



US006807524B1

(12) **United States Patent**
Bessette et al.

(10) **Patent No.:** US 6,807,524 B1
(45) **Date of Patent:** Oct. 19, 2004

(54) **PERCEPTUAL WEIGHTING DEVICE AND METHOD FOR EFFICIENT CODING OF WIDEBAND SIGNALS**

6,006,174 A * 12/1999 Lin et al. 704/201
6,064,962 A * 5/2000 Oshikiri et al. 704/262
6,192,334 B1 2/2001 Nomura
6,449,590 B1 * 9/2002 Gao 704/219

(75) Inventors: **Bruno Bessette**, Rock Forest (CA);
Redwan Salami, Sherbrooke (CA);
Roch Lefebvre, Canton de Magog (CA)

FOREIGN PATENT DOCUMENTS

EP 0 465 057 A1 1/1992
EP 0465057 A 1/1992
EP 0732686 A 9/1996
EP 0 732 686 A2 9/1996
JP 02-012300 A 1/1990
JP 03-116199 A 5/1991
JP 6-348300 A 12/1994
JP 10-282997 A 10/1998
WO WO 96/21220 7/1996

(73) Assignee: **Voiceage Corporation**, Quebec (CA)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

OTHER PUBLICATIONS

“Predictive Coding of Speech Signals and Subjective Error Criteria” by Bishnus S. Atal et al., IEEE Transaction ASSP, vol. 27, No. 3, pp. 247–254 Jun. 1979.

(21) Appl. No.: **09/830,276**

(22) PCT Filed: **Oct. 27, 1999**

(86) PCT No.: **PCT/CA99/01010**

§ 371 (c)(1),
(2), (4) Date: **Jun. 20, 2001**

* cited by examiner

(87) PCT Pub. No.: **WO00/25304**

PCT Pub. Date: **May 4, 2000**

Primary Examiner—Tāivaldis Ivars Šmits
Assistant Examiner—James S. Wozniak
(74) *Attorney, Agent, or Firm*—Birch, Stewart, Kolasch & Birch, LLP

Foreign Application Priority Data

Oct. 27, 1998 (CA) 2252170

(51) **Int. Cl.**⁷ **G10L 19/04**

(52) **U.S. Cl.** **704/200.1; 704/219; 704/262; 704/224**

(58) **Field of Search** 704/222, 201, 704/219, 262, 224

(57) **ABSTRACT**

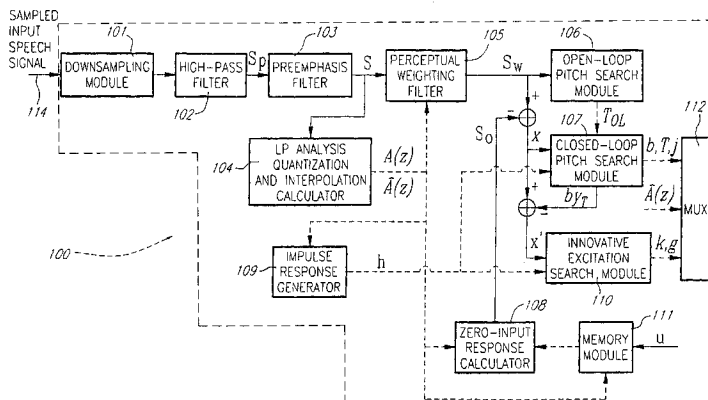
A perceptual weighting device for producing a perceptually weighted signal in response to a wideband signal comprises a signal pre-emphasis filter, a synthesis filter calculator, and a perceptual weighting filter. The signal pre-emphasis filter enhances the high frequency content of the wideband signal to thereby produce a pre-emphasized signal. The signal pre-emphasis filter has a transfer function of the form: $P(z)=1-\mu z^{-1}$, wherein μ is a pre-emphasis factor having a value located between 0 and 1. The synthesis filter calculator is responsive to the pre-emphasized signal for producing synthesis filter coefficients. Finally, the perceptual weighting filter processes the pre-emphasized signal in relation to the synthesis filter coefficients to produce the perceptually weighted signal. The perceptual weighting filter has a transfer function, with fixed denominator, of the form: $W(z)=A(z/\gamma_1)/(1-\gamma_2 z^{-1})$ where $0<\gamma_2<\gamma_1\leq 1$.

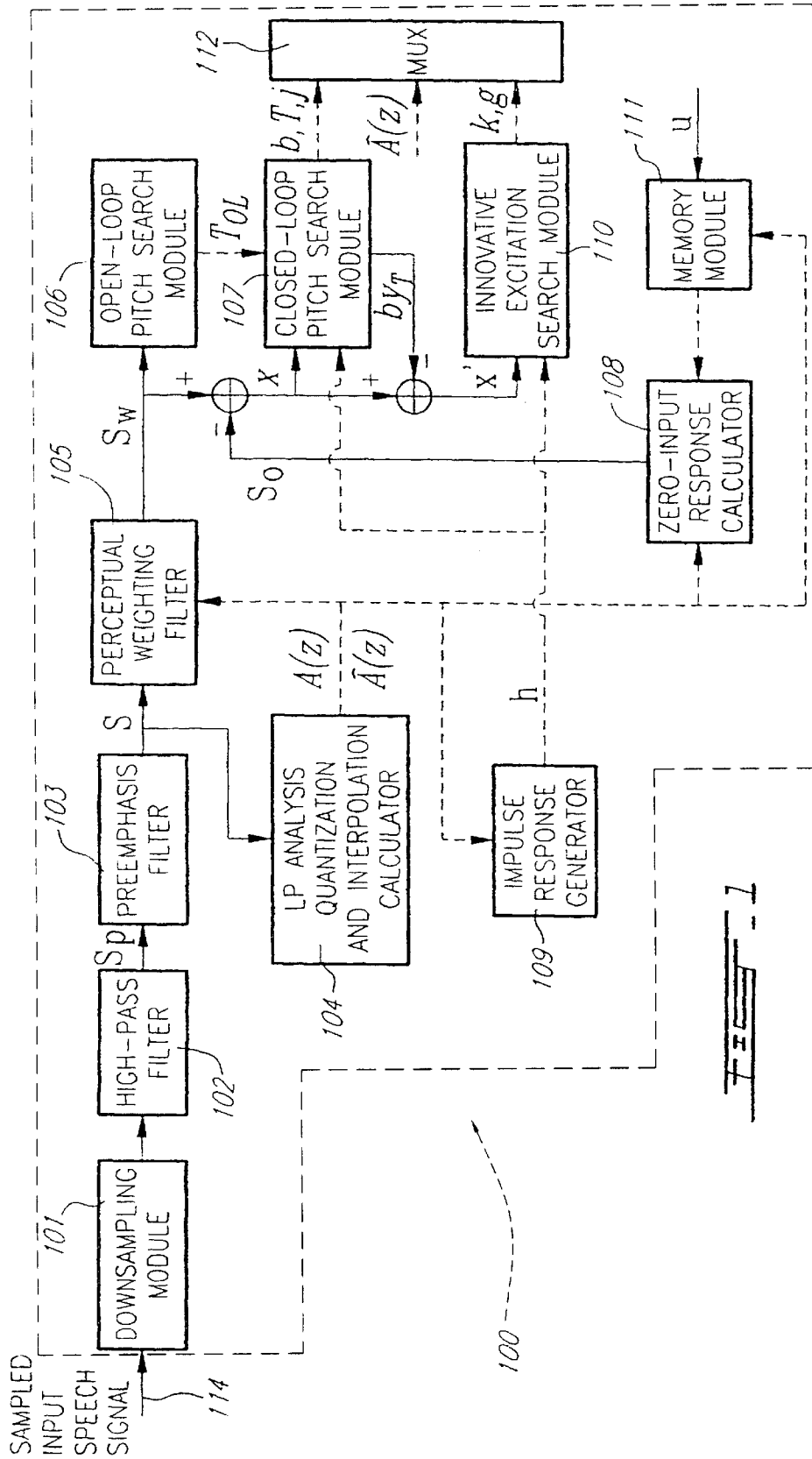
(56) **References Cited**

U.S. PATENT DOCUMENTS

4,932,061 A 6/1990 Kroon et al.
5,307,441 A * 4/1994 Tzeng 704/222
5,359,696 A * 10/1994 Gerson et al. 704/223
5,444,816 A 8/1995 Adoul et al.
5,519,807 A 5/1996 Cellario et al.
5,664,055 A * 9/1997 Kroon 704/223
5,699,482 A 12/1997 Adoul et al.
5,701,392 A 12/1997 Adoul et al.
5,754,976 A 5/1998 Adoul et al.
5,963,898 A 10/1999 Navarro et al.

