



(11) **EP 1 686 565 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
02.05.2007 Bulletin 2007/18

(51) Int Cl.:
G10L 21/02 (2006.01)

(21) Application number: **05001960.3**

(22) Date of filing: **31.01.2005**

(54) **Bandwidth extension of bandlimited speech data**

Erweiterung der Bandbreite eines schmalbandigen Sprachsignals

Extension de la largeur de bande d'un signal vocal à bande étroite

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR

(43) Date of publication of application:
02.08.2006 Bulletin 2006/31

(73) Proprietor: **Harman Becker Automotive Systems GmbH**
76307 Karlsbad (DE)

(72) Inventors:
• **Iser, Bernd**
89077 Ulm (DE)
• **Schmidt, Gerhard Uwe**
89081 Ulm (DE)

(74) Representative: **Grünecker, Kinkeldey, Stockmair & Schwanhäusser**
Anwaltssozietät
Maximilianstrasse 58
80538 München (DE)

(56) References cited:
WO-A-03/003350

- **EPPS J ET AL: "A new technique for wideband enhancement of coded narrowband speech" SPEECH CODING PROCEEDINGS, 1999 IEEE WORKSHOP ON PORVOO, FINLAND 20-23 JUNE 1999, PISCATAWAY, NJ, USA,IEEE, US, 20 June 1999 (1999-06-20), pages 174-176, XP010345554 ISBN: 0-7803-5651-9**
- **CHEUNG-FAT CHAN ET AL: "Wideband re-synthesis of narrowband CELP-coded speech using multiband excitation model" SPOKEN LANGUAGE, 1996. ICSLP 96. PROCEEDINGS., FOURTH INTERNATIONAL CONFERENCE ON PHILADELPHIA, PA, USA 3-6 OCT. 1996, NEW YORK, NY, USA,IEEE, US, vol. 1, 3 October 1996 (1996-10-03), pages 322-325, XP010237710 ISBN: 0-7803-3555-4**

EP 1 686 565 B1

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

Field of Invention

[0001] The present invention relates to bandwidth extension of transmitted speech data by synthesizing frequency ranges that are not transmitted and, in particular, to bandwidth extension of speech signals transmitted by telephone systems using speaker-dependent information.

Prior Art

[0002] The quality of transmitted audio signals often suffers from some bandwidth limitations. Different from natural face-to-face speech communication, that covers a frequency range from approximately 20 Hz to 20 kHz, communication by telephones or cellular phones is characterized by a limited bandwidth. Common telephone audio signals, in particular, speech signals show a limited bandwidth of only 300 Hz - 3.4 kHz. Speech signals with lower and higher frequencies are simply not transmitted thereby resulting in degradation in speech quality, in particular, manifested in a reduced intelligibility.

[0003] Possible solutions for the problem of enhancing telephone bandwidths consist in the combination of two or more bandlimited speech channels or the utilization of so-called wideband speech codes. Both methods demand for significant service modifications and result in an undesirably increase of costs.

[0004] Thus, it is highly preferable to provide an enhanced bandwidth at the receiver side of the communication. Due to the very nature of the human vocal tract, there is some correlation between a bandlimited speech signal and those frequency parts of the original utterance that are missing due to band restrictions. Consequently, promising methods of bandwidth extension comprise the synthesizing of wideband speech signals from bandlimited speech signals.

[0005] Usually, some speech signal analysis precedes the generation of wideband speech signals from bandlimited ones as, e.g., telephone speech signals. Generally, at least two processing steps have to be performed. In the first step, the wideband spectral envelope is estimated from the determined bandlimited envelope extracted from the bandlimited speech signal.

[0006] In general, lookup tables or code books (see "A New Technique for Wideband Enhancement of Coded Bandlimited Speech," by J. Epps and W.H. Holmes, IEEE Workshop on Speech Coding, Conf. Proc., p. 174, 1999) have to be generated, which define correspondences between bandlimited and wideband spectral envelope representations of speech signals. The closest wideband spectral envelope representation of the extracted bandlimited spectral envelope representation of the received speech signal has to be identified in the code book and has subsequently to be used to synthesize the required wideband speech signal. The synthesizing process in-

cludes the generation of highband and lowband signals in the respective frequency ranges above and below the frequency range of the bandlimited signals.

