(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 159 738 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
05.04.2006 Bulletin 2006/14

(21) Application number: **00914511.1**

(22) Date of filing: **04.02.2000**

(51) Int Cl.:
G10L 19/00 *(2006.01)*

(86) International application number:
PCT/US2000/002900

(87) International publication number:
WO 2000/046795 (10.08.2000 Gazette 2000/32)

(54) **SPEECH SYNTHESIZER BASED ON VARIABLE RATE SPEECH CODING**

SPRACHSYNTHETISIERER AUF DER BASIS VON SPRACHKODIERUNG MIT VERÄNDERLICHER BIT-RATE

SYNTHETISEUR VOCAL BASE SUR UN CODAGE VOCAL A DEBIT VARIABLE

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **08.02.1999 US 246605**

(43) Date of publication of application:
05.12.2001 Bulletin 2001/49

(73) Proprietor: **QUALCOMM INCORPORATED**
San Diego, California 92121-1714 (US)

(72) Inventor: **CHANG, Chienchung**
California 92067-3874 (US)

(74) Representative: **Dunlop, Hugh Christopher et al**
R G C Jenkins & Co.
26 Caxton Street
London SW1H 0RJ (GB)

(56) References cited:
EP-A- 0 762 711
US-A- 5 657 420
WO-A-99/17516

EP 1 159 738 B1

Printed by Jouve, 75001 PARIS (FR)

Description

BACKGROUND OF THE INVENTION

I. Field of the Invention

[0001] The present invention relates to speech synthesis. More particularly, the present invention relates to synthesis of speech encoded by a variable rate vocoder. The invention further relates to use of speech synthesis with wireless communication devices.

II. Description of the Related Art

[0002] Electronic speech synthesis is useful in a number of applications. More and more, computers and other electronic equipment are providing the option of voiced prompts as a user interface. For example, speech may be utilized for reading electronic mail messages, for generating spoken prompts in a voice response system, or for providing directions to a driver in a vehicle.

[0003] There are two general types of speech synthesizers or techniques used to generate speech. The first type is referred to as a text-to-speech (TTS) speech synthesizer, and is grammar based. A TTS based system converts ordinary text into intelligible and natural sounding speech. It is useful for applications needing an automatic conversion of arbitrary input text into intelligible and natural sounding speech output. It is especially useful where large vocabularies and/or dynamically changing data are involved. The TTS system is useful in applications such as providing automatic voice alerts and prompts, proofreading, telephone access to databases, and conversion of electronic mail to voice mail or audio output. Because TTS is flexible and powerful, it offers utility in many applications. However, implementation of a TTS system may require tremendous memory and processing power resources. It may also contain a machine tone if synthesizer doesn't simulate human speech intonation closely. Accordingly, TTS is not a practical choice for applications with limited memory and processing resources, such as in small portable wireless devices, remotely located communication devices or computers, and the like.

[0004] A second type of speech synthesizer is Voice Coder (Vocoder) based. A vocoder compresses voiced speech, or audio signals, by extracting parameters that relate to a model of human speech generation. Vocoders have been developed to compress input speech that has been digitally converted to a rate of 64 kilo bits per second (kbps) down to 13 kbps, 8 kbps, or even lower rates. A vocoder based speech synthesizer generates certain parameters of or for the speech to be synthesized. The parameters are stored in some type of memory, preferably flash type, and are decoded upon speech synthesis. Because the parameters of all words to be synthesized need to be stored in memory, vocoder based speech synthesizers are more suitable for applications that do not require large vocabularies. They are especially suitable for systems having limited memory and processing resources.

[0005] EP-A-0 762 711 describes speech storage in a portable cellular telephone. Speech may be input via a microphone and compressed and stored for subsequent play back, or transmission over a channel. The compression may be effected using existing circuitry, such as a digital

[0006] For vocoder based speech synthesizers, there is a need to optimize memory usage while maintaining acceptable speech quality. For some applications, it may be desirable to maximize the size of the vocabulary for a given size of memory. Furthermore, it may also be desirable to use signal processing resources already available within a given communication system design for accomplishing speech synthesis. A speech synthesizer possessing these and other characteristics is provided by the present invention in the manner described below.

SUMMARY OF THE INVENTION

[0007] The present invention is an apparatus and method for speech synthesis as set forth in claims 1 and 11, respectively, and based on variable rate vocoding. The speech to be synthesized is encoded by a variable rate vocoder. A variable rate vocoder encodes a frame of speech at one of a set of predetermined rates based on the speech activity taking place within the frame of speech. In one embodiment, the variable rate vocoder is a code excited linear prediction (CELP) encoder having four bit rates. Thus, an input speech signal is encoded into speech parameters at one of the four rates using a CELP encoding scheme for the selected rate. The speech parameters are generally provided to a decoder which performs a variable rate decoding scheme corresponding to the variable rate encoding scheme utilized. The decoder produces speech samples, which are provided to a coder-decoder or codec for digital-to-analog conversion. The resulting analog signal generated by the codec is then broadcast through a speaker or other known audio output device as synthesized speech.

[0008] The speech synthesizer of the present invention is especially suitable for use in wireless communication systems in which variable rate vocoding is already implemented. In these systems, the existing vocoding resources may be employed for speech synthesis. Alternatively, DSP elements, already present or easily incorporated, can be used in conjunction with a small amount of memory to provide the speech synthesizer function. In addition, a speech synthesizer based on variable rate vocoding is able to provide good speech quality without requiring a large amount of memory. The level of compression provided by a variable rate vocoder makes it suitable for applications with limited memory.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The features, objects, and advantages of the present invention will become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

FIG. 1 is a block diagram of a variable rate vocoder; and

FIG. 2 is a block diagram of the speech synthesizer of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0010] The present invention provides an apparatus and method for synthesizing speech which is very useful when used with wireless communication equipment. The invention can take advantage of existing signal processing resources in wireless communication equipment or a minimum of additional hardware to synthesize speech in a manner that provides high speech quality and requires a small memory size.