49 Claims, 4 Drawing Sheets





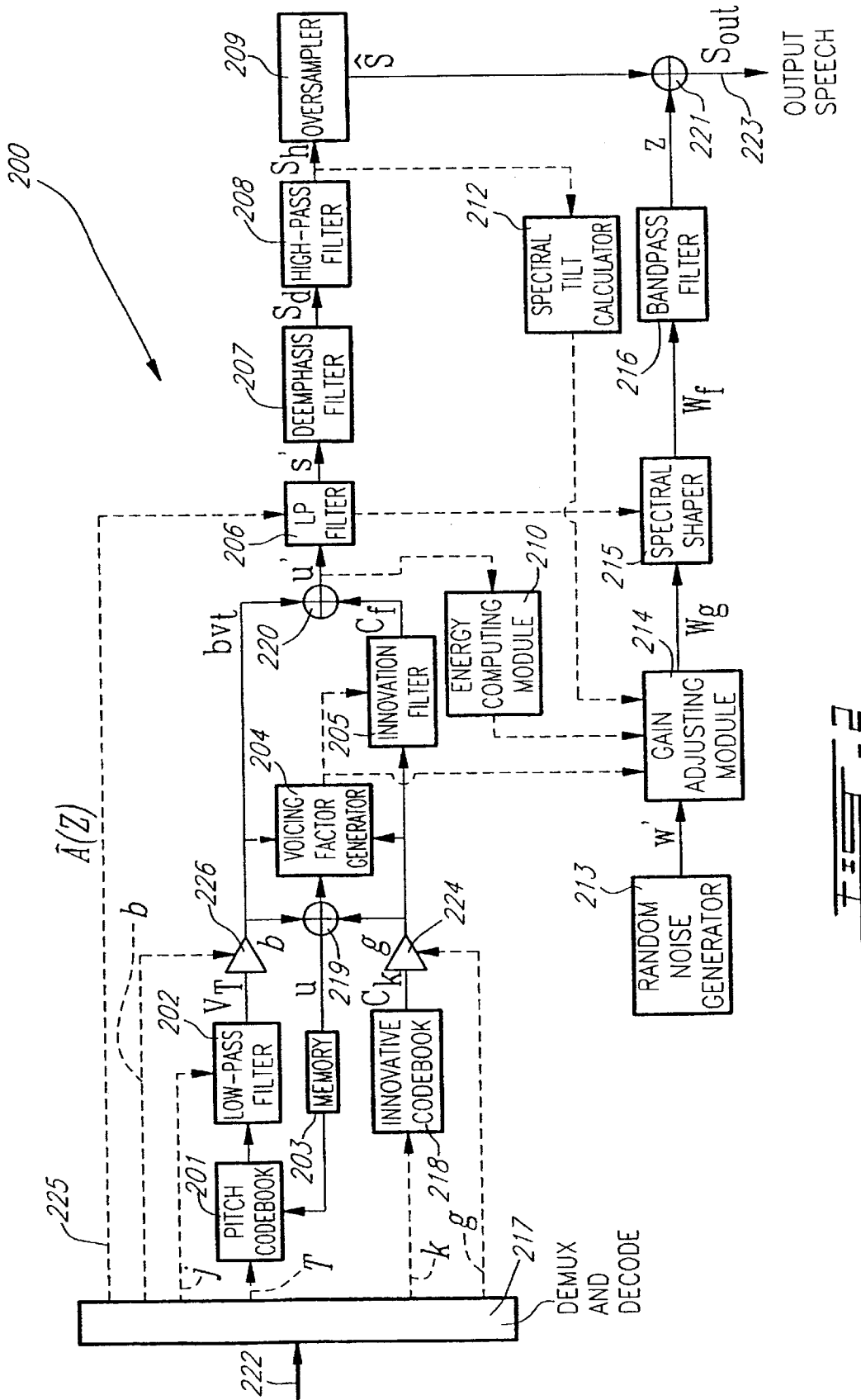
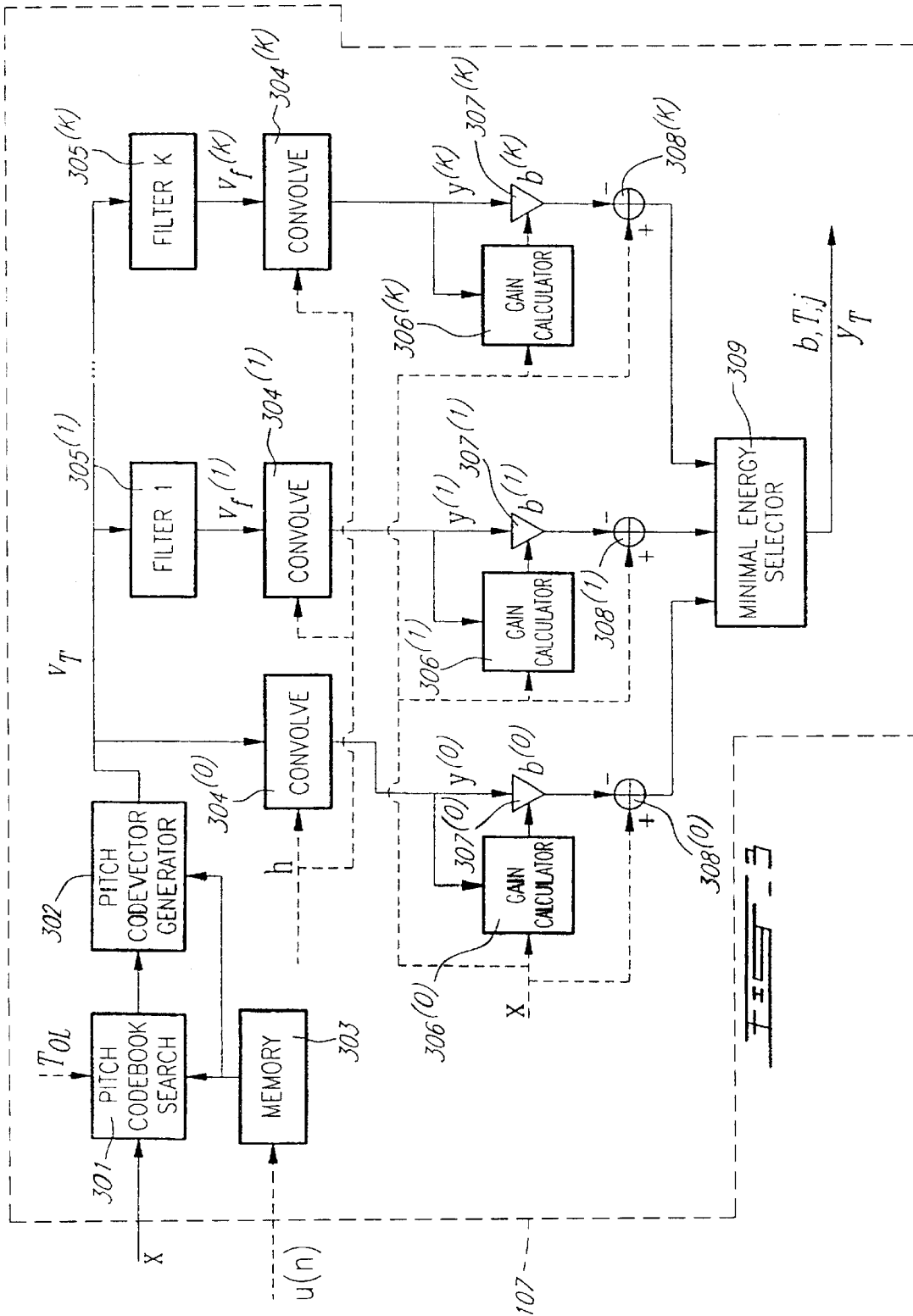
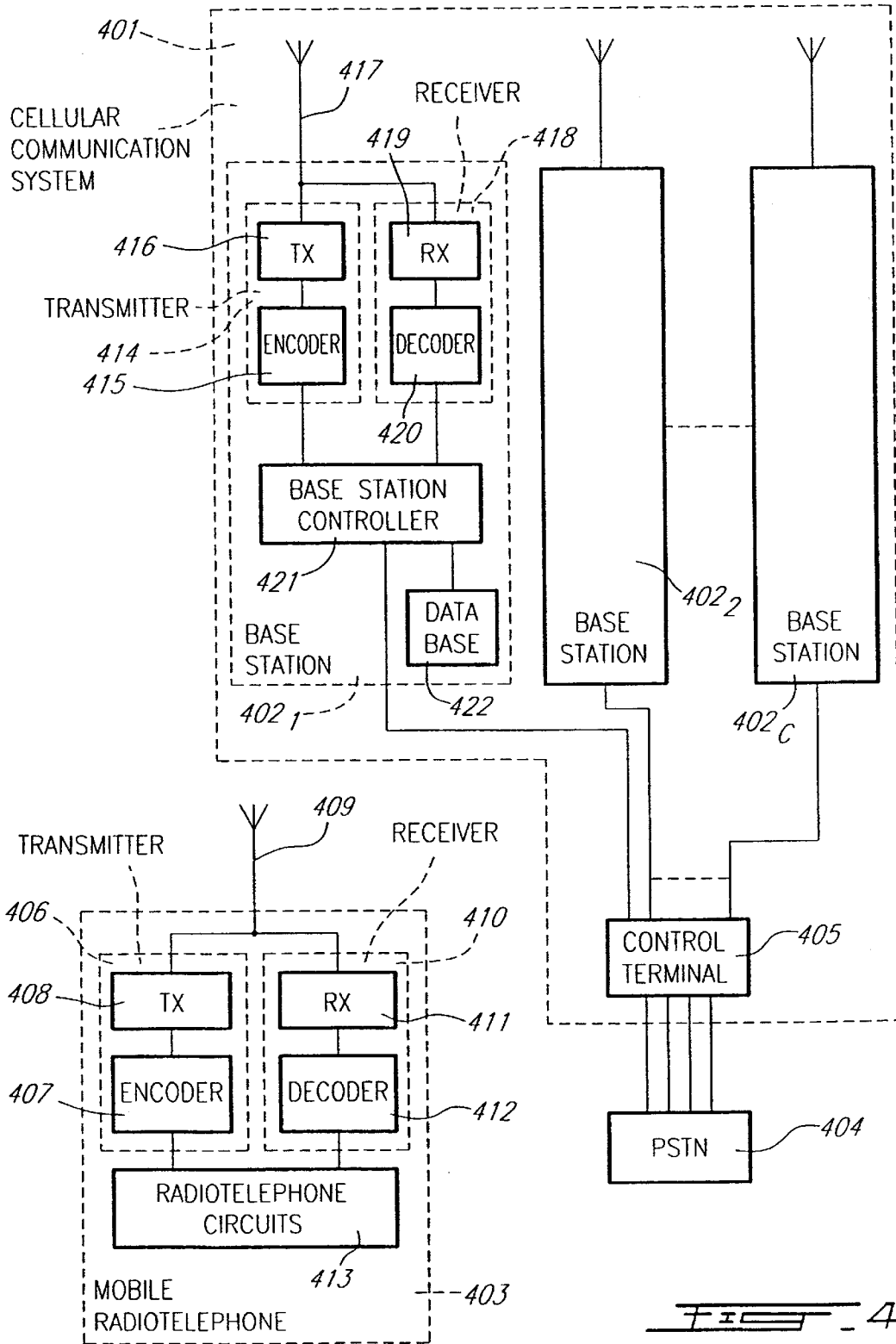


FIG. 2





PERCEPTUAL WEIGHTING DEVICE AND METHOD FOR EFFICIENT CODING OF WIDEBAND SIGNALS

This application is the national phase under 35 U.S.C. §371 of PCT International Application No. PCT/CA99/01010 which has an International filing date of Oct. 27, 1999, which designated the United States of America and was published in English.