[0007] The generation of the code books may be achieved by means of a Linear Predictive Coding (LPC) analysis. According to this method, LPC coefficients are extracted from wideband training signals. These signals are band-pass filtered and the LPC coefficients of the resulting bandlimited signals are also extracted thereby allowing to establishing a correspondence between the LPC representations of the bandlimited and the wideband signals.

[0008] Alternatively or complementing the utilization of code books, artificial neural networks can be employed for the non-linear mapping of bandlimited spectral envelope representations of speech signals to the respective wideband representations (see, e.g., "Bandwidth Extension of Narrowband Speech for Low Bit-Rate Wideband Coding", by J.-M. Valin and R. Lefebvre, IEEE Workshop on Speech Coding, Conf. Proc., p. 130, 2000).

[0009] In the second step, a wideband excitation signal is to be generated from the received bandlimited speech signal. The excitation signal ideally represents the signal that would be detected immediately at the vocal chords. The excitation signal may be generated, e.g. by non-linear characteristic curves (see "Spectral Widening of the Excitation Signal for Telephone-Band Speech Enhancement", by U. Kornagel, IWANEC 2001, Conf. Proc., p. 215, 2001), or on the basis of the pitch and power of the bandlimited excitation signal. In order to extend the bandwidth of the telephone band the modeled excitation signal is then shaped with the estimated wideband spectral envelope and added to the bandlimited signal.

[0010] However, code books are generated during a training phase that is generally very time consuming. The weights of artificial neural networks have usually to be extensively trained off-line before usage. Moreover, the training is usually to be performed in a speaker-independent way, since the user is not known a priori. This implies that large databases have to be processed and generated which makes the training procedure rather time-consuming.

[0011] Nevertheless, the achievable quality is not the highest possible, since individual speaker-dependent features cannot be taken into account. Moreover, the training results received in some studio environment are usually not sufficiently compatible with real life applications, in particular, in the case of noisy environments as in vehicular cabins.

[0012] It is therefore the problem underlying the present invention to overcome the above-mentioned drawbacks and to provide a system and a method for speech processing of bandlimited speech communication with an effectively extended bandwidth synthesized reliably at the receiver side.

Description of the Invention

[0013] The above-mentioned problem is solved by the method according to claim 1 and the system according to claim 11. According to claim 1 it is provided a method for generating wideband speech signals from bandlimited speech signals transmitted and received by a first party and by a second party, comprising generating and transmitting first speaker-dependent data by the first party; receiving the first speaker-dependent data by the second party; and generating first wideband speech signals on the basis of the first speaker-dependent data by the second party.

[0014] Speaker-dependent data are generated by the first party from the utterances by a first speaker or communication partner. The utterances, i.e. the speech signals, are detected and analyzed in order to build a database that subsequently can be used by the other party, i.e. the second party, for determining appropriate wideband signals for the received bandlimited signals that are transmitted by the first party. A second communication partner can therefore listen to synthesized wideband signals. Note, that by the expression first or second "party" herein it is referred to the corresponding side, in particular, the technical means of the telecommunication system.

[0015] The speaker-dependent data may comprise bandlimited speech parameters and the associated wideband speech parameters. The bandlimited speech parameters can be obtained by the first party utilizing a band pass filter that allows the frequency range to be passed that corresponds to the frequency range available for the data channel used for transmitting the speech signals detected by the first party to the second party.

[0016] The bandlimited parameters may comprise characteristic parameters for the determination of bandlimited spectral envelopes and/or the pitch and/or the short-time power and/or the highband-pass-to-lowband-pass power ratio and/or the signal-to-noise ratio, and the wideband parameters may comprise wideband spectral envelopes and/or characteristic parameters for the determination of wideband spectral envelopes and/or wideband excitation signals.

