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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 July 2004 (29.07.2004)

PCT

(10) International Publication Number
WO 2004/064041 A1

(51) International Patent Classification[7]: G10L 19/12, H03M 7/00

(21) International Application Number: PCT/AU2004/000014

(22) International Filing Date: 9 January 2004 (09.01.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/439,420 9 January 2003 (09.01.2003) US

(71) Applicant *(for all designated States except US)*: **DILITHIUM NETWORKS PTY LIMITED** [AU/AU]; Level 7, 3 Smail Street, Broadway, NSW 2007 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **JABRI, Marwan** [AU/AU]; Level 7, 3 Smail Street, Broadway, NSW 2007 (AU). **WANG, Jianwei** [AU/AU]; 104 Killarney Drive, Killarney Heights, NSW 2087 (AU). **CHONG-WHITE, Nicola** [AU/AU]; 364 Penshurst Street, Chatswood, NSW 2067 (AU). **IBRAHIM, Michael** [AU/AU]; 1 Salerwong Place, Ryde, NSW 2112 (AU).

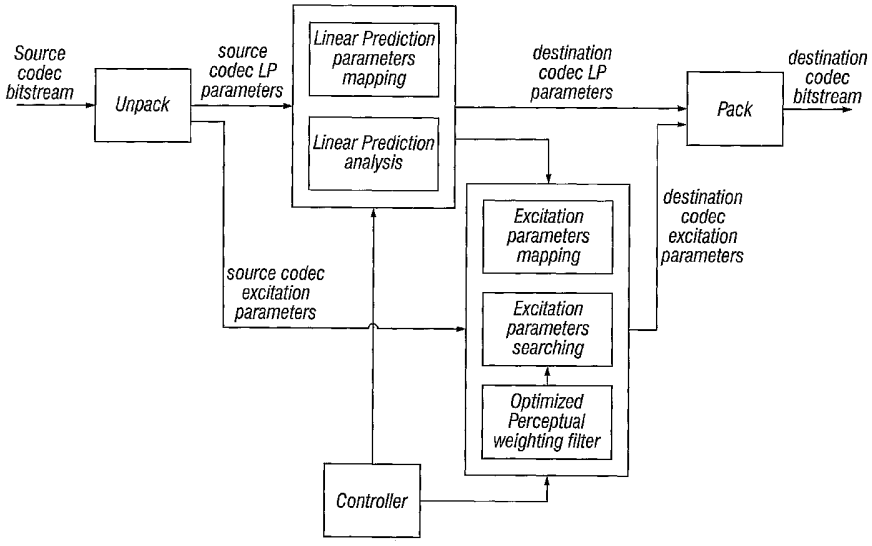
(74) Agent: **SPRUSON & FERGUSON**; GPO Box 3898, Sydney, NSW 2001 (AU).

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

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

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR IMPROVED QUALITY VOICE TRANSCODING



(57) Abstract: A method and apparatus for a voice transcoder that converts a bitstream representing frames of data encoded according to a first voice compression standard to a bitstream representing frames of data according to a second voice compression standard using perceptual weighting that uses tuned weighting factors, such that the bitstream of a second voice compression standard to produce a higher quality decoded voice signal than a comparable tandem transcoding solution. The method includes pre-computing weighting factors for a perceptual weighting filter optimized to a specific source and destination codec pair, pre-configuring the transcoding strategies, mapping CELP parameters in the CELP parameter space according to the selected coding strategy, performing Linear Prediction analysis if specified by the transcoding strategy, perceptually weighting the speech using with tuned weighting factors, and searching for adaptive codebook and fixed-codebook parameters to obtain a quantized set of destination codec parameters.

WO 2004/064041 A1

Published:
— *with international search report*

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

METHOD AND APPARATUS FOR IMPROVED QUALITY VOICE TRANSCODING

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Serial No. 60/439,420 (Attorney Docket Number 021318-001900US) titled "High Quality Audio Transcoding" filed January 9, 2003, which is incorporated by reference herein for all purposes.

BACKGROUND OF THE INVENTION

[0002] The present invention relates generally to processing telecommunication signals. More particularly, the invention relates to a method and apparatus for improving the output signal quality of a transcoder that translates digital packets from one compression format to another compression format. Merely by way of example, the invention has been applied to voice transcoding between Code-Excited Linear Prediction (CELP) codecs, but it would be recognized that the invention has a much broader range of applicability. To this end, the class of applicable codecs is designated as being "common" codecs.

[0003] The process of converting from one voice compression format to another voice compression format can be performed using various techniques. The tandem coding approach is to fully decode the compressed signal back to a Pulse-Code Modulation (PCM) representation and then re-encode the signal. This requires a large amount of processing and incurs increased delays. More efficient approaches include transcoding methods where the compressed parameters are converted from one compression format to the other while remaining in the parameter space.

[0004] Many of the current standardized low bit rate speech coders are based on the Code-Excited Linear Prediction (CELP) model. Common parameters of a CELP coder are the linear prediction parameters, adaptive codebook lag and gain parameters, and fixed codebook index and gain parameters.

[0005] The similarities between CELP-based codecs allow one to take advantage of the processing redundancies inherent in them. Figure 1 shows a block diagram for a typical prior art CELP decoder. The decoder receives as input a bitstream consisting of several parameters,

commonly representing the fixed codebook index, fixed codebook gain, adaptive codebook gain, adaptive codebook (pitch) lag and the linear prediction (LP) parameters. The decoder constructs the fixed codeword, which is then scaled by the codebook gain. The adaptive codeword, which is a previous excitation segment that has been delayed by the pitch lag and scaled by the adaptive gain, is added to the fixed codebook contribution. The resulting excitation signal is then filtered by a short term predictor producing synthesized speech. This speech is then post-filtered in order to reduce the perceptual significance of any synthesis artifacts and improve speech quality.

[0006] Figure 2 shows a block diagram for a typical prior art CELP encoder. The incoming speech signal is first pre-processed, for example, high-pass filtered to get rid of any superfluous information such as very low frequency information. Next, the spectral shape information is extracted by linear prediction (LP) analysis. The LP parameters are often represented as Line Spectral Pairs (LSPs) and quantized. The speech signal is then filtered using the inverse LP synthesis filter to remove the spectral envelope contribution and produce the excitation signal. Both the pre-processed speech and excitation are filtered with a perceptual weighting filter. The perceptually weighted speech is analyzed for periodicity, often using both a open loop pitch lag search and a closed loop (analysis-by-synthesis) pitch lag and pitch gain search. The pitch contribution is subtracted from the perceptually weighted speech to create a target signal for the fixed codebook search. The fixed codebook search consists of an analysis-by-synthesis algorithm, in which various code words are evaluated to minimize the error between the synthesized codeword and target signal.

[0007] Transcoding addresses the problem that occurs when two incompatible standard coders need to interoperate. The conventional prior art tandem coding solution, illustrated in Figure 3, is to fully decode the signal from one compression format to PCM, and then to re-encode the PCM signal using the other compression format. This solution has the disadvantages of being computationally complex, it and introduces quality degradations due to the full decode and full encode. Alternatively a prior art transcoder, as shown in Figure 4, may be used which converts the bitstream from one compression format to a different compression format without fully decoding to PCM and then re-encoding the signal.

[0008] Some transcoding approaches involve converting parameters solely in the CELP domain. These methods have the advantage of reducing computational complexity. Figure 5 shows an example of one prior art transcoding approach in which the source codec LSPs are

directly translated and quantized to the destination codec format. The speech is then synthesized using the destination codec LSPs and the remaining CELP parameters are found using a searching algorithm. This technique does not improve the quality of the transcoded signal to the fullest extent and is not necessarily the best solution in some situations.

[0009] While smart transcoding techniques that map parameters from one CELP format to another in a fast manner have been developed, a transcoding solution that provides transcoded speech of a higher quality than the conventional tandem coding solution and that may be configured and tuned for specific source and destination codec pairs is highly desirable.

SUMMARY OF THE INVENTION

[0010] According to the invention, a method and apparatus are provided for improving the output signal quality of a transcoder that translates digital packets from one compression format to another compression format by including perceptually weighting of the speech using a weighting filter with tuned weighting factors. Merely by way of example, the invention has been applied to voice transcoding between Code-Excited Linear Prediction (CELP) codecs, but it would be recognized that the invention has a much broader range of applicability, as explained herein and hereinafter referred to as common codecs.