BACKGROUND OF THE INVENTION

1. Field of the invention

The present invention relates to a perceptual weighting device and method for producing a perceptually weighted signal in response to a wideband signal (0–7000 Hz) in order to reduce a difference between a weighted wideband signal and a subsequently synthesized weighted wideband signal.

2. Brief description of the prior art

The demand for efficient digital wideband speech/audio encoding techniques with a good subjective quality/bit rate trade-off is increasing for numerous applications such as audio/video teleconferencing, multimedia, and wireless applications, as well as Internet and packet network applications. Until recently, telephone bandwidths filtered in the range 200–3400 Hz were mainly used in speech coding applications. However, there is an increasing demand for wideband speech applications in order to increase the intelligibility and naturalness of the speech signals. A bandwidth in the range 50–7000 Hz was found sufficient for delivering a face-to-face speech quality. For audio signals, this range gives an acceptable audio quality, but is still lower than the CD quality which operates on the range 20–20000 Hz.

A speech encoder converts a speech signal into a digital bitstream which is transmitted over a communication channel (or stored in a storage medium). The speech signal is digitized (sampled and quantized with usually 16-bits per sample) and the speech encoder has the role of representing these digital samples with a smaller number of bits while maintaining a good subjective speech quality. The speech decoder or synthesizer operates on the transmitted or stored bit stream and converts it back to a sound signal.

One of the best prior art techniques capable of achieving a good quality/bit rate trade-off is the so-called Code Excited Linear Prediction (CELP) technique. According to this technique, the sampled speech signal is processed in successive blocks of L samples usually called frames where L is some predetermined number (corresponding to 10–30 ms of speech). In CELP, a linear prediction (LP) synthesis filter is computed and transmitted every frame. The L-sample frame is then divided into smaller blocks called subframes of size N samples, where $L=kN$ and k is the number of subframes in a frame (N usually corresponds to 4–10 ms of speech). An excitation signal is determined in each subframe, which usually consists of two components: one from the past excitation (also called pitch contribution or adaptive codebook) and the other from an innovative codebook (also called fixed codebook). This excitation signal is transmitted and used at the decoder as the input of the LP synthesis filter in order to obtain the synthesized speech.

An innovative codebook in the CELP context, is an indexed set of N-sample-long sequences which will be referred to as N-dimensional codevectors. Each codebook sequence is indexed by an integer k ranging from 1 to M where M represents the size of the codebook often expressed as a number of bits b, where $M=2^b$.

To synthesize speech according to the CELP technique, each block of N samples is synthesized by filtering an appropriate codevector from a codebook through time varying filters modelling the spectral characteristics of the

speech signal. At the encoder end, the synthesis output is computed for all, or a subset, of the codevectors from the codebook (codebook search). The retained codevector is the one producing the synthesis output closest to the original speech signal according to a perceptually weighted distortion measure. This perceptual weighting is performed using a so-called perceptual weighting filter, which is usually derived from the LP synthesis filter.

The CELP model has been very successful in encoding telephone band sound signals, and several CELP-based standards exist in a wide range of applications, especially in digital cellular applications. In the telephone band, the sound signal is band-limited to 200–3400 Hz and sampled at 8000 samples/sec. In wideband speech/audio applications, the sound signal is band-limited to 50–7000 Hz and sampled at 16000 samples/sec.

Some difficulties arise when applying the telephone-band optimized CELP model to wideband signals, and additional features need to be added to the model in order to obtain high quality wideband signals. Wideband signals exhibit a much wider dynamic range compared to telephone-band signals, which results in precision problems when a fixed-point implementation of the algorithm is required (which is essential in wireless applications). Furthermore, the CELP model will often spend most of its encoding bits on the low-frequency region, which usually has higher energy contents, resulting in a low-pass output signal. To overcome this problem, the perceptual weighting filter has to be modified in order to suit wideband signals, and pre-emphasis techniques which boost the high frequency regions become important to reduce the dynamic range, yielding a simpler fixed-point implementation, and to ensure a better encoding of the higher frequency contents of the signal.

In CELP-type encoders, the optimum pitch and innovative parameters are searched by minimizing the mean squared error between the input speech and synthesized speech in a perceptually weighted domain. This is equivalent to minimizing the error between the weighted input speech and weighted synthesis speech, where the weighting is performed using a filter having a transfer function $W(z)$ of the form:

$$W(z)=A(z/g_1)/A(z/g_2) \text{ where } 0 < \Gamma_2 < \Gamma_1 \leq 1.$$

In analysis-by-synthesis (AbS) coders, analysis show that the quantization error is weighted by the inverse of the weighting filter, $W^{-1}(z)$, which exhibits some of the formant structure in the input signal. Thus, the masking property of the human ear is exploited by shaping the error, so that it has more energy in the formant regions, where it will be masked by the strong signal energy present in those regions. The amount of weighting is controlled by the factors Γ_1 and Γ_2 .

This filter works well with telephone band signals. However, it was found that this filter is not suitable for efficient perceptual weighting when it was applied to wideband signals. It was found that this filter has inherent limitations in modelling the formant structure and the required spectral tilt concurrently. The spectral tilt is more pronounced in wideband signals due to the wide dynamic range between low and high frequencies. It was suggested to add a tilt filter into filter $W(z)$ in order to control the tilt and formant weighting separately.

OBJECT OF THE INVENTION

An object of the present invention is therefore to provide a perceptual weighting device and method adapted to wideband signals, using a modified perceptual weighting filter to obtain a high quality reconstructed signal, these device and method enabling fixed point algorithmic implementation.

SUMMARY OF THE INVENTION

More specifically, in accordance with the present invention, there is provided a perceptual weighting device

3

for producing a perceptually weighted signal in response to a wideband signal in order to reduce a difference between a weighted wideband signal and a subsequently synthesized weighted wideband signal. This perceptual weighting device comprises:

- a) a signal preemphasis filter responsive to the wideband signal for enhancing the high frequency content of the wideband signal to thereby produce a preemphasised signal;
- b) a synthesis filter calculator responsive to the preemphasised signal for producing synthesis filter coefficients; and
- c) a perceptual weighting filter, responsive to the preemphasised signal and the synthesis filter coefficients, for filtering the preemphasised signal in relation to the perceptually weighted signal. The perceptual weighting filter has a transfer function with fixed denominator whereby weighting of the wideband signal in a formant region is substantially decoupled from a spectral tilt of that wideband signal.

The present invention also relates to a method for producing a perceptually weighted signal in response to a wideband signal in order to reduce a difference between a weighted wideband signal and a subsequently synthesized weighted wideband signal. This method comprises: filtering the wideband signal to produce a preemphasised signal with enhanced high frequency content; calculating, from the preemphasised signal, synthesis filter coefficients; and filtering the preemphasised signal in relation to the synthesis filter coefficients to thereby produce a perceptually weighted speech signal. The filtering comprises processing the preemphasis signal through a perceptual weighting filter having a transfer function with fixed denominator whereby weighting of the wideband signal in a formant region is substantially decoupled from a spectral tilt of the wideband signal.

In accordance with preferred embodiments of the subject invention:

reduction of the dynamic range comprises filtering the wideband signal through a transfer function of the form:

$$P(z)=1-\mu z^{-1}$$

wherein μ is a preemphasis factor having a value located between 0 and 1;

the preemphasis factor μ is 0.7;

the perceptual weighting filter has a transfer function of the form:

$$W(z)=A(z/\gamma_1)/(1-\gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values; and

the variable γ_2 is set equal to μ .

Therefore, the overall perceptual weighting of the quantization error is obtained by a combination of a preemphasis filter and a modified weighting filter to enable high subjective quality of the decoded wideband sound signal into filter $W(z)$ in order to control the tilt and formant weighting separately.

The solution to the problem exposed in the brief description of the prior art is accordingly to introduce a preemphasis filter at the input, compute the synthesis filter coefficients based on the preemphasized signal, and use a modified perceptual weighting filter by fixing its denominator. By reducing the dynamic range of the wideband signal, the preemphasis filter renders the wideband signal more suitable

4

for fixed-point implementation, and improves the encoding of the high frequency contents of the spectrum.

The present invention further relates to an encoder for encoding a wideband signal, comprising: a) a perceptual weighting device as described herein above; b) an pitch codebook search device responsive to the perceptually weighted signal for producing pitch codebook parameters and an innovative search target vector; c) an innovative codebook search device, responsive to the synthesis filter coefficients and to the innovative search target vector, for producing innovative codebook parameters; and d) a signal forming device for producing an encoded wideband signal comprising the pitch codebook parameters, the innovative codebook parameters, and the synthesis filter coefficients.