[0017] The second party is enabled to synthesize wideband speech signals received from the first party using the speaker-dependent data. In particular, the received signals are analyzed and bandlimited speech parameters are determined. Appropriate speaker-dependent bandlimited parameters included in the speaker-dependent data may be assigned to the analyzed bandlimited parameters. The speaker-dependent bandlimited parameters can subsequently be mapped to the according wideband parameters.

[0018] Consequently, individual speaker-dependent features can be used for the bandwidth extension. Moreover, the speaker-dependent data are automatically adapted to the environment of the speaker including the

room acoustics and, in particular, the microphone characteristics. Thus, an improved quality and reduced complexity as well as almost no artifacts can be achieved by the disclosed method for bandwidth extension.

[0019] A further data channel may be needed to transmit the speaker-dependent data in addition to and concurrently with bandlimited speech signal. The data transfer rate of the additional channel may be relatively low. In addition, no synchronization is necessary and relatively long delay times are tolerable.

[0020] The disclosed method may be used for one party only or may be used for both parties in which case the same additional data channel for the respective speaker-dependent speech data may be used. Accordingly, the disclosed method may also comprise generating and transmitting second speaker-dependent data by the second party, receiving the second speaker-dependent data by the first party and generating second wideband speech signals on the basis of the second speaker-dependent data by the first party. Thereby, both communication partners can profit from the increased quality of the speech signals.

[0021] Furthermore, an embodiment of the method may further comprise providing a database for the second party that is not transmitted by the first party and the first wideband speech signals may be generated on the basis of the first speaker-dependent data and the database that is not transmitted by the first party.

[0022] If bandwidth extension is provided for both the second and the first party, the method for generating wideband speech signals may further comprise providing a database for the second party that is not transmitted by the first party and/or providing another database for the first party that is not transmitted by the second party, and the first wideband speech signals may be generated on the basis of the first speaker-dependent data and the database that is not transmitted by the first party and/or the second wideband speech signals may be generated on the basis of the second speaker-dependent data and the other database that is not transmitted by the second party.

[0023] The two further provided and not transmitted databases may or may not be identical copies of the same database that may comprise speaker-independent offline training results.

[0024] Usage of both kinds of databases in parallel allows for a decision about which one gives better results, e.g., in terms of distance measures. Accordingly, some of the wideband speech signals may be generated using the transmitted data and some other may be generated using databases that are not transmitted. Moreover, some weighted average of the wideband parameters of the different data sets may be used for synthesizing the wideband speech signals.

[0025] In generating the wideband speech signals priority may be given to the transmitted speaker-dependent data, in particular, if it is to be expected, e.g., on the grounds of distance measures, that with the help of these

data better results can be achieved.

[0026] To give priority to the transmitted speaker-dependent data may, in particular, mean that at first the speaker-dependent data, e.g. speaker-dependent code books, are used in order to determine appropriate wideband speech parameters and only, if some distance measures show no satisfying values, the database that is not transmitted and possibly comprises speaker-independent data is taken into account for wideband speech synthesizing. It may also mean that in case of significantly different estimates for the wideband signal, the estimate obtained by means of the speaker-dependent data is chosen for the synthesizing of the wideband speech signal.

[0027] It may be preferred that the transmitting of speaker-dependent data starts after the generation of the entire speaker-dependent data is completed. The generation of valuable speaker-dependent data may take some time. In the case of transmission of the speaker-dependent data after completion of the training process, the bandwidth extension may be performed exclusively on the basis of the not transmitted database until the speaker-dependent data are transmitted to allow for a further increase of the quality of the speech signals.

[0028] The speaker-dependent data may comprise speaker-dependent code books and/or weights for artificial neural networks.

[0029] Artificial neural networks may be employed that are composed of many computing elements, usually denoted neurons, and working in parallel. The elements are connected by synaptic weights, which are allowed to adapt through learning or training processes.

[0030] Different network types may advantageously be employed, e.g. a model including supervised learning in a feed-forward (signal transfer) network. The neural network is given an input signal, which is transferred forward through the network. Eventually, an output signal is produced. The neural network can be understood as a means for mapping from the input space to the output space, and this mapping is defined by the free parameters of the model, which are the synaptic weights connecting the neurons.

