example, if a value that is used as a pitch lag is different between the encoding end and the decoding end, a time lag occurs between the adaptive codebook vectors. Because the adaptive codebook is updated with those adaptive codebook vectors, even if the next frame is correctly received, the adaptive codebook vector calculated at the encoding end and the adaptive codebook vector calculated at the decoding end do not coincide, and the synchronization of the adaptive codebooks is not recovered. Due to such inconsistency of the adaptive codebooks, the degradation of the audio quality occurs for several frames after the frame where packet loss has happened.

[0016] In the packet loss concealment in CELP encoding, a more advanced technique is described in Patent Literature 2. According to Patent Literature 2, an index of a transition mode codebook is transmitted instead of a pitch lag or an adaptive codebook gain in a specific frame that is largely affected by packet loss. The technique of Patent Literature 2

45 focuses attentions on a transition frame (transition from a silent audio segment to a sound audio segment, or transition between two vowels) as the frame that is largely affected by packet loss. By generating an excitation signal using the transition mode codebook in this transition frame, it is possible to generate an excitation signal that is not dependent on the past adaptive codebook and thereby recover from the inconsistency of the adaptive codebooks due to the past packet loss.

[0017] However, because the method of Patent Literature 2 does not use the transition frame codebook in a frame where a long vowel continues, for example, it is not possible to recover from the inconsistency of the adaptive codebooks in such a frame. Further, in the case where the packet containing the transition frame codebook is lost, packet loss affects the frames after the loss. This is the same when the next packet after the packet containing the transition frame codebook is lost.

55 **[0018]** Although it is feasible to apply a codebook to all frames that is not dependent on the past frames, such as the transition frame codebook, because the encoding efficiency is significantly degraded, it is not possible to achieve a low bit rate and high audio quality under these circumstances.

Citation List

Patent Literature

5 [0019]

Patent Literature 1: Japanese Unexamined Patent Application Publication No. 2003-533916

Patent Literature 2: Japanese Unexamined Patent Application Publication No. 2010-507818

10 **Non Patent Literature**

[0020]

Non Patent Literature 1: ITU-T G.711 Appendix I

15 Non Patent Literature 2: 3GPP TS26-191

Non Patent Literature 3: 3GPP TS26-190

Non Patent Literature 4: ITU-T G.718

Summary of Invention

20

Technical Problem

[0021] With use of the method of Patent Literature 1, after the arrival of a packet to be decoded, decoding is not started before the arrival of the next packet. Therefore, although the audio quality is improved by packet loss concealment, the algorithmic delay increases, which causes the degradation of the voice communication quality.

[0022] In the event of packet loss in CELP encoding, the degradation of the audio quality occurs due to the inconsistency of the adaptive codebooks between the encoding unit and the decoding unit. Although the method as described in Patent Literature 2 can allow for recovery from the inconsistency of the adaptive codebooks, the method is not sufficient to allow recovery when a frame different from the frame immediately before the transition frame is lost.

[0023] The present invention has been accomplished to solve the above problems and an object of the present invention is thus to provide an audio encoding device, an audio encoding method, an audio encoding program, an audio decoding device, an audio decoding method, and an audio decoding program that recover audio quality without increasing algorithmic delay in the event of packet loss in audio encoding.

35 **Solution to Problem**

[0024] To achieve the above object, an audio encoding device according to one aspect of the present invention is for encoding an audio signal, which includes an audio encoding unit configured to encode an audio signal, and a side information encoding unit configured to calculate side information from a look-ahead signal and encode the side information.

[0025] The side information may be related to a pitch lag in a look-ahead signal, related to a pitch gain in a look-ahead signal, or related to a pitch lag and a pitch gain in a look-ahead signal. Further, the side information may contain information related to availability of the side information.

[0026] The side information encoding unit may calculate side information for a look-ahead signal part and encode the side information, and also generate a concealment signal, and the audio encoding device may further include an error signal encoding unit configured to encode an error signal between an input audio signal and a concealment signal output from the side information encoding unit, and a main encoding unit configured to encode an input audio signal.

[0027] Further, an audio decoding device according to one aspect of the present invention is for decoding an audio code and outputting an audio signal, which includes an audio code buffer configured to detect packet loss based on a received state of an audio packet, an audio parameter decoding unit configured to decode an audio code when an audio packet is correctly received, a side information decoding unit configured to decode a side information code when an audio packet is correctly received, a side information accumulation unit configured to accumulate side information obtained by decoding a side information code, an audio parameter missing processing unit configured to output an audio parameter when audio packet loss is detected, and an audio synthesis unit configured to synthesize a decoded audio from an audio parameter.

[0028] The side information may be related to a pitch lag in a look-ahead signal, related to a pitch gain in a look-ahead signal, or related to a pitch lag and a pitch gain in a look-ahead signal. Further, the side information may contain information related to the availability of side information.

5 [0029] The side information decoding unit may decode a side information code and output side information, and may further output a concealment signal related to a look-ahead part by using the side information, and the audio decoding device may further include an error decoding unit configured to decode a code related to an error signal between an audio signal and a concealment signal, a main decoding unit configured to decode a code related to an audio signal, and a concealment signal accumulation unit configured to accumulate a concealment signal output from the side information decoding unit.

10 [0030] When an audio packet is correctly received, a part of a decoded signal may be generated by adding a concealment signal read from the concealment signal accumulation unit and a decoded error signal output from the error decoding unit, and the concealment signal accumulation unit may be updated with a concealment signal output from the side information decoding unit.

[0031] When audio packet loss is detected, a concealment signal read from the concealment signal accumulation unit may be used as a part, or a whole, of a decoded signal.

15 [0032] When audio packet loss is detected, a decoded signal may be generated by using an audio parameter predicted by the audio parameter missing processing unit, and the concealment signal accumulation unit may be updated by using a part of the decoded signal.

[0033] When audio packet loss is detected, the audio parameter missing processing unit may use side information read from the side information accumulation unit as a part of a predicted value of an audio parameter.

[0034] When audio packet loss is detected, the audio synthesis unit may correct an adaptive codebook vector, which is one of the audio parameters, by using side information read from the side information accumulation unit.

20 [0035] An audio encoding method according to one aspect of the present invention is an audio encoding method by an audio encoding device for encoding an audio signal, which includes an audio encoding step of encoding an audio signal, and a side information encoding step of calculating side information from a look-ahead signal and encoding the side information.

25 [0036] An audio decoding method according to one aspect of the present invention is an audio decoding method by an audio decoding device for decoding an audio code and outputting an audio signal, which includes an audio code buffer step of detecting packet loss based on a received state of an audio packet, an audio parameter decoding step of decoding an audio code when an audio packet is correctly received, a side information decoding step of decoding a side information code when an audio packet is correctly received, a side information accumulation step of accumulating side information obtained by decoding a side information code, an audio parameter missing processing step of outputting an audio parameter when audio packet loss is detected, and an audio synthesis step of synthesizing a decoded audio from an audio parameter.

30 [0037] An audio encoding program according to one aspect of the present invention causes a computer to function as an audio encoding unit to encode an audio signal, and a side information encoding unit to calculate side information from a look-ahead signal and encode the side information.

35 [0038] An audio decoding program according to one aspect of the present invention causes a computer to function as an audio code buffer to detect packet loss based on a received state of an audio packet, an audio parameter decoding unit to decode an audio code when an audio packet is correctly received, a side information decoding unit to decode a side information code when an audio packet is correctly received, a side information accumulation unit to accumulate side information obtained by decoding a side information code, an audio parameter missing processing unit to output an audio parameter when audio packet loss is detected, and an audio synthesis unit to synthesize a decoded audio from an audio parameter.

Advantageous Effects of Invention

45 [0039] It is possible to recover audio quality without increasing algorithmic delay in the event of packet loss in audio encoding. Particularly, in CELP encoding, it is possible to reduce degradation of an adaptive codebook that occurs when packet loss happens and thereby improve audio quality in the event of packet loss.

Brief Description of Drawings

50 [0040]

- Fig. 1 is a view showing a temporal relationship between packets and a decoded signal according to related art described in Patent Literature 1.
- 55 Fig. 2 is a view showing a temporal relationship between an LP analysis target signal and a look-ahead signal in CELP encoding.
- Fig. 3 is a view showing a temporal relationship between packets and a decoded signal according to an embodiment of the present invention.

- Fig. 4 is a view showing a functional configuration example of an audio signal transmitting device in an example 1 (first example) of the present invention.
- Fig. 5 is a view showing a functional configuration example of an audio signal receiving device in the example 1 of the present invention.
- 5 Fig. 6 is a view showing a procedure of the audio signal transmitting device in the example 1 of the present invention.
- Fig. 7 is a view showing a procedure of the audio signal receiving device in the example 1 of the present invention.
- Fig. 8 is a view showing a functional configuration example of a side information encoding unit in the example 1 of the present invention.
- Fig. 9 is a view showing a procedure of the side information encoding unit in the example 1 of the present invention.
- 10 Fig. 10 is a view showing a procedure of an LP coefficient calculation unit in the example 1 of the present invention.
- Fig. 11 is a view showing a procedure of a target signal calculation unit in the example 1 of the present invention.
- Fig. 12 is a view showing a functional configuration example of an audio parameter missing processing unit in the example 1 of the present invention.
- Fig. 13 is a view showing a procedure of audio parameter prediction in the example 1 of the present invention.
- 15 Fig. 14 is a view showing a procedure of an excitation vector synthesis unit in an alternative example 1-1 of the example 1 of the present invention.
- Fig. 15 is a view showing a functional configuration example of an audio synthesis unit in the example 1 of the present invention.
- Fig. 16 is a view showing a procedure of the audio synthesis unit in the example 1 of the present invention.
- 20 Fig. 17 is a view showing a functional configuration example of a side information encoding unit (when a side information output determination unit is included) in an alternative example 1-2 of the example 1 of the present invention.
- Fig. 18 is a view showing a procedure of the side information encoding unit (when the side information output determination unit is included) in the alternative example 1-2 of the example 1 of the present invention.
- 25 Fig. 19 is a view showing a procedure of audio parameter prediction in the alternative example 1-2 of the example 1 of the present invention.
- Fig. 20 is a view showing a functional configuration example of an audio signal transmitting device in an example 2 of the present invention.
- Fig. 21 is a view showing a functional configuration example of a main encoding unit in the example 2 of the present invention.
- 30 Fig. 22 is a view showing a procedure of the audio signal transmitting device in the example 2 of the present invention.
- Fig. 23 is a view showing a functional configuration example of an audio signal receiving device in the example 2 of the present invention.
- Fig. 24 is a view showing a procedure of the audio signal receiving device in the example 2 of the present invention.
- 35 Fig. 25 is a view showing a functional configuration example of an audio synthesis unit in the example 2 of the present invention.
- Fig. 26 is a view showing a functional configuration example of an audio parameter decoding unit in the example 2 of the present invention.
- Fig. 27 is a view showing a functional configuration example of a side information encoding unit in an example 3 of the present invention.
- 40 Fig. 28 is a view showing a procedure of the side information encoding unit in the example 3 of the present invention.
- Fig. 29 is a view showing a procedure of a pitch lag selection unit in the example 3 of the present invention.
- Fig. 30 is a view showing a procedure of a side information decoding unit in the example 3 of the present invention.
- Fig. 31 is a view showing a configuration of an audio encoding program and a storage medium according to an embodiment of the present invention.
- 45 Fig. 32 is a view showing a configuration of an audio decoding program and a storage medium according to an embodiment of the present invention.
- Fig. 33 is a view showing a functional configuration example of a side information encoding unit in an example 4 of the present invention.
- 50 Fig. 34 is a view showing a procedure of the side information encoding unit in the example 4 of the present invention.
- Fig. 35 is a view showing a procedure of a pitch lag prediction unit in the example 4 of the present invention.
- Fig. 36 is another view showing a procedure of the pitch lag prediction unit in the example 4 of the present invention.
- Fig. 37 is another view showing a procedure of the pitch lag prediction unit in the example 4 of the present invention.
- Fig. 38 is a view showing a procedure of an adaptive codebook calculation unit in the example 4 of the present invention.
- 55 Fig. 39 is a view showing a functional configuration example of a side information encoding unit in an example 5 of the present invention.
- Fig. 40 is a view showing a procedure of a pitch lag encoding unit in the example 5 of the present invention.

Fig. 41 is a view showing a procedure of a side information decoding unit in the example 5 of the present invention.
 Fig. 42 is a view showing a procedure of a pitch lag prediction unit in the example 5 of the present invention.
 Fig. 43 is a view showing a procedure of an adaptive codebook calculation unit in the example 5 of the present invention.

5

Description of Embodiments

[0041] Embodiments of the present invention are described hereinafter with reference to the attached drawings. Note that, where possible, the same elements are denoted by the same reference numerals and redundant description thereof is omitted.

10

[0042] An embodiment of the present invention relates to an encoder and a decoder that implement "packet loss concealment technology using side information" that encodes and transmits side information calculated on the encoder side for use in packet loss concealment on the decoder side.

15

[0043] In the embodiments of the present invention, the side information that is used for packet loss concealment is contained in a previous packet. Fig. 3 shows a temporal relationship between an audio code and a side information code contained in a packet. As illustrated in Fig. 3, the side information in the embodiments of the present invention is parameters (pitch lag, adaptive codebook gain, etc.) that are calculated for a look-ahead signal in CELP encoding.

20

[0044] Because the side information is contained in a previous packet, it is possible to perform decoding without waiting for a packet that arrives after a packet to be decoded. Further, when packet loss is detected, because the side information for a frame to be concealed is obtained from the previous packet, it is possible to implement highly accurate packet loss concealment without waiting for the next packet.

[0045] In addition, by transmitting parameters for CELP encoding in a look-ahead signal as the side information, it is possible to reduce the inconsistency of adaptive codebooks even in the event of packet loss.

25

[0046] The embodiments of the present invention can be composed of an audio signal transmitting device (audio encoding device) and an audio signal receiving device (audio decoding device). A functional configuration example of an audio signal transmitting device is shown in Fig. 4, and an example procedure of the same is shown in Fig. 6. Further, a functional configuration example of an audio signal receiving device is shown in Fig. 5, and an example procedure of the same is shown in Fig. 7.

30

[0047] As shown in Fig. 4, the audio signal transmitting device includes an audio encoding unit 111 and a side information encoding unit 112. As shown in Fig. 5, the audio signal receiving device includes an audio code buffer 121, an audio parameter decoding unit 122, an audio parameter missing processing unit 123, an audio synthesis unit 124, a side information decoding unit 125, and a side information accumulation unit 126.

[0048] The audio signal transmitting device encodes an audio signal for each frame and can transmit the audio signal by the example procedure shown in Fig. 6.

35

[0049] The audio encoding unit 111 can calculate audio parameters for a frame to be encoded and output an audio code (Step S131 in Fig. 6).

[0050] The side information encoding unit 112 can calculate audio parameters for a look-ahead signal and output a side information code (Step S 132 in Fig. 6).

40

[0051] It is determined whether the audio signal ends, and the above steps can be repeated until the audio signal ends (Step S133 in Fig. 6).

[0052] The audio signal receiving device decodes a received audio packet and outputs an audio signal by the example procedure shown in Fig. 7.

45

[0053] The audio code buffer 121 waits for the arrival of an audio packet and accumulates an audio code. When the audio packet has correctly arrived, the processing is switched to the audio parameter decoding unit 122. On the other hand, when the audio packet has not correctly arrived, the processing is switched to the audio parameter missing processing unit 123 (Step S141 in Fig. 7).

<When audio packet is correctly received>

50

[0054] The audio parameter decoding unit 122 decodes the audio code and outputs audio parameters (Step S142 in Fig. 7).

[0055] The side information decoding unit 125 decodes the side information code and outputs side information. The outputted side information is sent to the side information accumulation unit 126 (Step S 143 in Fig. 7).

55

[0056] The audio synthesis unit 124 synthesizes an audio signal from the audio parameters output from the audio parameter decoding unit 122 and outputs the synthesized audio signal (Step S144 in Fig. 7).

[0057] The audio parameter missing processing unit 123 accumulates the audio parameters output from the audio parameter decoding unit 122 in preparation for packet loss (Step S 145 in Fig. 7).

[0058] The audio code buffer 121 determines whether the transmission of audio packets has ended, and when the

transmission of audio packets has ended, stops the processing. While the transmission of audio packets continues, the above Steps S141 to S 146 are repeated (Step S 147 in Fig. 7).

<When audio packet is lost>

5

[0059] The audio parameter missing processing unit 123 reads the side information from the side information accumulation unit 126 and carries out prediction for the parameter(s) not contained in the side information and thereby outputs the audio parameters (Step S 146 in Fig. 7).

10

[0060] The audio synthesis unit 124 synthesizes an audio signal from the audio parameters output from the audio parameter missing processing unit 123 and outputs the synthesized audio signal (Step S144 in Fig. 7).

[0061] The audio parameter missing processing unit 123 accumulates the audio parameters output from the audio parameter missing processing unit 123 in preparation for packet loss (Step S145 in Fig. 7).

15

[0062] The audio code buffer 121 determines whether the transmission of audio packets has ended, and when the transmission of audio packets has ended, stops the processing. While the transmission of audio packets continues, the above Steps S141 to S 146 are repeated (Step S147 in Fig. 7).

[Example 1]

20

[0063] In this example of a case where a pitch lag is transmitted as the side information, the pitch lag can be used for generation of a packet loss concealment signal at the decoding end.

[0064] The functional configuration example of the audio signal transmitting device is shown in Fig. 4, and the functional configuration example of the audio signal receiving device is shown in Fig. 5. An example of the procedure of the audio signal transmitting device is shown in Fig. 6, and an example of the procedure of the audio signal receiving device is shown in Fig. 7.

25

<Transmitting end>

[0065] In the audio signal transmitting device, an input audio signal is sent to the audio encoding unit 111.

30

[0066] The audio encoding unit 111 encodes a frame to be encoded by CELP encoding (Step 131 in Fig. 6). For the details of CELP encoding, the method described in Non Patent Literature 3 is used, for example. The details of the procedure of CELP encoding are omitted. Note that, in the CELP encoding, local decoding is performed at the encoding end. The local decoding is to decode an audio code also at the encoding end and obtain parameters (ISP parameter and corresponding ISF parameter, pitch lag, long-term prediction parameter, adaptive codebook, adaptive codebook gain, fixed codebook gain, fixed codebook vector, etc.) required for audio synthesis. The parameters obtained by the local decoding include: at least one or both of the ISP parameter and the ISF parameter, the pitch lag, and the adaptive codebook, which are sent to the side information encoding unit 112. In the case where the audio encoding as described in Non Patent Literature 4 is used in the audio encoding unit 111, an index representing the characteristics of a frame to be encoded may also be sent to the side information encoding unit 112. In embodiments, encoding different from CELP encoding may be used in the audio encoding unit 111. In embodiments using different encoding, at least one or both of the ISP parameter and the ISF parameter, the pitch lag, and the adaptive codebook can be separately calculated from an input signal, or a decoded signal obtained by the local decoding, and sent to the side information encoding unit 112.

35

[0067] The side information encoding unit 112 calculates a side information code using the parameters calculated by the audio encoding unit 111 and the look-ahead signal (Step 132 in Fig. 6). As shown in the example of Fig. 8, the side information encoding unit 112 includes an LP coefficient calculation unit 151, a target signal calculation unit 152, a pitch lag calculation unit 153, an adaptive codebook calculation unit 154, an excitation vector synthesis unit 155, an adaptive codebook buffer 156, a synthesis filter 157, and a pitch lag encoding unit 158. An example procedure in the side information encoding unit is shown in Fig. 9.

45

[0068] The LP coefficient calculation unit 151 calculates an LP coefficient using the ISF parameter calculated by the audio encoding unit 111 and the ISF parameter calculated in the past several frames (Step 161 in Fig. 9). The procedure of the LP coefficient calculation unit 151 is shown in Fig. 10.

50

[0069] First, the buffer is updated using the ISF parameter obtained from the audio encoding unit 111 (Step 171 in Fig.10). Next, the ISF parameter $\hat{\omega}_i$ in the look-ahead signal is calculated. The ISF parameter $\hat{\omega}_i$ is calculated by the following equation (Step 172 in Fig.10).

55

$$\hat{\omega}_i = \alpha \omega_i^{(-1)} + (1 - \alpha) \bar{\omega}_i \quad \text{Equation 2}$$

$$\bar{\omega}_i = \beta \omega_i^C + (1 - \beta) \frac{\omega_i^{(-3)} + \omega_i^{(-2)} + \omega_i^{(-1)}}{3} \quad \text{Equation 3}$$

5 where $\omega_i^{(-j)}$ is the ISF parameter, stored in the buffer, which is for the frame preceding by j-number of frames. Further, ω_i^C is the ISF parameter during the speech period that is calculated in advance by learning or the like. β is a constant, and it may be a value such as 0.75, for example, though not limited thereto. Further, α is also constant, and it may be a value such as 0.9, for example, though not limited thereto. ω_i^C , α and β may be varied by the index representing the characteristics of the frame to be encoded as in the ISF concealment described in Non Patent Literature 4, for example.

10 **[0070]** In addition, the values of i are arranged so that $\bar{\omega}_i$ satisfies $0 < \bar{\omega}_0 < \bar{\omega}_1 < \dots < \bar{\omega}_{14}$, and the values of $\bar{\omega}_i$ can be adjusted so that the adjacent $\bar{\omega}_i$ is not too close. As a procedure to adjust the value of $\bar{\omega}_i$, Non Patent Literature 4 (Equation 151) may be used, for example (Step 173 in Fig. 10).

15 **[0071]** After that, the ISF parameter $\bar{\omega}_i$ is converted into an ISP parameter and interpolation can be performed for each sub-frame. As a method of calculating the ISP parameter from the ISF parameter, the method described in the section 6.4.4 in Non Patent Literature 4 may be used, and as a method of interpolation, the procedure described in the section 6.8.3 in Non Patent Literature 4 may be used (Step 174 in Fig. 10).

20 **[0072]** Then, the ISP parameter for each sub-frame is converted into an LP coefficient $\hat{\alpha}_j^i (0 < i \leq P, 0 \leq j < M_{ia})$. The number of sub-frames contained in the look-ahead signal is M_{ia} . For the conversion from the ISP parameter to the LP coefficient, the procedure described in the section 6.4.5 in Non Patent Literature 4 may be used (Step 175 in Fig. 10).

[0073] The target signal calculation unit 152 calculates a target signal $x(n)$ and an impulse response $h(n)$ by using the LP coefficient $\hat{\alpha}_j^i$ (Step 162 in Fig. 9). As described in the section 6.8.4.1.3 in Non Patent Literature 4, the target signal is obtained by applying an perceptual weighting filter to a linear prediction residual signal (Fig. 11).

25 **[0074]** First, a residual signal $r(n)$ of the look-ahead signal $S_{pre}^l(n) (0 \leq n < L')$ is calculated using the LP coefficient according to the following equation (Step 181 in Fig. 11).

$$r(n) = s_{pre}^l(n) + \sum_{i=1}^P \hat{\alpha}_i^j \cdot s_{pre}^l(n-i) \quad \text{Equation 4}$$

30 **[0075]** Note that L' indicates the number of samples of a sub-frame, and L indicates the number of samples of a frame to be encoded $S_{pre}(n) (0 \leq n < L)$. Then, $s_{pre}^l(n-p) = s_{pre}^l(n+L-p)$ is satisfied.

35 **[0076]** In addition, the target signal $x(n) (0 \leq n < L')$ is calculated by the following equations (Step 182 in Fig. 11).

$$e(n) = r(n) - \sum_{i=1}^P \hat{\alpha}_i^j \cdot e(n-i) (0 \leq n < L') \quad \text{Equation 5}$$

$$e(n) = s(n+L-1) - \hat{s}(n+L-1) (-P \leq n < 0) \quad \text{Equation 6}$$

$$\dot{e}(n) = r(n) + \sum_{i=1}^P \hat{\alpha}_i^j \cdot \dot{e}(n-i) \quad \text{Equation 7}$$

$$x(n) = e(n) + \gamma \cdot e(n-1) \quad \text{Equation 8}$$

50 where an perceptual weighting filter $\gamma=0.68$. The value of the perceptual weighting filter may be a different value according to the design policy of audio encoding.

[0077] Then, the impulse response $h(n) (0 \leq n < L')$ is calculated by the following equations (Step 183 in Fig. 11).

$$\dot{h}(n) = \hat{\alpha}_i^j + \sum_{i=1}^P \hat{\alpha}_i^j \cdot \dot{h}(n-i) \quad \text{Equation 9}$$

$$h(n) = \hat{h}(n) + \gamma \cdot \hat{h}(n-1) \quad \text{Equation 10}$$

5 [0078] The pitch lag calculation unit 153 calculates a pitch lag for each sub-frame by calculating k that maximizes the following equation (Step 163 in Fig. 9). Note that, in order to reduce the amount of calculations, the above-described target signal calculation (Step 182 in Fig. 11) and the impulse response calculation (Step 183 in Fig. 11) may be omitted, and the residual signal may be used as the target signal.

$$10 \quad T_p = \operatorname{argmax} T_k$$

$$15 \quad T_k = \frac{\sum_{n=0}^{L'-1} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{L'-1} y_k(n)y_k(n)}} \quad \text{Equation 11}$$

$$20 \quad y_k(n) = \sum_{i=0}^n v'(i) \cdot h(n-i) \quad \text{Equation 12}$$

$$25 \quad v'(n) = \sum_{i=-1}^l \operatorname{Int}(i) \cdot u(n + N_{\text{adapt}} - T_p + i) \quad \text{Equation 13}$$

Note that $y_k(n)$ is obtained by convoluting the impulse response with the linear prediction residual. $\operatorname{Int}(i)$ indicates an interpolation filter. The details of the interpolation filter are as described in the section 6.8.4.1.4.1 in Non Patent Literature 4. As a matter of course, $v'(n)=u(n+N_{\text{adapt}}-T_p+i)$ may be employed without using the interpolation filter.

30 [0079] Although the pitch lag can be calculated as an integer by the above-described calculation method, the accuracy of the pitch lag may be increased to after the decimal point accuracy by interpolating the above T_k . For the details of the procedure to calculate the pitch lag after the decimal point by interpolation, the processing method described in the section 6.8.4.1.4.1 in Non Patent Literature 4 may be used.

35 [0080] The adaptive codebook calculation unit 154 calculates an adaptive codebook vector $v'(n)$ and a long-term prediction parameter from the pitch lag T_p and the adaptive codebook $u(n)$ stored in the adaptive codebook buffer 156 according to the following equation (Step 164 in Fig. 9).

$$40 \quad v'(n) = \sum_{i=-1}^l \operatorname{Int}(i) \cdot u(n + N_{\text{adapt}} - T_p + i) \quad \text{Equation 14}$$

For the details of the procedure to calculate the long-term parameter, the method described in the section 5.7 in Non Patent Literature 3 may be used.

45 [0081] The excitation vector synthesis unit 155 multiplies the adaptive codebook vector $v'(n)$ by a predetermined adaptive codebook gain g_p^C and outputs an excitation signal vector according to the following equation (Step 165 in Fig. 9).

$$e(n) = g_p^C \cdot v'(n) \quad \text{Equation 15}$$

50 Although the value of the adaptive codebook gain g_p^C may be 1.0 or the like, for example, a value obtained in advance by learning may be used, or it may be varied by the index representing the characteristics of the frame to be encoded.

[0082] Then, the state of the adaptive codebook $u(n)$ stored in the adaptive codebook buffer 156 is updated by the excitation signal vector according to the following equations (Step 166 in Fig. 9).

$$55 \quad u(n)=u(n+L) \quad (0 \leq n < N-L) \quad \text{Equation 16}$$

$$u(n+N-L)=e(n) \quad (0 \leq n < L)$$

Equation 17

5 [0083] The synthesis filter 157 synthesizes a decoded signal according to the following equation by linear prediction inverse filtering using the excitation signal vector as an excitation source (Step 167 in Fig. 9).

$$\hat{s}(n) = e(n) - \sum_{i=1}^p \hat{a}_i \cdot \hat{s}(n-i)$$

Equation 18

10

[0084] The above-described Steps 162 to 167 in Fig. 9 are repeated for each sub-frame until the end of the look-ahead signal (Step 168 in Fig. 9).