Still further in accordance with the present invention, there is provided:

a cellular communication system for servicing a large geographical area divided into a plurality of cells, comprising: a) mobile transmitter/receiver units; b) cellular base stations respectively situated in the cells; c) a control terminal for controlling communication between the cellular base stations; d) a bidirectional wireless communication sub-system between each mobile unit situated in one cell and the cellular base station of this cell, this bidirectional wireless communication sub-system comprising, in both the mobile unit and the cellular base station:

- i) a transmitter including an encoder as described hereinabove for encoding a wideband signal and a transmission circuit for transmitting the encoded wideband signal; and
- ii) a receiver including a receiving circuit for receiving a transmitted encoded wideband signal and a decoder for decoding the received encoded wideband signal.

a cellular mobile transmitter/receiver unit comprising:

- a) a transmitter including an encoder as described hereinabove for encoding a wideband signal and a transmission circuit for transmitting the encoded wideband signal; and
- b) a receiver including a receiving circuit for receiving a transmitted encoded wideband signal and a decoder for decoding the received encoded wideband signal;

a cellular network element comprising:

- a) a transmitter including an encoder as described hereinabove for encoding a wideband signal and a transmission circuit for transmitting the encoded wideband signal; and
- b) a receiver including a receiving circuit for receiving a transmitted encoded wideband signal and a decoder for decoding the received encoded wideband signal; and

a bidirectional wireless communication sub-system between each mobile unit situated in one cell and the cellular base station of this cell, this bidirectional wireless communication sub-system comprising, in both the mobile unit and the cellular base station:

- a) a transmitter including an encoder as described hereinabove for encoding a wideband signal and a transmission circuit for transmitting the encoded wideband signal; and
- b) a receiver including a receiving circuit for receiving a transmitted encoded wideband signal and a decoder for decoding the received encoded wideband signal.

The objects, advantages and other features of the present invention will become more apparent upon reading of the following non restrictive description of preferred embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

In the appended drawings:

FIG. 1 is a schematic block diagram of a preferred embodiment of wideband encoding device;

FIG. 2 is a schematic block diagram of a preferred embodiment of wideband decoding device;

FIG. 3 is a schematic block diagram of a preferred embodiment of pitch analysis device; and

FIG. 4 is a simplified, schematic block diagram of a cellular communication system in which the wideband encoding device of FIG. 1 and the wideband decoding device of FIG. 2 can be used.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

As well known to those of ordinary skill in the art, a cellular communication system such as 401 (see FIG. 4) provides a telecommunication service over a large geographic area by dividing that large geographic area into a number C of smaller cells. The C smaller cells are serviced by respective cellular base stations 402₁, 402₂ . . . 402_C, to provide each cell with radio signalling, audio and data channels.

Radio signalling channels are used to page mobile radiotelephones (mobile transmitter/receiver units) such as 403 within the limits of the coverage area (cell) of the cellular base station 402, and to place calls to other radiotelephones 403 located either inside or outside the base station's cell or to another network such as the Public Switched Telephone Network (PSTN) 404.

Once a radiotelephone 403 has successfully placed or received a call, an audio or data channel is established between this radiotelephone 403 and the cellular base station 402 corresponding to the cell in which the radiotelephone 403 is situated, and communication between the base station 402 and radiotelephone 403 is conducted over that audio or data channel. The radiotelephone 403 may also receive control or timing information over a signalling channel while a call is in progress.

If a radiotelephone 403 leaves a cell and enters another adjacent cell while a call is in progress, the radiotelephone 403 hands over the call to an available audio or data channel of the new cell base station 402. If a radiotelephone 403 leaves a cell and enters another adjacent cell while no call is in progress, the radiotelephone 403 sends a control message over the signalling channel to log into the base station 402 of the new cell. In this manner mobile communication over a wide geographical area is possible.

The cellular communication system 401 further comprises a control terminal 405 to control communication between the cellular base stations 402 and the PSTN 404, for example during a communication between a radiotelephone 403 and the PSTN 404, or between a radiotelephone 403 located in a first cell and a radiotelephone 403 situated in a second cell.

Of course, a bidirectional wireless radio communication subsystem is required to establish an audio or data channel between a base station 402 of one cell and a radiotelephone 403 located in that cell. As illustrated in very simplified form in FIG. 4, such a bidirectional wireless radio communication subsystem typically comprises in the radiotelephone 403:

- a transmitter 406 including:
 - an encoder 407 for encoding the voice signal; and
 - a transmission circuit 408 for transmitting the encoded voice signal from the encoder 407 through an antenna such as 409; and
- a receiver 410 including:

- a receiving circuit 411 for receiving a transmitted encoded voice signal usually through the same antenna 409; and

- a decoder 412 for decoding the received encoded voice signal from the receiving circuit 411.

The radiotelephone further comprises other conventional radiotelephone circuits 413 to which the encoder 407 and decoder 412 are connected and for processing signals therefrom, which circuits 413 are well known to those of ordinary skill in the art and, accordingly, will not be further described in the present specification.

Also, such a bidirectional wireless radio communication subsystem typically comprises in the base station 402:

- a transmitter 414 including:

- an encoder 415 for encoding the voice signal; and
- a transmission circuit 416 for transmitting the encoded voice signal from the encoder 415 through an antenna such as 417; and

- a receiver 418 including:

- a receiving circuit 419 for receiving a transmitted encoded voice signal through the same antenna 417 or through another antenna (not shown); and
- a decoder 420 for decoding the received encoded voice signal from the receiving circuit 419.

The base station 402 further comprises, typically, a base station controller 421, along with its associated database 422, for controlling communication between the control terminal 405 and the transmitter 414 and receiver 418.

As well known to those of ordinary skill in the art, voice encoding is required in order to reduce the bandwidth necessary to transmit sound signal, for example voice signal such as speech, across the bidirectional wireless radio communication subsystem, i.e., between a radiotelephone 403 and a base station 402.

LP voice encoders (such as 415 and 407) typically operating at 13 kbits/second and below such as Code-Excited Linear Prediction (CELP) encoders typically use a LP synthesis filter to model the short-term spectral envelope of the voice signal. The LP information is transmitted, typically, every 10 or 20 ms to the decoder (such 420 and 412) and is extracted at the decoder end.

The novel techniques disclosed in the present specification may apply to different LP-based coding systems. However, a CELP-type coding system is used in the preferred embodiment for the purpose of presenting a non-limitative illustration of these techniques. In the same manner, such techniques can be used with sound signals other than voice and speech as well with other types of wideband signals.

FIG. 1 shows a general block diagram of a CELP-type speech encoding device 100 modified to better accommodate wideband signals.

The sampled input speech signal 114 is divided into successive L-sample blocks called "frames". In each frame, different parameters representing the speech signal in the frame are computed, encoded, and transmitted. LP parameters representing the LP synthesis filter are usually computed once every frame. The frame is further divided into smaller blocks of N samples (blocks of length N), in which excitation parameters (pitch and innovation) are determined. In the CELP literature, these blocks of length N are called "subframes" and the N-sample signals in the subframes are referred to as N-dimensional vectors. In this preferred embodiment, the length N corresponds to 5 ms while the length L corresponds to 20 ms, which means that a frame contains four subframes (N=80 at the sampling rate of 16 kHz and 64 after down-sampling to 12.8 kHz). Various N-dimensional vectors occur in the encoding procedure. A list of the vectors which appear in FIGS. 1 and 2 as well as a list of transmitted parameters are given herein below:

List of the Main N-dimensional Vectors

- s Wideband signal input speech vector (after down-sampling, pre-processing, and preemphasis);
 s_w Weighted speech vector;
 s_o Zero-input response of weighted synthesis filter;
 s_p Down-sampled pre-processed signal; Oversampled synthesized speech signal;
 s' Synthesis signal before deemphasis;
 s_d Deemphasized synthesis signal;
 s_n Synthesis signal after deemphasis and postprocessing;
 x Target vector for pitch search;
 x' Target vector for innovation search;
 h Weighted synthesis filter impulse response;
 v_T Adaptive (pitch) codebook vector at delay T;
 y_T Filtered pitch codebook vector (v_T convolved with h);
 c_k Innovative codevector at index k (k-th entry from the innovation codebook);
 c_i Enhanced scaled innovation codevector;
 u Excitation signal (scaled innovation and pitch codevectors);
 u' Enhanced excitation;
 z Band-pass noise sequence;
 w' White noise sequence; and
 w Scaled noise sequence.

List of Transmitted Parameters

- STP Short term prediction parameters (defining $A(z)$);
 T Pitch lag (or pitch codebook index);
 b Pitch gain (or pitch codebook gain);
 j Index of the low-pass filter used on the pitch codevector;
 k Codevector index (innovation codebook entry); and
 g Innovation codebook gain.

In this preferred embodiment, the STP parameters are transmitted once per frame and the rest of the parameters are transmitted four times per frame (every subframe).

Encoder Side

The sampled speech signal is encoded on a block by block basis by the encoding device **100** of FIG. 1 which is broken down into eleven modules numbered from **101** to **111**.

The input speech is processed into the above mentioned L-sample blocks called frames.

Referring to FIG. 1, the sampled input speech signal is down-sampled in a down-sampling module **101**. For example, the signal is down-sampled from 16 kHz down to 12.8 kHz, using techniques well known to those of ordinary skill in the art. Down-sampling down to another frequency can of course be envisaged. Down-sampling increases the coding efficiency, since a smaller frequency bandwidth is encoded. This also reduces the algorithmic complexity since the number of samples in a frame is decreased. The use of down-sampling becomes significant when the bit rate is reduced below 16 kbit/s, although down-sampling is not essential above 16 kbit/s.

After down-sampling, the 320-sample frame of 20 ms is reduced to 256-sample frame (down-sampling ratio of 4/5).

The input frame is then supplied to the optional pre-processing block **102**. Pre-processing block **102** may consist of a high-pass filter with a 50 Hz cut-off frequency. High-pass filter **102** removes the unwanted sound components below 50 Hz.

The down-sampled pre-processed signal is denoted by $s_p(n)$, $n=0, 1, 2, \dots, L-1$, where L is the length of the frame (256 at a sampling frequency of 12.8 kHz). In a preferred embodiment of the preemphasis filter **103**, the signal $s_p(n)$ is

preemphasized using a filter having the following transfer function:

$$P(z)=1-\mu z^{-1}$$

where μ is a preemphasis factor with a value located between 0 and 1 (a typical value is $\mu=0.7$). A higher-order filter could also be used. It should be pointed out that high-pass filter **102** and preemphasis filter **103** can be interchanged to obtain more efficient fixed-point implementations.

The function of the preemphasis filter **103** is to enhance the high frequency contents of the input signal. It also reduces the dynamic range of the input speech signal, which renders it more suitable for fixed-point implementation. Without preemphasis, LP analysis in fixed-point using single-precision arithmetic is difficult to implement.