[0031] According to embodiments of the inventive method for bandwidth extension smaller code books or neural networks with a reduced number of neurons as compared to the art are sufficiently reliable.

[0032] The speaker-dependent data can be generated using data sampled before the first and/or the second parties transmit and receive the bandlimited speech signals. This previously sampled data may be generated by a speech recognizing means.

[0033] It may be preferred not to use any stored speech data but rather the extracted parameters of the speech analysis performed by the recognizing means, as, e.g., coefficients for a Linear Predictive Coding (LPC). Alternatively, also the recorded speech during previous telephone calls, respectively extracted parameters, can be utilized for this purpose. However, speaker identification

is necessary to make sure that the person currently speaking is equal to the one who has spoken during the recordings.

[0034] Providing previously sampled data, i.e. data generated in the past before an actual communication has started, allows for immediately transmitting speaker-dependent data at the very beginning of the actual communication.

[0035] The synthesizing of wideband speech signals may comprise generating highband and/or lowband speech signals and adaptation of parameters that are needed to generate the highband and/or lowband speech signals.

[0036] By 'highband' and 'lowband' those parts of the frequency spectrum are meant, that are synthesized in addition to the received limited band.

[0037] Furthermore, the present invention provides a computer program product, comprising one or more computer readable media having computer-executable instructions for performing the steps of the above describe embodiments of the inventive method for speech processing of generating wideband speech signals from bandlimited speech signals.

[0038] It is further provided a system for bandwidth extension of bandlimited speech signals transmitted and received by a first party and by a second party, comprising

a first database generated by the first party comprising first speaker-dependent data;

a first transmitting means for transmitting the first database to the second party;

a first analyzing means configured to analyze the bandlimited speech signals received by the second party to determine at least one first analyzed bandlimited speech parameter;

a first mapping means configured to determine at least one first wideband speech parameter from the at least one first analyzed bandlimited speech parameter on the basis of the first database; and

a first wideband synthesizing means for synthesizing first wideband speech signals on the basis of the at least one first wideband speech parameter.

[0039] The wideband synthesizing means may comprise an audio signal generating means configured to generate a highband and/or lowband audio signal on the basis of the at least one wideband parameter.

[0040] Preferably, a second data channel is provided for transmitting and receiving the database comprising the speaker-dependent data.

[0041] The system may also comprise a second database generated by the second party comprising second speaker-dependent data, a second transmitting means for transmitting the second database to the first party, a second analyzing means configured to analyze the bandlimited speech signals received by the first party to determine at least one second analyzed bandlimited speech parameter, a second mapping means configured to determine at least one second wideband speech pa-

parameter from the at least one second analyzed bandlimited speech parameter on the basis of the second database and a second wideband synthesizing means for synthesizing second wideband speech signals on the basis of the at least one second wideband speech parameter.

[0042] Furthermore, the system may comprise a third database provided for the second party and not transmitted by the first party and the first mapping means can be configured to determine the at least one first wideband speech parameter on the basis of the first database and on the basis of the third database.

[0043] If bandwidth extension is provided for both the second and the first party, the system may also comprise a third database provided for the second party and not transmitted by the first party, and the first mapping means may be configured to determine the at least one first wideband speech parameter on the basis of the first database and on the basis of the third database and/or a fourth database provided for the first party and not transmitted by the second party, and the second mapping means may be configured to determine the at least one second wideband speech parameter on the basis of the second database and on the basis of the fourth database.

[0044] In the first (second) and third (forth) databases the same bandlimited parameter(s) or characterizing vector(s), may be identified in correspondence to the analyzed bandlimited parameter(s) or characterizing vector(s). Mapping to both the associated wideband parameter(s) in the first (second) and third (forth) databases may be performed.

[0045] The first and/or second mapping means can be configured to give priority to the speaker-dependent data when determining the at least one wideband speech parameter.

[0046] Moreover, the first and/or second transmitting means can be configured to start the transmission of speaker-dependent data after the generation of the entire speaker-dependent data is completed.

[0047] The first and/or second databases may preferably comprise speaker-dependent code books and/or weights for artificial neural networks.