15

[0085] The pitch lag encoding unit 158 encodes the pitch lag $T_p^{(j)}$ ($0 \leq j < M_{la}$) that is calculated in the look-ahead signal (Step 169 in Fig. 9). The number of sub-frames contained in the look-ahead signal is M_{la} .

[0086] Encoding may be performed by a method such as one of the following methods, for example, although any method may be used for encoding.

20

1. A method that performs binary encoding, scalar quantization, vector quantization or arithmetic encoding on a part or the whole of the pitch lag $T_p^{(j)}$ ($0 \leq j < M_{la}$) and transmits the result.

25

2. A method that performs binary encoding, scalar quantization, vector quantization or arithmetic encoding on a part or the whole of a difference $T_p^{(j)} - T_p^{(j-1)}$ ($0 \leq j < M_{la}$) from the pitch lag of the previous sub-frame and transmits the result, where $T_p^{(-1)}$ is the pitch lag of the last sub-frame in the frame to be encoded.

30

3. A method that performs vector quantization or arithmetic encoding on either of a part, or the whole, of the pitch lag $T_p^{(j)}$ ($0 \leq j < M_{la}$) and a part or the whole of the pitch lag calculated for the frame to be encoded and transmits the result.

4. A method that selects one of a number of predetermined interpolation methods based on a part or the whole of the pitch lag $T_p^{(j)}$ ($0 \leq j < M_{la}$) and transmits an index indicative of the selected interpolation method. At this time, the pitch lag of a plurality of sub-frames used for audio synthesis in the past also may be used for selection of the interpolation method.

35

[0087] For scalar quantization and vector quantization, a codebook determined empirically or a codebook calculated in advance by learning may be used. Further, a method that performs encoding after adding an offset value to the above pitch lag may also be included in the scope of the embodiment of the present invention as a matter of course.

<Decoding end>

40

[0088] As shown in Fig. 5, an example of the audio signal receiving device includes the audio code buffer 121, the audio parameter decoding unit 122, the audio parameter missing processing unit 123, the audio synthesis unit 124, the side information decoding unit 125, and the side information accumulation unit 126. The procedure of the audio signal receiving device is as shown in the example of Fig. 7.

45

[0089] The audio code buffer 121 determines whether a packet is correctly received or not. When the audio code buffer 121 determines that a packet is correctly received, the processing is switched to the audio parameter decoding unit 122 and the side information decoding unit 125. On the other hand, when the audio code buffer 121 determines that a packet is not correctly received, the processing is switched to the audio parameter missing processing unit 123 (Step 141 in Fig. 7).

50

<When packet is correctly received>

55

[0090] The audio parameter decoding unit 122 decodes the received audio code and calculates audio parameters required to synthesize the audio for the frame to be encoded (ISP parameter and corresponding ISF parameter, pitch lag, long-term prediction parameter, adaptive codebook, adaptive codebook gain, fixed codebook gain, fixed codebook vector etc.) (Step 142 in Fig. 7).

[0091] The side information decoding unit 125 decodes the side information code, calculates a pitch lag

$\hat{T}_p^{(j)}$ ($0 \leq j < M_{ia}$) and stores it in the side information accumulation unit 126. The side information decoding unit 125 decodes the side information code by using the decoding method corresponding to the encoding method used at the encoding end (Step 143 in Fig. 7).

5 **[0092]** The audio synthesis unit 124 synthesizes the audio signal corresponding to the frame to be encoded based on the parameters output from the audio parameter decoding unit 122 (Step 144 in Fig. 7). The functional configuration example of the audio synthesis unit 124 is shown in Fig. 15, and an example procedure of the audio synthesis unit 124 is shown in Fig. 16. Note that, although the audio parameter missing processing unit 123 is illustrated to show the flow of the signal, the audio parameter missing processing unit 123 is not included in the functional configuration of the audio synthesis unit 124.

10 **[0093]** An LP coefficient calculation unit 1121 converts an ISF parameter into an ISP parameter and then performs interpolation processing, and thereby obtains an ISP coefficient for each sub-frame. The LP coefficient calculation unit 1121 then converts the ISP coefficient into a linear prediction coefficient (LP coefficient) and thereby obtains an LP coefficient for each sub-frame (Step 11301 in Fig. 16). For the interpolation of the ISP coefficient and the ISP-LP coefficient, the method described in, for example, section 6.4.5 in Non Patent Literature 4 may be used. The procedure of parameter conversion is not the essential part of the embodiment of the present invention and thus not described in detail.

15 **[0094]** An adaptive codebook calculation unit 1123 calculates an adaptive codebook vector by using the pitch lag, a long-term prediction parameter and an adaptive codebook 1122 (Step 11302 in Fig. 16). An adaptive codebook vector $v'(n)$ is calculated from the pitch lag $\hat{T}_p^{(j)}$ and the adaptive codebook $u(n)$ according to the following equation.

$$v'(n) = \sum_{i=-1}^l Int(i) \cdot u(n + N_{adapt} - \hat{T}_p^{(j)} + i) (0 \leq n < L') \quad \text{Equation 19}$$

25

The adaptive codebook vector is calculated by interpolating the adaptive codebook $u(n)$ using FIR filter $Int(i)$. The length of the adaptive codebook is N_{adapt} . The filter $Int(i)$ that is used for the interpolation is the same as the interpolation filter of

30

$$v'(n) = \sum_{i=-1}^l Int(i) \cdot u(n + N_{adapt} - T_p + i) . \quad \text{Equation 20}$$

This is the FIR filter with a predetermined length $2l+1$. L' is the number of samples of the sub-frame. It is not necessary to use a filter for the interpolation, whereas at the encoder end a filter is used for the interpolation.

35

[0095] The adaptive codebook calculation unit 1123 carries out filtering on the adaptive codebook vector according to the value of the long-term prediction parameter (Step 11303 in Fig. 16). When the long-term prediction parameter has a value indicating the activation of filtering, filtering is performed on the adaptive codebook vector by the following equation.

40

$$v'(n) = 0.18v'(n-1) + 0.64v'(n) + 0.18v'(n+1) \quad \text{Equation 21}$$

[0096] On the other hand, when the long-term prediction parameter has a value indicating no filtering is needed, filtering is not performed, and $v(n) = v'(n)$ is established.

45

[0097] An excitation vector synthesis unit 1124 multiplies the adaptive codebook vector by an adaptive codebook gain g_p (Step 11304 in Fig. 16). Further, the excitation vector synthesis unit 1124 multiplies a fixed codebook vector $c(n)$ by a fixed codebook gain g_c (Step 11305 in Fig. 16). Furthermore, the excitation vector synthesis unit 1124 adds the adaptive codebook vector and the fixed codebook vector together and outputs an excitation signal vector (Step 11306 in Fig. 16).

50

$$e(n) = g_p \cdot v'(n) + g_c \cdot c(n) \quad \text{Equation 22}$$

55

[0098] A post filter 1125 performs post processing such as pitch enhancement, noise enhancement and low-frequency enhancement, for example, on the excitation signal vector. The details of techniques such as pitch enhancement, noise enhancement and low-frequency enhancement are described in the section 6.1 in Non Patent Literature 3. The processing in the post filter is not significantly related to the essential part of the embodiment of the present invention and thus not described in detail (Step 11307 in Fig. 16).

[0099] The adaptive codebook 1122 updates the state by an excitation signal vector according to the following equations (Step 11308 in Fig. 16).

$$u(n) = u(n+L) \quad (0 \leq n < N-L) \quad \text{Equation 23}$$

$$u(n+N-L) = e(n) \quad (0 \leq n < L) \quad \text{Equation 24}$$

[0100] A synthesis filter 1126 synthesizes a decoded signal according to the following equation by linear prediction inverse filtering using the excitation signal vector as an excitation source (Step 11309 in Fig. 16).

$$\hat{s}(n) = e(n) - \sum_{i=1}^P \hat{a}(i) \cdot \hat{s}(n-i) \quad \text{Equation 25}$$

[0101] An perceptual weighting inverse filter 1127 applies an perceptual weighting inverse filter to the decoded signal according to the following equation (Step 11310 in Fig. 16).

$$\hat{s}(n) = \hat{s}(n) + \beta \cdot \hat{s}(n-1) \quad \text{Equation 26}$$

The value of β is typically 0.68 or the like, though not limited to this value.

[0102] The audio parameter missing processing unit 123 stores the audio parameters (ISF parameter, pitch lag, adaptive codebook gain, fixed codebook gain) used in the audio synthesis unit 124 into the buffer (Step 145 in Fig. 7).

<When packet loss is detected>

[0103] The audio parameter missing processing unit 123 reads a pitch lag $\hat{T}_p^{(j)}$ ($0 \leq j < M_{ia}$) from the side information accumulation unit 126 and predicts audio parameters. The functional configuration example of the audio parameter missing processing unit 123 is shown in the example of Fig. 12, and an example procedure of audio parameter prediction is shown in Fig. 13.

[0104] An ISF prediction unit 191 calculates an ISF parameter using the ISF parameter for the previous frame and the ISF parameter calculated for the past several frames (Step 1101 in Fig. 13). The procedure of the ISF prediction unit 191 is shown in Fig. 10.

[0105] First, the buffer is updated using the ISF parameter of the immediately previous frame (Step 171 in Fig. 10). Next, the ISF parameter $\dot{\omega}_i$ is calculated according to the following equation (Step 172 in Fig. 10).

$$\dot{\omega}_i = \alpha \omega_i^{(-1)} + (1 - \alpha) \bar{\omega}_i \quad \text{Equation 27}$$

$$\bar{\omega}_i = \beta \omega_i^C + (1 - \beta) \frac{\omega_i^{(-3)} + \omega_i^{(-2)} + \omega_i^{(-1)}}{3} \quad \text{Equation 28}$$

where $\omega_i^{(j)}$ is the ISF parameter, stored in the buffer, which is for the frame preceding by j-number of frames. Further, ω_i^C , α and β are the same values as those used at the encoding end.

[0106] In addition, the values of i are arranged so that $\dot{\omega}_i$ satisfies $0 < \dot{\omega}_0 < \dot{\omega}_1 < \dots < \dot{\omega}_{14}$, and values of $\dot{\omega}_i$ are adjusted so that the adjacent $\dot{\omega}_i$ is not too close. As an example procedure to adjust the value of $\dot{\omega}_i$, Non Patent Literature 4 (Equation 151) may be used (Step 173 in Fig. 10).

[0107] A pitch lag prediction unit 192 decodes the side information code from the side information accumulation unit 126 and thereby obtains a pitch lag $\hat{T}_p^{(i)}$ ($0 \leq i < M_{ia}$). Further, by using a pitch lag $\hat{T}_p^{(-j)}$ ($0 \leq j < J$) used for the past decoding, the pitch lag prediction unit 192 outputs a pitch lag $\hat{T}_p^{(i)}$ ($M_{ia} \leq i < M$). The number of subframes contained in one frame is M , and the number of pitch lags contained in the side information is M_{ia} . For the prediction of the pitch lag $\hat{T}_p^{(i)}$ ($M_{ia} \leq i < M$), the procedure described in, for example, section 7.11.1.3 in Non Patent

Literature 4 may be used (Step 1102 in Fig. 13).

[0108] An adaptive codebook gain prediction unit 193 outputs an adaptive codebook gain $g_p^{(i)} (M_{la} \leq i < M)$ by using a predetermined adaptive codebook gain g_p^C and an adaptive codebook gain $g_p^{(j)} (0 \leq j < J)$ used in the past decoding. The number of sub-frames contained in one frame is M, and the number of pitch lags contained in the side information is M_{la} . For the prediction of the adaptive codebook gain $g_p^{(i)} (M_{la} \leq i < M)$, the procedure described in, for example, section 7.11.2.5.3 in Non Patent Literature 4 may be used (Step 1103 in Fig. 13).

[0109] A fixed codebook gain prediction unit 194 outputs a fixed codebook gain $g_c^{(i)} (0 \leq i < M)$ by using a fixed codebook gain $g_c^{(j)} (0 \leq j < J)$ used in the past decoding. The number of sub-frames contained in one frame is M. For the prediction of the fixed codebook gain $g_c^{(i)} (0 \leq i < M)$, the procedure described in the section 7.11.2.6 in Non Patent Literature 4 may be used, for example (Step 1104 in Fig. 13).

[0110] A noise signal generation unit 195 outputs a noise vector, such as a white noise, with a length of L (Step 1105 in Fig. 13). The length of one frame is L.

[0111] The audio synthesis unit 124 synthesizes a decoded signal based on the audio parameters output from the audio parameter missing processing unit 123 (Step 144 in Fig. 7). The operation of the audio synthesis unit 124 is the same as the operation of the audio synthesis unit <When audio packet is correctly received> and not redundantly described in detail (Step 144 in Fig. 7).

[0112] The audio parameter missing processing unit 123 stores the audio parameters (ISF parameter, pitch lag, adaptive codebook gain, fixed codebook gain) used in the audio synthesis unit 124 into the buffer (Step 145 in Fig. 7).

[0113] Although the case of encoding and transmitting the side information for all sub-frames contained in the look-ahead signal is described in the above example, the configuration that transmits only the side information for a specific sub-frame may be employed.

[Alternative example 1-1]

[0114] As an alternative example of the previously discussed example 1, an example that adds a pitch gain to the side information is described hereinafter. A difference between the alternative example 1-1 and the example 1 is only the operation of the excitation vector synthesis unit 155, and therefore description of the other parts is omitted.

<Encoding end>

[0115] The procedure of the excitation vector synthesis unit 155 is shown in the example of Fig. 14.

[0116] An adaptive codebook gain g_p^C is calculated from the adaptive codebook vector $v'(n)$ and the target signal $x(n)$ according to the following equation (Step 1111 in Fig. 14).

$$g_p = \frac{\sum_{n=0}^{L-1} x(n)y(n)}{\sum_{n=0}^{L-1} y(n)y(n)}, \text{ bounded by } 0 \leq g_p \leq 1.2, \quad \text{Equation 29}$$

where $y(n)$ is a signal $y(n)=v(n)*h(n)$ that is obtained by convoluting the impulse response with the adaptive codebook vector.

[0117] The calculated adaptive codebook gain is encoded and contained in the side information code (Step 1112 in Fig. 14). For the encoding, scalar quantization using a codebook obtained in advance by learning may be used, although any other technique may be used for the encoding.

[0118] By multiplying the adaptive codebook vector by an adaptive codebook gain \hat{g}_p obtained by decoding the code calculated in the encoding of the adaptive codebook gain, an excitation vector is calculated according to the following equation (Step 1113 in Fig. 14).

$$e(n) = \hat{g}_p \cdot v'(n) \quad \text{Equation 30}$$

<Decoding end>

[0119] The excitation vector synthesis unit 155 multiplies the adaptive codebook vector $v'(n)$ by an adaptive codebook gain \hat{g}_p obtained by decoding the side information code and outputs an excitation signal vector according to the following equation (Step 165 in Fig. 9).

$$e(n) = \hat{g}_p \cdot v'(n) \quad \text{Equation 31}$$

[Alternative example 1-2]

[0120] As an alternative example of the example 1, an example that adds a flag for determination of use of the side information to the side information is described hereinafter.

<Encoding end>

[0121] The functional configuration example of the side information encoding unit is shown in Fig. 17, and the procedure of the side information encoding unit is shown in the example of Fig. 18. A difference from the example 1 is only a side information output determination unit 1128 (Step 1131 in Fig. 18), and therefore description of the other parts is omitted.

[0122] The side information output determination unit 1128 calculates segmental SNR of the decoded signal and the look-ahead signal according to the following equation, and only when segmental SNR exceeds a threshold, sets the value of the flag to ON and adds it to the side information.

$$segSNR = \frac{\sum_{n=0}^{L'-1} \hat{s}^2(n)}{\sum_{n=0}^{L'-1} (s(n) - \hat{s}(n))^2} \quad \text{Equation 32}$$

On the other hand, when segmental SNR does not exceed a threshold, the side information output determination unit 1128 sets the value of the flag to OFF and adds it to the side information (Step 1131 in Fig. 18). Note that, the amount of bits of the side information may be reduced by adding the side information such as a pitch lag and a pitch gain to the flag and transmitting the added side information only when the value of the flag is ON, and transmitting only the value of the flag when the value of the flag is OFF.

<Decoding end>

[0123] The side information decoding unit decodes the flag contained in the side information code. When the value of the flag is ON, the audio parameter missing processing unit calculates a decoded signal by the same procedure as in the example 1. On the other hand, when the value of the flag is OFF, it calculates a decoded signal by the packet loss concealment technique without using side information (Step 1151 in Fig. 19).

[Example 2]

[0124] In this example, the decoded audio of the look-ahead signal part is also used when a packet is correctly received. For purposes of this discussion, the number of sub-frames contained in one frame is M sub-frames, and the length of the look-ahead signal is M' sub-frame(s).

<Encoding end>

[0125] As shown in the example of Fig. 20, the audio signal transmitting device includes a main encoding unit 211, a side information encoding unit 212, a concealment signal accumulation unit 213, and an error signal encoding unit 214. The procedure of the audio signal transmitting device is shown in Fig. 22.

[0126] The error signal encoding unit 214 reads a concealment signal for one sub-frame from the concealment signal accumulation unit 213, subtracts it from the audio signal and thereby calculates an error signal (Step 221 in Fig. 22).

[0127] The error signal encoding unit 214 encodes the error signal. As a specific example procedure, AVQ described

in the section 6.8.4.1.5 in Non Patent Literature 4, can be used. In the encoding of the error signal, local decoding is performed, and a decoded error signal is output (Step 222 in Fig. 22).

[0128] By adding the decoded error signal to the concealment signal, a decoded signal for one sub-frame is output (Step 223 in Fig. 22).

5 **[0129]** The above Steps 221 to 223 are repeated for M' sub-frames until the end of the concealment signal.

[0130] An example functional configuration of the main encoding unit 211 is shown in Fig. 21. The main encoding unit 211 includes an ISF encoding unit 2011, a target signal calculation unit 2012, a pitch lag calculation unit 2013, an adaptive codebook calculation unit 2014, a fixed codebook calculation unit 2015, a gain calculation unit 2016, an excitation vector calculation unit 2017, a synthesis filter 2018, and an adaptive codebook buffer 2019.

10 **[0131]** The ISF encoding unit 2011 obtains an LP coefficient by applying the Levinson-Durbin method to the frame to be encoded and the look-ahead signal. The ISF encoding unit 2011 then converts the LP coefficient into an ISF parameter and encodes the ISF parameter. The ISF encoding unit 2011 then decodes the code and obtains a decoded ISF parameter. Finally, the ISF encoding unit 2011 interpolates the decoded ISF parameter and obtains a decoded LP coefficient for each sub-frame. The procedures of the Levinson-Durbin method and the conversion from the LP coefficient to the ISF parameter are the same as in the example 1. Further, for the encoding of the ISF parameter, the procedure described in, for example, section 6.8.2 in Non Patent Literature 4 can be used. An index obtained by encoding the ISF parameter, the decoded ISF parameter, and the decoded LP coefficient (which is obtained by converting the decoded ISF parameter into the LP coefficient) can be obtained by the ISF encoding unit 2011 (Step 224 in Fig. 22).

15 **[0132]** The detailed procedure of the target signal calculation unit 2012 is the same as in Step 162 in Fig. 9 in the example 1 (Step 225 in Fig. 22).

[0133] The pitch lag calculation unit 2013 refers to the adaptive codebook buffer and calculates a pitch lag and a long-term prediction parameter by using the target signal. The detailed procedure of the calculation of the pitch lag and the long-term prediction parameter is the same as in the example 1 (Step 226 in Fig. 22).

20 **[0134]** The adaptive codebook calculation unit 2014 calculates an adaptive codebook vector by using the pitch lag and the long-term prediction parameter calculated by the pitch lag calculation unit 2013. The detailed procedure of the adaptive codebook calculation unit 2014 is the same as in the example 1 (Step 227 in Fig. 22).

[0135] The fixed codebook calculation unit 2015 calculates a fixed codebook vector and an index obtained by encoding the fixed codebook vector by using the target signal and the adaptive codebook vector. The detailed procedure is the same as the procedure of AVQ used in the error signal encoding unit 214 (Step 228 in Fig. 22).

25 **[0136]** The gain calculation unit 2016 calculates an adaptive codebook gain, a fixed codebook gain and an index obtained by encoding these two gains using the target signal, the adaptive codebook vector and the fixed codebook vector. A detailed procedure which can be used is described in, for example, section 6.8.4.1.6 in Non Patent Literature 4 (Step 229 in Fig. 22).

30 **[0137]** The excitation vector calculation unit 2017 calculates an excitation vector by adding the adaptive codebook vector and the fixed codebook vector to which the gain is applied. The detailed procedure is the same as in example 1. Further, the excitation vector calculation unit 2017 updates the state of the adaptive codebook buffer 2019 by using the excitation vector. The detailed procedure is the same as in the example 1 (Step 2210 in Fig. 22).

[0138] The synthesis filter 2018 synthesizes a decoded signal by using the decoded LP coefficient and the excitation vector (Step 2211 in Fig. 22).

35 **[0139]** The above Steps 224 to 2211 are repeated for $M-M'$ sub-frames until the end of the frame to be encoded.

[0140] The side information encoding unit 212 calculates the side information for the look-ahead signal M' sub-frame. A specific procedure is the same as in the example 1 (Step 2212 in Fig. 22).

[0141] In addition to the procedure of the example 1, the decoded signal output by the synthesis filter 157 of the side information encoding unit 212 is accumulated in the concealment signal accumulation unit 213 in the example 2 (Step 2213 in Fig. 22).

<Decoding unit>

40 **[0142]** As shown in Fig. 23, an example of the audio signal receiving device includes an audio code buffer 231, an audio parameter decoding unit 232, an audio parameter missing processing unit 233, an audio synthesis unit 234, a side information decoding unit 235, a side information accumulation unit 236, an error signal decoding unit 237, and a concealment signal accumulation unit 238. An example procedure of the audio signal receiving device is shown in Fig. 24. An example functional configuration of the audio synthesis unit 234 is shown in Fig. 25.

45 **[0143]** The audio code buffer 231 determines whether a packet is correctly received or not. When the audio code buffer 231 determines that a packet is correctly received, the processing is switched to the audio parameter decoding unit 232, the side information decoding unit 235 and the error signal decoding unit 237. On the other hand, when the audio code buffer 231 determines that a packet is not correctly received, the processing is switched to the audio parameter missing processing unit 233 (Step 241 in Fig. 24).

<When packet is correctly received>

5 [0144] The error signal decoding unit 237 decodes an error signal code and obtains a decoded error signal. As a specific example procedure, a decoding method corresponding to the method used at the encoding end, such as AVQ described in the section 7.1.2.1.2 in Non Patent Literature 4 can be used (Step 242 in Fig. 24).

[0145] A look-ahead excitation vector synthesis unit 2318 reads a concealment signal for one sub-frame from the concealment signal accumulation unit 238 and adds the concealment signal to the decoded error signal, and thereby outputs a decoded signal for one sub-frame (Step 243 in Fig. 24).

[0146] The above Steps 241 to 243 are repeated for M' sub-frames until the end of the concealment signal.

10 [0147] The audio parameter decoding unit 232 includes an ISF decoding unit 2211, a pitch lag decoding unit 2212, a gain decoding unit 2213, and a fixed codebook decoding unit 2214. The functional configuration example of the audio parameter decoding unit 232 is shown in Fig. 26.

[0148] The ISF decoding unit 2211 decodes the ISF code and converts it into an LP coefficient and thereby obtains a decoded LP coefficient. For example, the procedure described in the section 7.1.1 in Non Patent Literature 4 is used (Step 244 in Fig. 24).

15 [0149] The pitch lag decoding unit 2212 decodes a pitch lag code and obtains a pitch lag and a long-term prediction parameter (Step 245 in Fig. 24).

[0150] The gain decoding unit 2213 decodes a gain code and obtains an adaptive codebook gain and a fixed codebook gain. An example detailed procedure is described in the section 7.1.2.1.3 in Non Patent Literature 4 (Step 246 in Fig. 24).

20 [0151] An adaptive codebook calculation unit 2313 calculates an adaptive codebook vector by using the pitch lag and the long-term prediction parameter. The detailed procedure of the adaptive codebook calculation unit 2313 is as described in the example 1 (Step 247 in Fig. 24).

[0152] The fixed codebook decoding unit 2214 decodes a fixed codebook code and calculates a fixed codebook vector. The detailed procedure is as described in the section 7.1.2.1.2 in Non Patent Literature 4 (Step 248 in Fig. 24).

25 [0153] An excitation vector synthesis unit 2314 calculates an excitation vector by adding the adaptive codebook vector and the fixed codebook vector to which the gain is applied. Further, an excitation vector calculation unit updates the adaptive codebook buffer by using the excitation vector (Step 249 in Fig. 24). The detailed procedure is the same as in the example 1.

30 [0154] A synthesis filter 2316 synthesizes a decoded signal by using the decoded LP coefficient and the excitation vector (Step 2410 in Fig. 24). The detailed procedure is the same as in the example 1.

[0155] The above Steps 244 to 2410 are repeated for $M-M'$ sub-frames until the end of the frame to be encoded.

[0156] The functional configuration of the side information decoding unit 235 is the same as in the example 1. The side information decoding unit 235 decodes the side information code and calculates a pitch lag (Step 2411 in Fig. 24).

[0157] The functional configuration of the audio parameter missing processing unit 233 is the same as in the example 1.

35 [0158] The ISF prediction unit 191 predicts an ISF parameter using the ISF parameter for the previous frame and converts the predicted ISF parameter into an LP coefficient. The procedure is the same as in Steps 172, 173 and 174 of the example 1 shown in Fig. 10 (Step 2412 in Fig. 24).

[0159] The adaptive codebook calculation unit 2313 calculates an adaptive codebook vector by using the pitch lag output from the side information decoding unit 235 and an adaptive codebook 2312 (Step 2413 in Fig. 24). The procedure is the same as in Steps 11301 and 11302 in Fig. 16.

40 [0160] The adaptive codebook gain prediction unit 193 outputs an adaptive codebook gain. A specific procedure is the same as in Step 1103 in Fig. 13 (Step 2414 in Fig. 24).

[0161] The fixed codebook gain prediction unit 194 outputs a fixed codebook gain. A specific procedure is the same as in Step 1104 in Fig. 13 (Step 2415 in Fig. 24).

45 [0162] The noise signal generation unit 195 outputs a noise, such as a white noise as a fixed codebook vector. The procedure is the same as in Step 1105 in Fig. 13 (Step 2416 in Fig. 24).

[0163] The excitation vector synthesis unit 2314 applies gain to each of the adaptive codebook vector and the fixed codebook vector and adds them together and thereby calculates an excitation vector. Further, the excitation vector synthesis unit 2314 updates the adaptive codebook buffer using the excitation vector (Step 2417 in Fig. 24).

50 [0164] The synthesis filter 2316 calculates a decoded signal using the above-described LP coefficient and the excitation vector. The synthesis filter 2316 then updates the concealment signal accumulation unit 238 using the calculated decoded signal (Step 2418 in Fig. 24).

[0165] The above steps are repeated for M' sub-frames, and the decoded signal is output as the audio signal.

55 <When a packet is lost>

[0166] A concealment signal for one sub-frame is read from the concealment signal accumulation unit and is used as the decoded signal (Step 2419 in Fig. 24).

[0167] The above is repeated for M' sub-frames.

[0168] The ISF prediction unit 191 predicts an ISF parameter (Step 2420 in Fig. 24). As the procedure, Step 1101 in Fig. 13 can be used.

[0169] The pitch lag prediction unit 192 outputs a predicted pitch lag by using the pitch lag used in the past decoding (Step 2421 in Fig. 24). The procedure used for the prediction is the same as in Step 1102 in Fig. 13.

[0170] The operations of the adaptive codebook gain prediction unit 193, the fixed codebook gain prediction unit 194, the noise signal generation unit 195 and the audio synthesis unit 234 are the same as in the example 1 (Step 2422 in Fig. 24).

[0171] The above steps are repeated for M sub-frames, and the decoded signal for M-M' sub-frames is output as the audio signal, and the concealment signal accumulation unit 238 is updated by the decoded signal for the remaining M' sub-frames.

[Example 3]

[0172] A case of using glottal pulse synchronization in the calculation of an adaptive codebook vector is described hereinafter.

<Encoding end>

[0173] The functional configuration of the audio signal transmitting device is the same as in example 1. The functional configuration and the procedure are different only in the side information encoding unit, and therefore only the operation of the side information encoding unit is described below.