Preemphasis also plays an important role in achieving a proper overall perceptual weighting of the quantization error, which contributes to improved sound quality. This will be explained in more detail herein below.

The output of the preemphasis filter **103** is denoted $s(n)$. This signal is used for performing LP analysis in calculator module **104**. LP analysis is a technique well known to those of ordinary skill in the art. In this preferred embodiment, the autocorrelation approach is used. In the autocorrelation approach, the signal $s(n)$ is first windowed using a Hamming window (having usually a length of the order of 30–40 ms). The autocorrelations are computed from the windowed signal, and Levinson-Durbin recursion is used to compute LP filter coefficients, a_i , where $i=1, \dots, p$, and where p is the LP order, which is typically 16 in wideband coding. The parameters a_i are the coefficients of the transfer function of the LP filter, which is given by the following relation:

$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i}$$

LP analysis is performed in calculator module **104**, which also performs the quantization and interpolation of the LP filter coefficients. The LP filter coefficients are first transformed into another equivalent domain more suitable for quantization and interpolation purposes. The line spectral pair (LSP) and immittance spectral pair (ISP) domains are two domains in which quantization and interpolation can be efficiently performed. The 16 LP filter coefficients, a_i , can be quantized in the order of 30 to 50 bits using split or multi-stage quantization, or a combination thereof. The purpose of the interpolation is to enable updating the LP filter coefficients every subframe while transmitting them once every frame, which improves the encoder performance without increasing the bit rate. Quantization and interpolation of the LP filter coefficients is believed to be otherwise well known to those of ordinary skill in the art and, accordingly, will not be further described in the present specification.

The following paragraphs will describe the rest of the coding operations performed on a subframe basis. In the following description, the filter $A(z)$ denotes the unquantized interpolated LP filter of the subframe, and the filter $\hat{A}(z)$ denotes the quantized interpolated LP filter of the subframe.

Perceptual Weighting:

In analysis-by-synthesis encoders, the optimum pitch and innovation parameters are searched by minimizing the mean squared error between the input speech and synthesized speech in a perceptually weighted domain. This is equivalent to minimizing the error between the weighted input speech and weighted synthesis speech.

The weighted signal $s_w(n)$ is computed in a perceptual weighting filter **105**. Traditionally, the weighted signal $s_w(n)$

is computed by a weighting filter having a transfer function $W(z)$ in the form:

$$W(z)=A(z/\gamma_1)/A(z/\gamma_2) \text{ where } 0<\gamma_2<\gamma_1\leq 1$$

As well known to those of ordinary skill in the art, in prior art analysis-by-synthesis (AbS) encoders, analysis shows that the quantization error is weighted by a transfer function $W^{-1}(z)$, which is the inverse of the transfer function of the perceptual weighting filter **105**. This result is well described by B. S. Atal and M. R. Schroeder in "Predictive coding of speech and subjective error criteria", IEEE Transaction ASSP, vol. 27, no. 3, pp. 247-254, June 1979. Transfer function $W^{-1}(z)$ exhibits some of the formant structure of the input speech signal. Thus, the masking property of the human ear is exploited by shaping the quantization error so that it has more energy in the formant regions where it will be masked by the strong signal energy present in these regions. The amount of weighting is controlled by the factors γ_1 and γ_2 .

The above traditional perceptual weighting filter **105** works well with telephone band signals. However, it was found that this traditional perceptual weighting filter **105** is not suitable for efficient perceptual weighting of wideband signals. It was also found that the traditional perceptual weighting filter **105** has inherent limitations in modelling the formant structure and the required spectral tilt concurrently. The spectral tilt is more pronounced in wideband signals due to the wide dynamic range between low and high frequencies. The prior art has suggested to add a tilt filter into $W(z)$ in order to control the tilt and formant weighting of the wideband input signal separately.

A novel solution to this problem is, in accordance with the present invention, to introduce the preemphasis filter **103** at the input, compute the LP filter $A(z)$ based on the preemphasized speech $s(n)$, and use a modified filter $W(z)$ by fixing its denominator.

LP analysis is performed in module **104** on the preemphasized signal $s(n)$ to obtain the LP filter $A(z)$. Also, a new perceptual weighting filter **105** with fixed denominator is used. An example of transfer function for the perceptual weighting filter **104** is given by the following relation:

$$W(z)=A(z/\gamma_1)/(1-\gamma_2z^{-1}) \text{ where } 0<\gamma_2<\gamma_1\leq 1$$

A higher order can be used at the denominator. This structure substantially decouples the formant weighting from the tilt.

Note that because $A(z)$ is computed based on the preemphasized speech signal $s(n)$, the tilt of the filter $1/A(z/\gamma_1)$ is less pronounced compared to the case when $A(z)$ is computed based on the original speech. Since deemphasis is performed at the decoder end using a filter having the transfer function:

$$P^{-1}(z)=1/(1-\mu z^{-1}),$$

the quantization error spectrum is shaped by a filter having a transfer function $W^{-1}(z)P^{-1}(z)$. When γ_2 is set equal to μ , which is typically the case, the spectrum of the quantization error is shaped by a filter whose transfer function is $1/A(z/\gamma_1)$, with $A(z)$ computed based on the preemphasized speech signal. Subjective listening showed that this structure for achieving the error shaping by a combination of preemphasis and modified weighting filtering is very efficient for encoding wideband signals, in addition to the advantages of ease of fixed-point algorithmic implementation.

Pitch Analysis:

In order to simplify the pitch analysis, an open-loop pitch lag T_{OL} is first estimated in the open-loop pitch search module **106** using the weighted speech signal $s_w(n)$. Then

the closed-loop pitch analysis, which is performed in closed-loop pitch search module **107** on a subframe basis, is restricted around the open-loop pitch lag T_{OL} which significantly reduces the search complexity of the LTP parameters T and b (pitch lag and pitch gain). Open-loop pitch analysis is usually performed in module **106** once every 10 ms (two subframes) using techniques well known to those of ordinary skill in the art.

The target vector x for LTP (Long Term Prediction) analysis is first computed. This is usually done by subtracting the zero-input response S_0 of weighted synthesis filter $W(z)/\hat{A}(z)$ from the weighted speech signal $s_w(n)$. This zero-input response s_0 is calculated by a zero-input response calculator **108**. More specifically, the target vector x is calculated using the following relation:

$$x=s_w-s_0$$

where x is the N -dimensional target vector, S_w is the weighted speech vector in the subframe, and s_0 is the zero-input response of filter $W(z)/\hat{A}(z)$ which is the output of the combined filter $W(z)/\hat{A}(z)$ due to its initial states. The zero-input response calculator **108** is responsive to the quantized interpolated LP filter $\hat{A}(z)$ from the LP analysis, quantization and interpolation calculator **104** and to the initial states of the weighted synthesis filter $W(z)/\hat{A}(z)$ stored in memory module **111** to calculate the zero-input response s_0 (that part of the response due to the initial states as determined by setting the inputs equal to zero) of filter $W(z)/\hat{A}(z)$. This operation is well known to those of ordinary skill in the art and, accordingly, will not be further described.

Of course, alternative but mathematically equivalent approaches can be used to compute the target vector x .

A N -dimensional impulse response vector h of the weighted synthesis filter $W(z)/\hat{A}(z)$ is computed in the impulse response generator **109** using the LP filter coefficients $A(z)$ and $\hat{A}(z)$ from module **104**. Again, this operation is well known to those of ordinary skill in the art and, accordingly, will not be further described in the present specification.

The closed-loop pitch (or pitch codebook) parameters b , T and j are computed in the closed-loop pitch search module **107**, which uses the target vector x , the impulse response vector h and the open-loop pitch lag T_{OL} as inputs. Traditionally, the pitch prediction has been represented by a pitch filter having the following transfer function:

$$1/(1-bz^{-T})$$

where b is the pitch gain and T is the pitch delay or lag. In this case, the pitch contribution to the excitation signal $u(n)$ is given by $bu(n-T)$, where the total excitation is given by

$$u(n)=bu(n-T)+gc_k(n)$$

with g being the innovative codebook gain and $c_k(n)$ the innovative codevector at index k .

This representation has limitations if the pitch lag T is shorter than the subframe length N . In another representation, the pitch contribution can be seen as a pitch codebook containing the past excitation signal. Generally, each vector in the pitch codebook is a shift-by-one version of the previous vector (discarding one sample and adding a new sample). For pitch lags $T>N$, the pitch codebook is equivalent to the filter structure $(1/(1-bz^{-T}))$, and an pitch codebook vector v_n at pitch lag T is given by $v_T(n)=u(n-T)$, $n=0, \dots, N-1$.

For pitch lags T shorter than N , a vector $v_T(n)$ is built by repeating the available samples from the past excitation until the vector is completed (this is not equivalent to the filter structure).

In recent encoders, a higher pitch resolution is used which significantly improves the quality of voiced sound segments. This is achieved by oversampling the past excitation signal using polyphase interpolation filters. In this case, the vector $v_T(n)$ usually corresponds to an interpolated version of the past excitation, with pitch lag T being a non-integer delay (e.g. 50.25).

The pitch search consists of finding the best pitch lag T and gain b that minimize the mean squared weighted error E between the target vector x and the scaled filtered past excitation. Error E being expressed as:

$$E = \|x - by_T\|^2$$

where y_T is the filtered pitch codebook vector at pitch lag T:

$$y_T(n) = v_T(n) * h(n) = \sum_{i=0}^n v_T(i)h(n-i), n = 0, \dots, N-1.$$

It can be shown that the error E is minimized by maximizing the search criterion

$$C = \frac{x^t y_T}{\sqrt{y_T^t y_T}}$$

where t denotes vector transpose.

In the preferred embodiment of the present invention, a 1/3 subsample pitch resolution is used, and the pitch (pitch codebook) search is composed of three stages.