[0011] The present invention is very useful when employed in conjunction with a variety of known communication devices or systems, and it is described below in relation to a CDMA wireless communication system. In addition, it is contemplated that it is particularly well suited for specific applications, such as hands-free car kits used to mount and operate wireless devices in vehicles. However, those skilled in the art will readily understand that this is not a limitation of the present invention, and that it can be used with other types of communication devices including those communicating in wired, wire line, or optical cable, type systems, and those using other signal modulation techniques.

[0012] An exemplary wireless communication system makes use of code division multiple access (CDMA) modulation techniques. Although other techniques such as time division multiple access (TDMA), frequency division multiple access (FDMA), and amplitude modulation (AM) schemes such as amplitude companded single sideband (ACSSB) are known, CDMA has significant advantages over these other techniques. The use of CDMA techniques in a multiple access communication system is disclosed in U.S. Patent No. 4,901,307, entitled *"Sprend Spectrum Multiple Access Communication System Using Satellite Or Terrestrial Repeaters,"* assigned to the assignee of the present invention.

[0013] A speech synthesizer may be implemented in wireless communication devices or equipment for a number of reasons. For example, speech synthesis may be part of a voice recognition system in a wireless telephone or a "hands-free" carkit used to support operation in a vehicle. A speech synthesizer can provide information in audible form when a device user or operator cannot visually observe an output screen or indicators on the device. For example, information can be provided to allow device operation or output when a vehicle driver or machinery operator cannot safely look at the communication device, closely. The speech synthesizer would also allow for hands free operation of devices by providing voice prompts for operations to be performed. For example, the speech synthesizer may ask for the name of a person to be called, allowing the device to automatically dial a telephone number, or ask for a command to be implemented, such as dialing, storing, opening mail, terminating a call attempt, or shutting down.

[0014] In one embodiment the speech synthesizer of the present invention makes use of the vocoder circuitry already present in a number of wireless devices such as wireless telephones and other products used by communication service subscribers to generate voiced speech. Specifically, the speech synthesizer is based on a variable rate vocoder. A variable rate vocoder uses speech activity to vary its instantaneous data rate. During active speech, the vocoder encoder uses a large number of bits to encode the speech samples. During periods of silence, the vocoder encoder uses few or fewer bits to encode the background noise. Exemplary embodiments of variable rate vocoders are described in U.S. Patent No.s 5,414, 796 and 5 657,420, entitled *"Variable Rate Vocoder,"* and assigned to the assignee of the present invention.

[0015] Variable rate vocoders are commonly used in CDMA type communication systems to increase system capacity by decreasing the number of bits generally used by each communication signal. A variable rate vocoder may, for example, be implemented in the CDMA communication system of Patent No. 4,901,307 discussed above. In a CDMA communication system, different users communicate using the same bandwidth, but using different code channels. A variable rate vocoder in a CDMA communication system takes advantage of the fact that a user is only speaking actively about 40% of the time on any given channel. By sending fewer bits when a user is silent, the variable rate vocoder allows more users to share the same bandwidth.

[0016] A schematic block diagram of a typical variable rate vocoder is shown in FIG. 1 and is indicated generally by 100. The vocoder shown in FIG. 1 uses four different data rates, although it should be understood that a different number of data rates may be employed instead, as would be known in the art. In the set of four rates, if the peak rate is 13.2 kbps, then full rate corresponds to 13.2 kbps, 1/2 rate corresponds to approximately 6.2 kbps, 1/4 rate corresponds to approximately 2.7 kbps, and 1/8 rate corresponds to approximately 1.0 kbps. Note that the actual bit rate for rates other than the full rate are approximate because of the use of overhead bits, as is well understood in the art.

[0017] Referring still to FIG. 1, it can be seen that variable rate vocoder 100 includes an encoder 102 and a decoder 104. Encoder 102 receives speech samples for frames of speech data as an input, for example, as 8-bit

PCM samples at a 64 kbps data rate, in either mu-law or a-law format. Encoder 102 encodes these speech samples into speech parameters at one of four data rates, depending on the speech activity. The input speech samples are also provided to rate determination element 106.

[0018] Rate determination element 106 may implement any of a number of rate decision algorithms. In one embodiment, energy thresholds relative to the background noise energy level are used to determine the speech activity, and thereby the rate, at which the input samples are to be encoded. If the energy of the current frame of speech samples is far above the background noise energy, then the rate determination element 106 will determine that the frame is to be encoded at full rate. If the energy of the current frame is close to the background noise energy, then rate determination element 106 will determine that the frame is to be encoded at eighth rate, and so forth, as is known.

[0019] Another rate determination technique is disclosed in EP-A-1,339,044, entitled *"Method And Apparatus For Performing Reduced Rate Variable Rate Vocoding,"* assigned to the assignee of the present invention. This technique provides the set of rate decision criteria referred to as mode measures. A first mode measure is the target matching signal to noise ratio (TMSNR) from the previous encoding frame, which provides information on how well the encoding model is performing by comparing a synthesized speech signal with the input speech signal. A second mode measure is the normalized autocorrelation function (NACF), which measures periodicity in the speech frame. A third mode measure is the zero crossings (ZC) parameter, which measures high frequency content in an input speech frame. A fourth measure, the prediction gain differential (PGD), determines if the encoder is maintaining its prediction efficiency. A fifth measure is the energy differential (ED), which compares the energy in the current frame to an average frame energy.