[0174] The side information encoding unit includes an LP coefficient calculation unit 311, a pitch lag prediction unit 312, a pitch lag selection unit 313, a pitch lag encoding unit 314, and an adaptive codebook buffer 315. The functional configuration of an example of the side information encoding unit is shown in Fig. 27, and an example procedure of the side information encoding unit is shown in the example of Fig. 28.

[0175] The LP coefficient calculation unit 311 is the same as the LP coefficient calculation unit in example 1 and thus will not be redundantly described (Step 321 in Fig. 28).

[0176] The pitch lag prediction unit 312 calculates a pitch lag predicted value \hat{T}_p using the pitch lag obtained from the audio encoding unit (Step 322 in Fig. 28). The specific processing of the prediction is the same as the prediction of the pitch lag $\hat{T}_p^{(i)}$ ($M_{la} \leq i < M$) in the pitch lag prediction unit 192 in the example 1 (which is the same as in Step 1102 in Fig. 13).

[0177] Then, the pitch lag selection unit 313 determines a pitch lag to be transmitted as the side information (Step 323 in Fig. 28). The detailed procedure of the pitch lag selection unit 313 is shown in the example of Fig. 29.

[0178] First, a pitch lag codebook is generated from the pitch lag predicted value \hat{T}_p and the value of the past pitch lag $\hat{T}_p^{(-j)}$ ($0 \leq j < J$) according to the following equations (Step 331 in Fig. 29).

<When $\hat{T}_p - \hat{T}_p^{(-1)} \geq 0$ >

$$\hat{T}_C^j = \begin{cases} \hat{T}_p & (j = 0) \\ \hat{T}_p^{(-1)} - j \cdot \delta_j + \rho & (0 < j < I) \end{cases} \quad \text{Equation 33}$$

<When $\hat{T}_p - \hat{T}_p^{(-1)} < 0$ >

$$\hat{T}_C^j = \begin{cases} \hat{T}_p & (j = 0) \\ \hat{T}_p^{(-1)} + j \cdot \delta_j + \rho & (0 < j < I) \end{cases} \quad \text{Equation 34}$$

The value of the pitch lag for one sub-frame before is $\hat{T}_p^{(-1)}$. Further, the number of indexes of the codebook is I. δ_j is a predetermined step width, and ρ is a predetermined constant.

[0179] Then, by using the adaptive codebook and the pitch lag predicted value \hat{T}_p , an initial excitation vector $u_0(n)$ is

generated according to the following equation (Step 332 in Fig. 29).

$$u_0(n) = \begin{cases} 0.18u_0(n - \hat{T}_p - 1) + 0.64u_0(n - \hat{T}_p) + 0.18u_0(n - \hat{T}_p + 1) & (0 \leq n < \hat{T}_p) \\ u_0(n - \hat{T}_p) & (\hat{T}_p \leq n < L) \end{cases}$$

Equation 35

The procedure of calculating the initial excitation vector is the same as the equations (607) and (608) in Non Patent Literature 4.

[0180] Then, glottal pulse synchronization is applied to the initial excitation vector by using all candidate pitch lags \hat{T}_C^j ($0 \leq j < J$) in the pitch lag codebook to thereby generate a candidate adaptive codebook vector $u^j(n)$ ($0 \leq j < J$) (Step 333 in Fig. 29). For the glottal pulse synchronization, the same procedure can be used as in the case described in section 7.11.2.5 in Non Patent Literature 4 where a pulse position is not available. Note, however, that $u(n)$ in Non Patent Literature 4 corresponds to $u_0(n)$ in the embodiment of the present invention, and extrapolated pitch corresponds to \hat{T}_C^j

in the embodiment of the present invention, and the last reliable pitch (T_C) corresponds to $\hat{T}_p^{(-1)}$ in the embodiment of the present invention.

[0181] For the candidate adaptive codebook vector $u^j(n)$ ($0 \leq j < J$), a rate scale is calculated (Step 334 in Fig. 29). In the case of using segmental SNR as the rate scale, a signal is synthesized by inverse filtering using the LP coefficient, and segmental SNR is calculated with the input signal according to the following equation.

$$\hat{s}_j(n) = u^j(n) - \sum_{i=1}^P \hat{a}(i) \cdot \hat{s}_j(n-i) \tag{Equation 35}$$

$$segSNR_j = \frac{\sum_{n=0}^{L-1} \hat{s}_j^2(n)}{\sum_{n=0}^{L-1} (s(n) - \hat{s}_j(n))^2} \tag{Equation 36}$$

[0182] Instead of performing inverse filtering, segmental SNR may be calculated in the region of the adaptive codebook vector by using a residual signal according to the following equation.

$$r(n) = s(n) + \sum_{i=1}^P \hat{a}(i) \cdot s(n-i) \tag{Equation 37}$$

$$segSNR_j = \frac{\sum_{n=0}^{L-1} u^j(n)}{\sum_{n=0}^{L-1} (r(n) - u^j(n))^2} \tag{Equation 38}$$

In this case, a residual signal $r(n)$ of the look-ahead signal $s(n)$ ($0 \leq n < L$) is calculated by using the LP coefficient (Step 181 in Fig. 11).

[0183] An index corresponding to the largest rate scale calculated in Step 334 is selected, and a pitch lag corresponding to the index is calculated (Step 335 in Fig. 29).

$$\arg \max_j \lfloor \text{segSNR}_j \rfloor$$

Equation 39

5

<Decoding end>

10

[0184] The functional configuration of the audio signal receiving device is the same as in the example 1. Differences from the example 1 are the functional configuration and the procedure of the audio parameter missing processing unit 123, the side information decoding unit 125 and the side information accumulation unit 126, and only those are described hereinbelow.

<When packet is correctly received>

15

[0185] The side information decoding unit 125 decodes the side information code and calculates a pitch lag \hat{T}_C^{idx} and stores it into the side information accumulation unit 126. The example procedure of the side information decoding unit 125 is shown in Fig. 30.

20

[0186] In the calculation of the pitch lag, the pitch lag prediction unit 312 first calculates a pitch lag predicted value \hat{T}_p by using the pitch lag obtained from the audio decoding unit (Step 341 in Fig. 30). The specific processing of the prediction is the same as in Step 322 of Fig. 28 in the example 3.

[0187] Then, a pitch lag codebook is generated from the pitch lag predicted value \hat{T}_p , and the value of the past pitch lag $\hat{T}_p^{(-j)}$ ($0 \leq j < J$), according to the following equations (Step 342 in Fig. 30).

25

<When $\hat{T}_p - \hat{T}_p^{(-1)} \geq 0$ >

30

$$\hat{T}_C^j = \begin{cases} \hat{T}_p (j = 0) \\ \hat{T}_p^{(-1)} - j \cdot \delta_j + \rho (0 < j < I) \end{cases}$$

Equation 40

<When $\hat{T}_p - \hat{T}_p^{(-1)} < 0$ >

35

$$\hat{T}_C^j = \begin{cases} \hat{T}_p (j = 0) \\ \hat{T}_p^{(-1)} + j \cdot \delta_j + \rho (0 < j < I) \end{cases}$$

Equation 41

40

The procedure is the same as in Step 331 in Fig. 29. The value of the pitch lag for one sub-frame before is $\hat{T}_p^{(-1)}$. Further, the number of indexes of the codebook is I. δ_j is a predetermined step width, and ρ is a predetermined constant.

45

[0188] Then, by referring to the pitch lag codebook, a pitch lag \hat{T}_C^{idx} corresponding to the index idx transmitted as part of the side information is calculated and stored in the side information accumulation unit 126 (Step 343 in Fig. 30).

<When packet loss is detected>

50

[0189] Although the functional configuration of the audio synthesis unit is also the same as in the example 1 (which is the same as in Fig. 15), only the adaptive codebook calculation unit 1123 that operates differently from that in the example 1 is described hereinbelow.

[0190] The audio parameter missing processing unit 123 reads the pitch lag from the side information accumulation unit 126 and calculates a pitch lag predicted value according to the following equation, and uses the calculated pitch lag predicted value instead of the output of the pitch lag prediction unit 192.

55

$$\hat{T}_p = \hat{T}_p^{(-1)} + K \cdot (\hat{T}_C^{idx} - \hat{T}_p^{(-1)})$$

Equation 42

where κ is a predetermined constant.

[0191] Then, by using the adaptive codebook and the pitch lag predicted value \hat{T}_p , an initial excitation vector $u_0(n)$ is generated according to the following equation (Step 332 in Fig. 29).

$$u_0(n) = \begin{cases} 0.18u_0(n - \hat{T}_p^{(-1)} - 1) + 0.64u_0(n - \hat{T}_p^{(-1)}) + 0.18u_0(n - \hat{T}_p^{(-1)} + 1) & (0 \leq n < \hat{T}_p^{(-1)}) \\ u_0(n - \hat{T}_p^{(-1)}) & (\hat{T}_p^{(-1)} \leq n < L) \end{cases}$$

Equation 43

[0192] Then, glottal pulse synchronization is applied to the initial excitation vector by using the pitch lag \hat{T}_C^{idx} to thereby generate an adaptive codebook vector $u(n)$. For the glottal pulse synchronization, the same procedure as in Step 333 of Fig. 29 is used.

[0193] Hereinafter, an audio encoding program 70 that causes a computer to execute the above-described processing by the audio signal transmitting device is described. As shown in Fig. 31, the audio encoding program 70 is stored in a program storage area 61 formed in a recording medium 60 that is inserted into a computer and accessed, or included in a computer.

[0194] The audio encoding program 70 includes an audio encoding module 700 and a side information encoding module 701. The functions implemented by executing the audio encoding module 700 and the side information encoding module 701 are the same as the functions of the audio encoding unit 111 and the side information encoding unit 112 in the audio signal transmitting device described above, respectively.

[0195] Note that a part or the whole of the audio encoding program 70 may be transmitted through a transmission medium such as a communication line, received and stored (including being installed) by another device. Further, each module of the audio encoding program 70 may be installed not in one computer but in any of a plurality of computers. In this case, the above-described processing of the audio encoding program 70 is performed by a computer system composed of the plurality of computers.

[0196] Hereinafter, an audio decoding program 90 that causes a computer to execute the above-described processing by the audio signal receiving device is described. As shown in Fig. 32, the audio decoding program 90 is stored in a program storage area 81 formed in a recording medium 80 that is inserted into a computer and accessed, or included in a computer.

[0197] The audio decoding program 90 includes an audio code buffer module 900, an audio parameter decoding module 901, a side information decoding module 902, a side information accumulation module 903, an audio parameter missing processing module 904, and an audio synthesis module 905. The functions implemented by executing the audio code buffer module 900, the audio parameter decoding module 901, the side information decoding module 902, the side information accumulation module 903, an audio parameter missing processing module 904 and the audio synthesis module 905 are the same as the function of the audio code buffer 231, the audio parameter decoding unit 232, the side information decoding unit 235, the side information accumulation unit 236, the audio parameter missing processing unit 233 and the audio synthesis unit 234 described above, respectively.

[0198] Note that a part or the whole of the audio decoding program 90 may be transmitted through a transmission medium such as a communication line, received and stored (including being installed) by another device. Further, each module of the audio decoding program 90 may be installed not in one computer but in any of a plurality of computers. In this case, the above-described processing of the audio decoding program 90 is performed by a computer system composed of the plurality of computers.

[Example 4]

[0199] An example that uses side information for pitch lag prediction at the decoding end is described hereinafter.

<Encoding end>

[0200] The functional configuration of the audio signal transmitting device is the same as in the example 1. The functional configuration and the procedure are different only in the side information encoding unit 112, and therefore the operation of the side information encoding unit 112 only is described hereinbelow.

[0201] The functional configuration of an example of the side information encoding unit 112 is shown in Fig. 33, and an example procedure of the side information encoding unit 112 is shown in Fig. 34. The side information encoding unit

112 includes an LP coefficient calculation unit 511, a residual signal calculation unit 512, a pitch lag calculation unit 513, an adaptive codebook calculation unit 514, an adaptive codebook buffer 515, and a pitch lag encoding unit 516.

[0202] The LP coefficient calculation unit 511 is the same as the LP coefficient calculation unit 151 in example 1 shown in Fig. 8 and thus is not redundantly described.

5 [0203] The residual signal calculation unit 512 calculates a residual signal by the same processing as in Step 181 in example 1 shown in Fig. 11.

[0204] The pitch lag calculation unit 513 calculates a pitch lag for each sub-frame by calculating k that maximizes the following equation (Step 163 in Fig. 34). Note that $u(n)$ indicates the adaptive codebook, and L' indicates the number of samples contained in one sub-frame.

10

$$T_p = \arg_k \max T_k$$

15

$$T_k = \frac{\sum_{n=0}^{L'-1} r(n)u(n-k)}{\sqrt{\sum_{n=0}^{L'-1} u(n-k)u(n-k)}} \quad \text{Equation 43}$$

20

[0205] The adaptive codebook calculation unit 514 calculates an adaptive codebook vector $v'(n)$ from the pitch lag T_p and the adaptive codebook $u(n)$. The length of the adaptive codebook is N_{adapt} (Step 164 in Fig. 34).

25

$$v'(n) = u(n + N_{\text{adapt}} - T_p) \quad \text{Equation 44}$$

[0206] The adaptive codebook buffer 515 updates the state by the adaptive codebook vector $v'(n)$ (Step 166 in Fig. 34).

30

$$u(n) = u(n + L') \quad (0 \leq n < N - L') \quad \text{Equation 45}$$

$$u(n + N - L') = v'(n) \quad (0 \leq n < L') \quad \text{Equation 46}$$

35

[0207] The pitch lag encoding unit 516 is the same as that in example 1 and thus not redundantly described (Step 169 in Fig. 34).

<Decoding end>

40

[0208] The audio signal receiving device includes the audio code buffer 121, the audio parameter decoding unit 122, the audio parameter missing processing unit 123, the audio synthesis unit 124, the side information decoding unit 125, and the side information accumulation unit 126, just like in example 1. The procedure of the audio signal receiving device is as shown in Fig. 7.

[0209] The operation of the audio code buffer 121 is the same as in example 1.

45

<When packet is correctly received>

[0210] The operation of the audio parameter decoding unit 122 is the same as in the example 1.

[0211] The side information decoding unit 125 decodes the side information code, calculates a pitch lag

50

$\hat{T}_p^{(j)}$ ($0 \leq j < M_{ia}$) and stores it into the side information accumulation unit 126. The side information decoding unit 125 decodes the side information code by using the decoding method corresponding to the encoding method used at the encoding end.

[0212] The audio synthesis unit 124 is the same as that of example 1.

55

<When packet loss is detected>

[0213] The ISF prediction unit 191 of the audio parameter missing processing unit 123 (see Fig. 12) calculates an ISF

parameter the same way as in the example 1.

[0214] An example procedure of the pitch lag prediction unit 192 is shown in Fig. 35. The pitch lag prediction unit 192 reads the side information code from the side information accumulation unit 126 and obtains a pitch lag

5 $\hat{T}_p^{(i)} (0 \leq i < M_{la})$ in the same manner as in example 1 (Step 4051 in Fig. 35). Further, the pitch lag prediction unit 192 outputs the pitch lag $\hat{T}_p^{(i)} (M_{la} \leq i < M)$ by using the pitch lag $\hat{T}_p^{(-j)} (0 \leq j < J)$ used in the past decoding (Step 4052 in Fig. 35). The number of sub-frames contained in one frame is M, and the number of pitch lags contained

10 in the side information is M_{la} . In the prediction of the pitch lag $\hat{T}_p^{(i)} (M_{la} \leq i < M)$, the procedure as described in Non Patent Literature 4 can be used (Step 1102 in Fig. 13).

[0215] In the prediction of the pitch lag $\hat{T}_p^{(i)} (M_{la} \leq i < M)$, the pitch lag prediction unit 192 may predict the pitch

15 lag $\hat{T}_p^{(i)} (M_{la} \leq i < M)$ by using the pitch lag $\hat{T}_p^{(-j)} (1 \leq j < J)$ used in the past decoding and the pitch lag $\hat{T}_p^{(i)} (0 \leq i < M_{la})$. Further, $\hat{T}_p^{(i)} = \hat{T}_p^{(M_{la})}$ may be established. The procedure of the pitch lag prediction unit in this case is as shown in Fig. 36.

20 **[0216]** Further, the pitch lag prediction unit 192 may establish $\hat{T}_p^{(i)} = \hat{T}_p^{(M_{la})}$ only when the reliability of the pitch lag predicted value is low. The procedure of the pitch lag prediction unit in this case is shown in Fig. 37. Instruction information as to whether the predicated value is used, or the pitch lag $\hat{T}_p^{(M_{la})}$ obtained by the side information is used may be input

25 to the adaptive codebook calculation unit 154.

[0217] The adaptive codebook gain prediction unit 193 and the fixed codebook gain prediction unit 194 are the same as those of the example 1.

[0218] The noise signal generation unit 195 is the same as that of the example 1.

30 **[0219]** The audio synthesis unit 124 synthesizes, from the parameters output from the audio parameter missing processing unit 123, an audio signal corresponding to the frame to be encoded.

[0220] The LP coefficient calculation unit 1121 of the audio synthesis unit 124 (see Fig. 15) obtains an LP coefficient in the same manner as in example 1 (Step S11301 in Fig. 16).

35 **[0221]** The adaptive codebook calculation unit 1123 calculates an adaptive codebook vector in the same manner as in example 1. The adaptive codebook calculation unit 1123 may perform filtering on the adaptive codebook vector or may not perform filtering. Specifically, the adaptive codebook vector is calculated using the following equation. The filtering coefficient is f_i .

$$v(n) = f_{-1}v'(n-1) + f_0v'(n) + f_1v'(n+1) \quad \text{Equation 47}$$

40 In the case of decoding a value that does not indicate filtering, $v(n) = v'(n)$ is established (adaptive codebook calculation step A).

[0222] The adaptive codebook calculation unit 1123 may calculate an adaptive codebook vector in the following procedure (adaptive codebook calculation step B).

45 **[0223]** An initial adaptive codebook vector is calculated using the pitch lag and the adaptive codebook 1122.

$$v(n) = f_{-1}v'(n-1) + f_0v'(n) + f_1v'(n+1) \quad \text{Equation 48}$$

50 $v(n) = v'(n)$ may be established according to a design policy.

[0224] Then, glottal pulse synchronization is applied to the initial adaptive codebook vector. For the glottal pulse synchronization, the same procedure as in the case where a pulse position is not available in the section 7.11.2.5 in Non Patent Literature 4 is used. Note that, however, $u(n)$ in Non Patent Literature 4 corresponds to $v(n)$ in the embodiment

55 of the present invention, and extrapolated pitch corresponds to $\hat{T}_p^{(M-1)}$ in the embodiment of the present invention, and the last reliable pitch (T_c) corresponds to $\hat{T}_p^{(M_{la}-1)}$ in the embodiment of the present invention.

[0225] Further, in the case where the pitch lag prediction unit 192 outputs the above-described instruction information

for the predicated value, when the instruction information indicates that the pitch lag transmitted as the side information should not be used as the predicated value (NO in Step 4082 in Fig. 38), the adaptive codebook calculation unit 1123 may use the above-described adaptive codebook calculation step A, and if it is indicated that the pitch value should be used (YES in Step 4082 in Fig. 38), the adaptive codebook calculation unit 1123 may use the above-described adaptive codebook calculation step B. The procedure of the adaptive codebook calculation unit 1123 in this case is shown in the example of Fig. 38.

[0226] The excitation vector synthesis unit 1124 outputs an excitation vector in the same manner as in example 1 (Step 11306 in Fig. 16).

[0227] The post filter 1125 performs post processing on the synthesis signal in the same manner as in the example 1.

[0228] The adaptive codebook 1122 updates the state by using the excitation signal vector in the same manner as in the example 1 (Step 11308 in Fig. 16).

[0229] The synthesis filter 1126 synthesizes a decoded signal in the same manner as in the example 1 (Step 11309 in Fig. 16).

[0230] The perceptual weighting inverse filter 1127 applies an perceptual weighting inverse filter in the same manner as in the example 1.

[0231] The audio parameter missing processing unit 123 stores the audio parameters (ISF parameter, pitch lag, adaptive codebook gain, fixed codebook gain) used in the audio synthesis unit 124 into the buffer in the same manner as in the example 1 (Step 145 in Fig. 7).

[Example 5]

[0232] In this embodiment, a configuration is described in which a pitch lag is transmitted as side information only in a specific frame class, and otherwise a pitch lag is not transmitted.

<Transmitting end>

[0233] In the audio signal transmitting device, an input audio signal is sent to the audio encoding unit 111.

[0234] The audio encoding unit 111 in this example calculates an index representing the characteristics of a frame to be encoded and transmits the index to the side information encoding unit 112. The other operations are the same as in example 1.

[0235] In the side information encoding unit 112, a difference from the examples 1 to 4 is only with regard to the pitch lag encoding unit 158, and therefore the operation of the pitch lag encoding unit 158 is described hereinbelow. The configuration of the side information encoding unit 112 in the example 5 is shown in Fig. 39.

[0236] The procedure of the pitch lag encoding unit 158 is shown in the example of Fig. 40. The pitch lag encoding unit 158 reads the index representing the characteristics of the frame to be encoded (Step 5021 in Fig. 40) and, when the index representing the characteristics of the frame to be encoded is equal to a predetermined value, the pitch lag encoding unit 158 determines the number of bits to be assigned to the side information as B bits ($B > 1$). On the other hand, when the index representing the characteristics of the frame to be encoded is different from a predetermined value, the pitch lag encoding unit 158 determines the number of bits to be assigned to the side information as 1 bit (Step 5022 in Fig. 40).

[0237] When the number of bits to be assigned to the side information is 1 bit (No in Step 5022 in Fig. 40), a value indicating non-transmission of the side information, is used as the side information code, and is set to the side information index (Step 5023 in Fig. 40).

[0238] On the other hand, when the number of bits to be assigned to the side information is B bits (Yes in Step 5022 in Fig. 40), a value indicating transmission of the side information is set to the side information index (Step 5024 in Fig. 40), and further, a code of B-1 bits obtained by encoding the pitch lag by the method described in example 1 is added, for use as the side information code (Step 5025 in Fig. 40).

<Decoding end>

[0239] The audio signal receiving device includes the audio code buffer 121, the audio parameter decoding unit 122, the audio parameter missing processing unit 123, the audio synthesis unit 124, the side information decoding unit 125, and the side information accumulation unit 126, just like in example 1. The procedure of the audio signal receiving device is as shown in Fig. 7.

[0240] The operation of the audio code buffer 121 is the same as in example 1.

<When packet is correctly received>

[0241] The operation of the audio parameter decoding unit 122 is the same as in example 1.

5 [0242] The procedure of the side information decoding unit 125 is shown in the example of Fig. 41. The side information decoding unit 125 decodes the side information index contained in the side information code first (Step 5031 in Fig. 41). When the side information index indicates non-transmission of the side information, the side information decoding unit 125 does not perform any further decoding operations. Also, the side information decoding unit 125 stores the value of the side information index in the side information accumulation unit 126 (Step 5032 in Fig. 41).

10 [0243] On the other hand, when the side information index indicates transmission of the side information, the side information decoding unit 125 further performs decoding of B-1 bits and calculates a pitch lag $\hat{T}_p^{(j)}$ ($0 \leq j < M_{ia}$) and stores the calculated pitch lag in the side information accumulation unit 126 (Step 5033 in Fig. 41). Further, the side information decoding unit 125 stores the value of the side information index into the side information accumulation unit 126. Note that the decoding of the side information of B-1 bits is the same operation as the side information decoding unit 125 in example 1.

15 [0244] The audio synthesis unit 124 is the same as that of example 1.

<When packet loss is detected>

20 [0245] The ISF prediction unit 191 of the audio parameter missing processing unit 123 (see Fig. 12) calculates an ISF parameter the same way as in example 1.

[0246] The procedure of the pitch lag prediction unit 192 is shown in the example of Fig. 42. The pitch lag prediction unit 192 reads the side information index from the side information accumulation unit 126 (Step 5041 in Fig. 42) and checks whether it is the value indicating transmission of the side information (Step 5042 in Fig. 42).

25

<When the side information index is a value indicating transmission of side information >

[0247] In the same manner as in example 1, the side information code is read from the side information accumulation unit 126 to obtain a pitch lag $\hat{T}_p^{(i)}$ ($0 \leq i < M_{ia}$) (Step 5043 in Fig. 42). Further, the pitch lag $\hat{T}_p^{(i)}$ ($M_{ia} \leq i < M$)

30

is output by using the pitch lag $\hat{T}_p^{(-j)}$ ($0 \leq j < J$) used in the past decoding and $\hat{T}_p^{(i)}$ ($0 \leq i < M_{ia}$) obtained as the side information (Step 5044 in Fig. 42). The number of sub-frames contained in one frame is M, and the number of

35

pitch lags contained in the side information is M_{ia} . In the prediction of the pitch lag $\hat{T}_p^{(i)}$ ($M_{ia} \leq i < M$), the procedure as described in Non Patent Literature 4 can be used (Step 1102 in Fig. 13). Further, $\hat{T}_p^{(i)} = \hat{T}_p^{(M_{ia})}$ may be established.

40

[0248] Further, the pitch lag prediction unit 192 may establish $\hat{T}_p^{(i)} = \hat{T}_p^{(M_{ia})}$ only when the reliability of the pitch lag

predicted value is low, and otherwise set the predicted value to $\hat{T}_p^{(i)}$ (Step 5046 in Fig. 42). Further, pitch lag instruction

45

information indicating whether the predicted value is used, or the pitch lag $\hat{T}_p^{(M_{ia})}$ obtained by the side information is used, may be input into the adaptive codebook calculation unit 1123.

<When the side information index is a value indicating non-transmission of side information >

50 [0249] In the prediction of the pitch lag $\hat{T}_p^{(i)}$ ($M_{ia} \leq i < M$), the pitch lag prediction unit 192 predicts the pitch lag

$\hat{T}_p^{(i)}$ ($0 \leq i < M$) by using the pitch lag $\hat{T}_p^{(-j)}$ ($1 \leq j < J$) used in the past decoding (Step 5048 in Fig. 42).

55

[0250] Further, the pitch lag prediction unit 192 may establish $\hat{T}_p^{(i)} = \hat{T}_p^{(-1)}$ only when the reliability of the pitch lag predicted value is low (Step 5049 in Fig. 42), and the pitch lag prediction unit 192 can otherwise set the predicted value

to $\hat{T}_p^{(i)}$. Further, pitch lag instruction information indicating whether the predicted value is used, or the pitch lag $\hat{T}_p^{(-1)}$ used in past decoding is used, is input to the adaptive codebook calculation unit 1123 (Step 5050 in Fig. 42).

[0251] The adaptive codebook gain prediction unit 193 and the fixed codebook gain prediction unit 194 are the same as those of example 1.

[0252] The noise signal generation unit 195 is the same as that of the example 1.

[0253] The audio synthesis unit 124 synthesizes, from the parameters output from the audio parameter missing processing unit 123, an audio signal which corresponds to the frame to be encoded.

[0254] The LP coefficient calculation unit 1121 of the audio synthesis unit 124 (see Fig. 15) obtains an LP coefficient in the same manner as in example 1 (Step S11301 in Fig. 16).

[0255] The procedure of the adaptive codebook calculation unit 1123 is shown in the example of Fig. 43. The adaptive codebook calculation unit 1123 calculates an adaptive codebook vector in the same manner as in example 1. First, by referring to the pitch lag instruction information (Step 5051 in Fig. 43), when the reliability of the predicted value is low (YES in Step 5052 in Fig. 43), the adaptive codebook vector is calculated using the following equation (Step 5055 in Fig. 43). The filtering coefficient is f_i .

$$v(n)=f_{-1}v'(n-1)+f_0v'(n)+f_1v'(n+1) \quad \text{Equation 49}$$

Note that $v(n)=v'(n)$ may be established according to the design policy.

[0256] By referring to the pitch lag instruction information, when the reliability of the predicted value is high (NO in Step 5052 in Fig. 43), the adaptive codebook calculation unit 1123 calculates the adaptive codebook vector by the following procedure.

[0257] First, the initial adaptive codebook vector is calculated using the pitch lag and the adaptive codebook 1122 (Step 5053 in Fig. 43).

$$v(n)=f_{-1}v'(n-1)+f_0v'(n)+f_1v'(n+1) \quad \text{Equation 50}$$

$v(n)=v'(n)$ may be established according to the design policy.