[0011] In a specific embodiment, the present invention provides a method and apparatus for high quality voice transcoding between CELP-based voice codecs. The apparatus includes an input CELP parameters unpacking module that converts input bitstream packets to an input set of CELP parameters; a linear prediction parameters generation module for determining the destination codec Linear Prediction (LP) parameters, a perceptual weighting filter module that uses tuned weighting factors, an excitation parameter generation module for determining the excitation parameters for the destination codec, a packing module to pack the destination codec bitstream, and a control module that configures the transcoding strategies and controls the transcoding process. The linear prediction parameters generation module includes an LP analysis module and an LP parameter interpolation and mapping module. The excitation parameter generation module includes adaptive and fixed codebook parameter searching modules and adaptive and fixed codebook parameter interpolation and mapping modules.

[0012] The method includes pre-computing weighting factors for a perceptual weighting filter that are optimized to a specific source and destination codec pair and storing them to the

systems, pre-configuring the transcoding strategies, unpacking the source codec bitstream, reconstructing speech, mapping at least one but typically more than one CELP parameter in the CELP parameter space according to the selected coding strategy, performing LP analysis if specified by the transcoding strategy, perceptually weighting the speech using a weighting filter with tuned weighting factors, and searching for one or more of the adaptive codebook and fixed-codebook parameters to obtain the quantized set of destination codec parameters. Reconstructing speech does not involve any post-filtering processing. In addition, the reconstructed speech passed as input to the LP analysis and speech perceptual weighting does not undergo any pre-processing filtering or noise suppression. Mapping one or more CELP parameters includes interpolating parameters if there is a difference in frame size or subframe size between the source and destination codecs. The CELP parameters may include LP coefficients, adaptive codebook pitch lag, adaptive codebook gain, fixed codebook index, fixed codebook gain, excitation signals, and other parameters related to the source and destination codecs. Searching for adaptive codebook and fixed codebook parameters may be combined with mapping and conversion of CELP parameters to achieve high voice quality. This is controlled by the transcoding strategy. The algorithms within the searching module can be different to the algorithms used in the standard destination codec itself.

[0013] An advantage of the present invention is that it provides a transcoded voice signal with higher voice quality and lower complexity than that provided by a tandem coding solution. The processing strategy that combines both mapping and searching processes for determining parameter values can be adapted to suit different source and destination codec pairs.

[0014] The objects, features, and advantages of the present invention, which to the best of our knowledge are novel, are set forth with particularity in the appended claims. The present invention, both as to its organization and manner of operation, together with further objects and advantages, may best be understood by reference to the following description, taken in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] FIG 1 is a simplified block diagram illustrating an example of a prior art CELP decoder.

[0016] FIG 2 is a simplified block diagram illustrating an example of a prior art CELP encoder.

[0017] FIG 3 is a simplified block diagram illustrating a prior art tandem coding procedure.

[0018] FIG 4 is a simplified block diagram illustrating a transcoding procedure of the prior art which does not fully decode and re-encode the signal.

[0019] FIG 5 is a simplified block diagram of a prior-art transcoding approach.

[0020] FIG 6 is a diagram representation of high voice quality transcoder methods.

[0021] FIG 7 is a block diagram illustrating a high voice quality transcoder from one CELP-based codec to another CELP-based codec according to an embodiment of the present invention.

[0022] FIG 8 is a block diagram illustrating the processing options, controlled by the transcoding strategy, in the excitation parameter generation module of a high voice quality transcoder according to an embodiment of the present invention.

[0023] FIG 9 is an alternative representation of an excitation parameter searching module in a high voice quality transcoder according to an embodiment of the present invention.

[0024] FIG 10 is a flowchart of a high quality voice transcoding method according to an embodiment of the present invention.

[0025] FIG 11 is a flowchart of an excitation parameter searching method according to an embodiment of the present invention.

[0026] FIG 12 is a schematic diagram of the process to obtain weighting factors for a speech perceptual weighting filter for a specific source and destination codec pair according to an embodiment of the present invention.

[0027] FIG 13 is a flowchart illustrating the post-processing and pre-processing functions used in tandem transcoding from EVRC to SMV.

DETAILED DESCRIPTION OF THE INVENTION

[0028] In a specific embodiment of the invention, a Code-Excited Linear Prediction (CELP) based compression scheme is employed. Audio compression using a CELP-based compression scheme is a common technique used to reduce data bandwidth for audio transmission and storage. Hence, any common codec for which a common codec parameter space is defined may be used. In many situations, the ability to communicate across different

networks is desirable, for example from an Internet Protocol (IP) network to a cellular mobile network. These networks use different CELP compression schemes in order to communicate audio, and in particular voice. Different CELP coding standards, although incompatible with each other, generally utilize similar analysis and compression techniques.

[0029] Figure 6 shows a diagram illustrating several factors that contribute to a target or high voice quality resulting from transcoding according to the present invention. In addition to the removal of post-processing and pre-processing functions, the use of optimized perceptual weighting factors, configured transcoding strategies, mapping of parameters in the CELP domain and advanced searching functions contribute to higher quality transcoded signals.

[0030] Figure 7 shows a block diagram of a high quality transcoder according to the invention. The apparatus includes a unpacking module that converts input source codec bitstream packets to a set of common codec parameters, such as CELP parameters; a linear prediction parameters generation module for determining the destination codec parameters, such as linear prediction (LP) parameters, a perceptual weighting filter module that uses tuned or customized weighting factors, an excitation parameter generation module for determining the excitation parameters for the destination codec, a packing module to pack the destination codec bitstream, and a control module that configures the transcoding strategies and controls the transcoding process. The linear prediction parameters generation module includes a linear prediction (LP) analysis module, and an LP parameter interpolation and mapping module. The excitation parameter generation module includes adaptive and fixed codebook parameter searching modules and adaptive and fixed codebook parameter interpolation and mapping modules. The control module controls whether parameter mapping or searching is performed, according to the transcoding strategy.

[0031] The transcoding strategy is configured depending on the similarities of the source and destination codecs, in order to optimize mapping from source encoded CELP parameters into destination encoded CELP parameters. Figures 8 and 9 illustrate the excitation parameter generation modules in which one of several searching procedures, such as direct mapping, searching, or (in the case of identical source and destination codecs) pass-through, may be chosen to determine each of the excitation parameters, depending on the transcoding strategy. The algorithms for adaptive codebook searching and fixed codebook searching in the transcoder may differ from those of the conventional or standard destination CELP codec.

During searching, perceptual weighting filters are used to shape the quantization noise. The perceptual weighting factors are not necessarily the same as those defined in the destination standard. They can be further fine tuned or customized, for example, by empirical methods, taking into account the source codec characteristics. This operation can further improve audio quality.

[0032] The transcoding algorithm of the present invention can be made considerably more efficient than a conventional tandem solution by not using unneeded computationally intensive steps of source codec post-filtering, destination codec pre-filtering, destination codec LP analysis, or destination codec open loop pitch search. Further savings may be realized by directly mapping one or more excitation parameters rather than performing complex searches.

[0033] A flowchart of an embodiment of the inventive voice transcoding process is illustrated in Figure 10. If the source and destination codec type and bit-rate are the same, no (CELP) parameter searching is required, and the output bitstream is set to the input bitstream. Otherwise, the bitstream is unpacked. The excitation signal is reconstructed and the speech is synthesized. A choice is made between performing LP analysis on the synthesized speech or mapping the LP parameters from the source codec. The target and impulse response signals to determine the excitation parameters are generated using a perceptual weighting synthesis filter with weighting factors that are optimized to the specific source codec and destination codec pair. The remaining common codec (CELP) parameters are determined by searching, and then they are packed to the output bitstream.

[0034] Figure 11 shows a flowchart of an embodiment of the common codec (CELP) parameters searching method. For each of the common codec parameters of adaptive codebook lag, adaptive codebook gain, fixed codebook index and fixed codebook gain, a decision is made as to whether to directly map the parameter from the source codec (e.g., CELP) parameter set, or to perform a search for that parameter. The decision is controlled by the transcoding strategy selected, which is based on the source and destination codec pair.

[0035] Figure 12 is an illustration of the procedure used to optimize the weighting factors for the perceptual weighting filter used in searching for excitation parameters of the destination codec. The perceptual weighting filter can be expressed by the transfer function:

[0036] $$H_w(z) = \frac{A\left(\frac{z}{\gamma_1}\right)}{A\left(\frac{z}{\gamma_2}\right)},$$

where $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_N z^{-N}$, $a_1, ...$ represent the linear prediction coefficients for the current speech segment, and 1. 2 are the weighting factors. The quality of the transcoded output speech can be improved by tuning or customizing the weighting factors to best suit the source and destination codec pair. This can be done using automatically using feedback methods or using empirical methods by performing the transcoding on a set of test samples using different weighting factor combinations, evaluating the output voice quality by subjective or objective methods and retaining the weighting factors that result in the highest perceived or measured output voice quality for that specific source and destination codec pair.