[0020] Using the mode measures discussed above, rate determination logic selects an encoding rate for each frame of input speech data. The values for the various modes select one of say four or more modes in which to operate. That is, the values detected for each mode measure relative to a threshold or other criteria determines which encoding rate is selected, based on a preselected pattern or hiearchy. For example; if the value for NACF is less than a pre-selected threshold and ZC is greater than a second pre-selected threshold one rate could be selected. However, if these conditions are not met but ED is lower than a third threshold, then a quarter rate might be selected. If the value for TSNR is greater, PGD is less, and NACF is greater than fourth, fifth, and sixth thresholds, respectively, then a half rate might be selected. Various such combinations and thresholds may be employed by those skilled in the art to selected encoding rates.

[0021] It should be understood that still other rate determination techniques may be adopted by rate determination element 106.

[0022] Referring still to FIG. 1, a signal indicating the data rate determined by rate determination element 106 is provided to a switch 108. Switch 108 selects an element for encoding a frame of input speech samples from among a full rate encoding element 110, a half rate encoding element 112, a quarter rate encoding element **114,** and an eighth rate encoding element **116,** as designated by the data rate signal. The selected encoding element encodes the speech samples to produce a signal of an encoded data packet. Rate determination element 106 also provides a signal indicating the data rate to a switch **118,** which selects the same encoding element as switch **108** so that the signal of the encoded data packet generated by the selected encoding element can be provided to an output of the variable rate vocoder.

[0023] Each of the encoding elements **110, 112, 114,** and **116** is configured to encode speech using a predetermined encoding scheme. A linear-prediction-based encoding scheme, such as the Code Excited Linear Predictive (CELP) encoder, is used in a preferred embodiment. The CELP coder is described in the paper "*A 4.8 Kbps Code Excited Linear Predictive Coder,"* by Thomas E. Tremain *et al.,* Proceedings of the Mobile Satellite Conference, 1988. Linear-prediction-based encoders compress speech by removing the natural redundancies inherent in speech. Speech typically exhibits short term redundancies resulting from the mechanical action of the lips and tongue, and long term redundancies resulting from the vibration of the vocal cords. Linear predictive schemes model these operations as filters, remove the redundancies, and then model the resulting residual signal as white gaussian noise. Linear predictive coders, therefore, achieve a reduced bit rate by transmitting filter coefficients and quantized noise rather than a full bandwidth speech signal.

[0024] A linear predictive coding scheme that employs variable rates offers further reductions in bit rate without compromising the quality of speech. In FIG. 1, the full rate encoding element **110** encodes the parameters of the input speech signal using more bits to better preserve the characteristics of the input. For periods where no speech is detected, eighth rate encoding element **116** encodes the parameters using fewer bits since there is typically little detail or useful information to be captured. Transitions between periods of active speech and periods with no detected speech are encoded by half rate encoding element **112** and quarter rate encoding element **114.**

[0025] Referring now to the decoding element of the variable rate vocoder, decoder **104** receives a signal of the encoded speech parameters as well as a signal indicating the rate used to encode the speech. A rate extraction element **128** receives this input signal and determines the data rate of the speech. A signal of the data rate is also_provided to a switch **130,** which selects the decoding element from a set of decoding elements to properly decode the input parameters. In FIG. 1, four

decoding elements, full rate decoding element 120, half rate decoding element **122,** quarter rate decoding element **124,** and eighth rate decoding element **126** are provided for decoding the speech parameters at the four possible rates. The selected decoding element decodes the input parameters based on the data rate to produce a signal of decoded samples, which typically are 64 kbps pulse code modulated (PCM) samples. A signal of the data rate determined by rate extraction element 128 is also provided to a switch **132.** Switch 132 selects the same decoding element as switch **130** so that a signal of the decoded samples is provided to an output of the vocoder.

[0026] Referring now to FIG. 2, a block diagram of a speech synthesis system operating according to the principles of the present invention, which incorporates a variable rate vocoder, is shown. The speech synthesis system comprises a variable rate encoder **202** and a speech synthesizer **204.** An example of the variable rate encoder **202** is encoder 102 of FIG. 1. Variable rate encoder **202** receives a speech signal as input, and encodes the speech at one of a set of predetermined rates. In a preferred embodiment, variable rate encoder **202** is a CELP encoder that generates speech parameters at one of the rates based on the speech activity in the input segment of speech.

[0027] The present invention uses a variable rate vocoder as described in U.S. Patent No. 5,414,796, discussed above, which is commercially available, for example, as a 13 kbps vocoder product from Qualcomm Incorporated. In one preferred embodiment, the variable rate decoder is an enhanced variable rate decoder such as described in relation to the IS127 standard.

[0028] In one embodiment of the present invention, encoding rate decisions are based on "mode measures," as discussed above. The different combinations of criteria used to make rate selections are used to create what is termed "reduced rate mode" or "modes," and referred to more simply as mode 0, mode 1, mode 2, and so forth, as would be understood by those skilled in the art. The present invention can take advantage of such modes for purposes of speech synthesis.

[0029] The speech received by variable rate encoder **202** may be a word or a phrase from a pre-selected vocabulary that a communication device such as a wireless telephone, carkit, or other communication device is designed to synthesize. The vocabulary would include prompts and alerts to be given to a device user. For example, by extracting and synthesizing five individual vocabulary words: 'call', 'redial', 'program', 'or' and 'exit', the speech synthesizer may be designed to provide the prompts *"call, redial, program, or exit"* in solicitation of a response from the user. Alternatively, the speech synthesizer may be designed to provide previously stored information, such as in phone books, look-up tables, or databases, to a device user in response to various device inputs, including audio. The speech received by variable rate encoder **202** is encoded, and the encoded parameters are provided to a memory element or circuit **206** of the speech synthesizer 204 for storage.