[0258] Then, glottal pulse synchronization is applied to the initial adaptive codebook vector. For the glottal pulse synchronization, the same procedure can be used as in the case where a pulse position is not available in section 7.11.2.5 in Non Patent Literature 4 (Step 5054 in Fig. 43). Note that, however, $u(n)$ in Non Patent Literature 4 corresponds

to $v(n)$ in the embodiment of the present invention, extrapolated pitch corresponds to $\hat{T}_p^{(M-1)}$ in the embodiment of the present invention, and the last reliable pitch(T_c) corresponds to $\hat{T}_p^{(-1)}$ in the embodiment of the present invention.

[0259] The excitation vector synthesis unit 1124 outputs an excitation signal vector in the same manner as in the example 1 (Step 11306 in Fig. 16).

[0260] The post filter 1125 performs post processing on the synthesis signal in the same manner as in example 1.

[0261] The adaptive codebook 1122 updates the state using the excitation signal vector in the same manner as in the example 1 (Step 11308 in Fig. 16).

[0262] The synthesis filter 1126 synthesizes a decoded signal in the same manner as in example 1 (Step 11309 in Fig. 16).

[0263] The perceptual weighting inverse filter 1127 applies an perceptual weighting inverse filter in the same manner as in example 1.

[0264] The audio parameter missing processing unit 123 stores the audio parameters (ISF parameter, pitch lag, adaptive codebook gain, fixed codebook gain) used in the audio synthesis unit 124 into the buffer in the same manner as in example 1 (Step 145 in Fig. 7).

Reference Signs List

[0265] 60,80...storage medium, 61, 81...program storage area, 70...audio encoding program, 90...audio decoding program, 111...audio encoding unit, 112...side information encoding unit, 121, 231...audio code buffer, 122, 232...audio parameter decoding unit, 123, 233...audio parameter missing processing unit, 124, 234...audio synthesis unit, 125, 235...side information decoding unit, 126, 236...side information accumulation unit, 151, 511, 1121...LP coefficient calculation unit, 152, 2012...target signal calculation unit, 153, 513, 2013...pitch lag calculation unit, 154, 1123, 514, 2014, 2313...adaptive codebook calculation unit, 155, 1124, 2314...excitation vector synthesis unit, 156, 315, 515, 2019...adaptive codebook buffer, 157, 1126, 2018, 2316...synthesis filter, 158, 516...pitch lag encoding unit, 191...ISF prediction unit, 192...pitch lag prediction unit, 193...adaptive codebook gain prediction unit, 194...fixed codebook gain prediction

unit, 195...noise signal generation unit, 211...main encoding unit, 212...side information encoding unit, 213, 238...concealment signal accumulation unit, 214...error signal encoding unit, 237...error signal decoding unit, 311...LP coefficient calculation unit, 312...pitch lag prediction unit, 313...pitch lag selection unit, 314...pitch lag encoding unit, 512...residual signal calculation unit, 700...audio encoding module, 701...side information encoding module, 900...audio parameter decoding module, 901...audio parameter missing processing module, 902...audio synthesis module, 903...side information decoding module, 1128...side information output determination unit, 1122, 2312...adaptive codebook, 1125...post filter, 1127...perceptual weighting inverse filter, 2011...ISF encoding unit, 2015...fixed codebook calculation unit, 2016...gain calculation unit, 2017...excitation vector calculation unit, 2211...ISF decoding unit, 2212...pitch lag decoding unit, 2213...gain decoding unit, 2214...fixed codebook decoding unit, 2318...look-ahead excitation vector synthesis unit

10

Claims

1. An audio encoding device for encoding an audio signal, comprising:
 - an audio encoding unit configured to encode an audio signal; and
 - a side information encoding unit configured to calculate side information from a look-ahead signal and encode the side information.
2. The audio encoding device according to Claim 1, wherein the side information is related to a pitch lag in a look-ahead signal.
3. The audio encoding device according to Claim 1, wherein the side information is related to a pitch gain in a look-ahead signal.
4. The audio encoding device according to Claim 1, wherein the side information is related to a pitch lag and a pitch gain in a look-ahead signal.
5. The audio encoding device according to any one of Claims 1 to 4, wherein the side information contains information related to availability of side information.
6. The audio encoding device according to any one of Claims 1 to 5, wherein the side information encoding unit calculates side information for a look-ahead signal part and encodes the side information, and further generates a concealment signal, and the audio encoding device further comprises:
 - an error signal encoding unit configured to encode an error signal between an input audio signal and a concealment signal output from the side information encoding unit; and
 - a main encoding unit configured to encode the input audio signal.
7. An audio decoding device for decoding an audio code and outputting an audio signal, comprising:
 - an audio code buffer configured to detect packet loss based on a received state of an audio packet;
 - an audio parameter decoding unit configured to decode an audio code when an audio packet is correctly received;
 - a side information decoding unit configured to decode a side information code when an audio packet is correctly received;
 - a side information accumulation unit configured to accumulate side information obtained by decoding the side information code;
 - an audio parameter missing processing unit configured to output an audio parameter when audio packet loss is detected; and
 - an audio synthesis unit configured to synthesize a decoded audio from the audio parameter.
8. The audio decoding device according to Claim 7, wherein the side information is related to a pitch lag in a look-ahead signal.
9. The audio decoding device according to Claim 7, wherein the side information is related to a pitch gain in a look-ahead signal.

10. The audio decoding device according to Claim 7, wherein
the side information is related to a pitch lag and a pitch gain in a look-ahead signal.
- 5 11. The audio decoding device according to any one of Claims 7 to 10, wherein
the side information contains information related to availability of side information.
12. The audio decoding device according to any one of Claims 7 to 11, wherein
the side information decoding unit decodes the side information code and outputs the side information, and further
outputs a concealment signal related to a look-ahead part by using the side information,
10 the audio decoding device further comprising:

an error decoding unit configured to decode a code related to an error signal between the audio signal and a
concealment signal;
a main decoding unit configured to decode a code related to the audio signal; and
15 a concealment signal accumulation unit configured to accumulate a concealment signal output from the side
information decoding unit.
13. The audio decoding device according to Claim 12, wherein
when the audio packet is correctly received, a part of a decoded signal is generated by adding the concealment
signal read from the concealment signal accumulation unit and a decoded error signal output from the error decoding
unit, and the concealment signal accumulation unit is updated with the concealment signal output from the side
information decoding unit.
20
14. The audio decoding device according to Claim 12 or 13, wherein
when audio packet loss is detected, a concealment signal read from the concealment signal accumulation unit is
used as a part, or whole, of a decoded signal.
25
15. The audio decoding device according to any one of Claims 12 to 14, wherein
when audio packet loss is detected, a decoded signal is generated by using an audio parameter predicted by the
audio parameter missing processing unit, and the concealment signal accumulation unit is updated by using a part
30 of the decoded signal.
16. The audio decoding device according to any one of Claims 7 to 11, wherein
when audio packet loss is detected, the audio parameter missing processing unit uses side information read from
the side information accumulation unit as a part of a predicted value of an audio parameter.
35
17. The audio decoding device according to any one of Claims 7 to 16, wherein
when audio packet loss is detected, the audio synthesis unit corrects an adaptive codebook vector by using side
information read from the side information accumulation unit, the audio parameter comprising the adaptive codebook
vector.
40
18. An audio encoding method by an audio encoding device for encoding an audio signal, comprising:

an audio encoding step of encoding an audio signal; and
45 a side information encoding step of calculating side information from a look-ahead signal and encoding the side
information.
19. An audio decoding method by an audio decoding device for decoding an audio code and outputting an audio signal,
comprising:
50

an audio code buffer step of detecting packet loss based on a received state of an audio packet;
an audio parameter decoding step of decoding an audio code when an audio packet is correctly received;
a side information decoding step of decoding a side information code when an audio packet is correctly received;
a side information accumulation step of accumulating side information obtained by decoding a side information
code;
55 an audio parameter missing processing step of outputting an audio parameter when audio packet loss is detected;
and
an audio synthesis step of synthesizing a decoded audio from an audio parameter.

20. An audio encoding program causing a computer to function as:

5 an audio encoding unit to encode an audio signal; and
a side information encoding unit to calculate side information from a look-ahead signal and encode the side information.

21. An audio decoding program causing a computer to function as:

10 an audio code buffer to detect packet loss based on a received state of an audio packet;
an audio parameter decoding unit to decode an audio code when an audio packet is correctly received;
a side information decoding unit to decode a side information code when an audio packet is correctly received;
a side information accumulation unit to accumulate side information obtained by decoding a side information code;
15 an audio parameter missing processing unit to output an audio parameter when audio packet loss is detected; and
an audio synthesis unit to synthesize a decoded audio from an audio parameter.