[0037] As an example, high quality voice transcoding is applied between GSM-AMR (all modes) and G.729. A person skilled in the relevant art will recognize that other steps, configurations and arrangements can be used without departing from the spirit and scope of the present invention.

[0038] The GSM-AMR standard utilizes a 20ms frame, divided into four 5ms subframes. For the highest GSM-AMR mode, LP analysis is performed twice per frame, and once per frame for all other modes. The open loop pitch estimate is obtained from the perceptually weighted speech signal. This is performed twice per frame for the 12.2kbps mode, and once per frame for the other modes. The closed loop pitch search and fixed codeword search are both performed once per subframe, and the fixed codebook is based on an interleaved single-pulse permutation (ISPP) design.

[0039] The G.729 standard utilizes a 10ms frame divided into two 5ms subframes. LP analysis is performed once per frame. The open loop pitch estimate is calculated on the perceptually weighted speech signal, once per frame. Like GSM-AMR, the closed loop pitch search and fixed codeword search are both performed once per subframe, and the fixed codebook is based on an interleaved single-pulse permutation (ISPP) design.

[0040] For the G.729 to GSM-AMR transcoder, two input G.729 frames produces one GSM-AMR output frame. The LP parameters, codebook index, gains and pitch lag are

unpacked and decoded from the input bitstream. Due to the differences in search procedures, codebooks, and quantization frequency of some parameters, the best transcoding strategy may differ depending on the AMR mode. In particular, the similarities associated with G.729 and AMR 7.95kbps may lead to the configuration of a transcoding strategy that selects more parameters for direct mapping and less parameters for searching than the G.729 to AMR 4.75kbps transcoder.

[0041] If the transcoding strategy specifies that some excitation parameters are found by searching methods, the synthesized reconstructed excitation signal is perceptually weighted to produce a target signal. The best weighting factors for the perceptual weighting filter for each mode and bit rate of the source and destination codecs of the transcoder are determined prior to transcoding. Typically, when transcoding from G.729 to AMR 12.2kbps, a different set of weighting factors will be used than for transcoding to other AMR modes, for example, from G.729 to AMR 7.95kbps or from G.729 to AMR 4.75kbps.

[0042] In a transcoding scenario, the upper quality limit is the lower of the source codec quality or destination codec quality. The high quality voice transcoding of the present invention is able to significantly reduce the quality gap between the upper quality limit and the quality obtained by the tandem coding solution.

[0043] In an alternative embodiment, voice transcoding is applied in a transcoder whereby the source codec is the Enhanced Variable Rate Codec (EVRC) and the destination codec is the Selectable Mode Vocoder (SMV). SMV and EVRC are both common codec parameters types that employ built-in noise suppression algorithms. A flowchart of the post-processing functions of EVRC and the pre-processing functions of SMV used in the tandem transcoding solution is illustrated in Figure 13. A transcoding solution with lower complexity and higher quality than the tandem transcoding solution can be achieved by removing one or more of the processes of EVRC postfiltering, SMV highpass filtering, SMV silence enhancement, SMV noise suppression, and SMV adaptive tilt filtering. Since EVRC already uses noise suppression, much of the background noise in the input has already been removed at the source encoder, hence a second noise suppression algorithm during transcoding causes further speech degradation with little change to the background noise level. Further complexity reductions and/or quality improvements can be realized using the optimization of perceptual weighting factors, and the mixed transcoding strategy of mapping some parameters in the CELP domain and determining some by searching.

[0044] The present invention for high voice quality transcoding is generic to all voice transcoding between CELP-based codecs and applies any voice transcoders among the existing codecs G.723.1, GSM-EFR, GSM-AMR, EVRC, G.728, G.729, SMV, QCELP, MPEG-4 CELP, AMR-WB, and all other future CELP based voice codecs that make use of voice transcoding. The foregoing common codec standards for each of which a common codec parameter space is defined are considered exemplary but not limiting.

[0045] The foregoing description of specific embodiments is provided to enable a person having ordinary skill in the art to make or use the present invention. The various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without the use of the inventive faculty. Thus, the present invention is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

WHAT IS CLAIMED IS:

1. An apparatus for a voice transcoder that produces a destination code bitstream in a destination codec format from a source code bitstream in a source codec format, the apparatus comprising:

an unpacking module operative to unpack the source codec bitstream and decode the information into at least one parameter of a common codec for which a common codec parameter space is defined;

a linear prediction parameters generation module operative to generate destination codec linear prediction parameters by mapping from source codec linear prediction parameters or by linear prediction analysis;

a perceptual weighting filter module operative to use weighting factors that have been optimized for transcoding between a specific source codec and destination codec pair;

an excitation parameter generation module for determining at least one common codec excitation parameter in the destination codec format, said parameter generation module operative to provide direct mapping processes and searching processes for each said common codec excitation parameter;

a packing module operative to pack the destination codec common codec parameters to the bitstream; and

a control module for selecting a transcoding strategy and to provide additional control information.

2. The apparatus of claim 1, wherein said linear prediction parameters generation module comprises:

a linear prediction parameters mapping and conversion module for interpolating the linear prediction parameters upon determination of a difference between source codec frame size and destination codec frame size, and for mapping the linear prediction parameters to the destination codec format; and

a linear prediction analysis module for generating linear prediction parameters from a reconstructed speech signal.

3. The apparatus of claim 1, wherein optimized weighting factors of said perceptual weighting filter module are pre-computed prior to transcoding and storing as part of the apparatus.

4. The apparatus of claim 1, wherein said excitation parameter generation module comprises:

first modules for direct mapping of the source codec excitation parameters format to the destination codec excitation parameters format;

second modules for searching for said source codec excitation parameters and said destination codec excitation parameters; and

pass-through modules for third excitation parameters, said third excitation parameters being used if the types of said source codec and said destination codec and respective bit-rates are the same.

5. The apparatus of claim 4, wherein said first modules for direct mapping of excitation parameters comprise an adaptive codebook pitch lag mapping module, an adaptive codebook pitch gain mapping module, a fixed codebook gain mapping module, and a fixed codebook index mapping module.

6. The apparatus of claim 4, wherein said second modules for searching for excitation parameters comprise an adaptive codebook pitch lag searching module, an adaptive codebook pitch gain searching module, a fixed codebook gain searching module, a fixed codebook index searching module, and an excitation reconstruction module.

7. The apparatus of claim 4, wherein said pass-through modules for excitation parameters comprise an adaptive codebook pitch lag searching module, an adaptive codebook pitch gain searching module, a fixed codebook gain searching module, a fixed codebook index searching module and an excitation reconstruction module.

8. The apparatus of claim 1, wherein said control module is operative to employ a transcoding strategy comprising a set of rules to determine a specific process of transcoding.

9. The apparatus of claim 1, wherein said linear prediction parameters generation module is controlled by said control module.

10. The apparatus of claim 1, wherein said excitation parameter generation module is controlled by said control module.

11. The apparatus of claim 1, wherein reconstructed speech of the source codec is not pre-processed.

12. The apparatus of claim 1 having no noise suppression functions.

13. The apparatus of claim 1 having no post-filtering and no gain adjustment.

14. A method for producing a destination code bitstream in a destination codec format from a source code bitstream in a source codec format in order to perform voice transcoding between common codec parameter-based voice codecs comprising:

determining and storing weighting factors for a perceptual weighting filter, said weighting factors being optimized for a specific source codec and destination codec pair;

configuring transcoding strategies for each preselected transcoding pair;

unpacking said source codec bitstream to produce source codec common codec parameters;

reconstructing a speech signal using source codec common codec parameters;

mapping one or more parameters in parameter space of the common codec parameters according to a selected transcoding strategy;

perceptually weighting voice signals using said perceptual weighting filter according to the selected transcoding strategy;

searching for one or more excitation parameters according to the selected transcoding strategy; and

packing the destination codec common codec parameters to the destination codec bitstream.

15. The method of claim 14, wherein said common codec parameters are defined by a linear code, further including the interim step of:

performing linear prediction analysis according to the selected transcoding strategy to determine linear prediction coefficients for further processing.

16. The method of claim 14, wherein said excitation parameters mapping comprises determining quantized values of at least one of adaptive codebook pitch lag, adaptive codebook pitch gain, fixed-codebook index and fixed-codebook gain by interpolating the source codec parameters upon determination of at least one of a difference

in frame size, subframe size, and mappable characteristics between the source codec and the destination codec; and

directly converting the excitation parameters to the destination codec format.

17. The method of claim 14, wherein said excitation parameters searching step comprises determining quantized values of at least one of adaptive codebook pitch lag, adaptive codebook pitch gain, fixed-codebook index, and fixed-codebook gain by minimizing the error between a reconstructed signal and a target signal.