[0030] Memory **206** is intended to hold or store the parameters over some time for operation of the desired device. However, it is also generally desirable to have the parameters stored in a manner that makes them updateable or replaceable, such as when the vocabulary needs to be changed for changing conditions or upgrades to device features. Therefore, memory **206** is configured in the form of non-volatile but re-writable memory, which can be accomplished using flash type memory elements, as is well known in the art.

[0031] As one would recognize, the operation of loading parameters may be performed during manufacture of a communication device for which the invention is to be used. Since the prompts and alerts to be synthesized may be predetermined, these may be encoded during manufacture and stored in flash memory **206** prior to use. The parameters can be changed or replaced during service of the device, or through newly developed over-the-air programming techniques for wireless devices.

[0032] Alternatively, variable rate encoder **202** may receive a speech signal input during operation of the communication device. For example, in response to a prompt from the speech synthesizer, the user may provide a spoken response. Variable rate encoder **202** will then encode the user's speech, and the encoded parameters may be provided to flash memory **206** for storage, and/or provided to a voice recognizer (not shown) for voice recognition purposes. In this manner, the parameters are input post manufacture such as immediately upon the device entering useful service or over time, such as by building a personal vocabulary library for each device (vocoder) user, related to that user's requirements.

[0033] Flash memory **206** should be of a size that is sufficient to store the parameters of the pre-selected vocabulary as well as the parameters of speech anticipated from the user. Thus, the size of flash memory **206** may vary based on the requirements of the specific application. Post manufacture storage may have an advantage of reducing memory requirements where each device user does not require as extensive a vocabulary as compared to what a manufacturer would have to install to cover an entire larger device market. The speech synthesizer can record names or other words, like 'Fred Smith' by detecting the endpoints of the target or desired phrase or speech, removing silence or redundancies, and encoding it. Therefore, speech can be recorded "on-line" and used later to synthesize speech output.

[0034] It should be noted that variable rate encoder 202 may be configured based on the available memory and the voice quality required. In the system having four rates wherein the full rate is 13 kbps, the average rate will generally be 5.88 kbps based on 40% voice activity. The use of the variable rates will provide high speech quality. If, however, the memory size is limited, variable rate encoder 202 may be configured to operate at, say, a fixed half-rate of approximately 800 bytes per second.

Otherwise, the rate may be selected from a subset of the predetermined set of rates instead of the whole set of rates. For example, the reduced rate modes discussed above can be used to select various rates. In one embodiment of the invention, the rates are divided into a set of four modes, labeled as modes 0, 1, 2, and 3. Using fixed rates according to the mode, rates on the order of 1800 bytes per second, 1540 bytes per second, 1400 bytes per second, and 1100 bytes per second, respectively, can be used. The use of such fixed reduced rates allows delivery of very high quality voice given a predefined data rate, approaching land-line quality. These four modes provide the best tradeoff between synthesized speech quality and memory requirement.

[0035] Furthermore, variable rate encoder **202** may switch between different modes of operation (variable rate, all half-rate, a subset of the variable rates, etc.) based on the instantaneous requirements of the application. Because there may be a trade off between voice quality and memory size, the configuration to be adopted will depend on the application being implemented.

[0036] The speech parameters stored in flash memory **206** will be provided to a variable rate decoder **208** when speech synthesis is desired. The variable rate decoder **208** is configured to decode the parameters generated by corresponding variable rate encoder 202. An example of variable rate decoder **208** is decoder **104** of FIG. 1. Generally, variable rate decoder **208** will be implemented as part of a digital signal processor (DSP)-used within the communication device. Such DSPs are used as or to form the processing elements for signal coding/decoding, combining, CDMA coding, power adjustment, and so forth. Since such elements are typically used in wireless devices, and many other devices in which the invention may find use, advantage can be taken of their presence to very cost effectively implement the present invention.

[0037] In order to implement the decoding functionality for the present invention, only a small amount of memory is required in or coupled to a DSP. A stand-alone decoder within or using a DSP requires a very small amount of memory (both program and data) to attain speech synthesis capability. The speech synthesizer can be implemented using well known DSP circuits and devices such as commercially available from Analog Devices and Qualcomm Incorporated.

[0038] The decoded parameters, typically in the form of pulse code modulated (PCM) samples, are then provided to a codec **210.** Codec **210** converts the PCM samples from a digital format to an analog signal. The analog signal is provided to speaker or other known audio output device **212,** which projects or broadcasts synthesized speech into the surrounding device environment where it can be heard.

[0039] Therefore, a speech synthesizer based on variable rate vocoding is provided by the present invention. The speech synthesizer is especially suitable for use in wireless communication devices that already comprise a variable rate vocoder. In other words, an existing variable rate vocoder that may be employed by the speech synthesizer, through the use of appropriate changes in program or operational instructions, or using control hardware. In addition, through the use of variable rate vocoding, the compression achieved may allow a pre-determined vocabulary to be stored in a memory of limited size associated with the wireless device or other equipment with which it interfaces. Furthermore, the trade off between voice quality and memory size may be considered in configuring the variable rate vocoder to provide a speech synthesizer with the desired voice quality and memory size.

[0040] The present invention can find application in a variety of communication devices and interface equipment. The above example embodiments were discussed in relation to wireless communication devices such as, but not limited to, cellular and satellite telephones, often referred to as user terminals, subscriber units, mobile stations, or simply "users," "mobiles," or "subscribers". In addition, other devices are also contemplated, such as message receivers and data transfer devices (e.g., portable computers, personal data assistants, modems, machinery controllers), or interfaces for public telephone networks or dedicated communications channels.