20

25

30

35

40

45

50

55

Fig.1

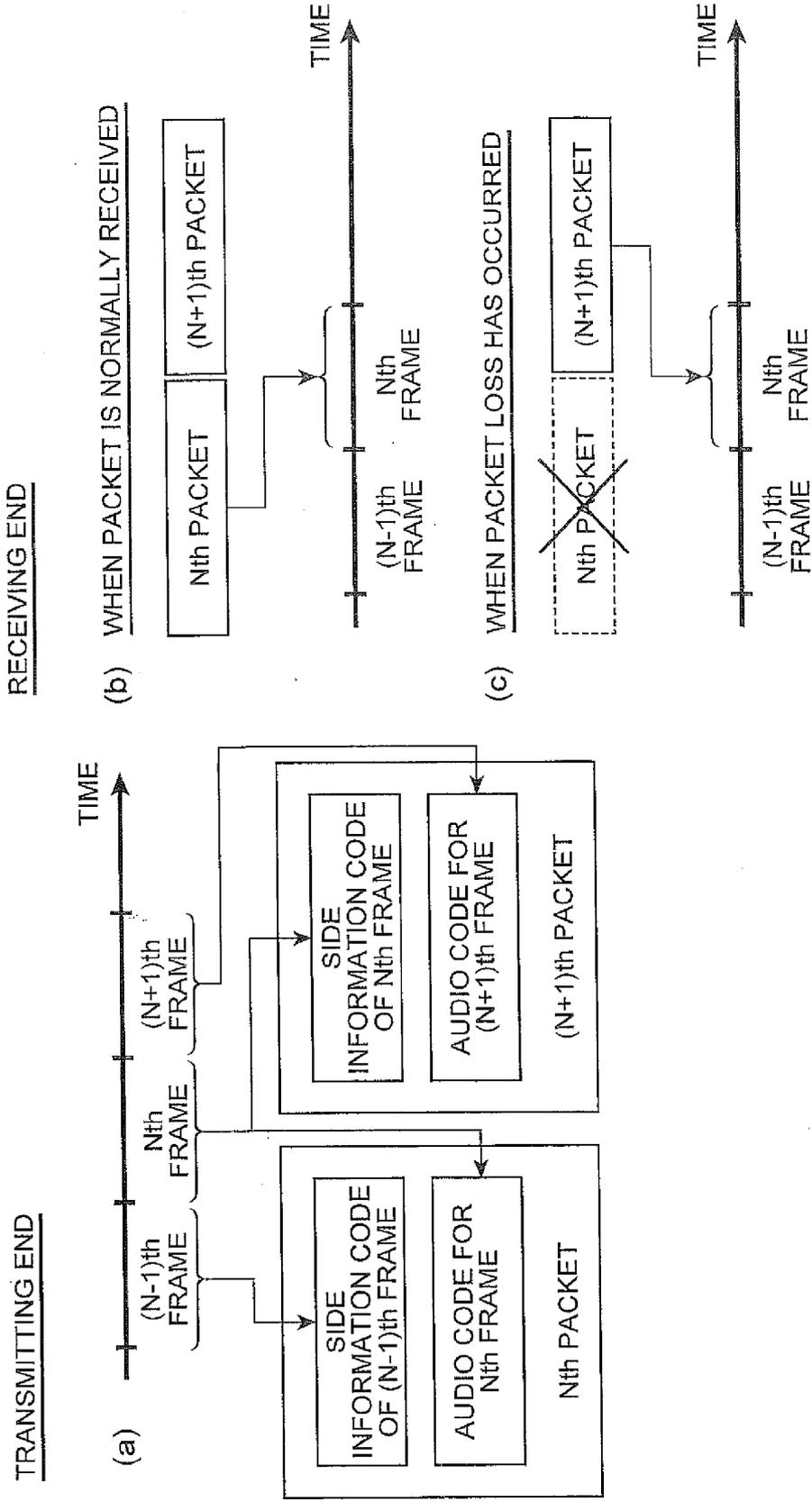


Fig.2

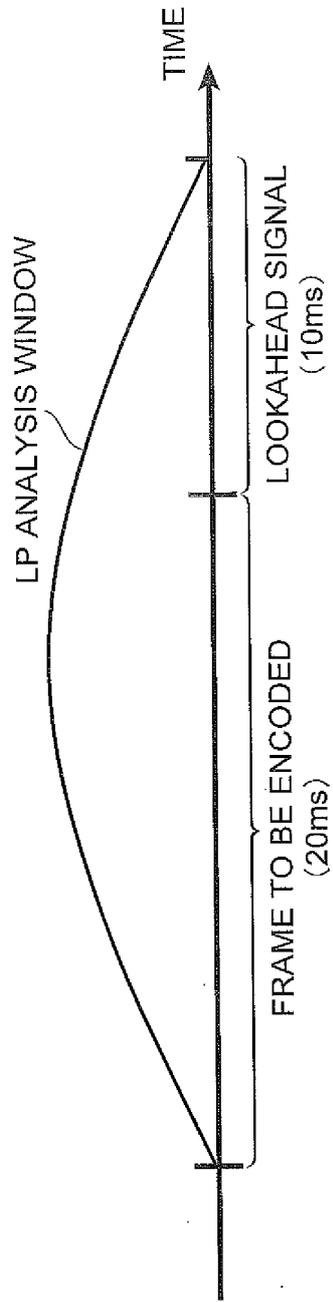


Fig.3

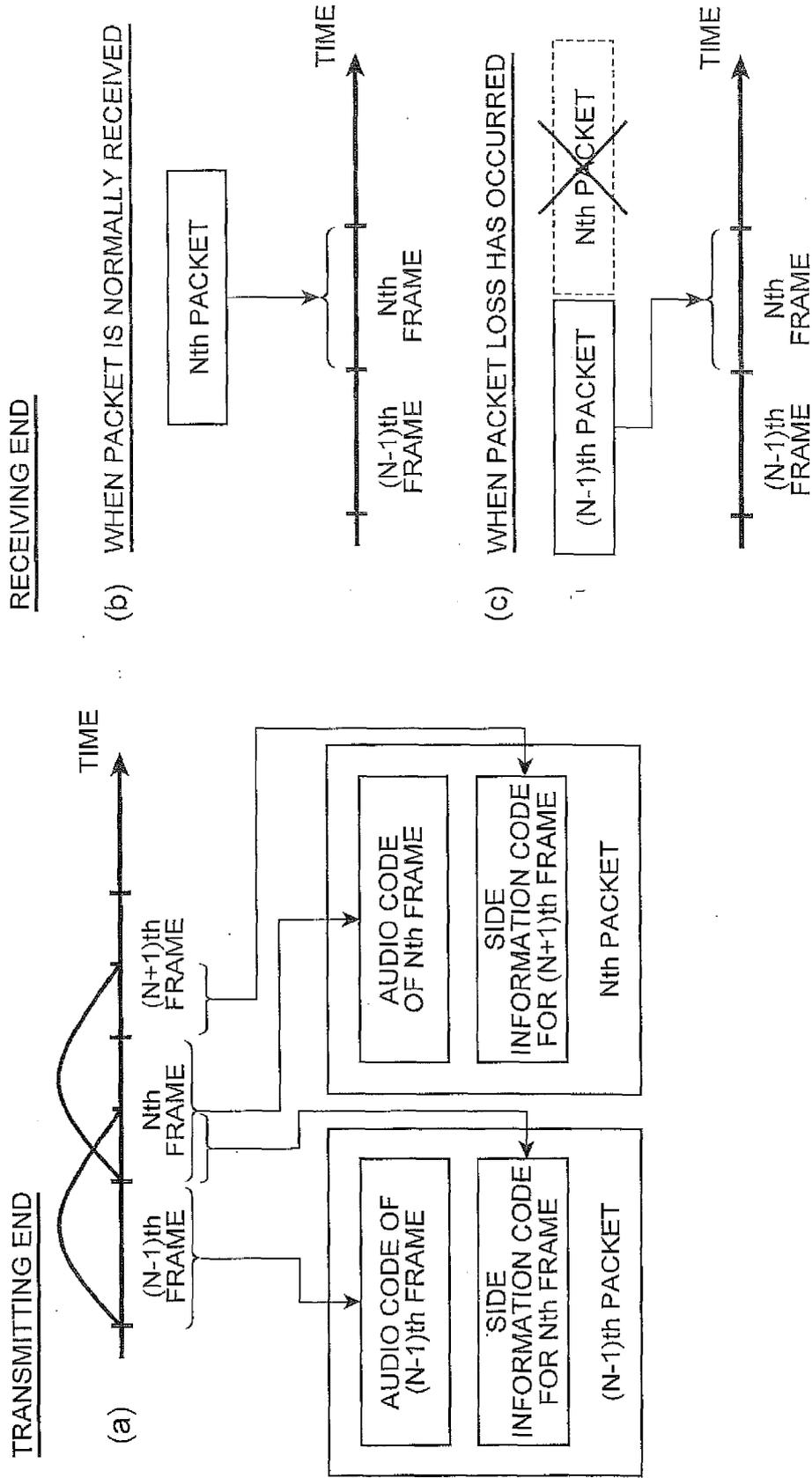


Fig.4

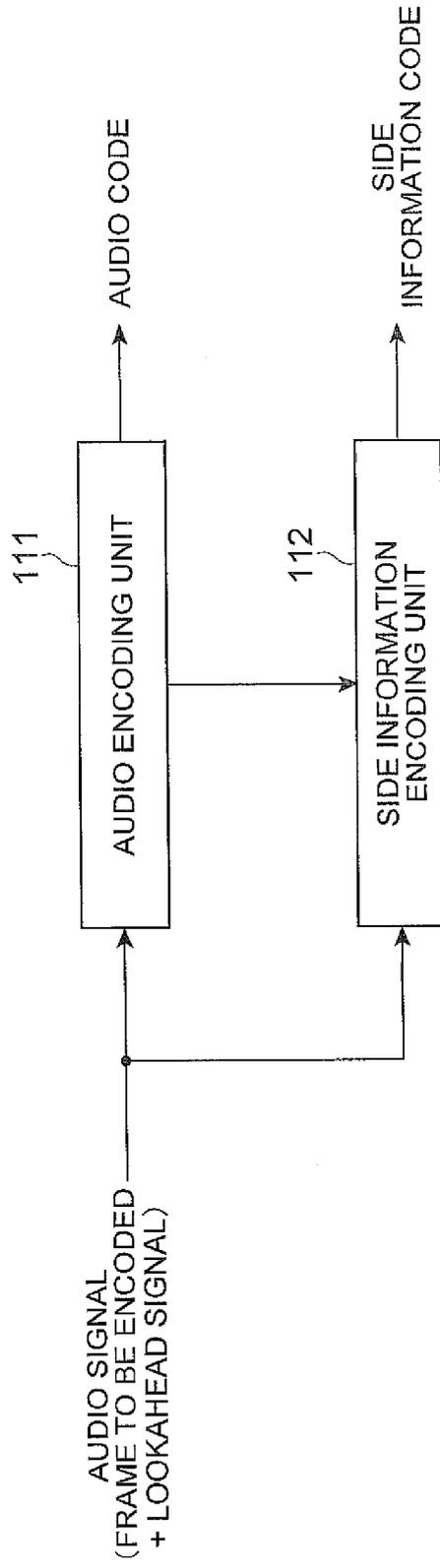


Fig.5

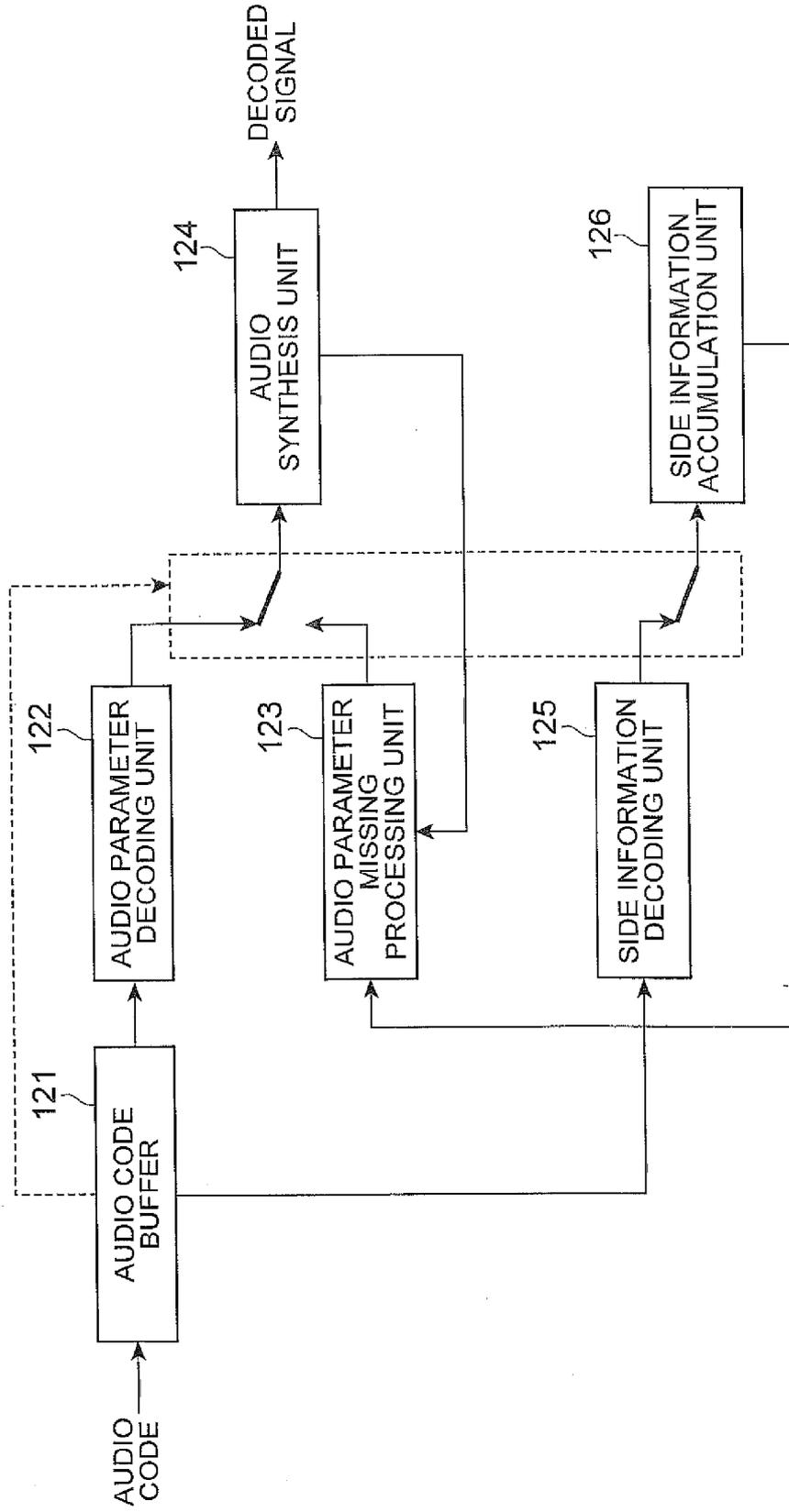


Fig.6

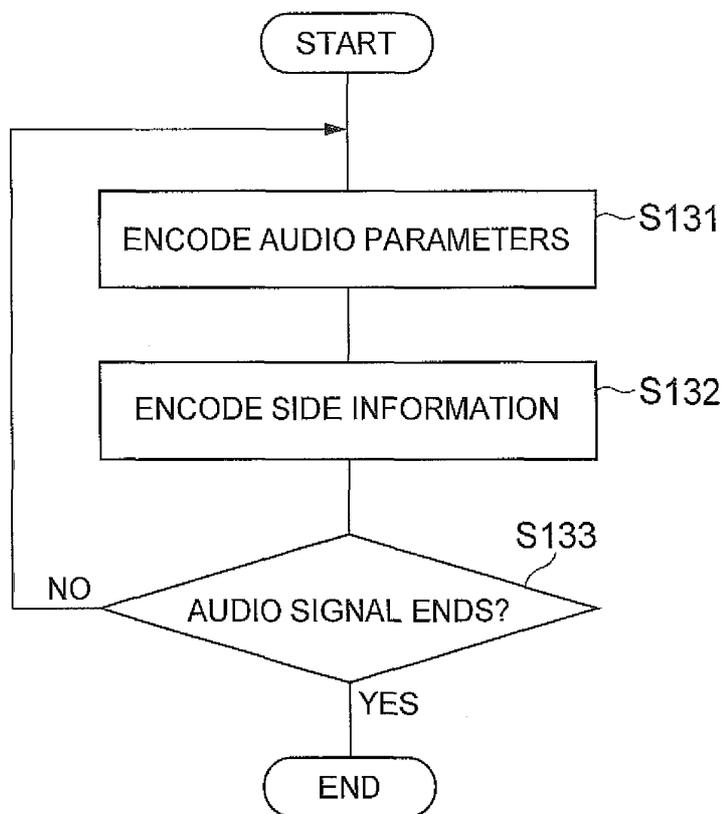


Fig.7

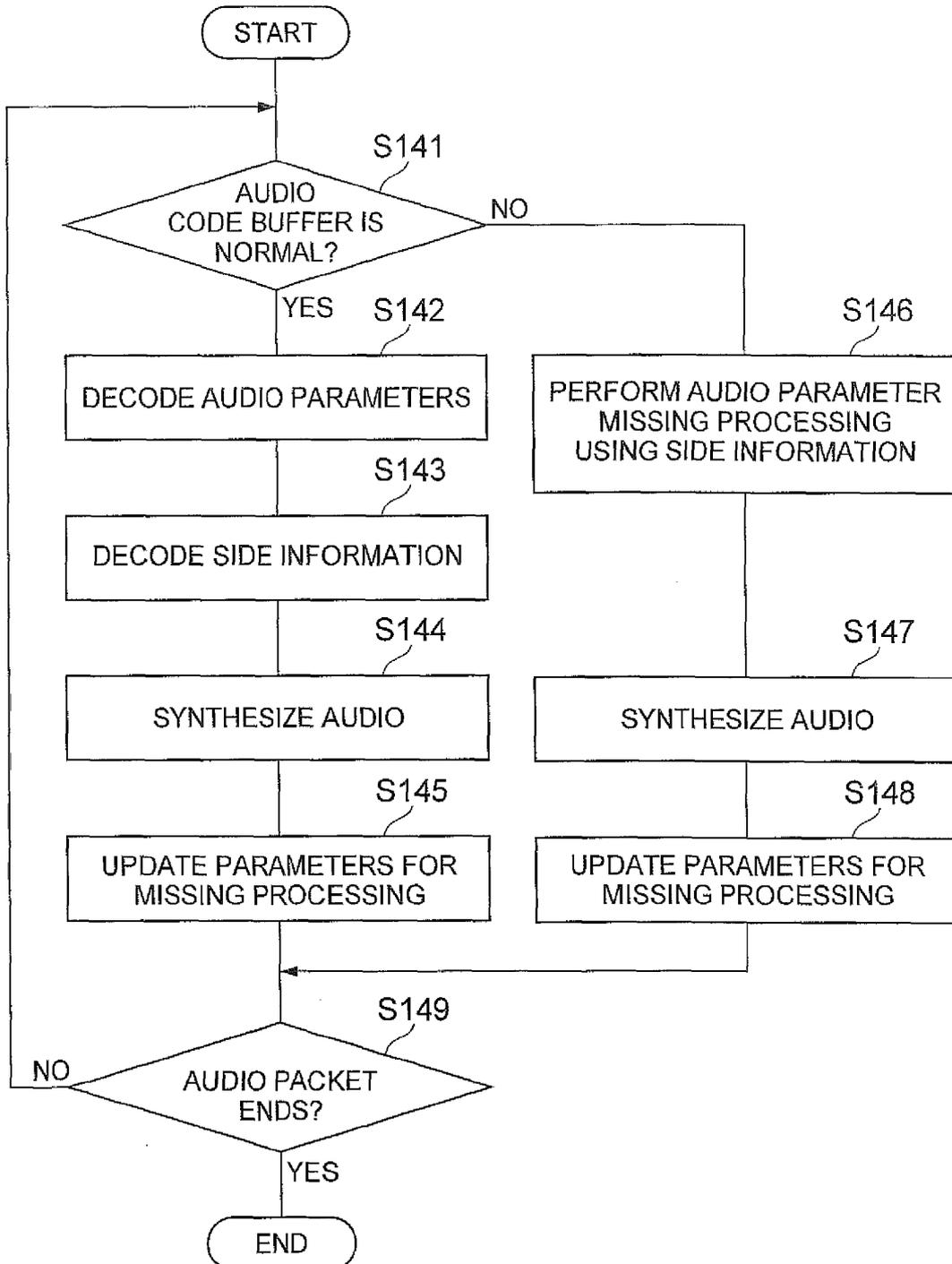


Fig.8

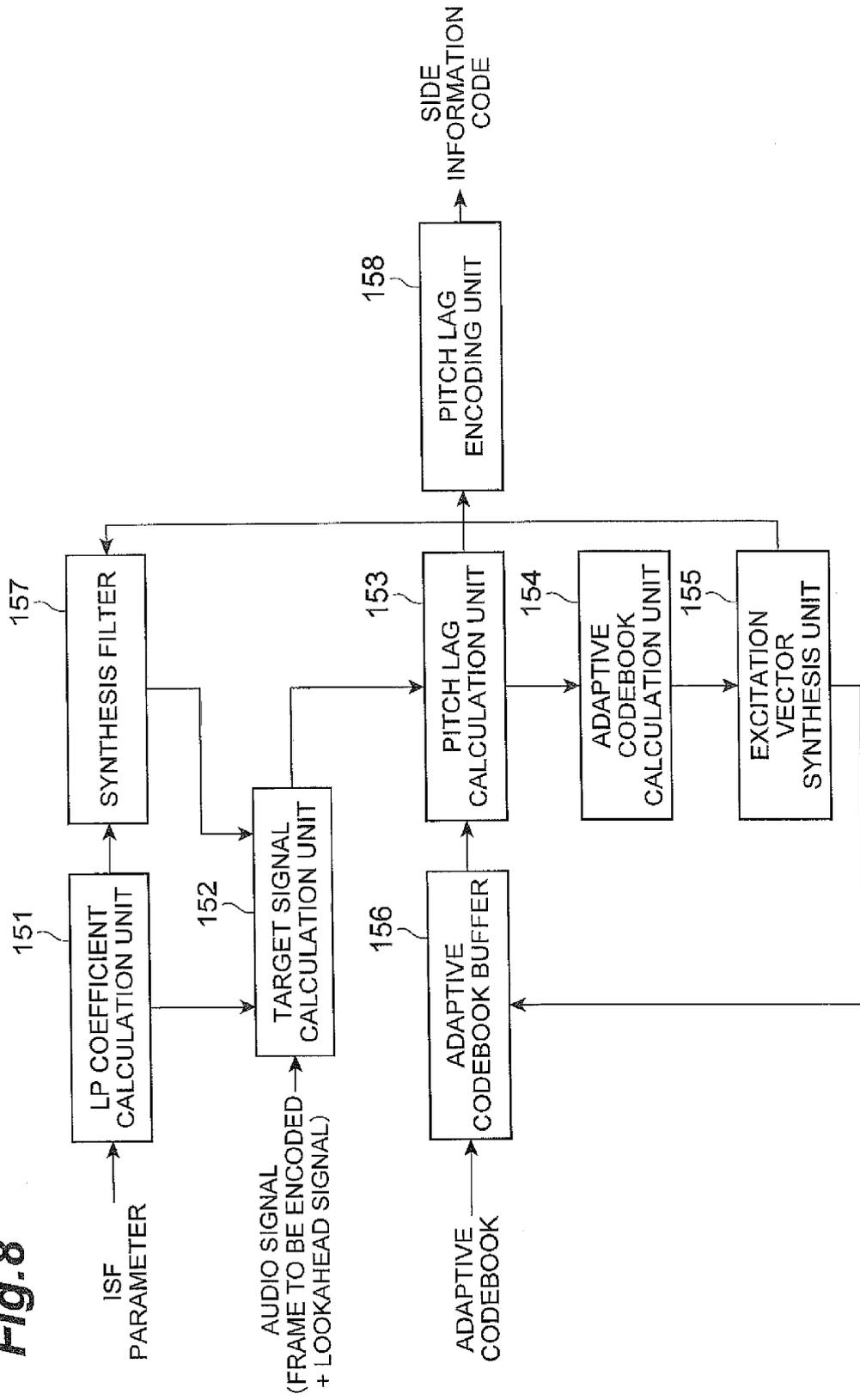


Fig.9

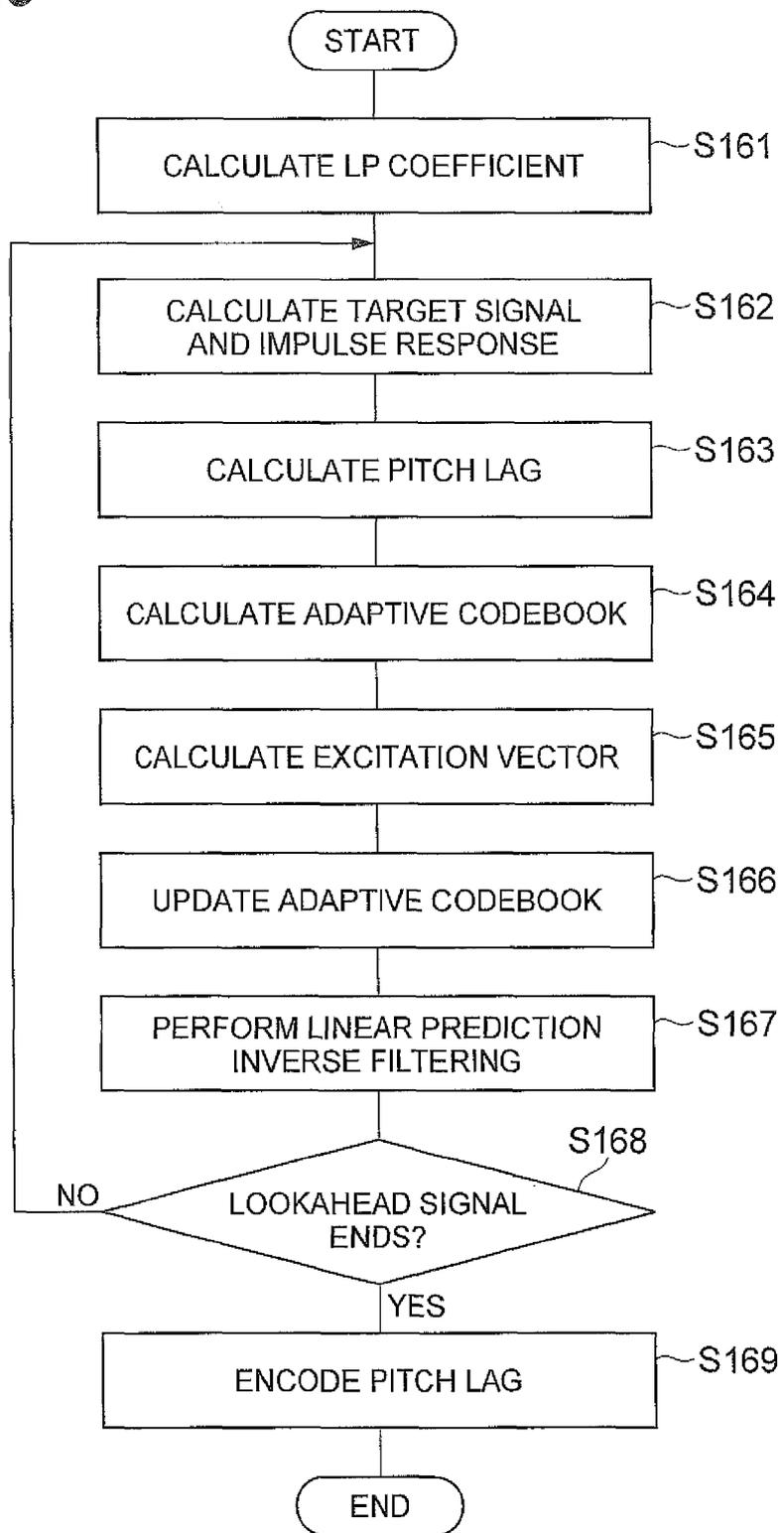


Fig.10

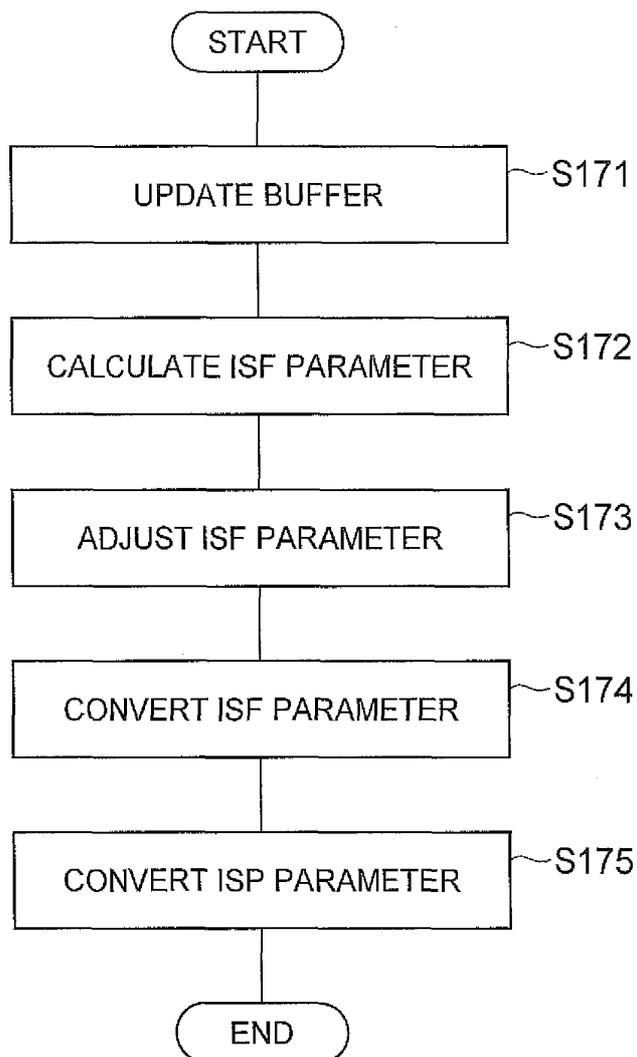


Fig.11

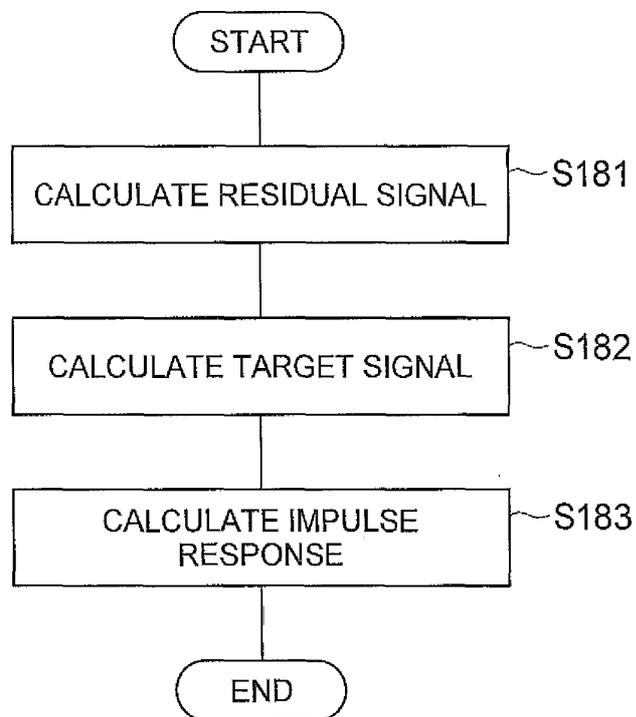