18. The method of claim 14, wherein transcoding strategies configuring step comprise selecting a number of respective mapping and searching options to determine signal processing flow.

19. The method of claim 14 wherein the transcoding strategy specifies a process whereby some parameters are first obtained from said common codec parameter mapping and remaining parameters are obtained through a searching procedure.

20. The method of claim 14, wherein the transcoding strategy specifies a process whereby all common codec parameters from the source codec are mapped to the destination codec without searching.

21. The method of claim 14, wherein reconstructing a speech signal involves no post-processing operations.

22. The method of claim 14, wherein no noise suppression or speech pre-processing is performed prior to speech perceptual weighting.

23. The method of claim 14, wherein said transcoding strategies comprise:

direct mapping of a code-excited linear prediction parameter upon determination of presence of a similar code-excited linear prediction parameter compression process between the source codec and destination codec of the transcoding pair;

--performing speech reconstruction and speech perceptual weighting if searching is required to determine code-excited linear prediction parameters for the destination codec;

performing linear prediction analysis if there are substantial differences in linear prediction parameter compression processes between the source codec and the

destination codec in a transcoding pair, and if the steps of linear prediction parameter interpolation, mapping, and conversion do not produce a target output voice quality in the transcoding;

searching the adaptive codebook, if LP analysis processing is required;

searching the adaptive codebook, 1) if the adaptive codebook parameter compression process has substantial differences between source codec and destination codec in a transcoding pair, and 2) the adaptive codebook parameter space mapping method does not produce the target output voice quality in the transcoding;

searching the fixed codebook, if adaptive codebook searching is required;

searching the fixed codebook, if the fixed codebook parameter compression process has substantial differences between source codec and destination codec in a transcoding pair, and if the fixed codebook parameter space mapping method does not produce the target output voice quality in the transcoding.

24. The method of claim 14, wherein said weighting factors obtaining step comprises transcoding a set of voice samples using different weighting factor values, performing voice quality tests on the transcoded voice signals, and selecting specific weighting factors for a specific source codec and destination codec pair in order to produce a target voice quality.

25. The method of claim 14, wherein said weighting factors obtaining step comprises finding best weighting factors for each possible mode and bit rate combination of the source codec and the destination codec.

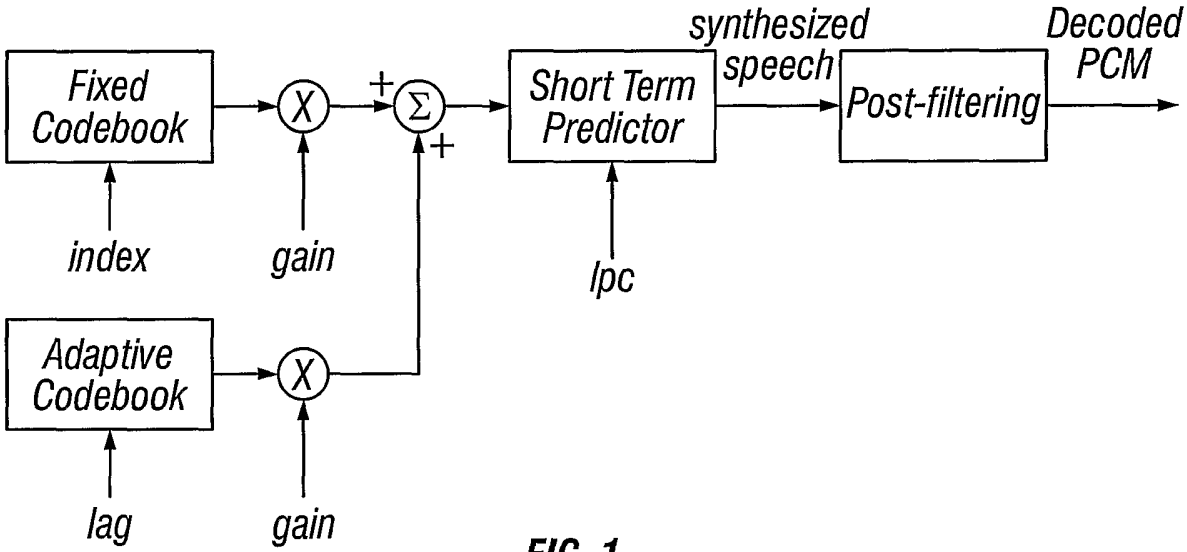


FIG. 1
(Prior Art)

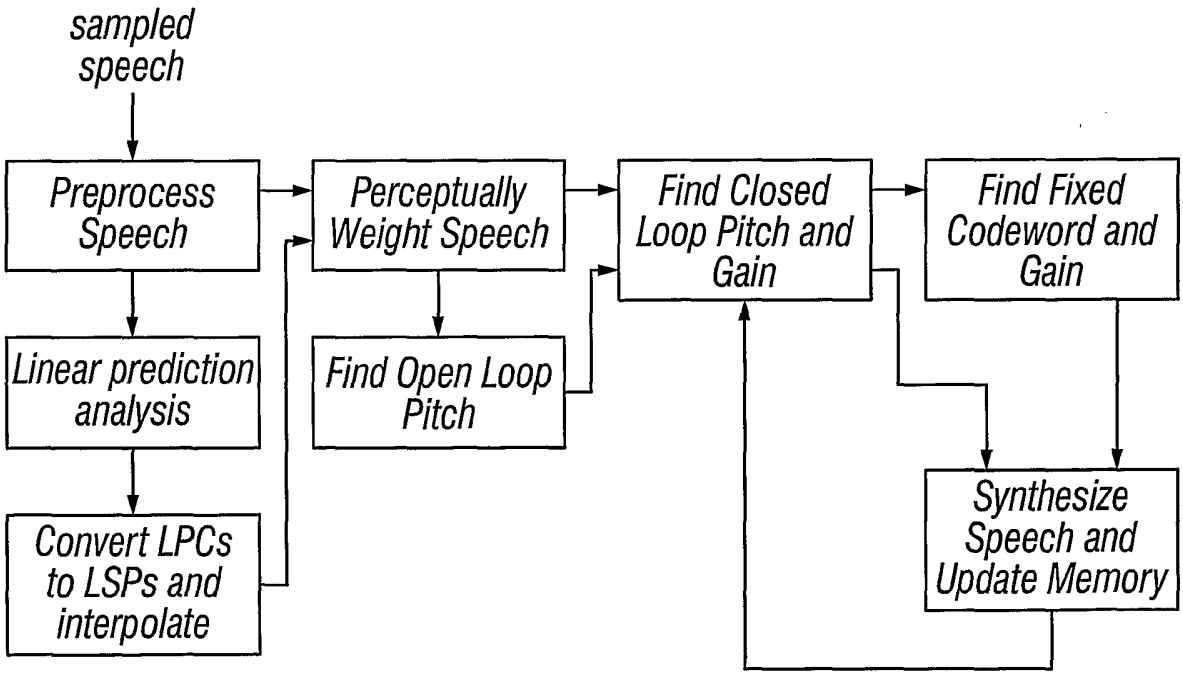


FIG. 2
(Prior Art)

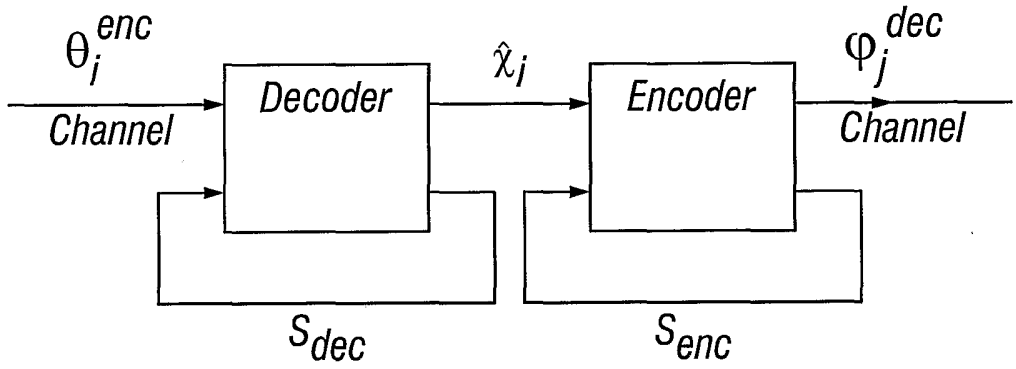


FIG. 3
(Prior Art)