[0041] The invention can be implemented using separate circuits in the form of dedicated components or application specific integrated circuits (ASIC) to form a speech synthesizer which is installed within a desired device. Alternatively, it can be incorporated within other ASICs and devices by using a small amount of additional memory to work with existing digital signal processing elements.

[0042] The previous description of the preferred embodiments is provided to enable any person skilled 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 as defined by the claims.

Claims

1. An apparatus (204) for synthesising a pre-selected vocabulary in a wireless communication system, wherein said pre-selected vocabulary has been encoded by a variable rate encoder (102,202) at a set of variable rates, the apparatus (204) comprising:

a memory (206) for storing a set of speech parameters;
a processor configured to accept a verbal input from a user;
a variable rate decoder (104,208) for decoding

speech parameters to generate decoded speech samples; and
a digital-to-analog converter (210) for converting said speech samples into an analog signal for broadcast as synthesized speech, **characterised in that**:

said set of speech parameters represents said encoded pre-selected vocabulary;
the processor is further configured to choose a subset of speech parameters from said set of speech parameters in accordance with said verbal input from said user; and
the variable rate decoder (104,208) is configured to decode said subset of speech parameters.

2. The apparatus (204) of claim 1, wherein said variable rate encoder (102,202) is linear-prediction-based.

3. The apparatus (204) of claim 1, wherein said variable rate decoder (104,208) is linear-prediction-based.

4. The apparatus (204) of claim 1, wherein said set of speech parameters are encoded at a set of variable rates comprising of a full rate, a half rate, a quarter rate, and an eighth rate.

5. The apparatus (204) of claim 4, wherein said full rate is 13.2 kbps, said half rate is approximately 6.2 kbps, and quarter rate is approximately 2.7 kbps, and said eighth rate is approximately 1.0 kbps.

6. The apparatus (204) of claim 4, wherein said set of speech parameters are encoded at a rate fixed in response to one or more measured mode criteria.

7. The apparatus (204) of claim 4, wherein said set of speech parameters are encoded at a rate fixed at said half rate.

8. The apparatus (204) of claim 4, wherein the encoding rate is selected in accordance with the requirements of the quality of voice and the size of said memory (206).

9. The apparatus (204) of claim 1, wherein said wireless communication system is a CDMA system.

10. The apparatus (204) of claim 1, wherein said variable rate encoder (102,202) comprises an enhanced variable rate encoder.

11. A method of synthesizing a pre-selected vocabulary in a wireless communication system, wherein said pre-selected vocabulary has been encoded by a variable rate encoder (102, 202) at a set of variable rates, the method comprising:

receiving a verbal user input;
retrieving a set of speech parameters stored in a memory (206);
decoding (104,208) said set of speech parameters using a variable rate decoding scheme to generate decoded speech samples; and
converting (210) said speech samples into an analog signal for broadcast as synthesized speech, **characterised in that**:
said set of speech parameters represents said encoded pre-selected vocabulary;
a subset of speech parameters is chosen from said set of speech parameters in accordance with said verbal user input ; and
said subset of speech parameters is decoded using the variable rate decoding scheme (104,208).

12. The method of claim 11, wherein said variable rate encoder (102, 202) performs a variable rate encoding scheme that is linear-prediction-based.

13. The method of claim 11, wherein said variable rate decoding scheme is linear-prediction-based.

14. The method of claim 11, wherein said set of speech parameters are encoded at a set of variable rates comprising of a full rate, a half rate, a quarter rate, and an eighth rate.

15. The method of claim 14, wherein said full rate is 13.2 kbps, said half rate is approximately 6.2 kbps, said quarter rate is approximately 2.7 kbps, and said eighth rate is approximately 1.0 kbps.

16. The method of claim 14, wherein said set of speech parameters are encoded at a rate fixed in response to one or more measured mode criteria.

17. The method of claim 14, wherein said set of speech parameters are encoded at a rate fixed at said half rate.

18. The method of claim 14, wherein the encoding rate is selected in accordance with the requirements of the quality of voice and the size of said memory (206).

19. The method of claim 11, wherein said wireless communication system comprises a CDMA system.

20. The method of claim 11, further comprising:

encoding (102, 202) said verbal user input; and adding said encoded verbal user input into said memory (206) as a part of said set of speech

parameters.

Patentansprüche

1. Eine Vorrichtung (204) zur Synthetisierung von vorselektiertem Vokabular in einem drahtlosen Kommunikationssystem, wobei das vorselektierte Vokabular von einem variablen Ratenkodierer (102, 202) mit einer Gruppe variabler Raten kodiert wurde, wobei die Vorrichtung (204) folgendes aufweist:

Einen Speicher (206) für das Speichern von einer Gruppe von Sprachparametern;
einen Prozessor, der darauf konfiguriert ist, eine sprachliche Eingabe von einem Anwender zu akzeptieren;
einen variablen Ratendekoder (104, 208) für das Dekodieren der Sprachparameter, um dekodierte Sprachsamples zu generieren; und
einen Digital-zu-Analog-Konverter (210) für das Konvertieren der Sprachsamples in ein analoges Signal für die Aussendung als synthetisierte Sprache,

dadurch gekennzeichnet, dass:

Die Gruppe von Sprachparametern repräsentiert das kodierte, vorselektierte Vokabular;
der Prozessor ist weiterhin so konfiguriert, eine Untergruppe von Sprachparametern dieser Sprachparameter gemäß der sprachlichen Eingabe des Anwenders auszuwählen; und
der variable Ratendekoder (104, 208) ist konfiguriert, um die Untergruppe von Sprachparametern zu dekodieren.