Fig. 12

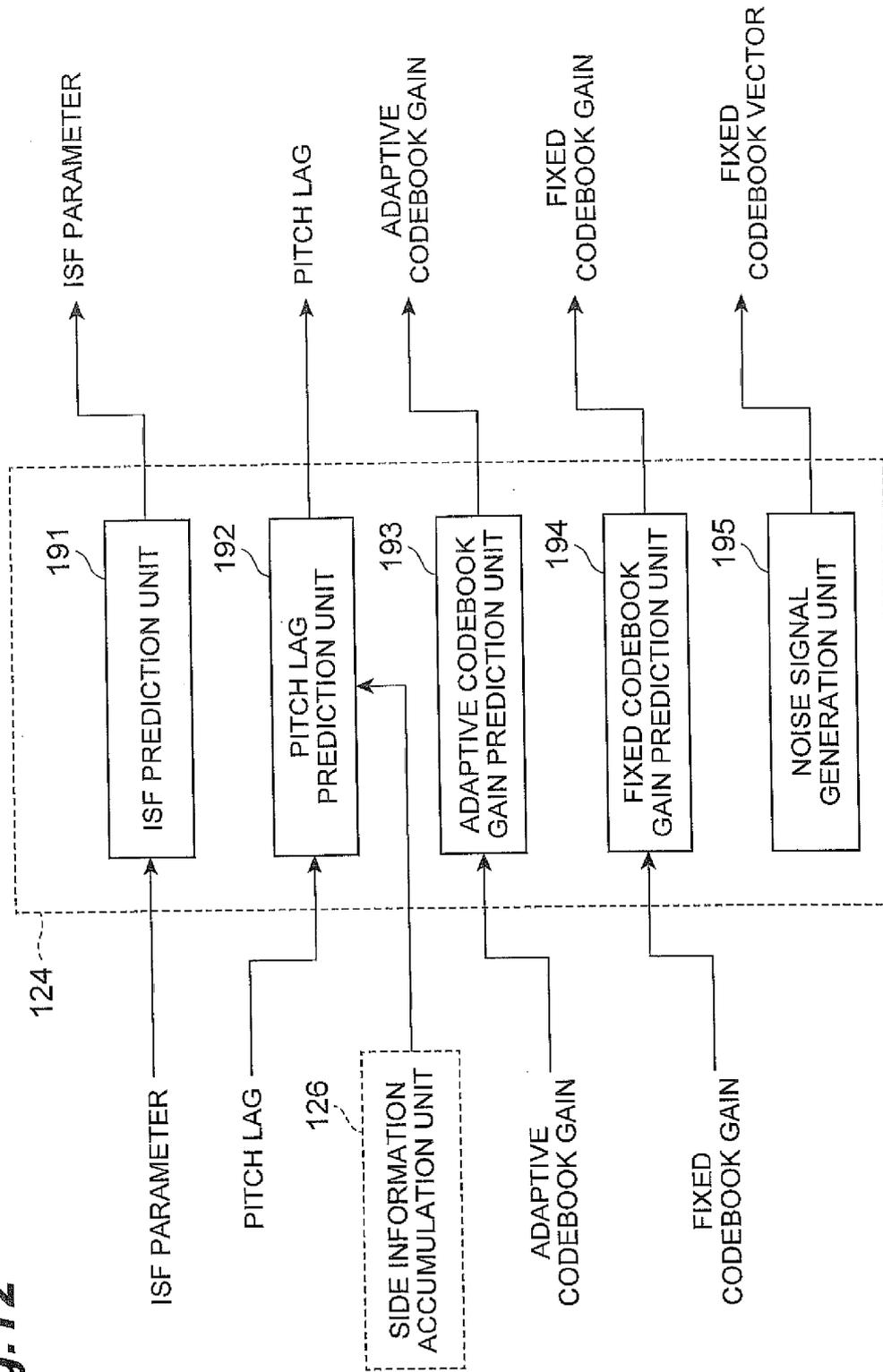


Fig.13

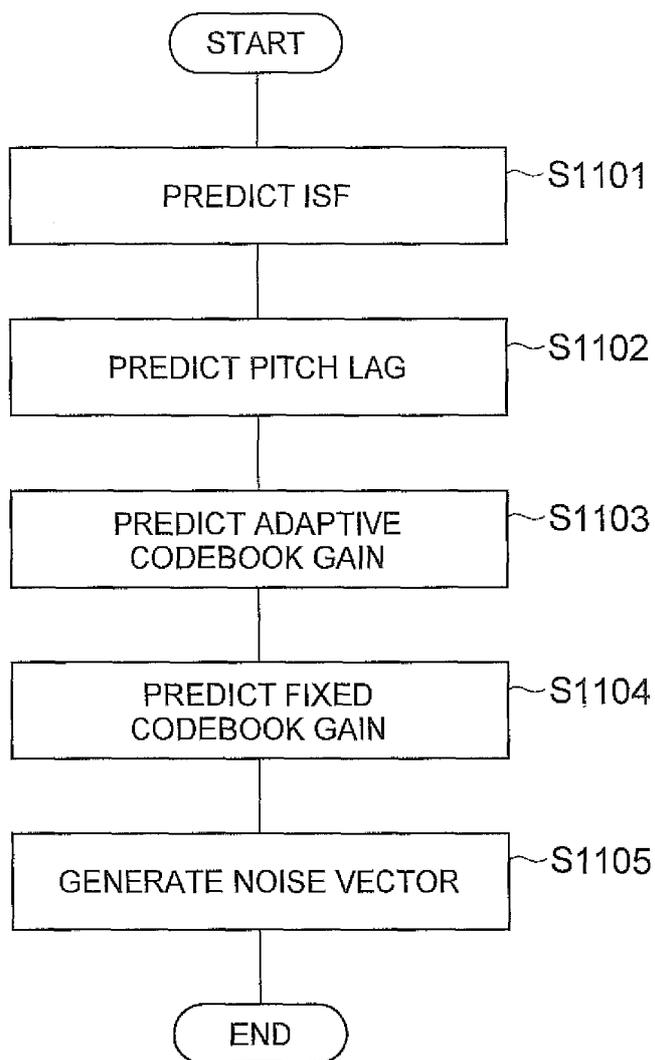


Fig.14

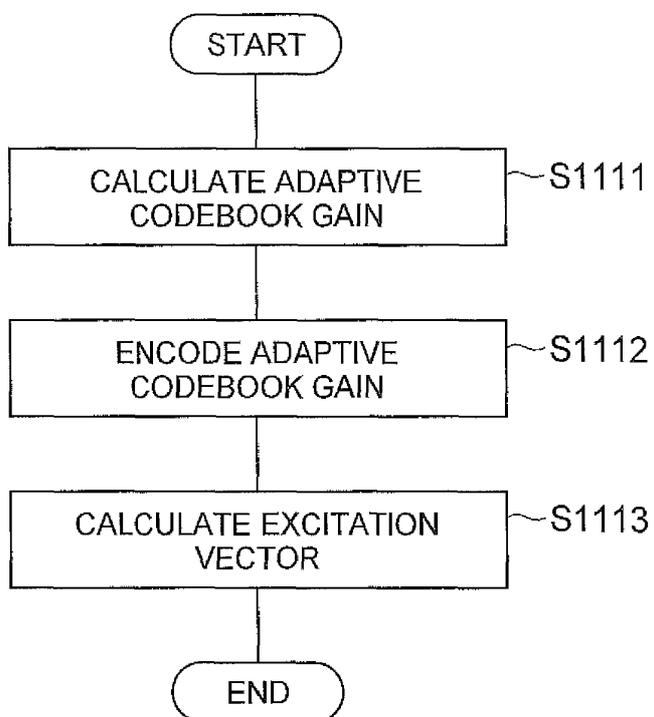


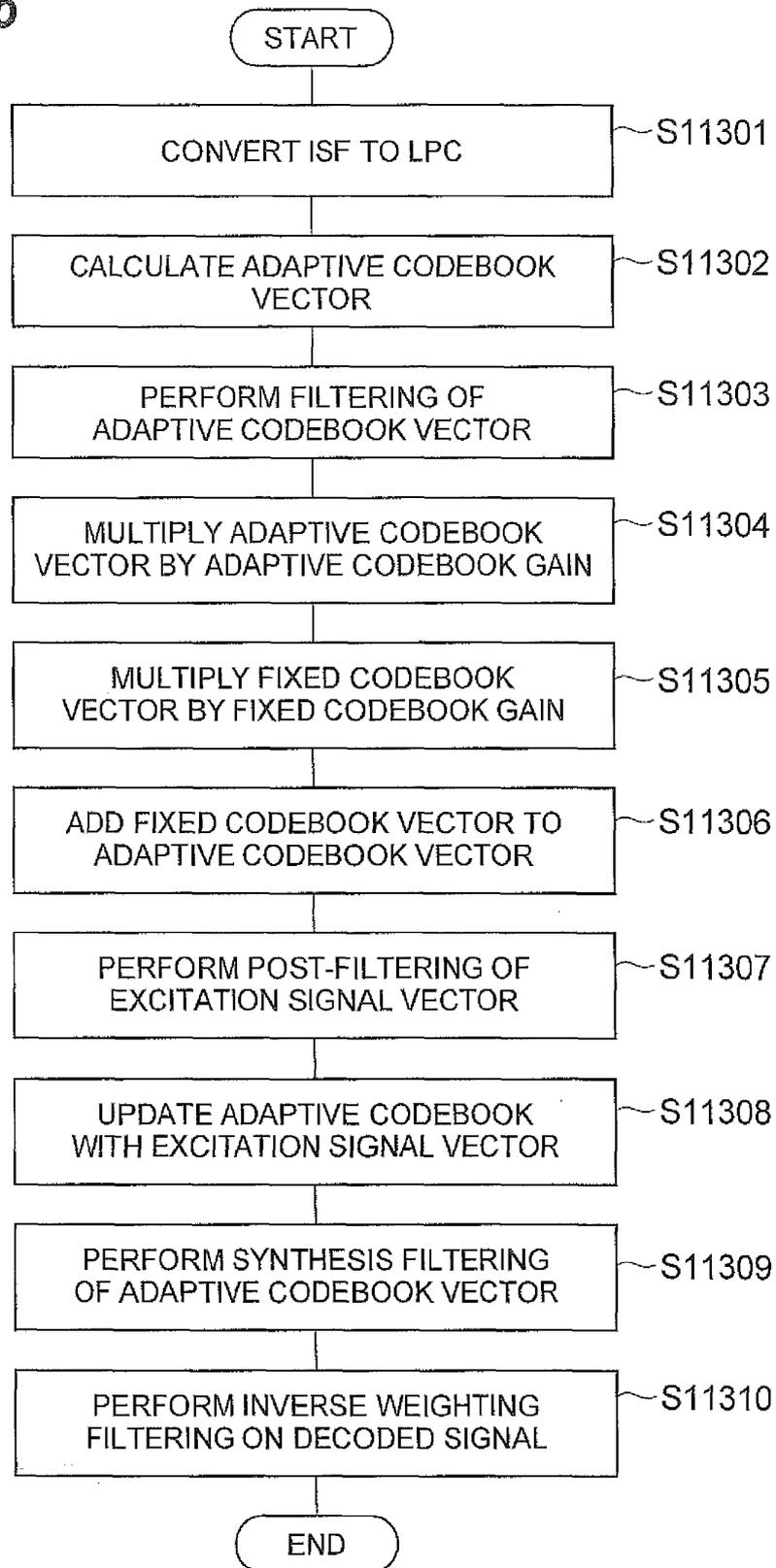
Fig.16

Fig. 17

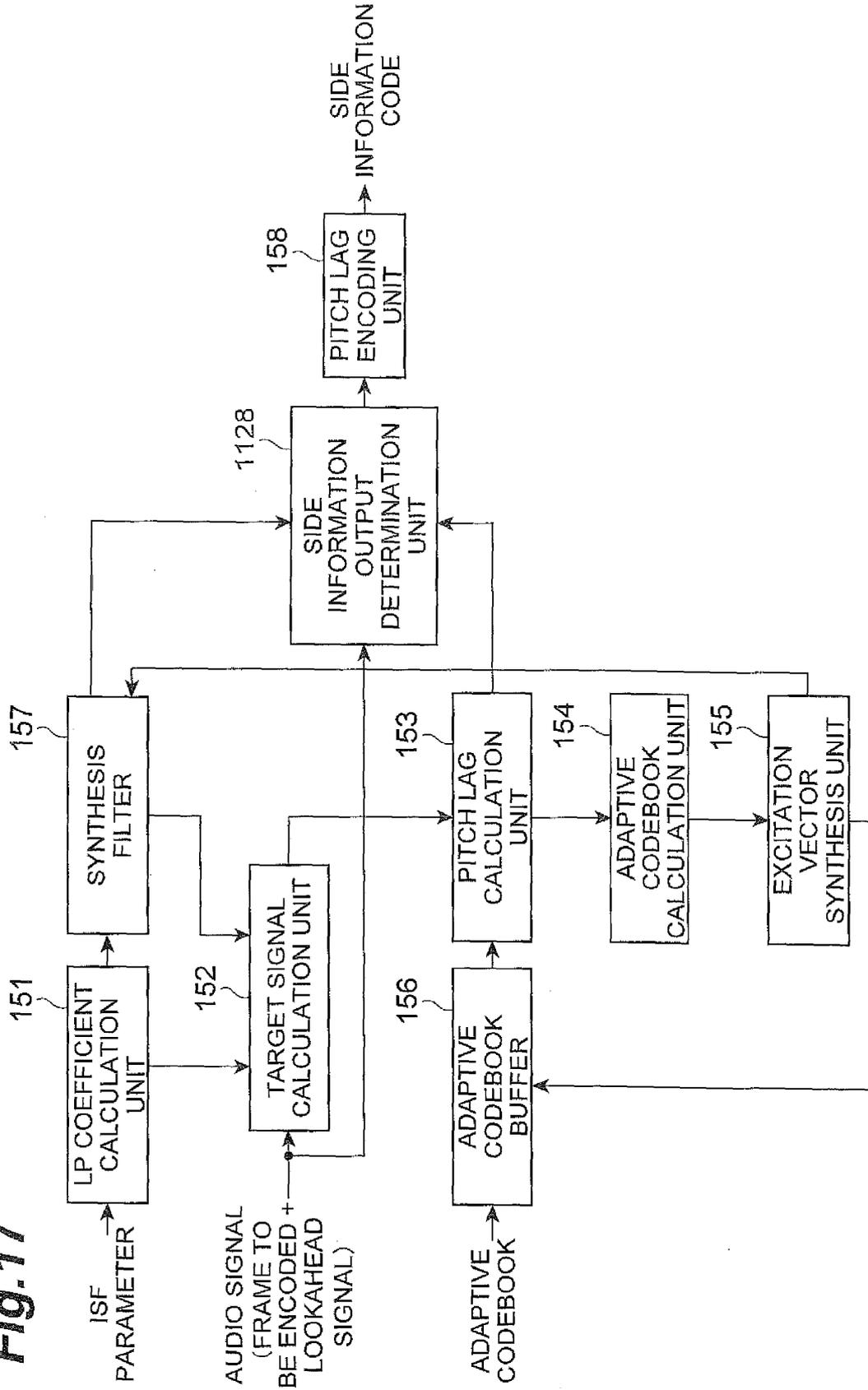


Fig.18

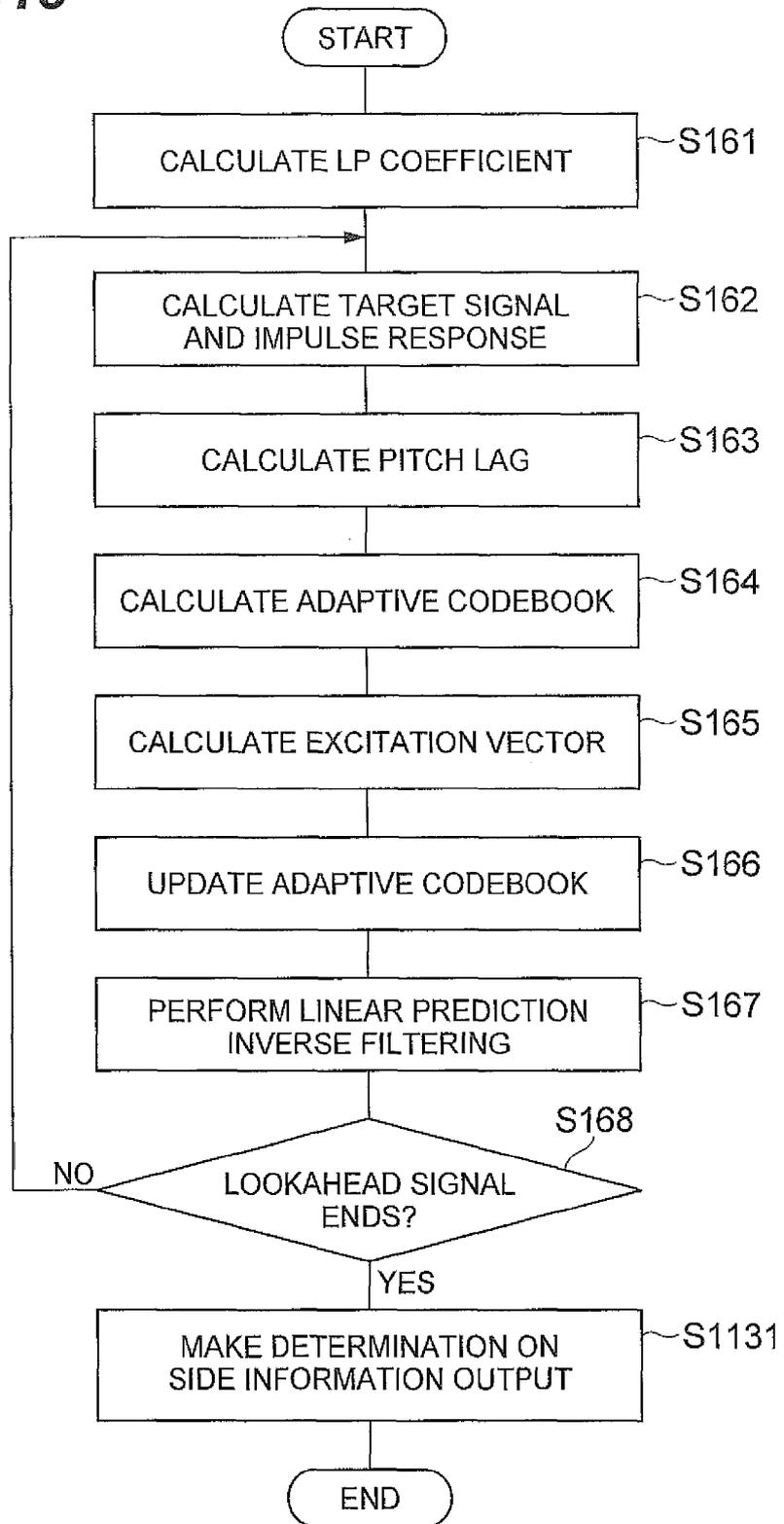


Fig.19

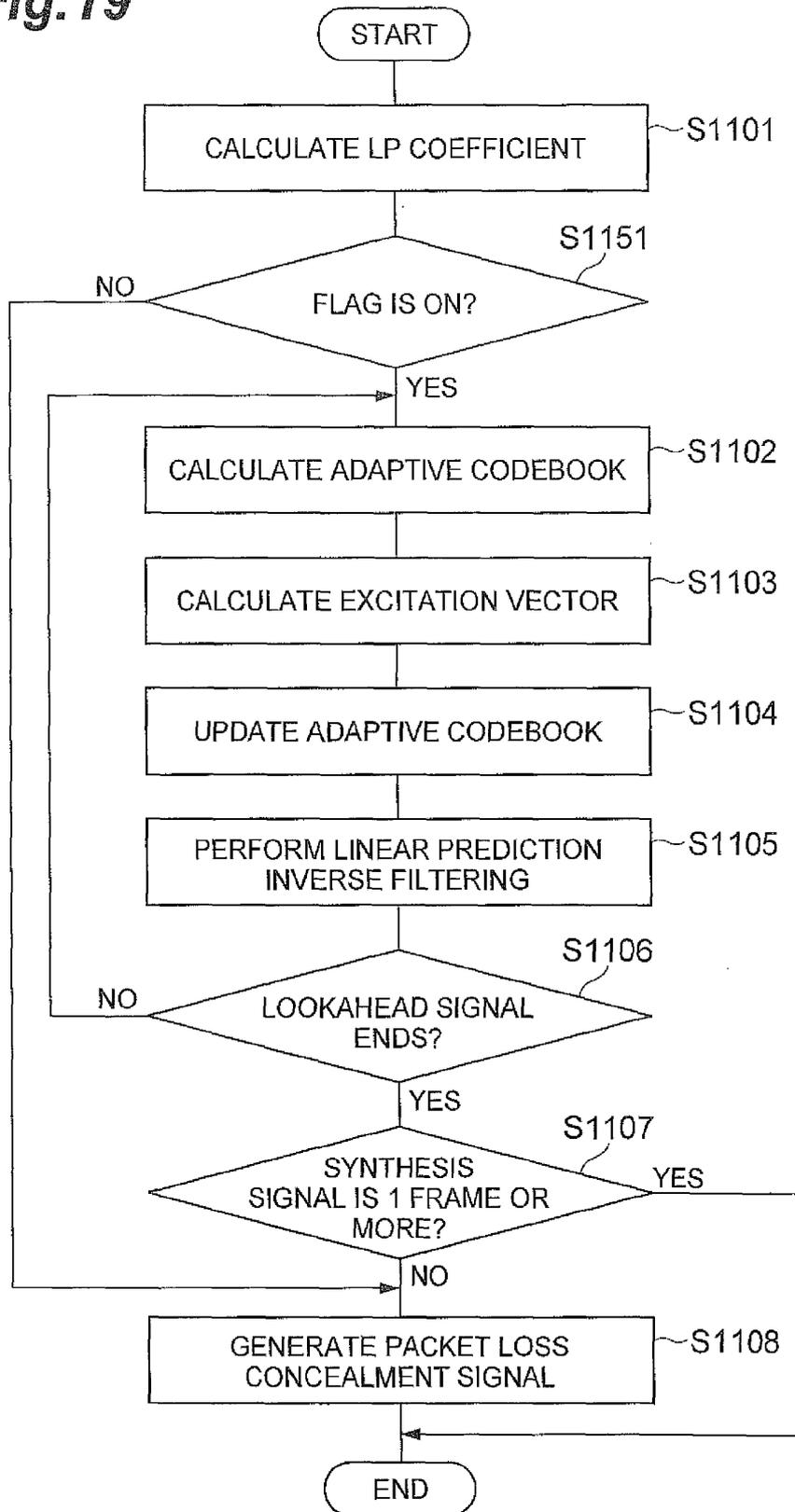


Fig.20

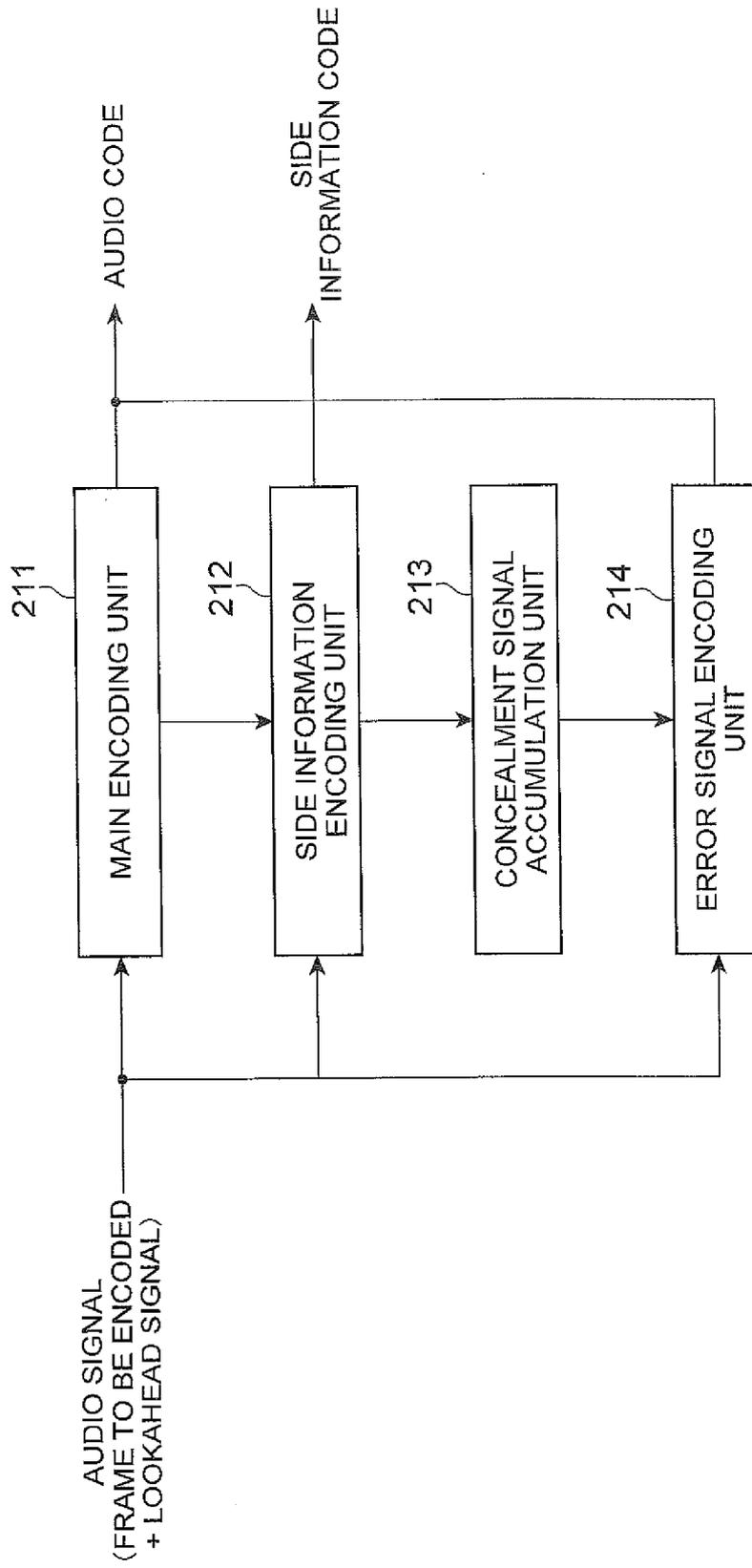


Fig. 21

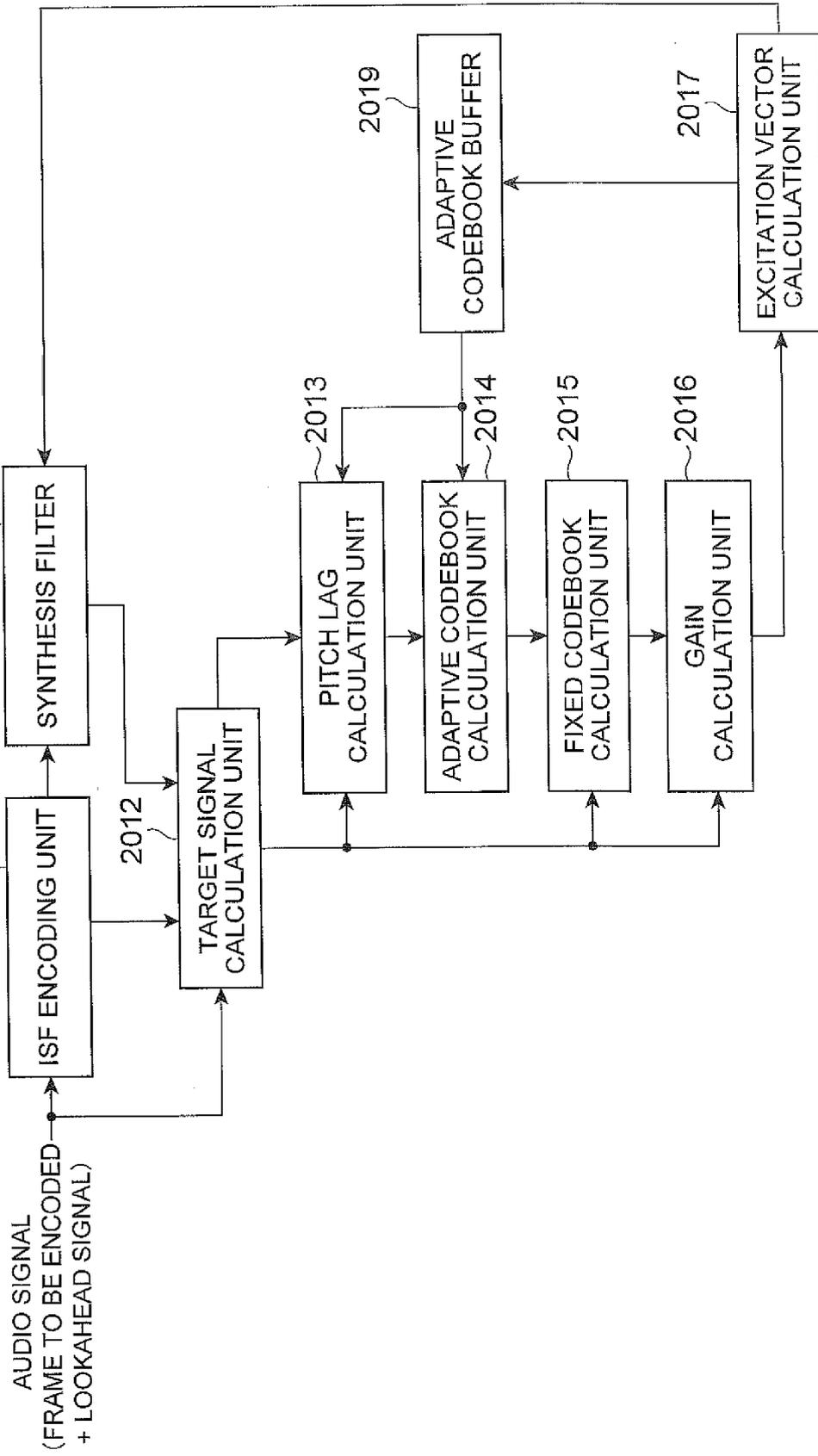


Fig.22

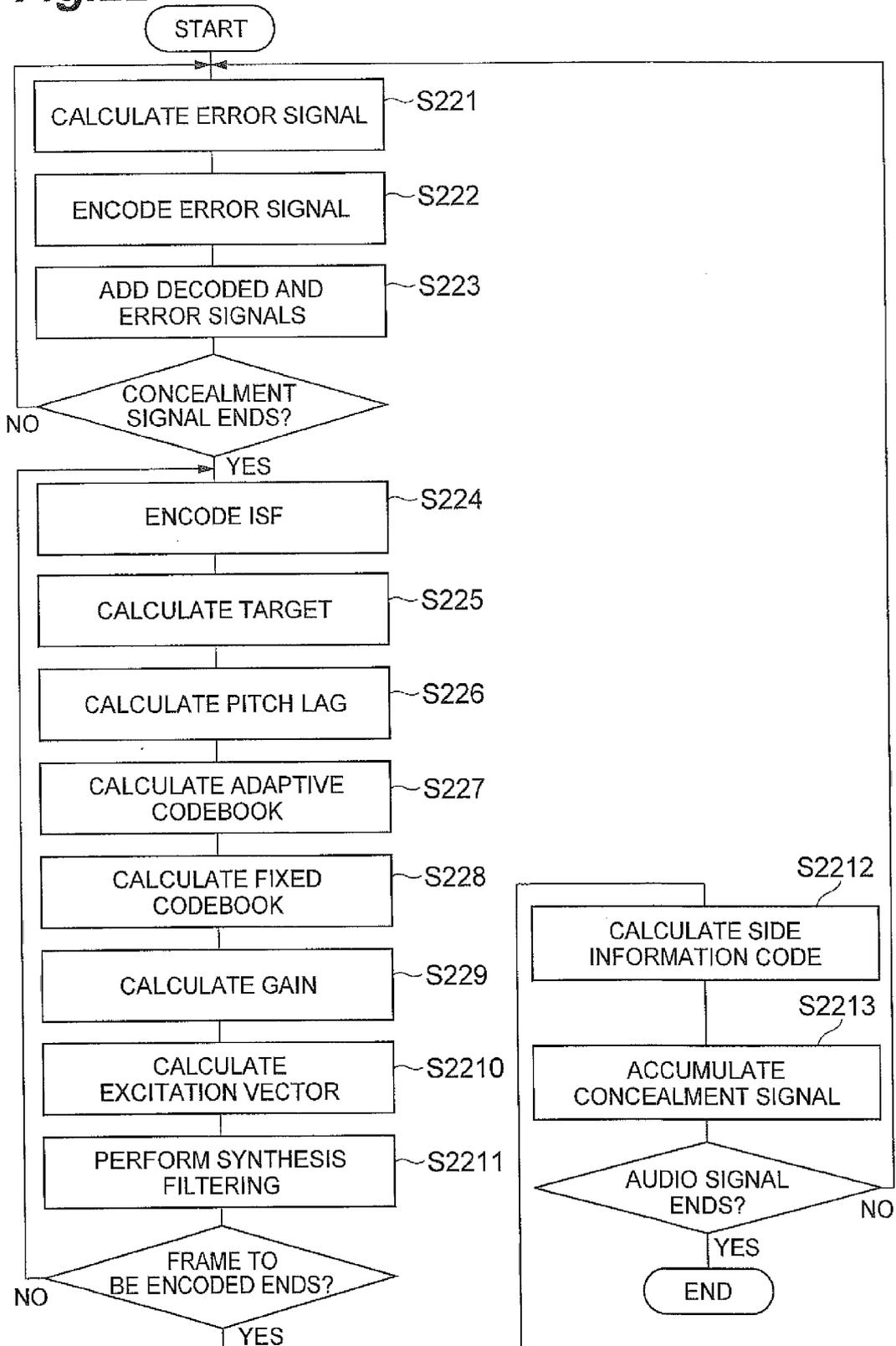
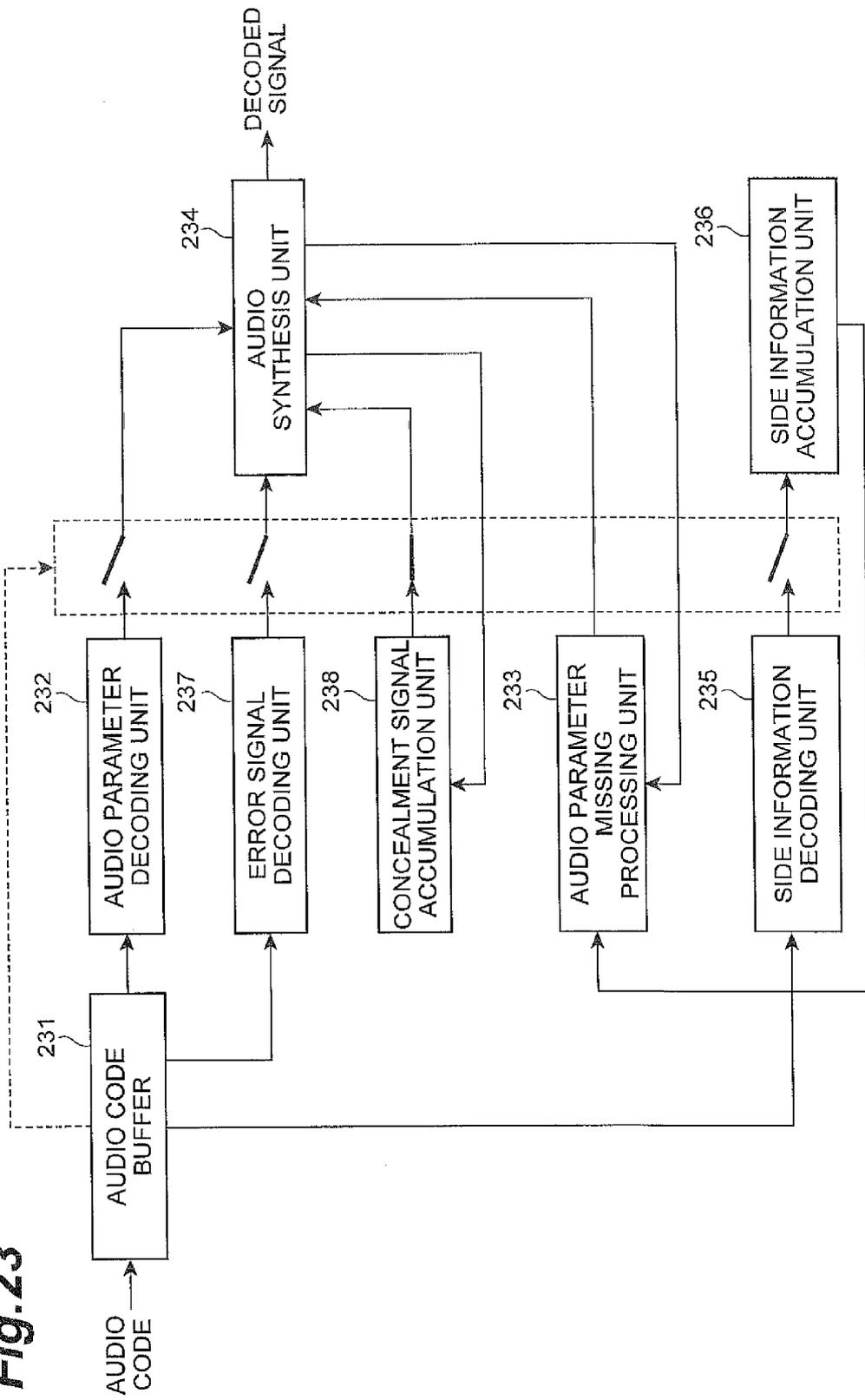