[0048] Furthermore, the first and/or second databases may comprise speaker-dependent data sampled before the first and/or the second parties transmit and receive the bandlimited speech signals.

[0049] According to an embodiment the inventive system may further comprise a speech recognizing means for generating speaker-dependent data. If the first and/or second databases may comprise speaker-dependent data sampled before the first and/or the second parties transmit and receive the bandlimited speech signals, this data may be generated by the speech recognizing means.

[0050] The disclosed system may comprise a control unit for controlling the determining of the at least one wideband speech parameter and the synthesizing of the wideband speech signals and the control unit may control

the synthesizing means to adapt parameters that are needed to generate highband and/or lowband speech.

[0051] Further provided are a hands-free set, in particular, for use in a vehicle, as well as a mobile phone comprising one of the above described embodiments of the inventive system. Employment of embodiments the inventive system in fixed-installed phones, mobile phones and hands-free sets improves the intelligibility of speech signals significantly. In the rather noise environment of vehicular cabins embodiments of the disclosed system is considered to be advantageous for the communication via hands-free sets.

[0052] Additional features and advantages of the present invention will be described with reference to the drawings. In the description, reference is made to the accompanying figures that are meant to illustrate preferred embodiments of the invention. It is understood that such embodiments do not represent the full scope of the invention that is defined by the claims given below.

[0053] Figure 1 shows elementary steps of an example for the disclosed method for bandwidth extension of bandlimited speech signals comprising training of speaker-dependent code books, transmitting and receiving speech signals as well as the code books and analyzing the speech signals and performing a mapping of the results of the speech analysis to the entries of the code books.

[0054] Figure 2 further illustrates steps of an example of the disclosed method comprising non-linear mapping by means of speaker-dependent and speaker-independent code books.

[0055] Figure 3 shows elements of the disclosed system for bandwidth extension of bandlimited speech signals comprising a control unit, a pair of code books and a wideband synthesizing means.

[0056] Consider two parties, a remote or first party and a near or second party, connected by a telephone line. With respect to Figure 1, one may consider that a remote speaker speaks at some given time. The verbal utterances by the first speaker 10 are detected and processed 11 by the first party. Detection of the utterances may be performed by a microphone or a microphone array. The processing can include noise reduction as well as beamforming of the speech signals that are converted to electrical signals.

[0057] The speech signals are limited to a bandwidth of 300 Hz - 3.4 kHz and they are transmitted 12 to the second/near party. Transmission may, e.g., be performed by radio using the Global System for Mobile Communications (GSM).

[0058] When the communication starts, at least at the first/remote party speaker-dependent code books 13 are trained. According to one embodiment, Linear Predictive Coding (LPC) analysis coefficients are extracted from wideband training signals. These signals are subsequently band-pass filtered and the LPC coefficients of the resulting bandlimited signals are extracted. Thereby a unique correspondence between the LPC representa-

tions of the bandlimited and the wideband signals can be established.

[0059] The training process will take some time. It might be supported by training results achieved at telephone calls by the same speaker in the past, i.e. a combined on-line and off-line training can be performed to increase performance. The off-line training can also be performed using stored analyzing parameters of a speech recognizing system. A decision means will decide when the training process is completed, e.g., after some hundred words have been learned.

[0060] After completion of the training process the speaker-dependent code books 13, i.e. one code book for the wideband speech signals and another for the bandlimited speech signals, are transmitted to the second party 14. For the transmission of the code books a data transmission channel different from the one used for the transmission of the bandlimited speech signals representing the verbal utterances of the first speaker must usually be available. The data transfer rate of this further data channel may be relatively low as compared to the one of the channel for the speech signals. Alternatively, the same channel that is used for the speech transmission can be utilized as well for the transmission of the code books by applying techniques such as watermarking, i.e. hiding the additional data within the speech signal by means of masking.

[0061] Thus, according to the present example for the inventive method of bandwidth extension of transmitted speech signals the second party receives both bandlimited speech signals and, after completion of the training process, a (remote) speaker-dependent pair of code books 15.

[0062] The received bandlimited speech signals are processed and, in particular, analyzed for the spectral features and based on the results of the speech analysis a non-linear mapping to the appropriate wideband signals is performed 16.