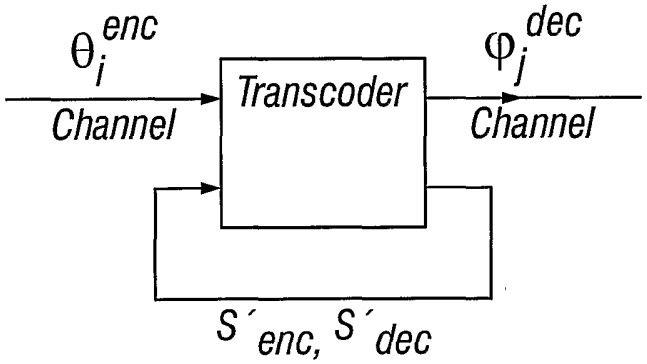


FIG. 4
(Prior Art)

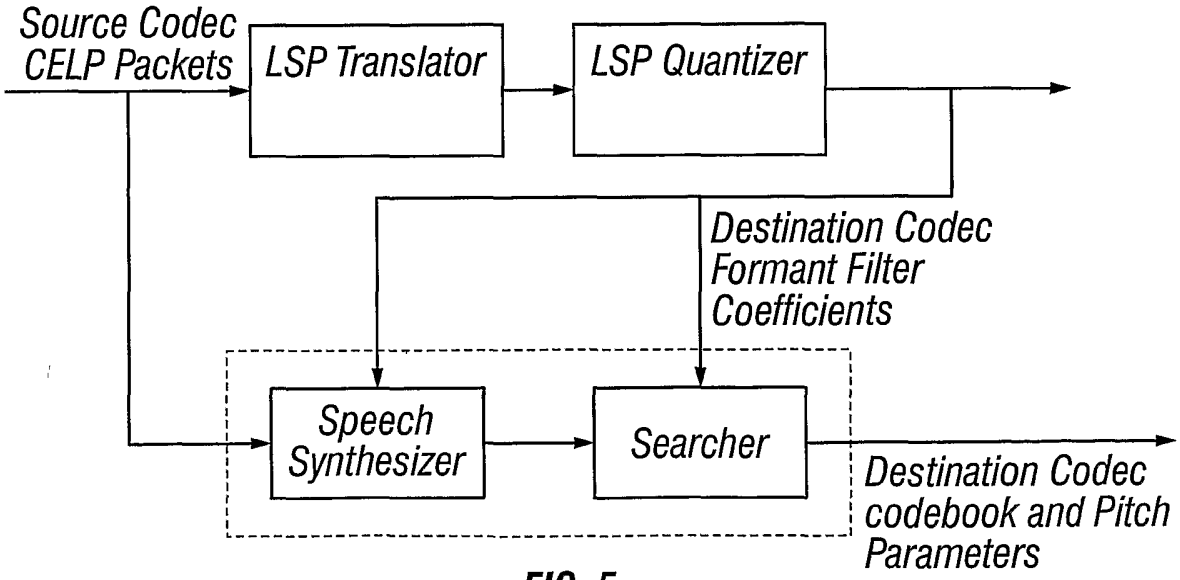


FIG. 5
(Prior Art)

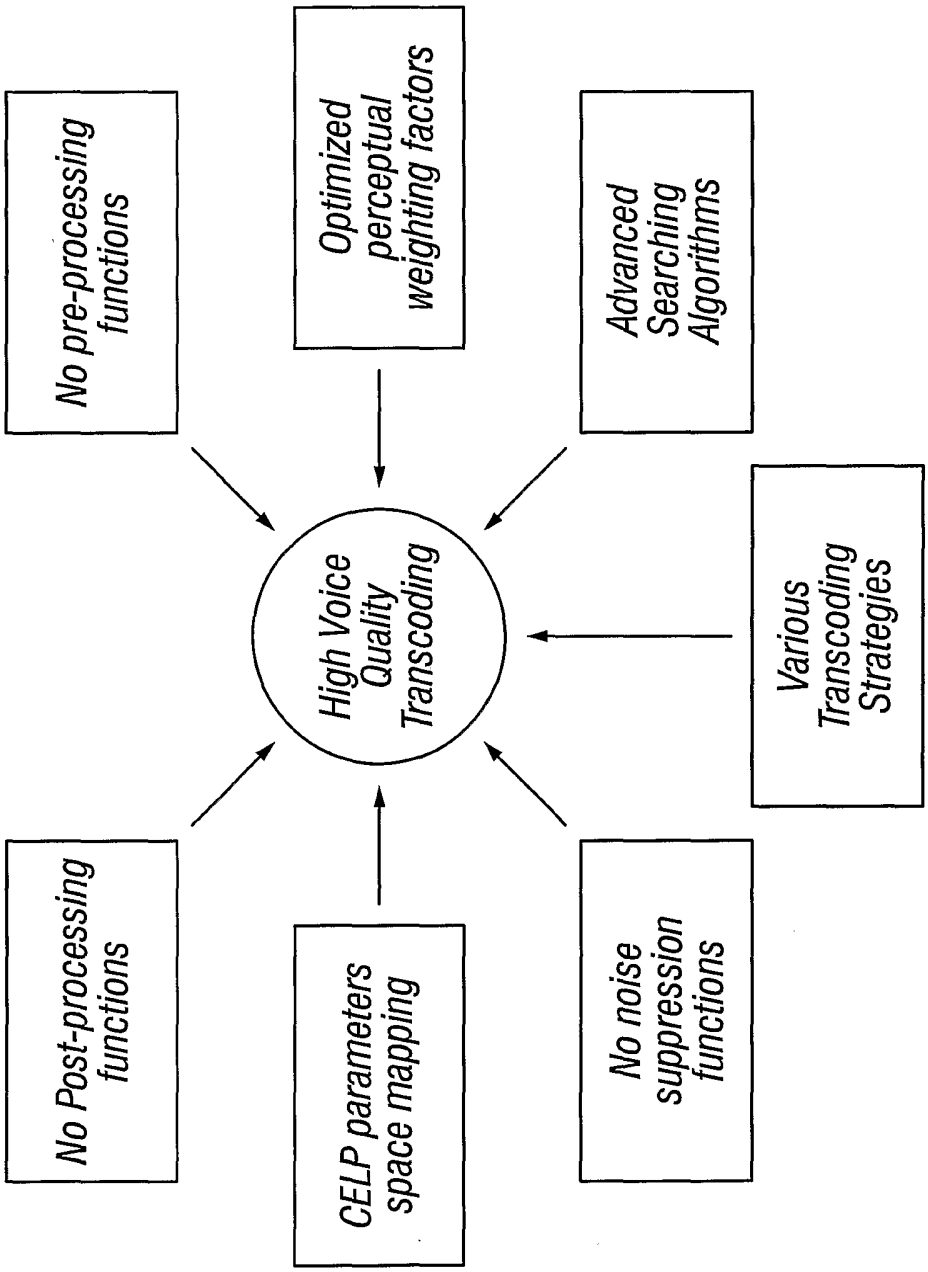

No pre-processing functions
Optimized perceptual weighting factors
Advanced Searching Algorithms
High Voice Quality Transcoding
Various Transcoding Strategies
No Post-processing functions
CELP parameters space mapping
No noise suppression functions