In the first stage, an open-loop pitch lag T_{OL} is estimated in open-loop pitch search module 106 in response to the weighted speech signal $s_w(n)$. As indicated in the foregoing description, this open-loop pitch analysis is usually performed once every 10 ms (two subframes) using techniques well known to those of ordinary skill in the art.

In the second stage, the search criterion C is searched in the closed-loop pitch search module 107 for integer pitch lags around the estimated open-loop pitch lag T_{OL} (usually ± 5), which significantly simplifies the search procedure. A simple procedure is used for updating the filtered codevector y_T without the need to compute the convolution for every pitch lag.

Once an optimum integer pitch lag is found in the second stage, a third stage of the search (module 107) tests the fractions around that optimum integer pitch lag.

When the pitch predictor is represented by a filter of the form $1/(1-bz^{-T})$, which is a valid assumption for pitch lags $T > N$, the spectrum of the pitch filter exhibits a harmonic structure over the entire frequency range, with a harmonic frequency related to $1/T$. In case of wideband signals, this structure is not very efficient since the harmonic structure in wideband signals does not cover the entire extended spectrum. The harmonic structure exists only up to a certain frequency, depending on the speech segment. Thus, in order to achieve efficient representation of the pitch contribution in voiced segments of wideband speech, the pitch prediction filter needs to have the flexibility of varying the amount of periodicity over the wideband spectrum.

A new method which achieves efficient modeling of the harmonic structure of the speech spectrum of wideband signals is disclosed in the present specification, whereby several forms of low pass filters are applied to the past excitation and the low pass filter with higher prediction gain is selected.

When subsample pitch resolution is used, the low pass filters can be incorporated into the interpolation filters used to obtain the higher pitch resolution. In this case, the third stage of the pitch search, in which the fractions around the

chosen integer pitch lag are tested, is repeated for the several interpolation filters having different low-pass characteristics and the fraction and filter index which maximize the search criterion C are selected.

A simpler approach is to complete the search in the three stages described above to determine the optimum fractional pitch lag using only one interpolation filter with a certain frequency response, and select the optimum low-pass filter shape at the end by applying the different predetermined low-pass filters to the chosen pitch codebook vector v_T and select the low-pass filter which minimizes the pitch prediction error. This approach is discussed in detail below.

FIG. 3 illustrates a schematic block diagram of a preferred embodiment of the proposed approach.

In memory module 303, the past excitation signal $u(n)$, $n < 0$, is stored. The pitch codebook search module 301 is responsive to the target vector x, to the open-loop pitch lag T_{OL} and to the past excitation signal $u(n)$, $n < 0$, from memory module 303 to conduct a pitch codebook (pitch codebook) search minimizing the above-defined search criterion C. From the result of the search conducted in module 301, module 302 generates the optimum pitch codebook vector v_T . Note that since a sub-sample pitch resolution is used (fractional pitch), the past excitation signal $u(n)$, $n < 0$, is interpolated and the pitch codebook vector v_T corresponds to the interpolated past excitation signal. In this preferred embodiment, the interpolation filter (in module 301, but not shown) has a low-pass filter characteristic removing the frequency contents above 7000 Hz.

In a preferred embodiment, K filter characteristics are used; these filter characteristics could be low-pass or band-pass filter characteristics. Once the optimum codevector v_T is determined and supplied by the pitch codevector generator 302, K filtered versions of v_T are computed respectively using K different frequency shaping filters such as 305^(j), where $j=1, 2, \dots, K$. These filtered versions are denoted $v_T^{(j)}$, where $j=1, 2, \dots, K$. The different vectors $v_T^{(j)}$ are convolved in respective modules 304^(j), where $j=0, 1, 2, \dots, K$, with the impulse response h to obtain the vectors $y^{(j)}$, where $j=0, 1, 2, \dots, K$. To calculate the mean squared pitch prediction error for each vector $y^{(j)}$, the value $y^{(j)}$ is multiplied by the gain b by means of a corresponding amplifier 307^(j) and the value $by^{(j)}$ is subtracted from the target vector x by means of a corresponding subtractor 308^(j). Selector 309 selects the frequency shaping filter 305^(j) which minimizes the mean squared pitch prediction error

$$e^{(j)} = \|x - b^{(j)}y^{(j)}\|^2, j=1, 2, \dots, K$$

To calculate the mean squared pitch prediction error $e^{(j)}$ for each value of $y^{(j)}$, the value $y^{(j)}$ is multiplied by the gain b by means of a corresponding amplifier 307^(j) and the value $b^{(j)}y^{(j)}$ is subtracted from the target vector x by means of subtractors 308^(j). Each gain $b^{(j)}$ is calculated in a corresponding gain calculator 306^(j) in association with the frequency shaping filter at index j, using the following relationship:

$$b^{(j)} = x^t y^{(j)} / \|y^{(j)}\|^2$$

In selector 309, the parameters b, T, and are j chosen based on v_T or $v_T^{(j)}$ which minimizes the mean squared pitch prediction error e.

Referring back to FIG. 1, the pitch codebook index T is encoded and transmitted to multiplexer 112. The pitch gain b is quantized and transmitted to multiplexer 112. With this new approach, extra information is needed to encode the index j of the selected frequency shaping filter in multiplexer 112. For example, if three filters are used ($j=0, 1, 2, 3$), then two bits are needed to represent this information. The filter index information j can also be encoded jointly with the pitch gain b.

Innovative codebook search:

Once the pitch, or LTP (Long Term Prediction) parameters b, T, and j are determined, the next step is to search for the optimum innovative excitation by means of search module **110** of FIG. 1. First, the target vector x is updated by subtracting the LTP contribution:

$$x' = x - by_T$$

where b is the pitch gain and y_T is the filtered pitch codebook vector (the past excitation at delay T filtered with the selected low pass filter and convolved with the impulse response h as described with reference to FIG. 3).

The search procedure in CELP is performed by finding the optimum excitation codevector c_k and gain g which minimize the mean-squared error between the target vector and the scaled filtered codevector

$$E = \|x' - gHc_k\|^2$$

where H is a lower triangular convolution matrix derived from the impulse response vector h.

In the preferred embodiment of the present invention, the innovative codebook search is performed in module **110** by means of an algebraic codebook as described in U.S. Pat. No. 5,444,816 (Adoul et al.) issued on Aug. 22, 1995; U.S. Pat. No. 5,699,482 granted to Adoul et al., on Dec. 17, 1997; U.S. Pat. No. 5,754,976 granted to Adoul et al., on May 19, 1998; and U.S. Pat. No. 5,701,392 (Adoul et al.) dated Dec. 23, 1997.

Once the optimum excitation codevector c_k and its gain g are chosen by module **110**, the codebook index k and gain g are encoded and transmitted to multiplexer **112**.

Referring to FIG. 1, the parameters b, T, j, $\hat{A}(z)$, k and g are multiplexed through the multiplexer **112** before being transmitted through a communication channel.

Memory Update:

In memory module **111** (FIG. 1), the states of the weighted synthesis filter $W(z)/\hat{A}(z)$ are updated by filtering the excitation signal $u = gc_k + by_T$ through the weighted synthesis filter. After this filtering, the states of the filter are memorized and used in the next subframe as initial states for computing the zero-input response in calculator module **108**.

As in the case of the target vector x, other alternative but mathematically equivalent approaches well known to those of ordinary skill in the art can be used to update the filter states.

Decoder Side

The speech decoding device **200** of FIG. 2 illustrates the various steps carried out between the digital input **222** (input stream to the demultiplexer **217**) and the output sampled speech **223** (output of the adder **221**).

Demultiplexer **217** extracts the synthesis model parameters from the binary information received from a digital input channel. From each received binary frame, the extracted parameters are:

- the short-term prediction parameters (STP) $\hat{A}(z)$ (once per frame);
- the long-term prediction (LTP) parameters T, b, and j (for each subframe); and
- the innovation codebook index k and gain g (for each subframe).

The current speech signal is synthesized based on these parameters as will be explained hereinbelow.

The innovative codebook **218** is responsive to the index k to produce the innovation codevector c_k , which is scaled by the decoded gain factor g through an amplifier **224**. In the preferred embodiment, an innovative codebook **218** as described in the above mentioned U.S. Pat Nos. 5,444,816; 5,699,482; 5,754,976; and 5,701,392 is used to represent the innovative codevector c_k .

The generated scaled codevector gc_k at the output of the amplifier **224** is processed through an innovation filter **205**. Periodicity Enhancement:

The generated scaled codevector at the output of the amplifier **224** is processed through a frequency-dependent pitch enhancer **205**.

Enhancing the periodicity of the excitation signal u improves the quality in case of voiced segments. This was done in the past by filtering the innovation vector from the innovative codebook (fixed codebook) **218** through a filter in the form $1/(1-\epsilon bz^{-T})$ where ϵ is a factor below 0.5 which controls the amount of introduced periodicity. This approach is less efficient in case of wideband signals since it introduces periodicity over the entire spectrum. A new alternative approach, which is part of the present invention, is disclosed whereby periodicity enhancement is achieved by filtering the innovative codevector c_k from the innovative (fixed) codebook through an innovation filter **205** ($F(z)$) whose frequency response emphasizes the higher frequencies more than lower frequencies. The coefficients of $F(z)$ are related to the amount of periodicity in the excitation signal u.

Many methods known to those skilled in the art are available for obtaining valid periodicity coefficients. For example, the value of gain b provides an indication of periodicity. That is, if gain b is close to 1, the periodicity of the excitation signal u is high, and if gain b is less than 0.5, then periodicity is low.

Another efficient way to derive the filter $F(z)$ coefficients used in a preferred embodiment, is to relate them to the amount of pitch contribution in the total excitation signal u. This results in a frequency response depending on the subframe periodicity, where higher frequencies are more strongly emphasized (stronger overall slope) for higher pitch gains. Innovation filter **205** has the effect of lowering the energy of the innovative codevector c_k at low frequencies when the excitation signal u is more periodic, which enhances the periodicity of the excitation signal u at lower frequencies more than higher frequencies. Suggested forms for innovation filter **205** are

$$F(z) = 1 - \sigma z^{-1}, \tag{1}$$

or

$$F(z) = -\alpha z + 1 - \alpha z^{-1} \tag{2}$$

where σ or α are periodicity factors derived from the level of periodicity of the excitation signal u.