2. Vorrichtung (204) nach Anspruch 1, wobei der variable Ratenkodierer (102, 202) auf lineare Prädiktion bzw. Linear-Prediction basiert.

3. Vorrichtung (204) nach Anspruch 1, wobei der variable Ratendekoder (104, 208) auf lineare Prädiktion basiert.

4. Vorrichtung (204) nach Anspruch 1, wobei die Gruppe von Sprachparametern mit einer Gruppe von variablen Raten kodiert wurde, welche eine Vollrate, eine halbe Rate, eine Viertelrate und eine Achtelrate aufweist.

5. Vorrichtung (204) nach Anspruch 4, wobei die Vollrate 13,2 kbps beträgt, die halbe Rate ungefähr 6,2 kbps und die Viertelrate ungefähr 2,7 kbps und die Achtelrate ungefähr 1,0 kbps.

6. Vorrichtung (204) nach Anspruch 4, wobei die Gruppe von Sprachparametern mit einer fixen Rate kodiert wurde, als Antwort auf eine oder mehrere gemessene Moduskriterien.

7. Vorrichtung (204) nach Anspruch 4, wobei die Gruppe von Sprachparametern mit einer Rate festgelegt auf der halben Rate kodiert werden.

8. Vorrichtung (204) nach Anspruch 4, wobei die Kodierrate gemäß den Anforderungen der Sprachqualität und der Größe des Speichers (206) selektiert wird.

9. Vorrichtung (204) nach Anspruch 1, wobei das drahtlose Kommunikationssystem ein CDMA-System ist.

10. Vorrichtung (204) nach Anspruch 1 wobei der variable Ratenkodierer (102, 202) einen erweiterten bzw. verbesserten variablen Ratenkodierer aufweist.

11. Ein Verfahren zur Synthetisierung eines vorselektierten Vokabulars in einem drahtlosen Kommunikationssystem, wobei das vorselektierte Vokabular von einem variablen Ratenkodierer (102, 202) mit einer Gruppe variabler Raten kodiert wurde, wobei das Verfahren Folgendes aufweist:

Empfangen einer sprachlichen Anwendereingabe;
Abrufen einer Gruppe von Sprachparametern, gespeichert in einem Speicher (206);
Dekodieren (104, 208) der Gruppe von Sprachparametern unter Verwendung von einem variablen Ratendekodierungsschema, um die dekodierten Sprachsamples zu generieren; und
Konvertieren (210) der Sprachsamples in ein analoges Signal für die Aussendung als synthetisierte Sprache, **gekennzeichnet dadurch, dass**:

Die Gruppe von Sprachparametern repräsentiert das kodierte vorselektierte Vokabular; eine Untergruppe von Sprachparametern wird aus der Gruppe von Sprachparametern gemäß der sprachlichen Anwendereingabe ausgewählt; und
die Untergruppe von Sprachparametern wird dekodiert unter Verwendung des variablen Ratendekodierungsschemas (104, 208).

12. Verfahren nach Anspruch 11, wobei der variable Ratenkodierer (102, 202) mit einem variablen Ratenkodierungsschema basierend auf lineare Prädiktion arbeitet.

13. Verfahren nach Anspruch 11, wobei das variable Ratendekodierungsschema auf linearer Prädiktion basiert.

14. Verfahren nach Anspruch 11, wobei die Gruppe von Sprachparametern mit einer Gruppe von variablen Raten kodiert wurde, welche eine Vollrate, eine halbe Rate, eine Viertelrate und eine Achtelrate aufweist.

15. Verfahren nach Anspruch 14, wobei die Vollrate 13,2 kbps beträgt, die halbe Rate ungefähr 6,2 kbps, die Viertelrate ungefähr 2,7 kbps und die Achtelrate ungefähr 1,0 kbps.

16. Verfahren nach Anspruch 14, wobei die Gruppe von Sprachparametern mit einer fixen Rate kodiert wurde, ansprechend auf eine oder mehrere gemessenen Moduskriterien.

17. Verfahren nach Anspruch 14, wobei die Gruppe von Sprachparametern mit einer Rate festgelegt auf die halbe Rate kodiert werden.

18. Verfahren nach Anspruch 14, wobei die Kodierrate gemäß den Anforderungen der Sprachqualität und der Größe des Speichers (206) selektiert wird.

19. Verfahren nach Anspruch 11, wobei das drahtlose Kommunikationssystem ein CDMA-System aufweist.

20. Verfahren nach Anspruch 11, welches weiterhin aufweist:

Kodieren (102, 202) der sprachlichen Anwendereingabe; und
Hinzufügen der kodierten sprachlichen Anwendereingabe zu dem Speicher (206) als ein Teil der Gruppe von Sprachparametern.

Revendications

1. Appareil (204) permettant de synthétiser un vocabulaire présélectionné dans un système de communication sans fil, dans lequel ledit vocabulaire présélectionné a été codé par un codeur à débit variable (102, 202) avec un ensemble de débits variables, l'appareil (204) comprenant :

une mémoire (206) pour stocker un ensemble de paramètres de parole ;
un processeur configuré pour accepter une entrée verbale d'un utilisateur ;
un décodeur à débit variable (104, 208) pour décoder des paramètres de parole afin de produire des échantillons de parole décodés ; et
un convertisseur numérique-analogique (210) pour convertir lesdits échantillons de parole en un signal analogique destiné à être diffusé en tant que voix synthétisée, **caractérisé en ce que** :

ledit ensemble de paramètres de parole représente ledit vocabulaire présélectionné codé ;
le processeur est en outre configuré pour choisir un sous-ensemble de paramètres de parole à partir dudit ensemble de paramètres de parole en fonction de ladite entrée verbale dudit utilisateur ; et
le décodeur à débit variable (104, 208) est configuré pour décoder ledit sous-ensemble de paramètres de parole.