FIG. 6

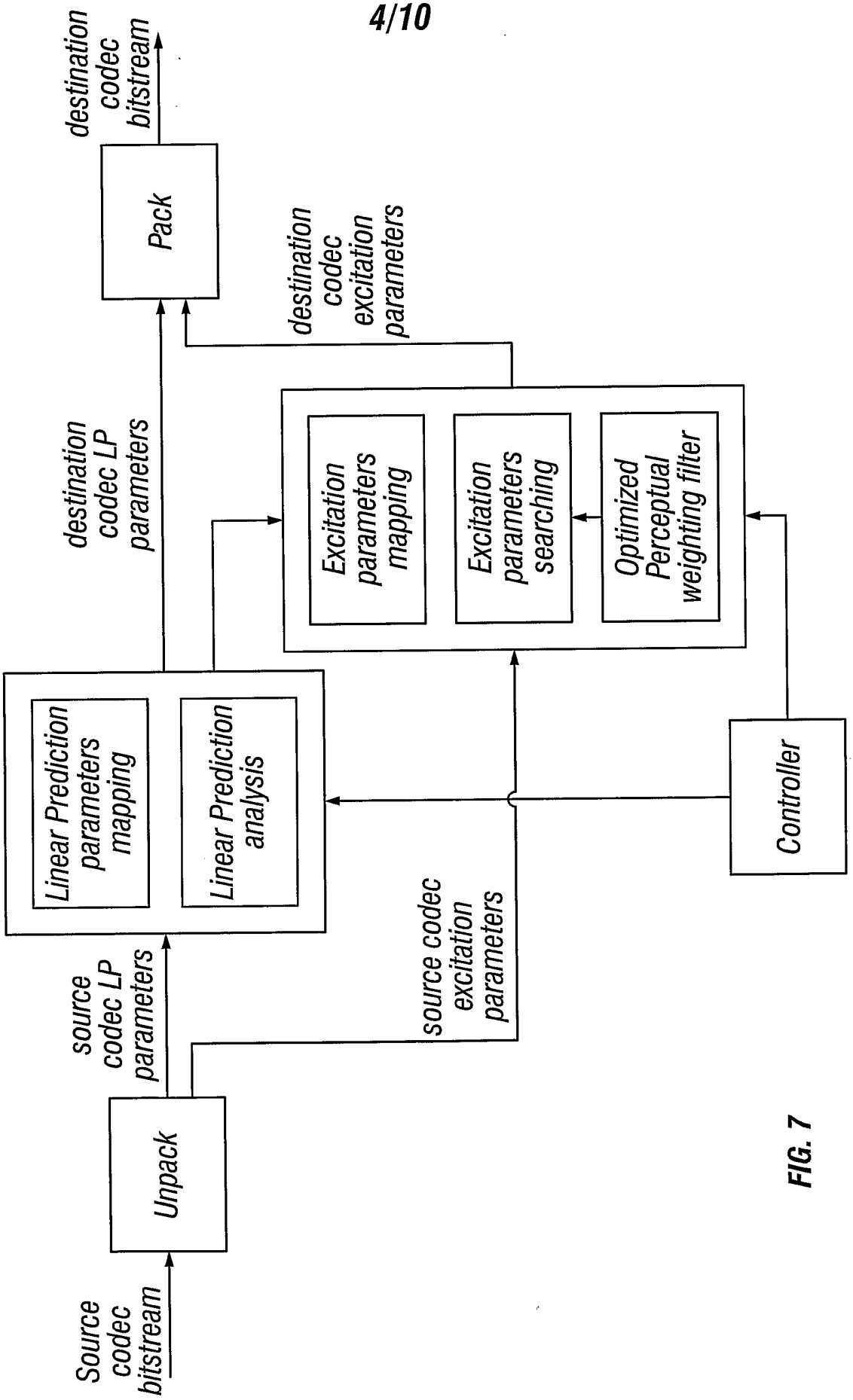

destination codec bitstream
Pack
destination codec excitation parameters
destination codec LP parameters
Excitation parameters mapping
Excitation parameters searching
Optimized Perceptual weighting filter
Linear Prediction parameters mapping
Linear Prediction analysis
Controller
source codec LP parameters
source codec excitation parameters
Unpack
Source codec bitstream

FIG. 7

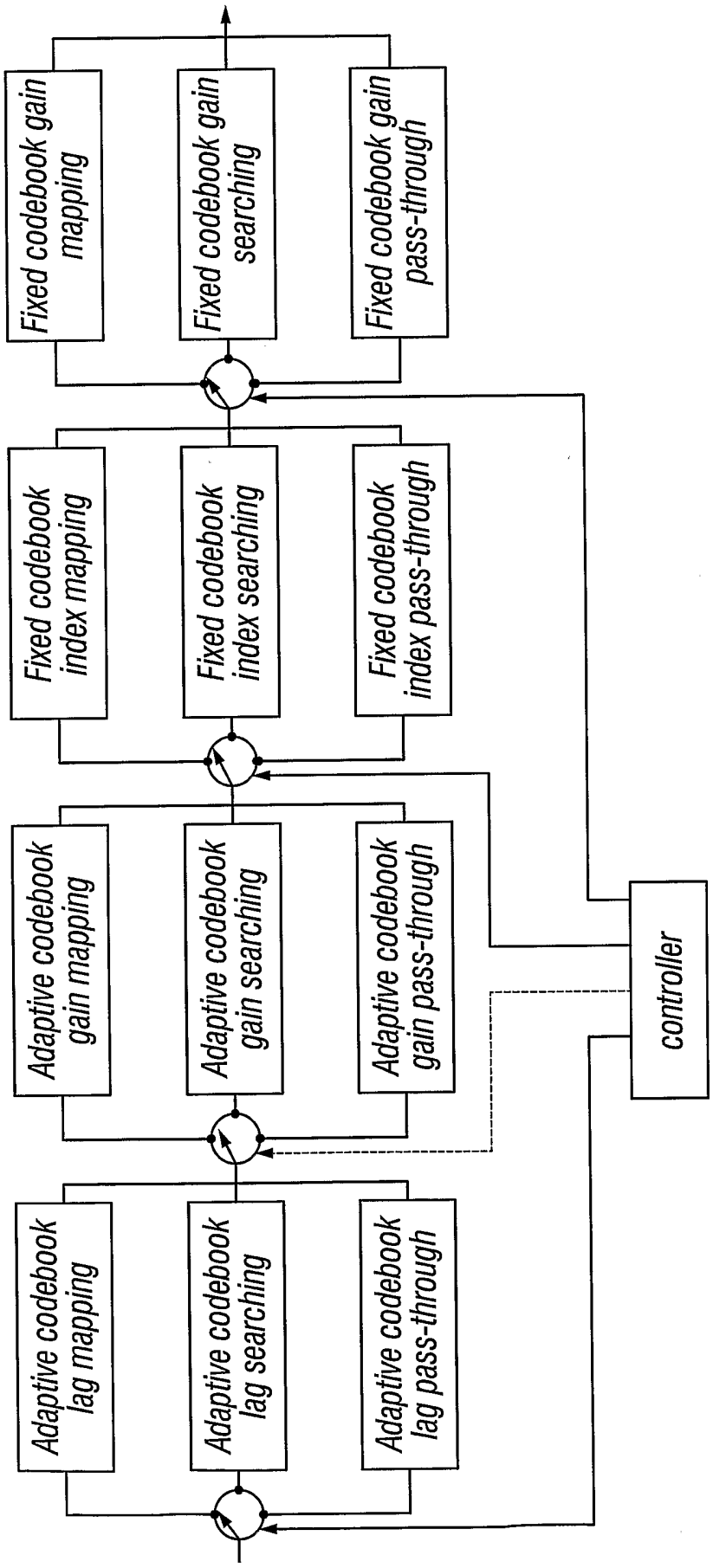

Fixed codebook gain mapping
Fixed codebook gain searching
Fixed codebook gain pass-through
Fixed codebook index mapping
Fixed codebook index searching
Fixed codebook index pass-through
Adaptive codebook gain mapping
Adaptive codebook gain searching
Adaptive codebook gain pass-through
Adaptive codebook lag mapping
Adaptive codebook lag searching
Adaptive codebook lag pass-through
controller