The second three-term form of $F(z)$ is used in a preferred embodiment. The periodicity factor α is computed in the voicing factor generator **204**. Several methods can be used to derive the periodicity factor α based on the periodicity of the excitation signal u. Two methods are presented below.

Method 1:

The ratio of pitch contribution to the total excitation signal u is first computed in voicing factor generator **204** by

$$R_p = \frac{b^2 v_T v_T}{u^T u} = \frac{b^2 \sum_{n=0}^{N-1} v_T^2(n)}{\sum_{n=0}^{N-1} u^2(n)}$$

where v_T is the pitch codebook vector, b is the pitch gain, and u is the excitation signal u given at the output of the adder **219** by

$$u = gc_k + by_T$$

Note that the term by_T has its source in the pitch codebook (pitch codebook) **201** in response to the pitch lag T and the

15

past value of u stored in memory 203. The pitch codevector v_T from the pitch codebook 201 is then processed through a low-pass filter 202 whose cut-off frequency is adjusted by means of the index j from the demultiplexer 217. The resulting codevector v_T is then multiplied by the gain b from the demultiplexer 217 through an amplifier 226 to obtain the signal bv_T .

The factor α is calculated in voicing factor generator 204 by

$$\alpha = qR_p \text{ bounded by } \alpha < q$$

where q is a factor which controls the amount of enhancement (q is set to 0.25 in this preferred embodiment).

Method 2:

Another method used in a preferred embodiment of the invention for calculating periodicity factor α is discussed below.

First, a voicing factor r_v is computed in voicing factor generator 204 by

$$r_v = (E_v - E_c) / (E_v + E_c)$$

where E_v is the energy of the scaled pitch codevector bv_T and E_c is the energy of the scaled innovative codevector gc_k . That is

$$E_v = b^2 v_T^2 = b^2 \sum_{n=0}^{N-1} v_T^2(n)$$

and

$$E_c = g^2 c_k^2 = g^2 \sum_{n=0}^{N-1} c_k^2(n)$$

Note that the value of r_v lies between -1 and 1 (1 corresponds to purely voiced signals and -1 corresponds to purely unvoiced signals).

In this preferred embodiment, the factor α is then computed in voicing factor generator 204 by

$$\alpha = 0.125 (1 + r_v)$$

which corresponds to a value of 0 for purely unvoiced signals and 0.25 for purely voiced signals.

In the first, two-term form of $F(z)$, the periodicity factor α can be approximated by using $\sigma = 2\alpha$ in methods 1 and 2 above. In such a case, the periodicity factor σ is calculated as follows in method 1 above:

$$\sigma = 2qR_p \text{ bounded by } \sigma < 2q$$

In method 2, the periodicity factor σ is calculated as follows:

$$\sigma = 0.25 (1 + r_v)$$

The enhanced signal c_p is therefore computed by filtering the scaled innovative codevector gc_k through the innovation filter 205 ($F(z)$).

The enhanced excitation signal u' is computed by the adder 220 as:

$$u' = c_p + bv_T$$

Note that this process is not performed at the encoder 100. Thus, it is essential to update the content of the pitch codebook 201 using the excitation signal u without enhancement to keep synchronism between the encoder 100 and

16

decoder 200. Therefore, the excitation signal u is used to update the memory 203 of the pitch codebook 201 and the enhanced excitation signal u' is used at the input of the LP synthesis filter 206.

5 Synthesis and Deemphasis

The synthesized signal s' is computed by filtering the enhanced excitation signal u' through the LP synthesis filter 206 which has the form $1/\hat{A}(z)$, where $\hat{A}(z)$ is the interpolated LP filter in the current subframe. As can be seen in FIG. 2, the quantized LP coefficients $\hat{A}(z)$ on line 225 from demultiplexer 217 are supplied to the LP synthesis filter 206 to adjust the parameters of the LP synthesis filter 206 accordingly. The deemphasis filter 207 is the inverse of the preemphasis filter 103 of FIG. 1. The transfer function of the deemphasis filter 207 is given by

$$D(z) = 1/(1 - \mu z^{-1})$$

where μ is a preemphasis factor with a value located between 0 and 1 (a typical value is $\mu = 0.7$). A higher-order filter could also be used.

The vector s' is filtered through the deemphasis filter $D(z)$ (module 207) to obtain the vector s_o , which is passed through the high-pass filter 208 to remove the unwanted frequencies below 50 Hz and further obtain s_h .

25 Oversampling and High-frequency Regeneration

The over-sampling module 209 conducts the inverse process of the down-sampling module 101 of FIG. 1. In this preferred embodiment, oversampling converts from the 12.8 kHz sampling rate to the original 16 kHz sampling rate, using techniques well known to those of ordinary skill in the art. The oversampled synthesis signal is denoted \hat{s} . Signal \hat{s} is also referred to as the synthesized wideband intermediate signal.

The oversampled synthesis \hat{s} signal does not contain the higher frequency components which were lost by the down-sampling process (module 101 of FIG. 1) at the encoder 100. This gives a low-pass perception to the synthesized speech signal. To restore the full band of the original signal, a high frequency generation procedure is disclosed. This procedure is performed in modules 210 to 216, and adder 221, and requires input from voicing factor generator 204 (FIG. 2).

In this new approach, the high frequency contents are generated by filling the upper part of the spectrum with a white noise property scaled in the excitation domain, then converted to the speech domain, preferably by shaping it with the same LP synthesis filter used for synthesizing the down-sampled signal \hat{s} .

The high frequency generation procedure in accordance with the present invention is described hereinbelow.

The random noise generator 213 generates a white noise sequence w' with a flat spectrum over the entire frequency bandwidth, using techniques well known to those of ordinary skill in the art. The generated sequence is of length N' which is the subframe length in the original domain. Note that N is the subframe length in the down-sampled domain. In this preferred embodiment, $N = 64$ and $N' = 80$ which correspond to 5 ms.

The white noise sequence is properly scaled in the gain adjusting module 214. Gain adjustment comprises the following steps. First, the energy of the generated noise sequence w' is set equal to the energy of the enhanced excitation signal u' computed by an energy computing module 210, and the resulting scaled noise sequence is given by

$$w(n) = w'(n) \sqrt{\frac{\sum_{n=0}^{N-1} u^2(n)}{\sum_{n=0}^{N-1} w'^2(n)}} , n = 0, \dots, N' - 1.$$

The second step in the gain scaling is to take into account the high frequency contents of the synthesized signal at the output of the voicing factor generator **204** so as to reduce the energy of the generated noise in case of voiced segments (where less energy is present at high frequencies compared to unvoiced segments). In this preferred embodiment, measuring the high frequency contents is implemented by measuring the tilt of the synthesis signal through a spectral tilt calculator **212** and reducing the energy accordingly. Other measurements such as zero crossing measurements can equally be used. When the tilt is very strong, which corresponds to voiced segments, the noise energy is further reduced. The tilt factor is computed in module **212** as the first correlation coefficient of the synthesis signal s_h and it is given by:

$$\text{tilt} = \frac{\sum_{n=1}^{N-1} s_h(n)s_h(n-1)}{\sum_{n=0}^{N-1} s_h^2(n)} , \text{conditioned by tilt} \geq 0 \text{ and tilt} \geq r_v -$$

where voicing factor r_v is given by

$$r_v = (E_v - E_c) / (E_v + E_c)$$

where E_v is the energy of the scaled pitch codevector bv_T and E_c is the energy of the scaled innovative codevector gc_k , as described earlier. Voicing factor r_v is most often less than 1 but this condition was introduced as a precaution against high frequency tones where the tilt value is negative and the value of r_v is high. Therefore, this condition reduces the noise energy for such tonal signals.

The tilt value is 0 in case of flat spectrum and 1 in case of strongly voiced signals, and it is negative in case of unvoiced signals where more energy is present at high frequencies.

Different methods can be used to derive the scaling factor g_r from the amount of high frequency contents. In this invention, two methods are given based on the tilt of signal described above.

Method 1:

The scaling factor g_r is derived from the tilt by

$$g_r = 1 - \text{tilt} \text{ bounded by } 0.2 \leq g_r \leq 1.0$$

For strongly voiced signal where the tilt approaches 1, g_r is 0.2 and for strongly unvoiced signals g_r becomes 1.0.

Method 2:

The tilt factor g_r is first restricted to be larger or equal to zero, then the scaling factor is derived from the tilt by

$$g_r = 10^{-0.6 \text{ tilt}}$$

The scaled noise sequence w_g produced in gain adjusting module **214** is therefore given by:

$$w_g = g_r w.$$

When the tilt is close to zero, the scaling factor g_r is close to 1, which does not result in energy reduction. When the tilt value is 1, the scaling factor g_r results in a reduction of 12 dB in the energy of the generated noise.

Once the noise is properly scaled (w_g), it is brought into the speech domain using the spectral shaper **215**. In the

preferred embodiment, this is achieved by filtering the noise w_g through a bandwidth expanded version of the same LP synthesis filter used in the down-sampled domain ($1/\hat{A}(z/0.8)$). The corresponding bandwidth expanded LP filter coefficients are calculated in spectral shaper **215**.

The filtered scaled noise sequence w_f is then band-pass filtered to the required frequency range to be restored using the band-pass filter **216**. In the preferred embodiment, the band-pass filter **216** restricts the noise sequence to the frequency range 5.6–7.2 kHz. The resulting band-pass filtered noise sequence z is added in adder **221** to the over-sampled synthesized speech signal \hat{s} to obtain the final reconstructed sound signal s_{out} on the output **223**.