Fig. 23



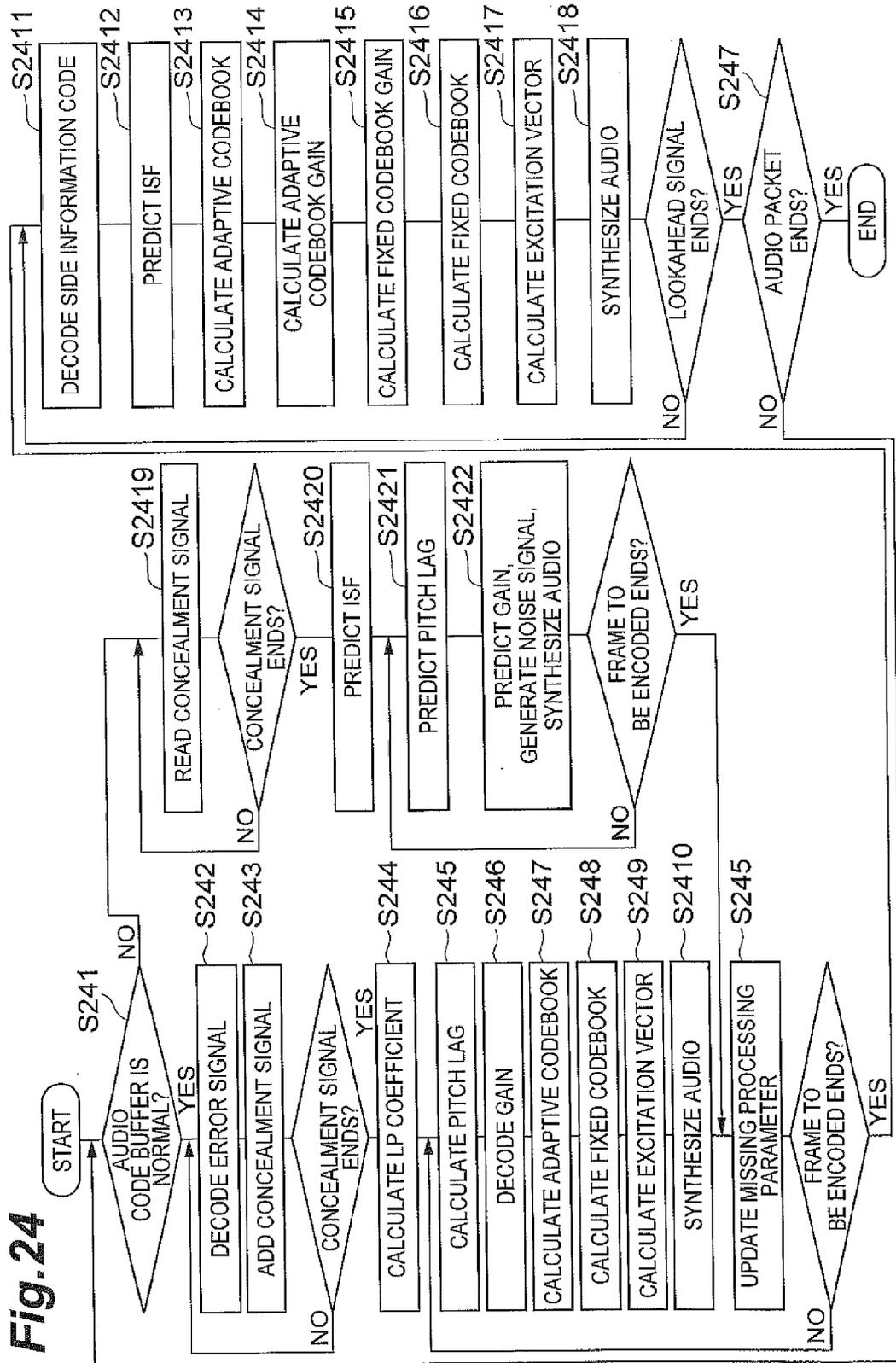


Fig. 24

Fig. 25

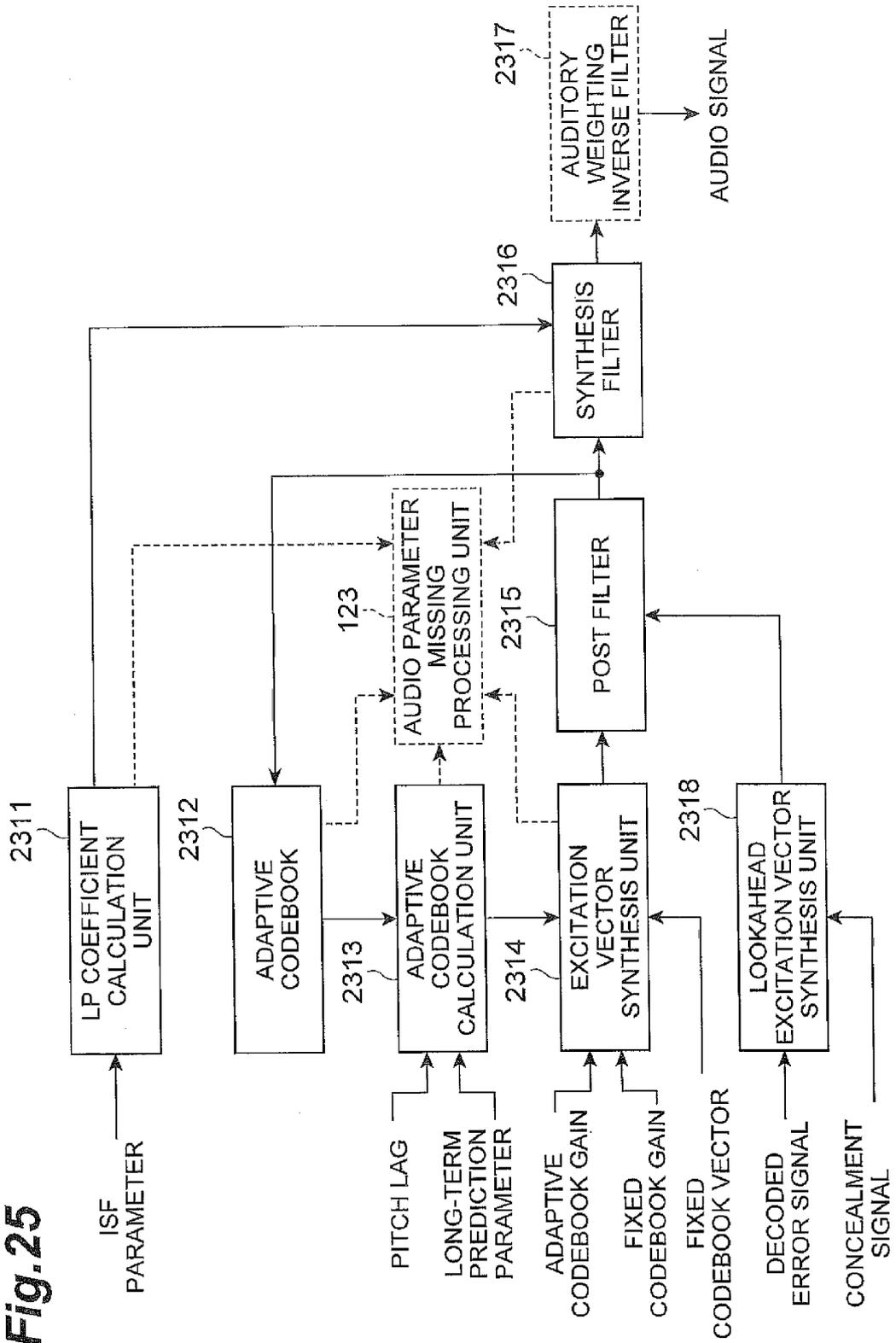


Fig. 26

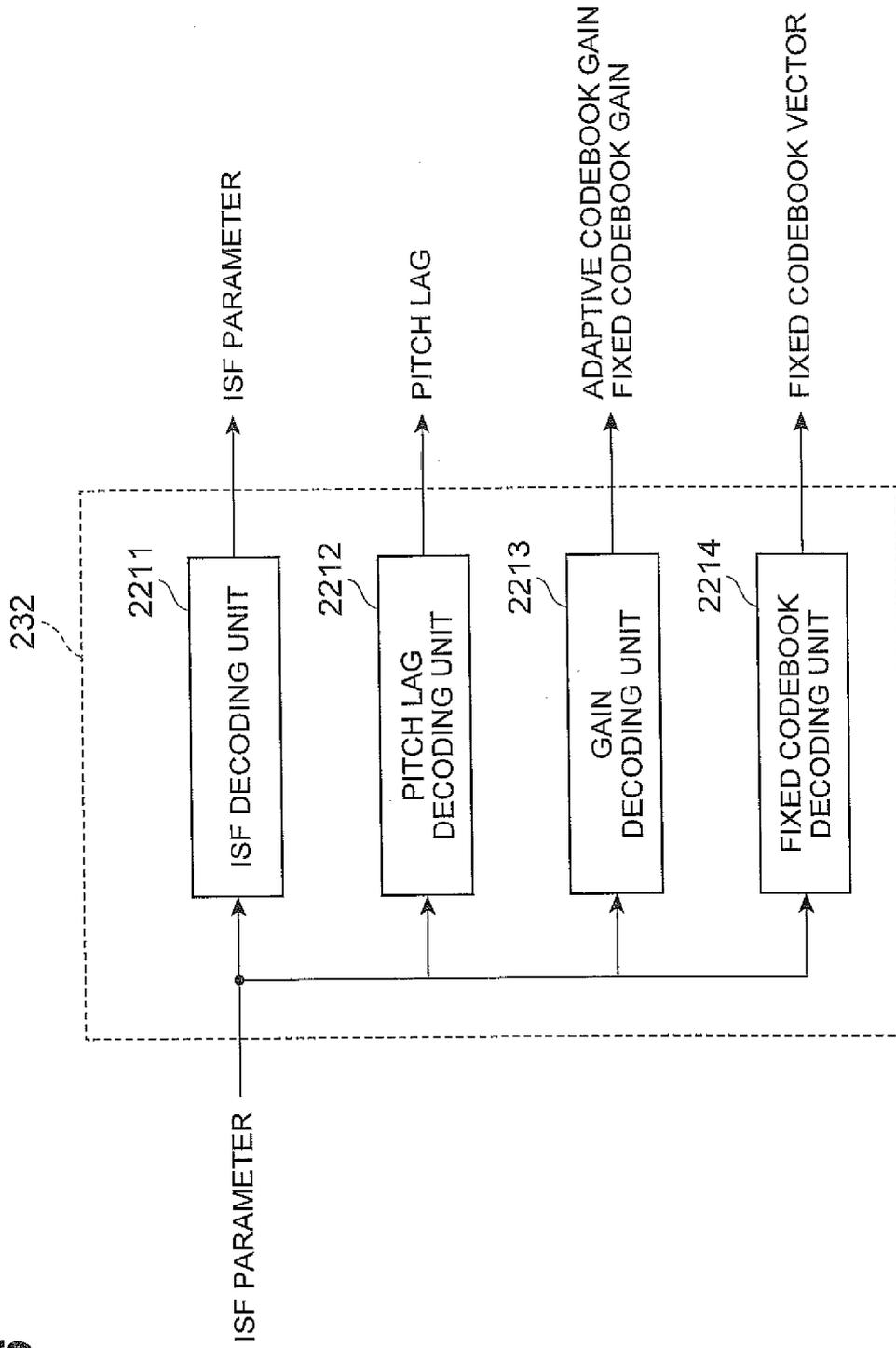


Fig.27

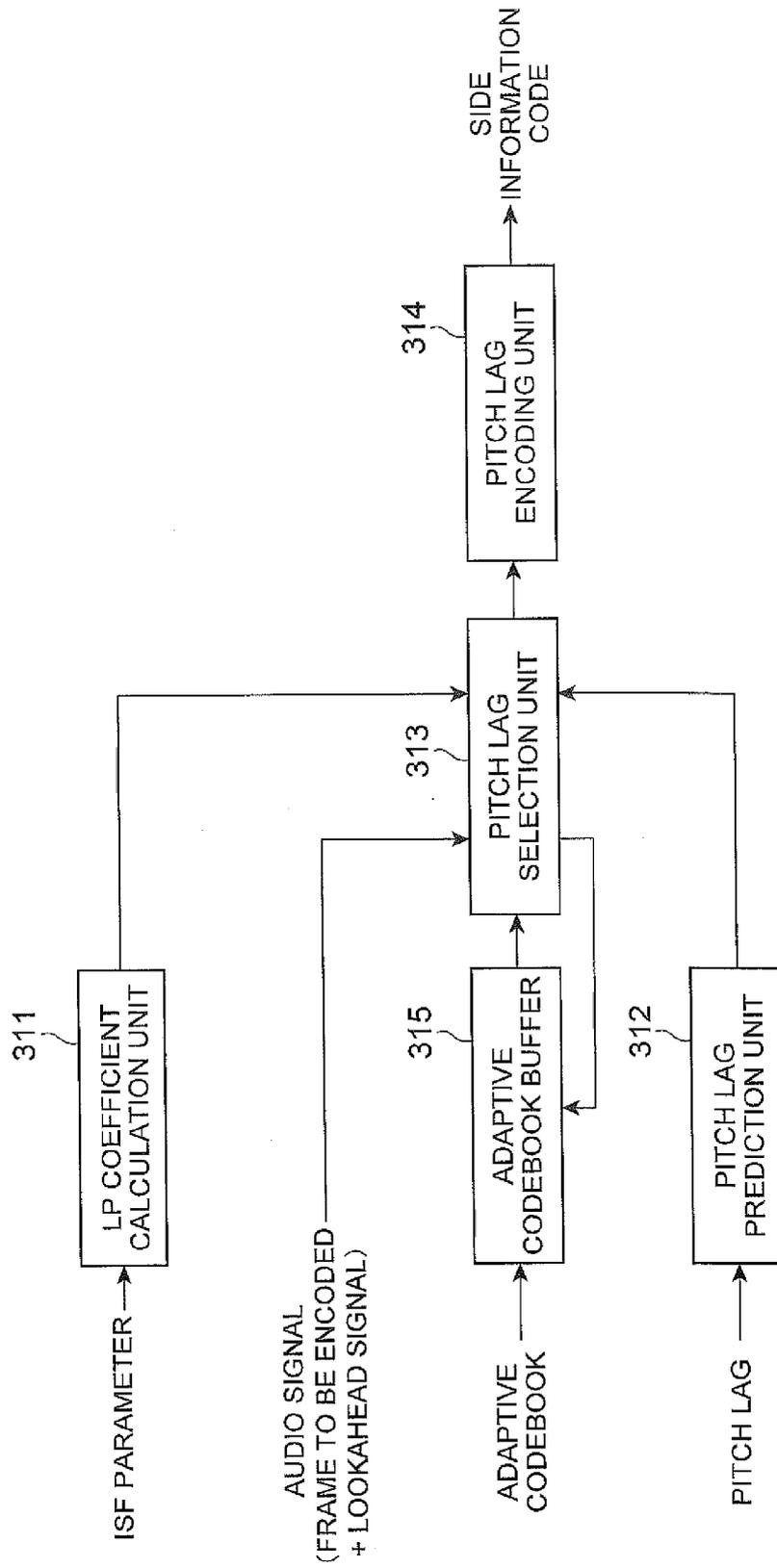


Fig.28

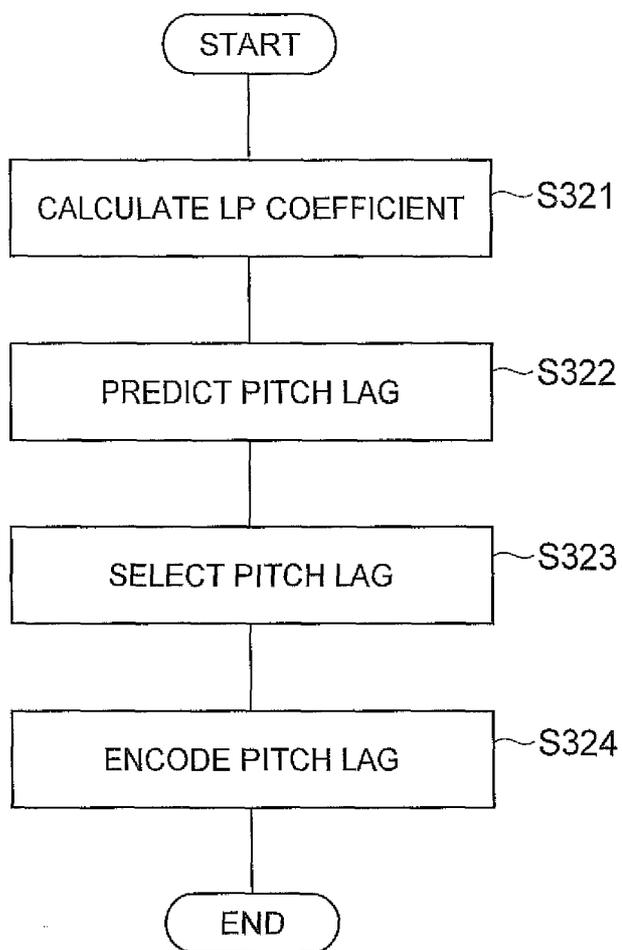


Fig.29

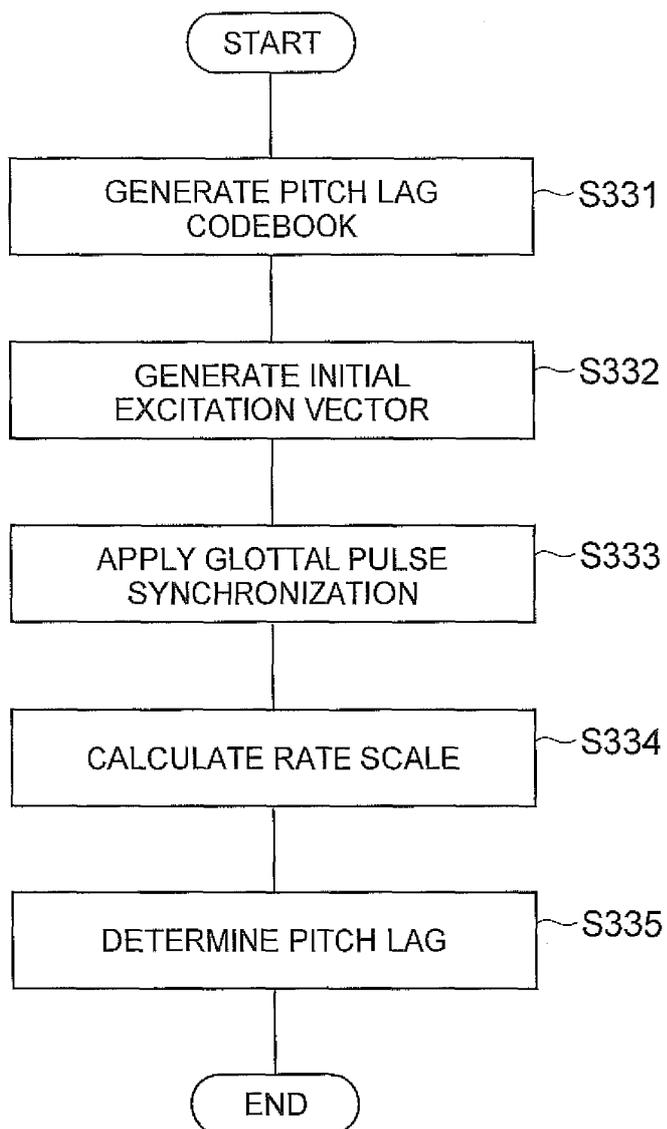


Fig.30

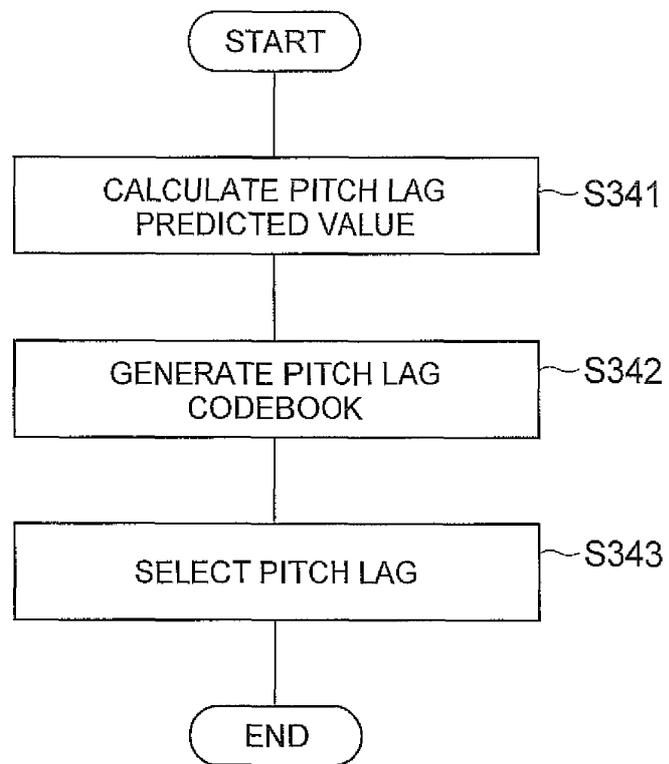


Fig.31

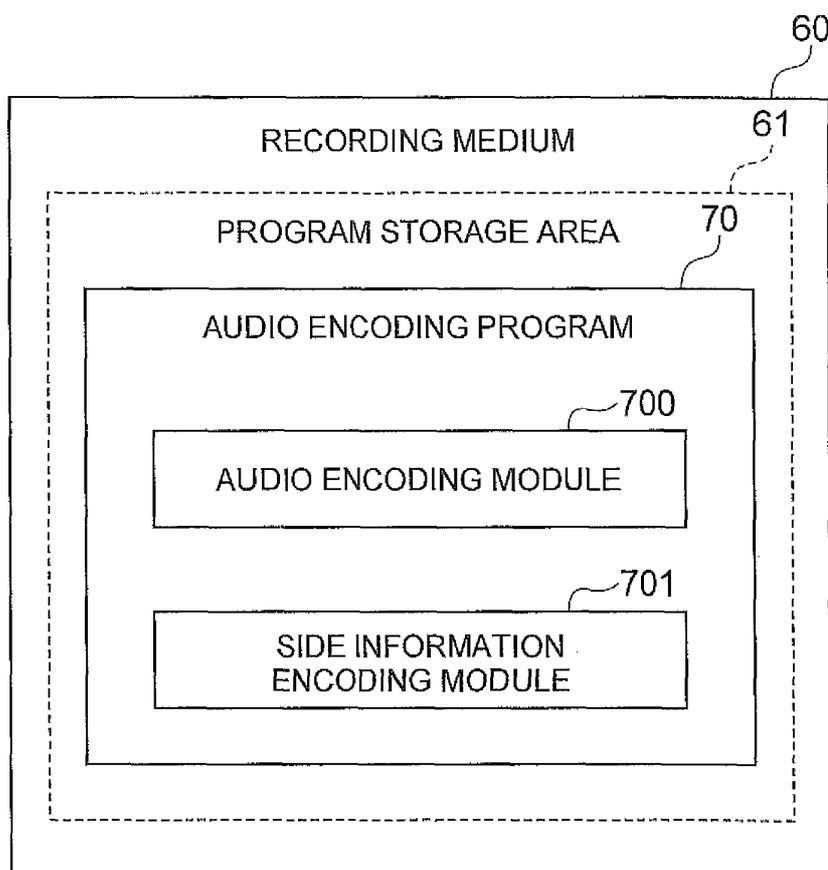


Fig.32

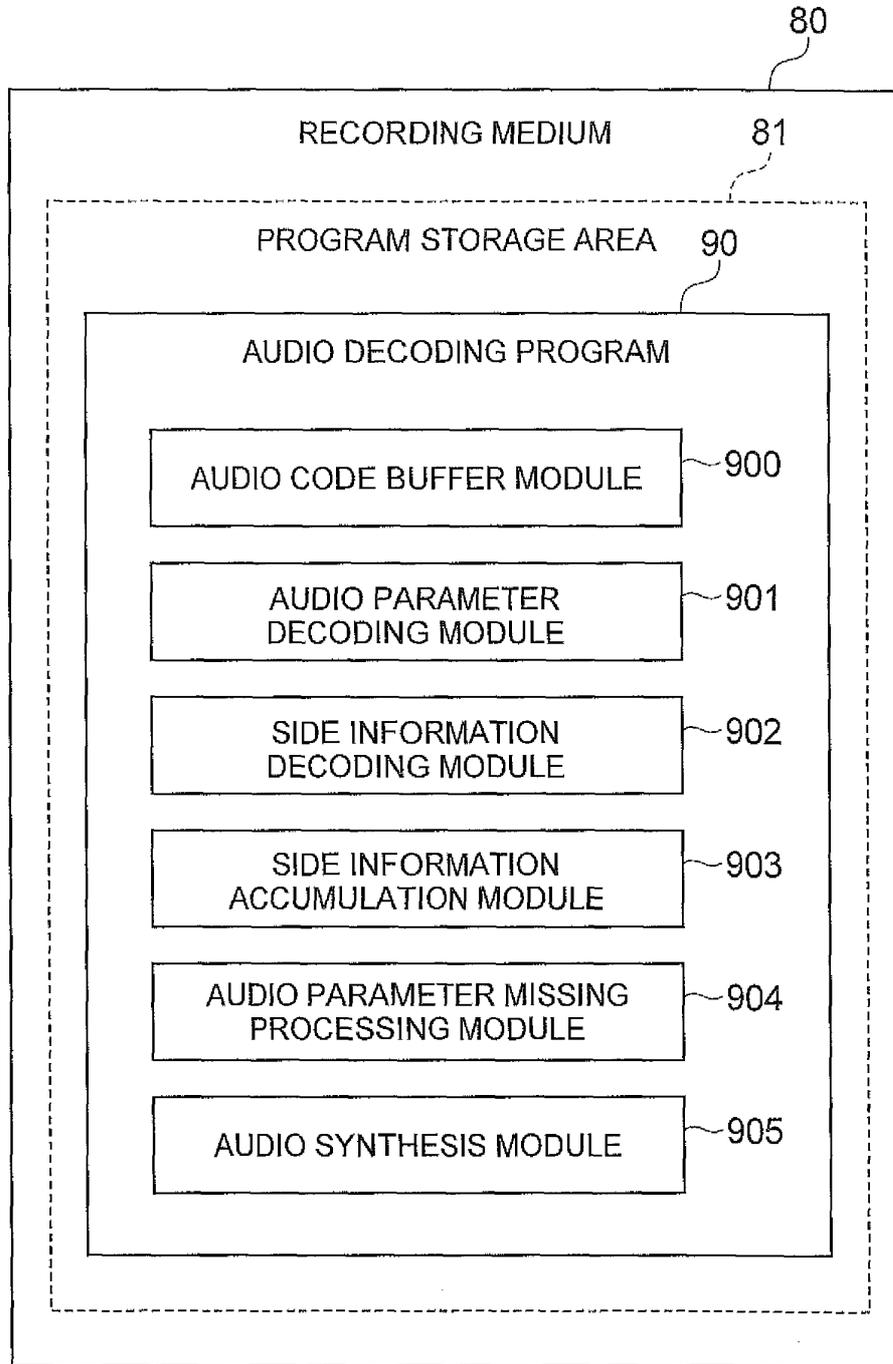


Fig. 33

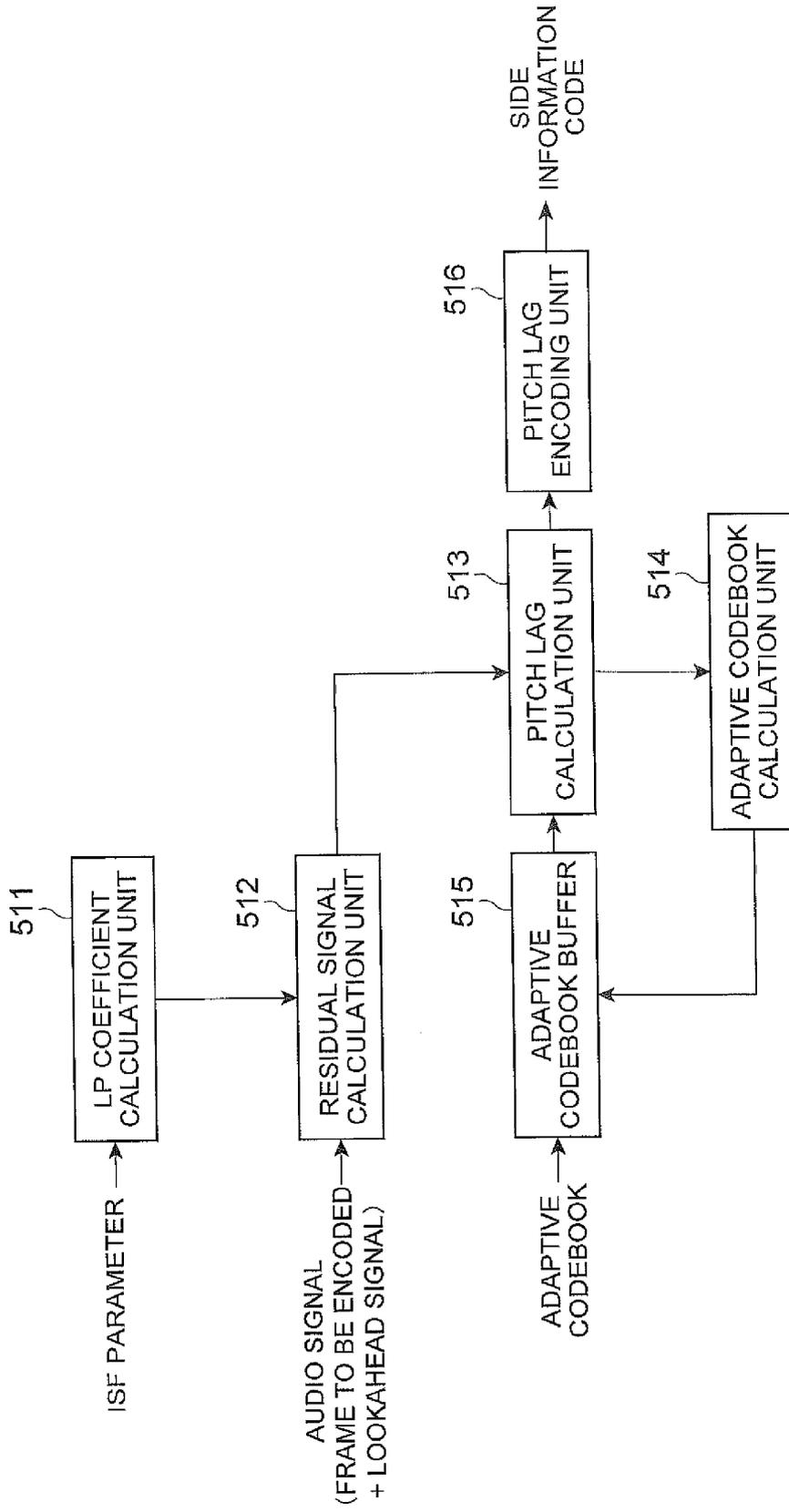


Fig.34

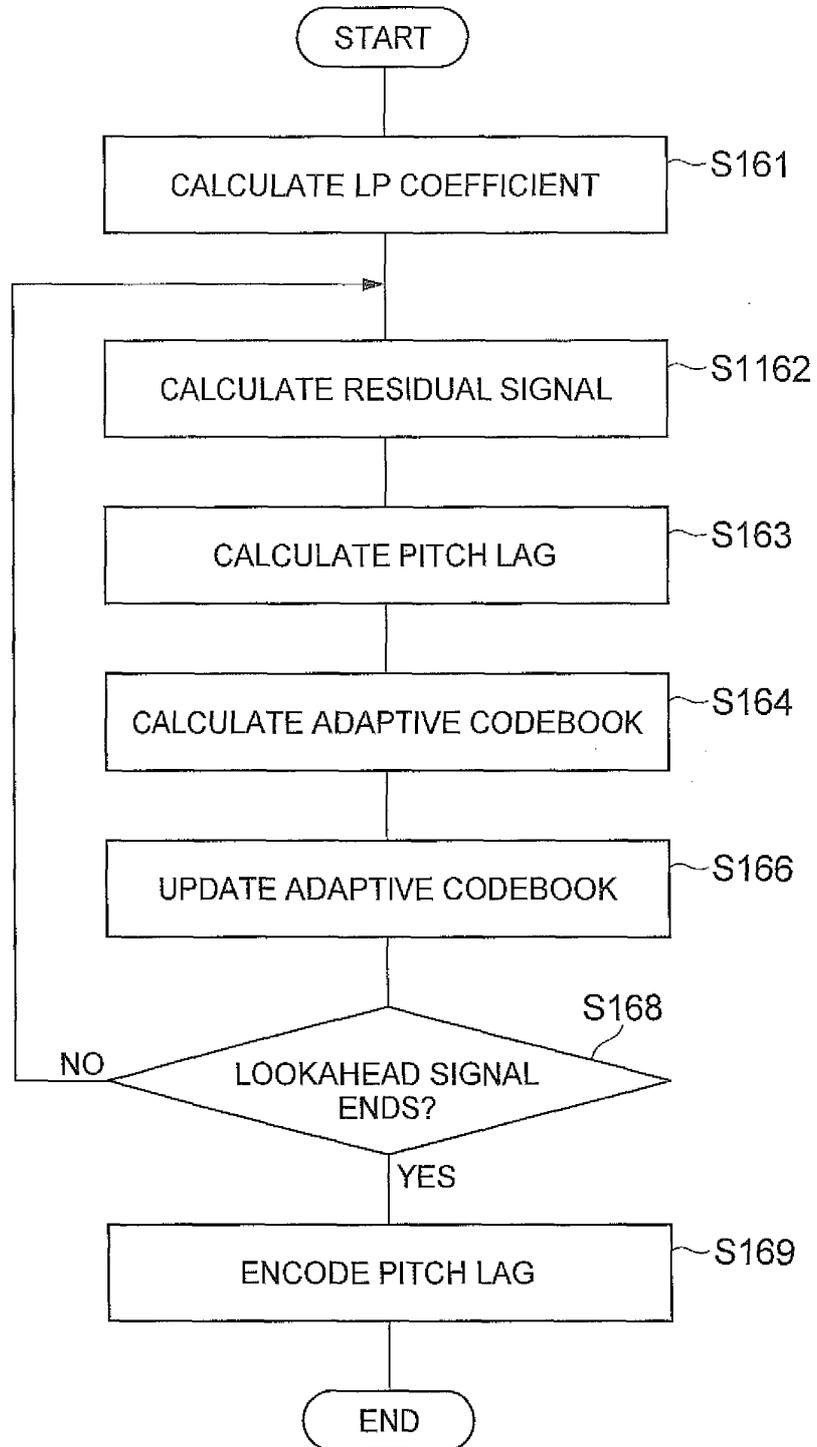


Fig.35

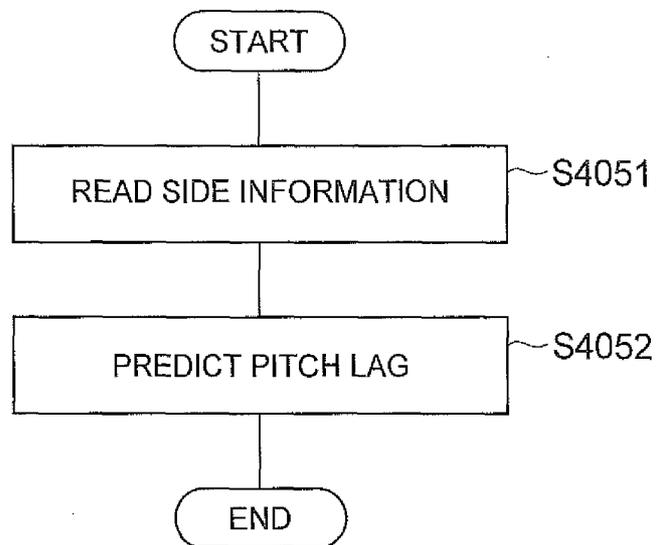


Fig.36

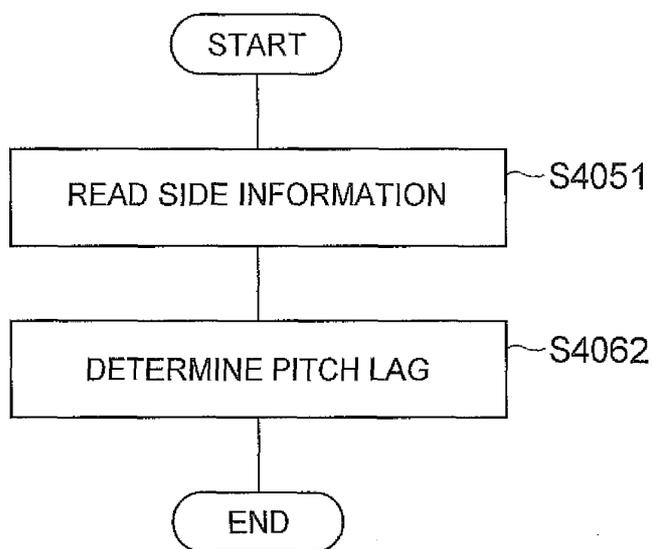


Fig.37

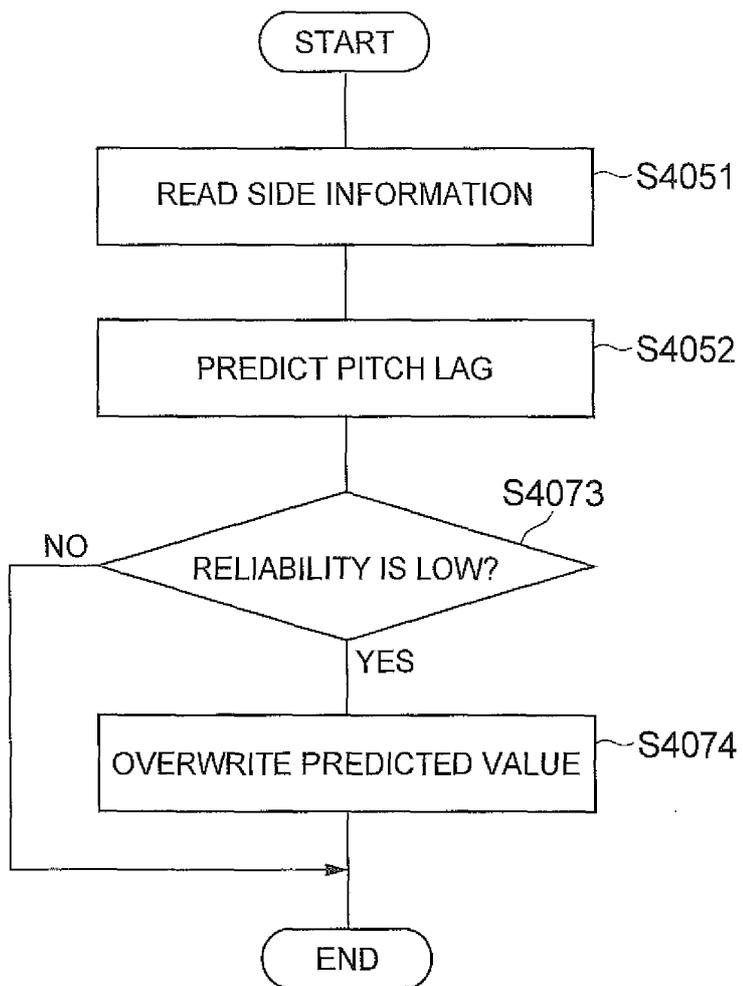


Fig.38

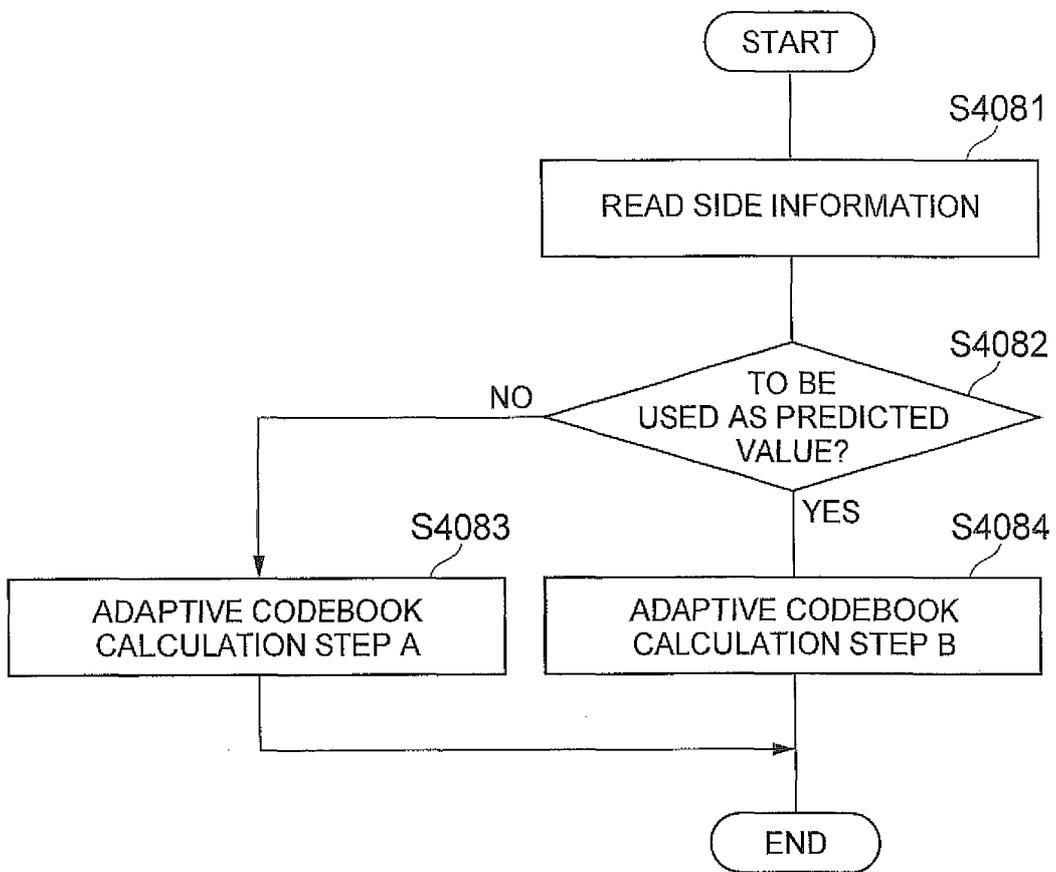


Fig.39

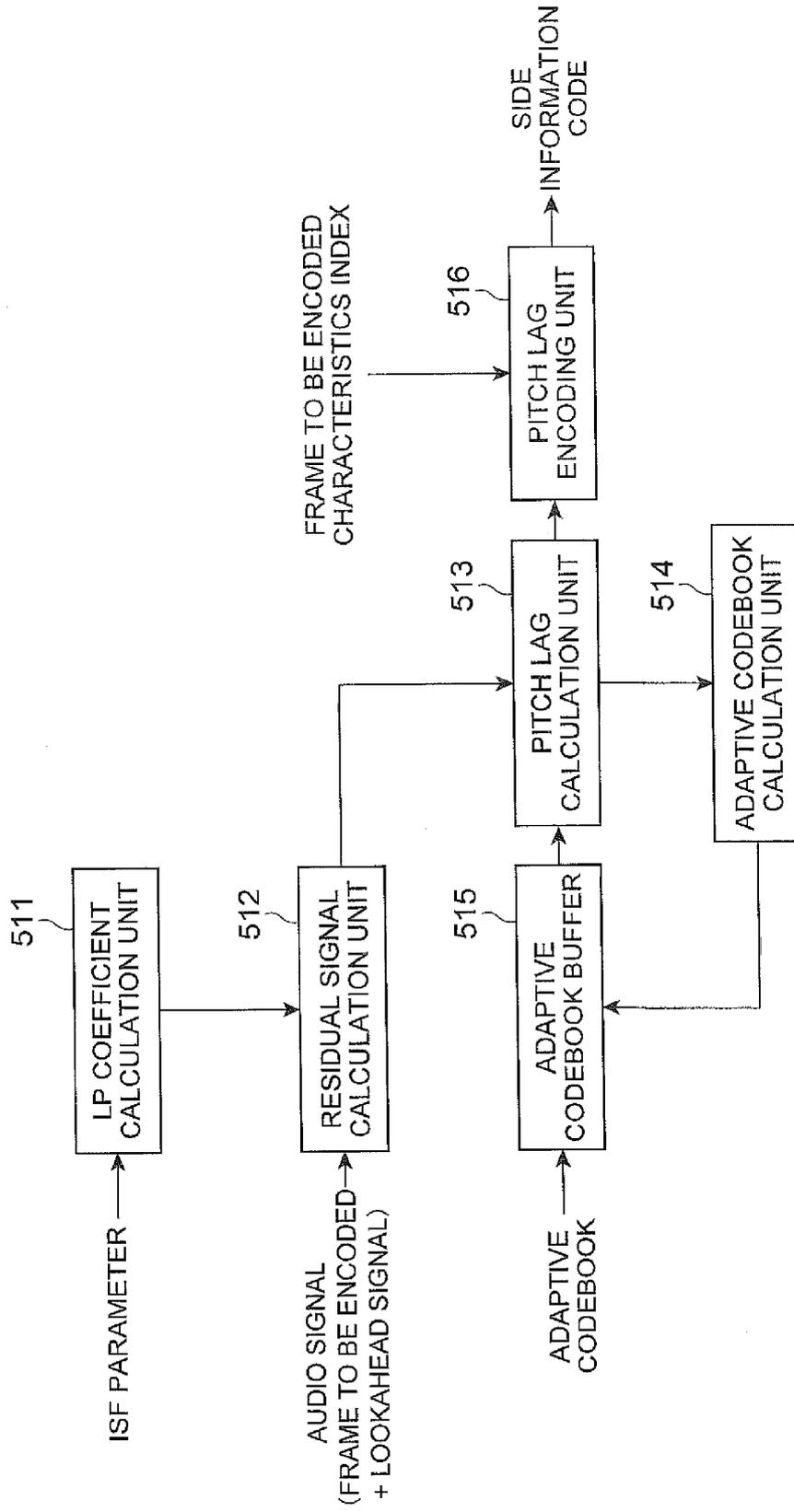


Fig.40

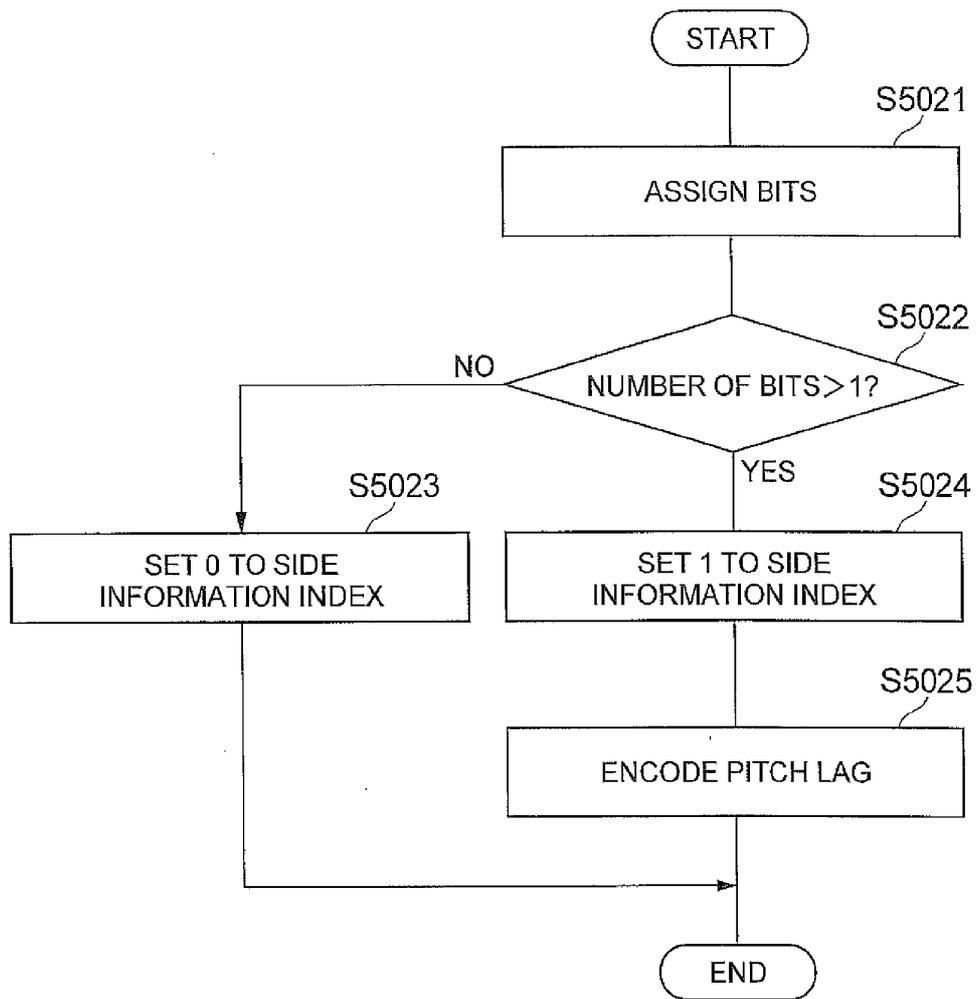


Fig.41

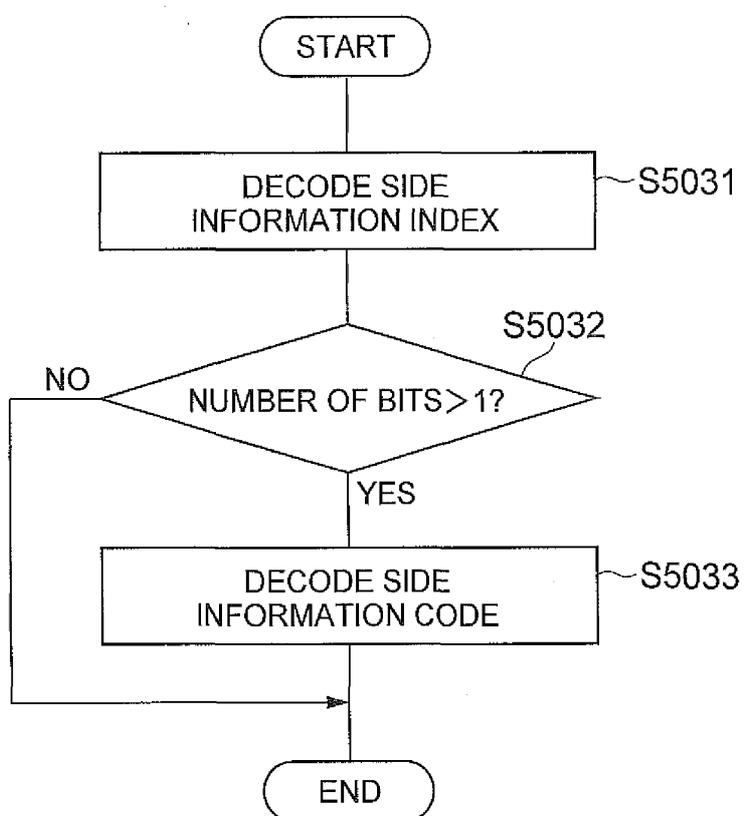


Fig.42

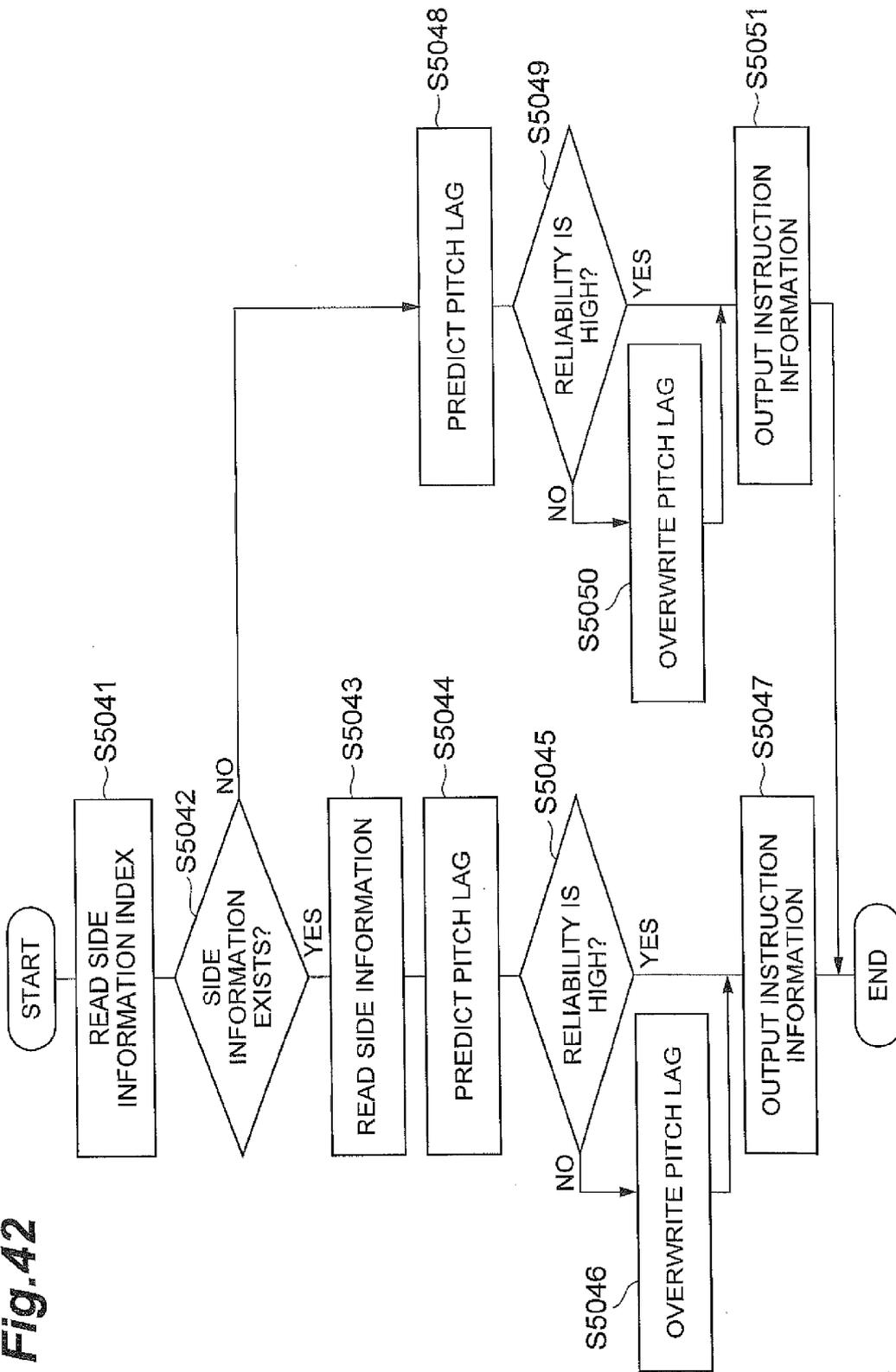
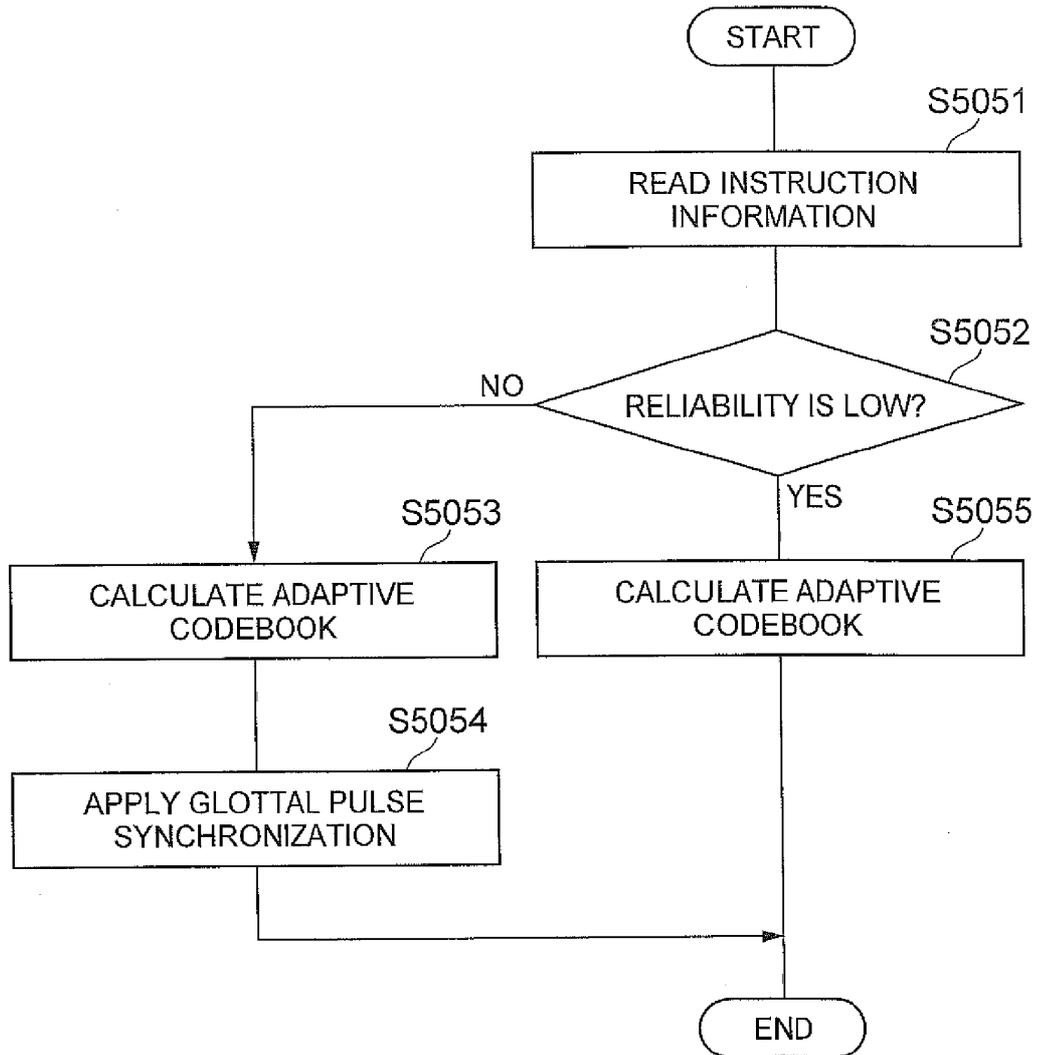


Fig.43



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2013/080589

5

A. CLASSIFICATION OF SUBJECT MATTER
G10L19/005(2013.01)i, G10L19/00(2013.01)i, G10L19/09(2013.01)i, H03M7/30
(2006.01)i

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

10

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L19/005, G10L19/00, G10L19/09, H03M7/30

15

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2014
Kokai Jitsuyo Shinan Koho 1971-2014 Toroku Jitsuyo Shinan Koho 1994-2014

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

20

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2004-138756 A (Matsushita Electric Industrial Co., Ltd.), 13 May 2004 (13.05.2004), paragraphs [0060] to [0063]; fig. 1, 3 (Family: none)	1-5, 7-11, 16-21 6, 12-15
A	WO 2008/007698 A1 (Matsushita Electric Industrial Co., Ltd.), 17 January 2008 (17.01.2008), paragraph [0023] & US 2009/0248404 A1	3-6, 9-17

25

30

35

40

Further documents are listed in the continuation of Box C. See patent family annex.

45

* Special categories of cited documents:
 "A" document defining the general state of the art which is not considered to be of particular relevance
 "E" earlier application or patent but published on or after the international filing date
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 "O" document referring to an oral disclosure, use, exhibition or other means
 "P" document published prior to the international filing date but later than the priority date claimed
 "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 "&" document member of the same patent family

50

Date of the actual completion of the international search