FIG. 8

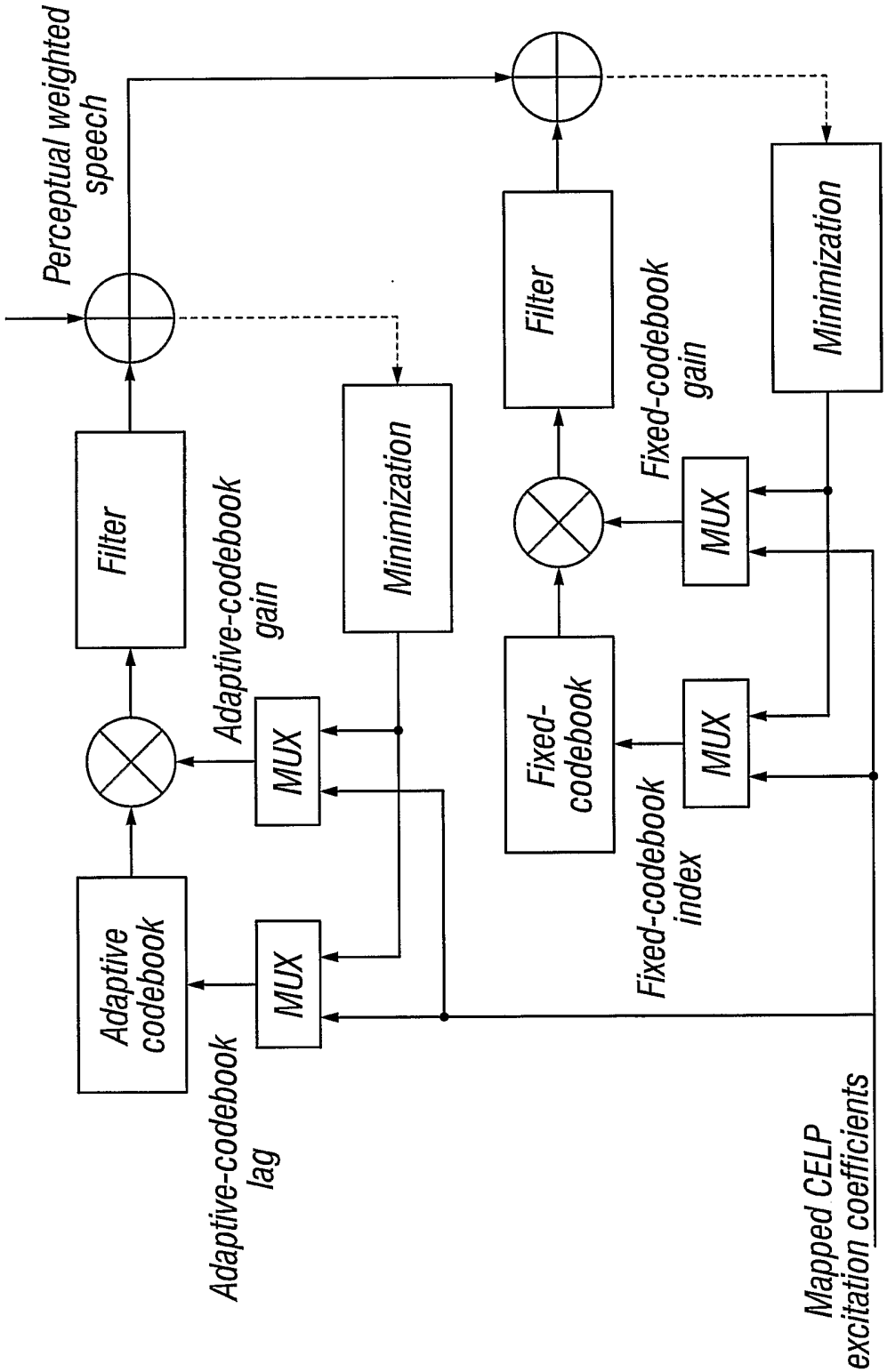

Perceptual weighted speech
Filter
Adaptive codebook
Adaptive-codebook lag
MUX
Adaptive-codebook gain
MUX
Minimization
Filter
Fixed-codebook
Fixed-codebook index
MUX
Fixed-codebook gain
MUX
Minimization
Mapped CELP excitation coefficients

FIG. 9

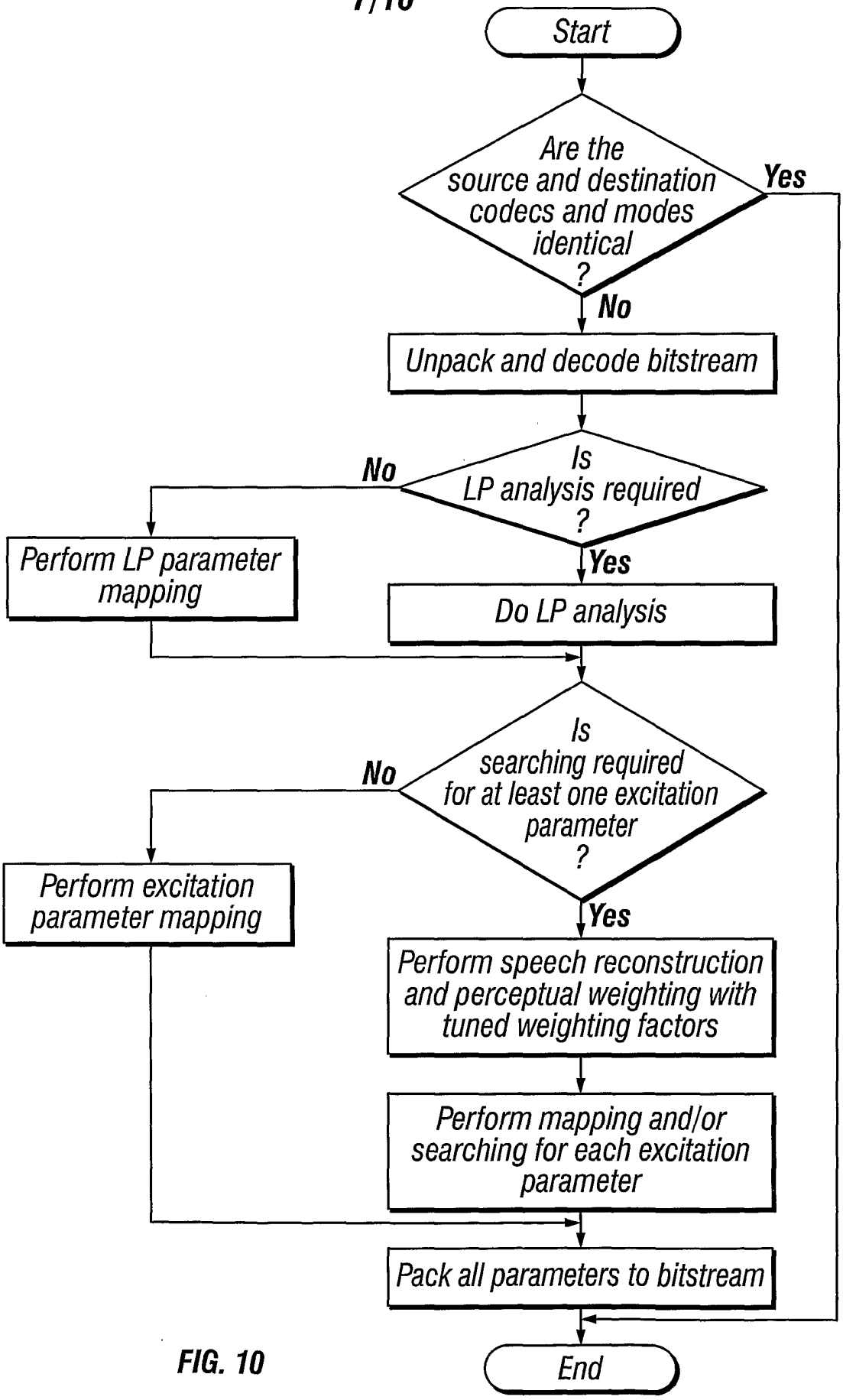

Start
Are the source and destination codecs and modes identical ?
Yes
No
Unpack and decode bitstream
Is LP analysis required ?
No
Yes
Perform LP parameter mapping
Do LP analysis
Is searching required for at least one excitation parameter ?
No
Yes
Perform excitation parameter mapping
Perform speech reconstruction and perceptual weighting with tuned weighting factors
Perform mapping and/or searching for each excitation parameter
Pack all parameters to bitstream
End