2. Appareil (204) selon la revendication 1, dans lequel ledit codeur à débit variable (102, 202) est basé sur une prédiction linéaire.

3. Appareil (204) selon la revendication 1, dans lequel ledit décodeur à débit variable (104, 208) est basé sur une prédiction linéaire.

4. Appareil (204) selon la revendication 1, dans lequel ledit ensemble de paramètres de parole est codé avec un ensemble de débits variables comprenant un plein débit, un demi débit, un quart de débit et un huitième de débit.

5. Appareil (204) selon la revendication 4, dans lequel ledit plein débit vaut 13,2 kbit/s, ledit demi débit vaut approximativement 6,2 kbit/s, ledit quart de débit vaut approximativement 2,7 kbit/s et ledit huitième de débit vaut approximativement 1,0 kbit/s.

6. Appareil (204) selon la revendication 4, dans lequel ledit ensemble de paramètres de parole est codé à un débit fixé en réponse à un ou plusieurs critère(s) de mode mesuré(s).

7. Appareil (204) selon la revendication 4, dans lequel ledit ensemble de paramètres de parole est codé à un débit fixé audit demi débit.

8. Appareil (204) selon la revendication 4, dans lequel le débit de codage est choisi selon les exigences de qualité de la voix et selon la taille de ladite mémoire (206).

9. Appareil (204) selon la revendication 1, dans lequel ledit système de communication sans fil est un système AMRC.

10. Appareil (204) selon la revendication 1, dans lequel ledit codeur à débit variable (102, 202) comprend un codeur à débit variable évolué.

11. Procédé de synthèse d'un vocabulaire présélectionné dans un système de communication sans fil, dans

lequel ledit vocabulaire présélectionné a été codé par un codeur à débit variable (102, 202) avec un ensemble de débits variables, le procédé comprenant les étapes consistant à :

recevoir une entrée d'utilisateur verbale ;
récupérer un ensemble de paramètres de parole stockés dans une mémoire (206) ;
décoder (104, 208) ledit ensemble de paramètres de parole en utilisant un programme de décodage afin de produire des échantillons de parole décodés ; et
convertir (210) lesdits échantillons de parole en un signal analogique destiné à être diffusé en tant que voix synthétisée,

caractérisé en ce que :

ledit ensemble de paramètres de parole représente ledit vocabulaire présélectionné codé ;
un sous-ensemble de paramètres de parole est choisi dans ledit ensemble de paramètres de parole en fonction de ladite entrée d'utilisateur verbale ; et
ledit sous-ensemble de paramètres de parole est décodé en utilisant le programme de décodage à débit variable (104, 208).

12. Procédé selon la revendication 11, dans lequel ledit codeur à débit variable (102, 202) exécute un programme de codage à débit variable qui est basé sur une prédiction linéaire.

13. Procédé selon la revendication 11, dans lequel ledit programme de décodage à débit variable est basé sur une prédiction linéaire.

14. Procédé selon la revendication 11, dans lequel ledit ensemble de paramètres de parole est codé avec un ensemble de débits variables comprenant un plein débit, un demi débit, un quart de débit et un huitième de débit.

15. Procédé selon la revendication 14, dans lequel ledit plein débit vaut 13,2 kbit/s, ledit demi débit vaut approximativement 6,2 kbit/s, ledit quart de débit vaut approximativement 2,7 kbit/s et ledit huitième de débit vaut approximativement 1,0 kbit/s.

16. Procédé selon la revendication 14, dans lequel ledit ensemble de paramètres de parole est codé à un débit fixé en réponse à un ou plusieurs critère(s) de mode mesuré(s).

17. Procédé selon la revendication 14, dans lequel dans lequel ledit ensemble de paramètres de parole est codé à un débit fixé audit demi débit.

18. Procédé selon la revendication 14, dans lequel le débit de codage est choisi selon les exigences de qualité de la voix et selon la taille de ladite mémoire (206).

19. Procédé selon la revendication 11, dans lequel ledit système de communication sans fil comprend un système AMRC.

20. Procédé selon la revendication 11, comprenant en outre les étapes consistant à :

coder (102, 202) ladite entrée d'utilisateur verbale ; et
ajouter ladite entrée d'utilisateur verbale dans ladite mémoire (206) en tant que partie dudit ensemble de paramètres de parole.