07 January, 2014 (07.01.14)

Date of mailing of the international search report
21 January, 2014 (21.01.14)

55

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/080589

5

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-526173 A (Nokia Corp.), 26 August 2004 (26.08.2004), paragraph [0004] & US 6968309 B1 & EP 1330818 A & WO 2002/037475 A1 & DE 60121201 D & DE 60121201 T & AU 1513802 A & CA 2424202 A & BR 115057 A & AT 332002 T & CN 1489762 A & ES 2266281 T & PT 1330818 E & ZA 200302556 A	3-6, 9-17
X	JP 2002-268696 A (Nippon Telegraph and Telephone Corp.), 20 September 2002 (20.09.2002), paragraph [0010] (Family: none)	1-4, 7-10, 18-21 5, 6, 11-17
A	WO 2012/070370 A1 (NTT Docomo Inc.), 31 May 2012 (31.05.2012), fig. 5 & CN 103229234 A & TW 201243825 A	1, 7, 18-21 2-6, 8-17
X	JP 2003-249957 A (Nippon Telegraph and Telephone Corp.), 05 September 2003 (05.09.2003), fig. 3 (Family: none)	1, 7, 18-21 2-6, 8-17
A	JP 2008-111991 A (NTT Docomo Inc.), 15 May 2008 (15.05.2008), entire text; all drawings (Family: none)	1-21

10

15

20

25

30

35

40

45

50

55

Form PCT/ISA/210 (continuation of second sheet) (July 2009)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- JP 2003533916 A [0019]
- JP 2010507818 A [0019]

Abstract

An objective of the present invention is, when packet loss occurs in audio encoding, to recover audio quality without increasing algorithmic delay. An audio signal transmission device for encoding an audio signal includes an audio encoding unit that encodes an audio signal and a side information encoding unit that calculates and encodes side information from a look-ahead signal. On the other hand, an audio signal receiving device for decoding an audio code and outputting an audio signal includes an audio code buffer that detects packet loss based on a received state of an audio packet, an audio parameter decoding unit that decodes an audio code when an audio packet is correctly received, a side information decoding unit that decodes a side information code when an audio packet is correctly received, a side information accumulation unit that accumulates side information obtained by decoding a side information code, an audio parameter missing processing unit that outputs an audio parameter when audio packet loss is detected, and an audio synthesis unit that synthesizes a decoded audio from an audio parameter.

摘要

本發明的目的在於，在音頻編碼中的分組丟失時，能夠在不增加算法延遲的情況下恢復音頻質量。對音頻信號進行編碼的音頻信號發送裝置具有：音頻編碼部，其對音頻信號進行編碼；以及輔助信息編碼部，其根據先行信號計算輔助信息而進行編碼。另一方面，對音頻碼進行解碼而輸出音頻信號的音頻信號接收裝置具有：音頻碼緩存器，其根據音頻分組的接收狀態檢測分組丟失；音頻參數解碼部，其在音頻分組正常接收時對音頻碼進行解碼；輔助信息解碼部，其在音頻分組正常接收時對輔助信息碼進行解碼；輔助信息蓄積部，其蓄積對輔助信息碼進行解碼而得到的輔助信息；音頻參數丟失處理部，其在音頻分組丟失檢測時輸出音頻參數；以及音頻合成部，其根據音頻參數合成解碼音頻。