FIG. 10

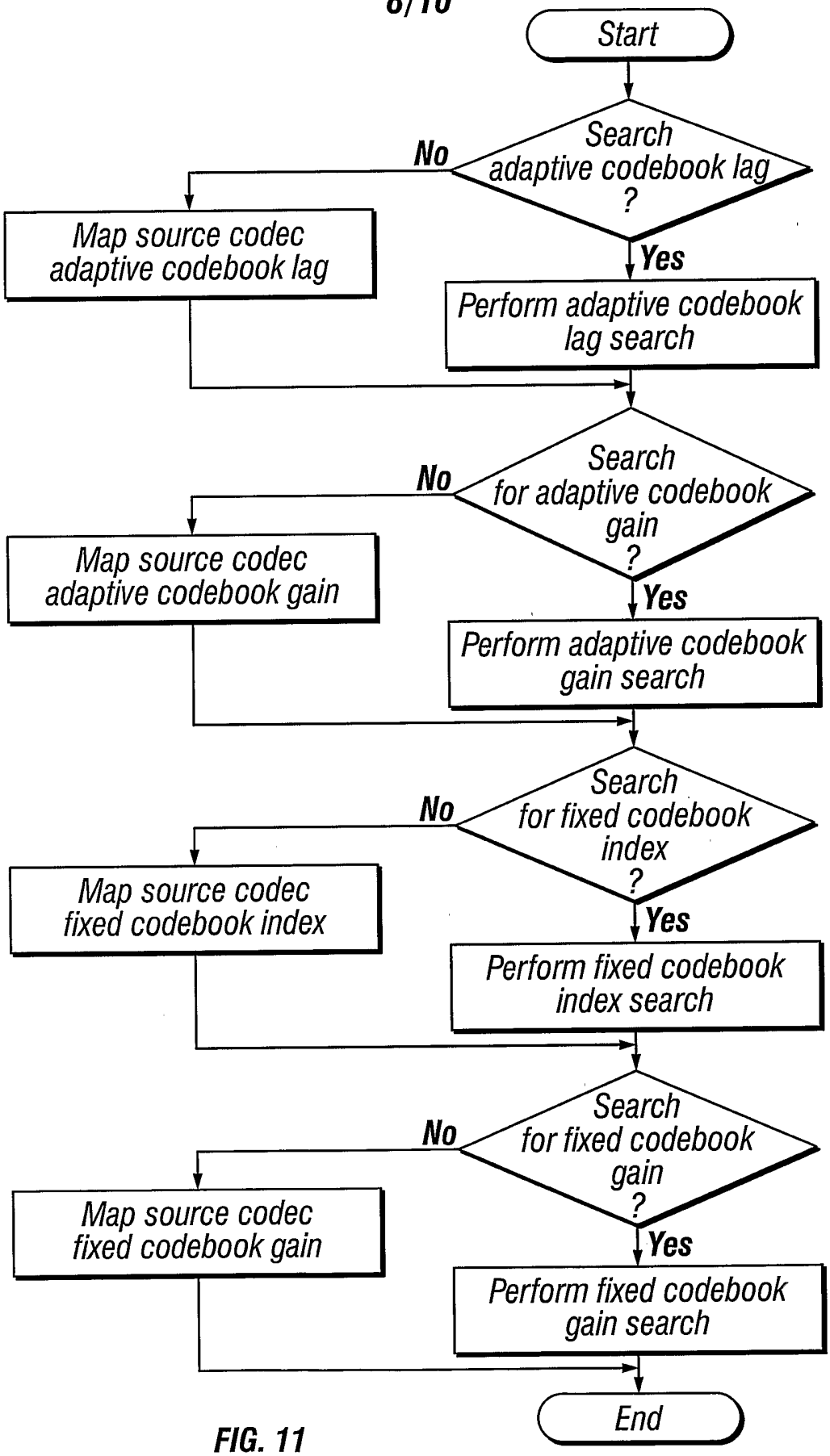

Start
Search adaptive codebook lag ?
No
Yes
Map source codec adaptive codebook lag
Perform adaptive codebook lag search
Search for adaptive codebook gain ?
No
Yes
Map source codec adaptive codebook gain
Perform adaptive codebook gain search
Search for fixed codebook index ?
No
Yes
Map source codec fixed codebook index
Perform fixed codebook index search
Search for fixed codebook gain ?
No
Yes
Map source codec fixed codebook gain
Perform fixed codebook gain search
End

FIG. 11

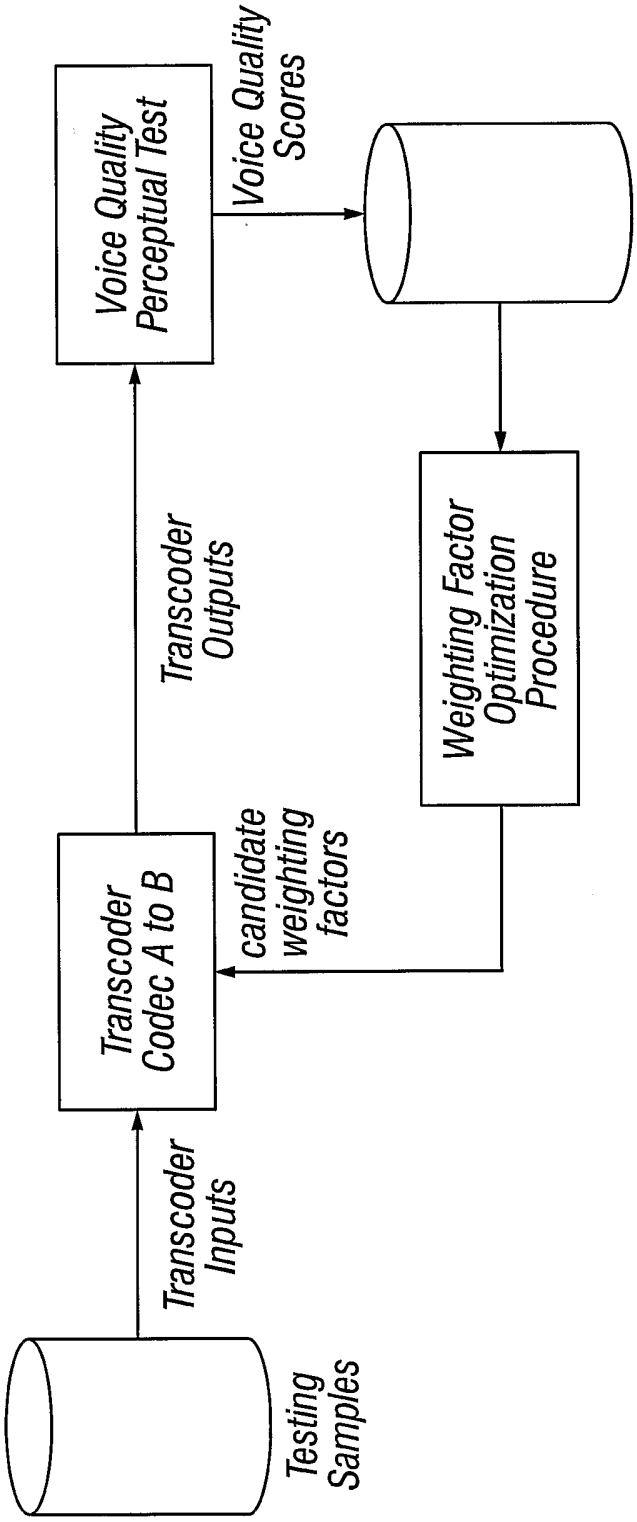

Voice Quality Perceptual Test
Voice Quality Scores
Transcoder Outputs
Weighting Factor Optimization Procedure
Transcoder Codec A to B
candidate weighting factors
Transcoder Inputs
Testing Samples

FIG. 12

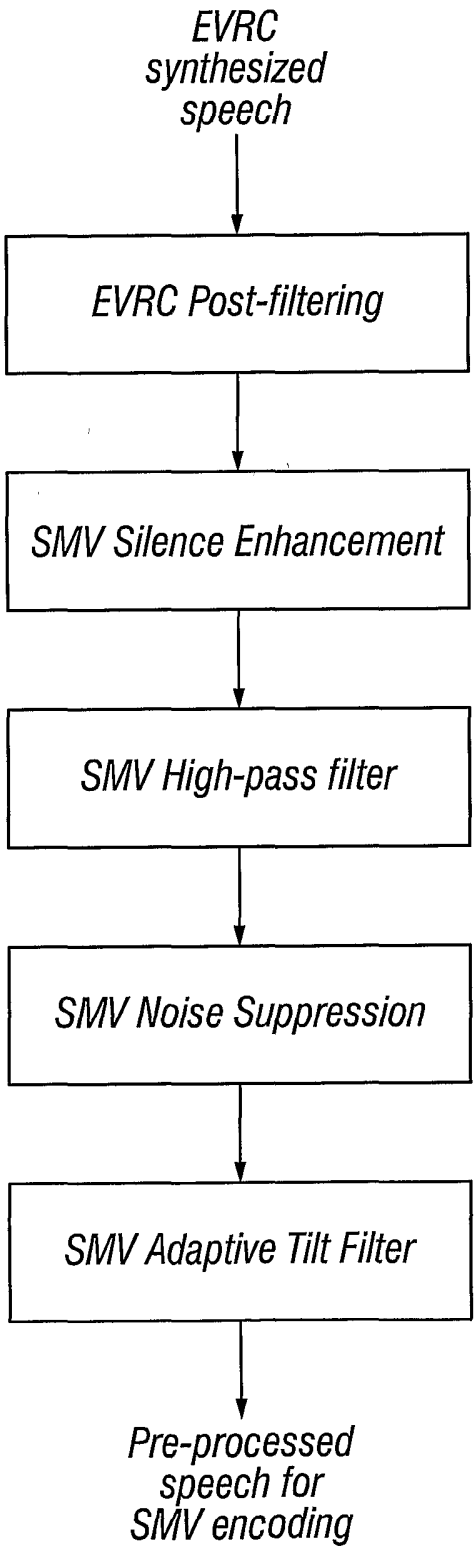

EVRC synthesized speech
EVRC Post-filtering
SMV Silence Enhancement
SMV High-pass filter
SMV Noise Suppression
SMV Adaptive Tilt Filter
Pre-processed speech for SMV encoding

FIG. 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2004/000014

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: G10L 19/12, H03M 7/00

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)

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
dwpi (transcode, map, format, parameter, weight, percept)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2003/058407 A2 (DILITHIUM NETWORKS, INC) 17 July 2003 The whole document	
A	WO 2001/069936 A2 (SONY CORPORATION) 20 September 2001 The whole document	

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 February 2004	03 MAR 2004
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer **J. LAW** Telephone No : (02) 6283 2179

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/AU2004/000014

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
WO	03058407	US	2003177004	US	2004002855	WO	03079330
WO	0169936	AU	41122/01	CA	2374067	EP	1177691
		US	2002157112				

END OF ANNEX