Although the present invention has been described hereinabove by way of a preferred embodiment thereof, this embodiment can be modified at will, within the scope of the appended claims, without departing from the spirit and nature of the subject invention. Even though the preferred embodiment discusses the use of wideband speech signals, it will be obvious to those skilled in the art that the subject invention is also directed to other embodiments using wideband signals in general and that it is not necessarily limited to speech applications.

What is claimed is:

1. A perceptual weighting device for producing a perceptually weighted signal in response to a wideband speech signal in order to reduce a difference between the wideband speech signal and a subsequently synthesized wideband speech signal, said perceptual weighting device comprising:

- a) a signal preemphasis filter responsive to the wideband speech signal for enhancing a high frequency content of the wideband speech signal to thereby produce a pre-emphasised signal;
- b) a synthesis filter calculator responsive to said pre-emphasised signal for producing synthesis filter coefficients; and
- c) a perceptual weighting filter, responsive to said pre-emphasised signal and said synthesis filter coefficients, for filtering said pre-emphasised signal in relation to said synthesis filter coefficients to thereby produce said perceptually weighted signal, said perceptual weighting filter having a transfer function with fixed denominator whereby weighting of said wideband speech signal in a formant region is substantially decoupled from a spectral tilt of said wideband speech signal.

2. A perceptual weighting device as defined in claim 1, wherein said signal preemphasis filter has a transfer function of the form:

$$P(z) = 1 - \mu z^{-1}$$

wherein μ is a preemphasis factor having a value located between 0 and 1.

3. A perceptual weighting device as defined in claim 2, wherein said preemphasis factor μ is 0.7.

4. A perceptual weighting device as defined in claim 2, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1) / (1 - \gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

5. A perceptual weighting device as defined in claim 4, wherein γ_2 is set equal to μ .

6. A perceptual weighting device as defined in claim 1, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1) / (1 - \gamma_2 z^{-1})$$

19

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

7. A perceptual weighting device as defined in claim 6, wherein γ_2 is set equal to μ .

8. A method for producing a perceptually weighted signal in response to a wideband speech signal in order to reduce a difference between the weighted wideband speech signal and a subsequently synthesized weighted wideband speech signal, said method comprising:

- a) filtering the wideband speech signal to produce a preemphasised signal with enhanced high frequency content;
- b) calculating, from said preemphasised signal, synthesis filter coefficients; and
- c) filtering said preemphasised signal in relation to said synthesis filter coefficients to thereby produce a perceptually weighted speech signal, wherein said filtering comprises processing the preemphasis signal through a perceptual weighting filter having a transfer function with fixed denominator whereby weighting of said wideband speech signal in a formant region is substantially decoupled from a spectral tilt of said wideband speech signal.

9. A method for producing a perceptually weighted signal as defined in claim 8, wherein filtering the wideband speech signal comprises filtering through a transfer function of the form:

$$P(z) = 1 - \mu z^{-1}$$

wherein μ is a preemphasis factor having a value located between 0 and 1.

10. A method for producing a perceptually weighted signal as defined in claim 9, wherein said preemphasis factor μ is 0.7.

11. A method for producing a perceptually weighted signal as defined in claim 9, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1)/(1 - \gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

12. A method for producing a perceptually weighted signal as defined in claim 11, wherein γ_2 is set equal to μ .

13. A method for producing a perceptually weighted signal as defined in claim 8, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1)/(1 - \gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

14. A method for producing a perceptually weighted signal as defined in claim 13, wherein γ_2 is set equal to μ .

15. An encoder for encoding a wideband speech signal, comprising:

- a) a perceptual weighting device as recited in claim 1;
- b) a pitch codebook search device responsive to said perceptually weighted signal for producing pitch codebook parameters and an innovative search target vector;
- c) an innovative codebook search device, responsive to said synthesis filter coefficients and to said innovative search target vector, for producing innovative codebook parameters; and
- d) a signal forming device for producing an encoded wideband speech signal comprising said pitch codebook parameters, said innovative codebook parameters, and said synthesis filter coefficients.

20

16. An encoder as defined in claim 15, wherein said signal preemphasis filter has a transfer function of the form:

$$P(z) = 1 - \mu z^{-1}$$

wherein μ is a preemphasis factor having a value located between 0 and 1.

17. An encoder as defined in claim 16, wherein said preemphasis factor μ is 0.7.

18. An encoder as defined in claim 16, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1)/(1 - \gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

19. An encoder as defined in claim 18, wherein γ_2 is set equal to μ .

20. An encoder as defined in claim 15, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1)/(1 - \gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

21. An encoder as defined in claim 20, wherein μ is set equal to γ_2 .

22. A cellular communication system for servicing a large geographical area divided into a plurality of cells, comprising:

- a) mobile transmitter/receiver units;
- b) cellular base stations respectively situated in said cells;
- c) a control terminal for controlling communication between the cellular base stations;
- d) a bidirectional wireless communication sub-system between each mobile unit situated in one cell and the cellular base station of said one cell, said bidirectional wireless communication sub-system comprising, in both the mobile unit and the cellular base station:
 - i) a transmitter including an encoder for encoding a wideband speech signal as recited in claim 15 and a transmission circuit for transmitting the encoded wideband speech signal; and
 - ii) a receiver including a receiving circuit for receiving a transmitted encoded wideband speech signal and a decoder for decoding the received encoded wideband speech signal.

23. A cellular communication system as defined in claim 22, wherein said signal preemphasis filter has a transfer function of the form:

$$P(z) = 1 - \mu z^{-1}$$

wherein μ is a preemphasis factor having a value located between 0 and 1.

24. A cellular communication system as defined in claim 23, wherein said preemphasis factor μ is 0.7.

25. A cellular communication system as defined in claim 23, wherein said perceptual weighting filter has a transfer function of the form:

$$W(z) = A (z/\gamma_1)/(1 - \gamma_2 z^{-1})$$

where $0 < \gamma_2 < \gamma_1 \leq 1$ and γ_2 and γ_1 are weighting control values.

26. A cellular communication system as defined in claim 25, wherein μ is set equal to γ_2 .

21

27. A cellular communication system as defined in claim 22, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1<=1 and gamma_2 and gamma_1 are weighting control values.

28. A cellular communication system as defined in claim 27, wherein gamma_2 is set equal to mu.

29. A cellular mobile transmitter/receiver unit comprising:

- a) a transmitter including an encoder for encoding a wideband speech signal as recited in claim 15 and a transmission circuit for transmitting the encoded wideband speech signal; and
b) a receiver including a receiving circuit for receiving a transmitted encoded wideband speech signal and a decoder for decoding the received encoded wideband speech signal.

30. A cellular mobile transmitter/receiver unit as defined in claim 29, wherein said signal preemphasis filter has a transfer function of the form:

P(z)=1-mu z^-1

wherein mu is a preemphasis factor having a value located between 0 and 1.

31. A cellular mobile transmitter/receiver unit as defined in claim 30, wherein said preemphasis factor mu is 0.7.

32. A cellular mobile transmitter/receiver unit as defined in claim 30, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1<=1 and gamma_2 and gamma_1 are weighting control values.

33. A cellular mobile transmitter/receiver unit as defined in claim 32, wherein gamma_2 is set equal to mu.

34. A cellular mobile transmitter/receiver unit as defined in claim 29, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1<=1 and gamma_2 and gamma_1 are weighting control values.

35. A cellular mobile transmitter/receiver unit as defined in claim 34, wherein gamma_2 is set equal to mu.

36. A cellular network element comprising:

- a) a transmitter including an encoder for encoding a wideband speech signal as defined in claim 15 and a transmission circuit for transmitting the encoded wideband speech signal; and
b) a receiver including a receiving circuit for receiving a transmitted encoded wideband speech signal and a decoder for decoding the received encoded wideband speech signal.

37. A cellular network element as defined in claim 36, wherein said signal preemphasis filter has a transfer function of the form:

P(z)=1-mu z^-1

wherein mu is a preemphasis factor having a value located between 0 and 1.

38. A cellular network element as defined in claim 37, wherein said preemphasis factor mu is 0.7.

22

39. A cellular network element as defined in claim 37, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1<=1 and gamma_2 and gamma_1 are weighting control values.

40. A cellular network element as defined in claim 39, wherein gamma_2 is set equal to mu.

41. A cellular network element as defined in claim 36, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1<=1 and gamma_2 and gamma_1 are weighting control values.

42. A cellular network element as defined in claim 41, wherein mu is set equal to y_2.

43. In a cellular communication system for servicing a large geographical area divided into a plurality of cells, comprising: mobile transmitter/receiver units; cellular base stations, respectively situated in said cells; and control terminal for controlling communication between the cellular base stations:

- a) a bidirectional wireless communication sub-system between each mobile unit situated in one cell and the cellular base station of said one cell, said bidirectional wireless communication sub-system comprising, in both the mobile unit and the cellular base station:

- a) a transmitter including an encoder for encoding a wideband speech signal as recited in claim 15 and a transmission circuit for transmitting the encoded wideband speech signal; and
b) a receiver including a receiving circuit for receiving a transmitted encoded wideband speech signal and a decoder for decoding the received encoded wideband speech signal.

44. A bidirectional wireless communication sub-system as defined in claim 43, wherein said signal preemphasis filter has a transfer function of the form:

P(z)=1-mu z^-1

wherein mu is a preemphasis factor having a value located between 0 and 1.

45. A bidirectional wireless communication sub-system as defined in claim 44, wherein said preemphasis factor mu is 0.7.

46. A bidirectional wireless communication sub-system as defined in claim 44, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1 and gamma_2 and gamma_1 are weighting control values.

47. A bidirectional wireless communication sub-system as defined in claim 46, wherein mu is set equal to gamma_2.

48. A bidirectional wireless communication sub-system as defined in claim 43, wherein said perceptual weighting filter has a transfer function of the form:

W(z)=A (z/gamma_1)/(1-gamma_2 z^-1)

where 0<gamma_2<gamma_1<=1 and gamma_2 and gamma_1 are weighting control values.

49. A bidirectional wireless communication subsystem as defined in claim 48, wherein gamma_2 is set equal to mu.