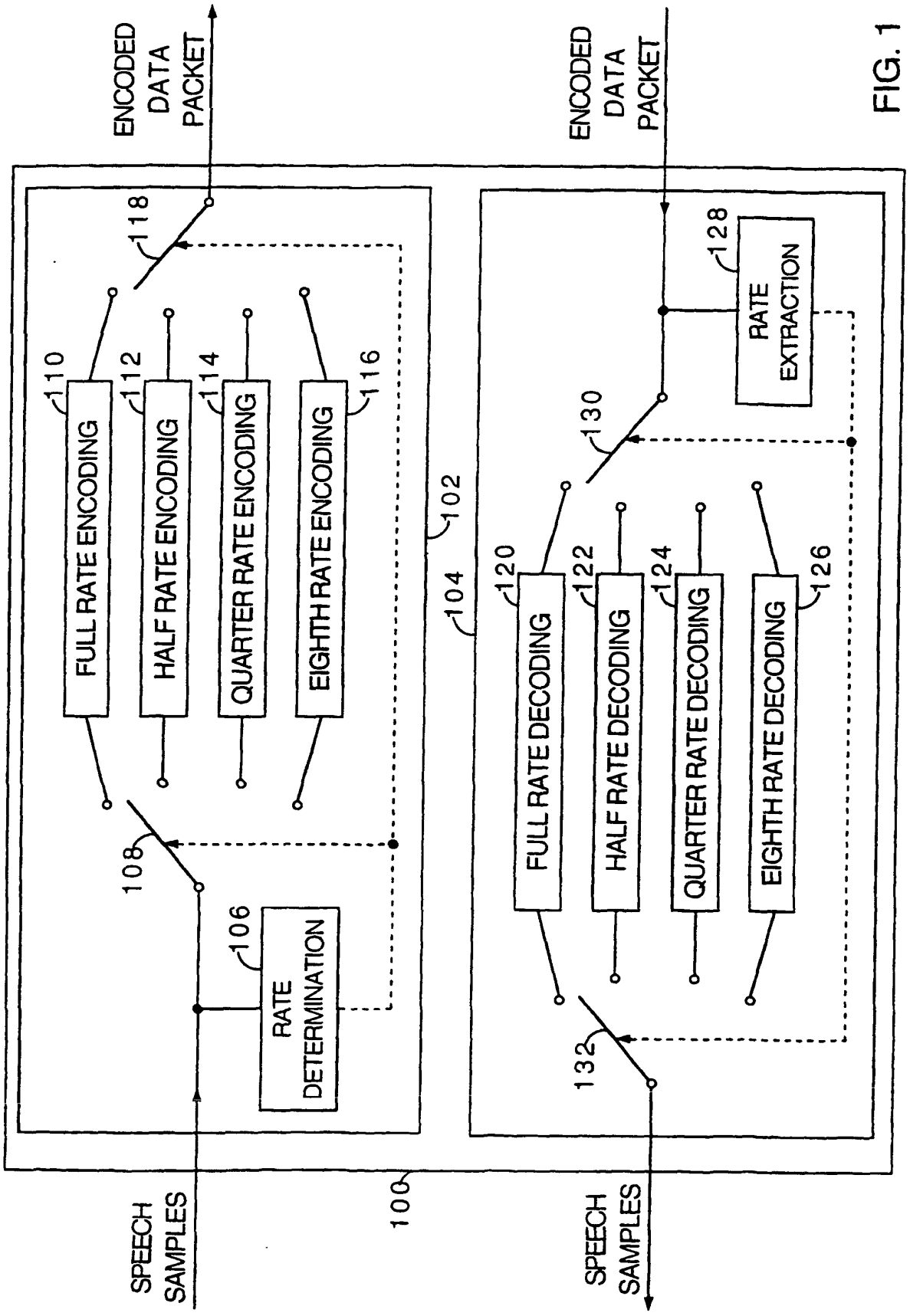
ENCODED DATA PACKET
ENCODED DATA PACKET
118
110
112
114
116
FULL RATE ENCODING
HALF RATE ENCODING
QUARTER RATE ENCODING
EIGHTH RATE ENCODING
108
106
RATE DETERMINATION
102
104
128
RATE EXTRACTION
130
120
122
124
126
FULL RATE DECODING
HALF RATE DECODING
QUARTER RATE DECODING
EIGHTH RATE DECODING
132
100
SPEECH SAMPLES
SPEECH SAMPLES

FIG. 1

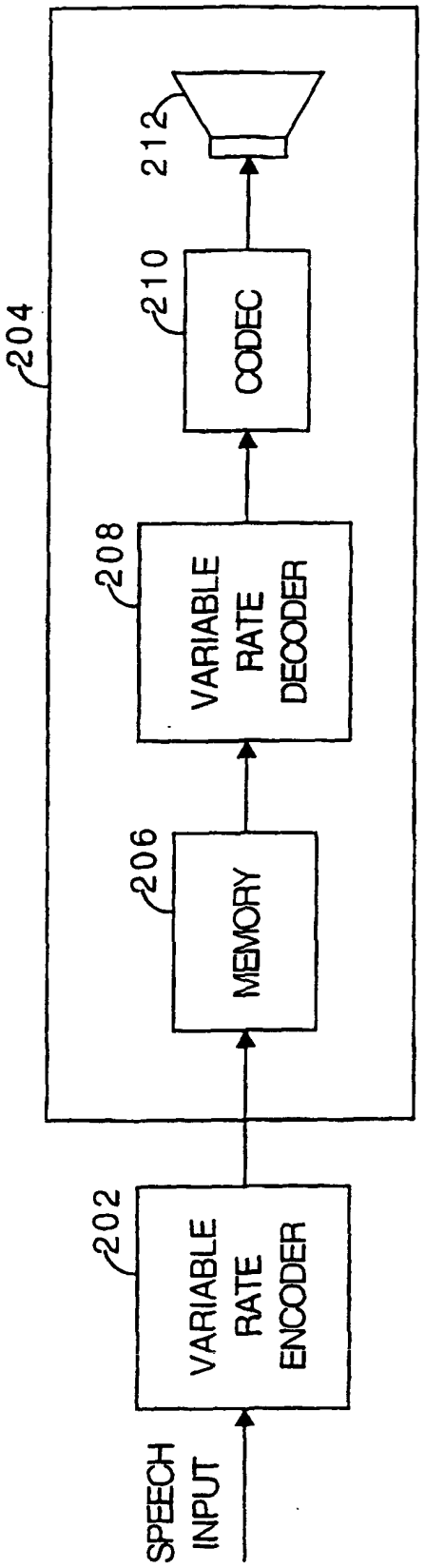
SPEECH INPUT
202
VARIABLE RATE ENCODER
206
MEMORY
208
VARIABLE RATE DECODER
210
CODEC
212
204

FIG. 2