[0063] The bandlimited speech signal is converted to the desired bandwidth, by increasing the sample rate, without generating additional frequency ranges. If, for example, a bandlimited signal is sampled at 8 kHz it may be processed to obtain the signal at a sampling frequency of 16 kHz.

[0064] Whereas the analyzing step may be performed according to any method known from the art, the mapping makes use of the received code books trained on the first side. The analyzing process provides the bandlimited spectral envelope and estimate for the narrowband excitation signal that ideally represents the signal that would be detected immediately at the vocal chords after appropriate bandpass filtering. The estimated wideband excitation signal can subsequently be shaped by the estimated wideband spectral envelope in order to obtain a synthesized wideband signal 17.

[0065] The wideband spectral envelope is assigned to the bandlimited one by a non-linear mapping means on the basis of the analyzed bandlimited parameters.

[0066] Wideband parameters contained in one of the code books are mapped to bandlimited parameters contained in another code book. According to the inventive method the code books are trained by the first speaker. Different from the art speaker-dependent code books can, thus, be used for the estimation of the wideband spectral envelope.

[0067] After the appropriate wideband spectral envelopes for the received bandlimited speech signals are obtained, the entire wideband speech signals may be synthesized 17. Instead of synthesizing the entire wideband signals, it may be preferred that the synthesized speech signal portion outside the bandwidth of the bandlimited signal, i.e. the highband and lowband speech signals, are added to the detected and analyzed bandlimited signal.

[0068] It is noted that instead of or in addition to code books artificial networks may be used. In this case, the weights of the neural network are trained during the telephone conversation at the first side. After completion of the training process, the weights are transmitted to the second party.

[0069] The most popular of all neural networks, the Multi-Layer Perceptron network, may advantageously be used. The basic unit (neuron) of the network is a perceptron. This is a computation unit, which produces its output by taking a linear combination of the input signals and by transforming this by a function called activity function. The output of the perceptron as a function of the input signals can thus be written:

$$y = \sigma(\sum w_i x_i + \theta),$$

where y is the output, x_i is the input signals ($i = 1, \dots, n$), w_i is the neuron weights, θ is the bias term (another neuron weight) and σ is the activity function. Possible forms of the activity function are linear function, step function, logistic function and hyperbolic tangent function. The kind of activity function may also be transmitted together with the weights. Usually, however, the activity function will be pre-determined in the neural networks the first and the second party may be provided with.

[0070] Whereas in the present example training is performed by the first/remote party and the trained code books and/or weights for the neural network are transmitted to and received by the second/near party in order to extend the bandlimited signals transmitted to the second party, the same operation may be carried out in the opposite direction. The same data channel can be used to transmit code books from the first party to the second party and vice versa.

[0071] Figure 2 illustrates a further example for the herein disclosed method for bandwidth extension of bandlimited speech signals. Speech signals 20 are received by a near party that are transmitted by a remote party. These speech signals are bandlimited due to some

restrictions of the data processing by the remote party and/or a limited capacity of the data channel used for the transmission of the speech signals.

[0072] The received speech signal are analyzed 21 as described above. Based on the results of the speech analysis a non-linear mapping means can perform assignment of wideband parameters to the analyzed bandlimited parameters 22. Different from the example discussed with respect to Figure 1, speaker-independent code books 23 and speaker-dependent ones 24 are available for the mapping 22. To be more specific, the analyzed bandlimited parameters can be compared with respective ones in bandlimited speaker-independent 23 or speaker-dependent 24 code books, and the best matching bandlimited parameter is identified in the code books 23 and 24. This bandlimited parameter can be the same in both code books 23 and 24.

[0073] At the beginning of the communication with the first party, the assignment of the appropriate wideband parameters, that are necessary for synthesizing wideband signals 25, to the analyzed 21 bandlimited parameters is exclusively done by using the speaker-independent code books 23. The risk of producing artifacts is relatively high at this stage of modeling wideband speech signals.

[0074] However, after some time of communication has elapsed, speaker-dependent code books 24 are generated and transmitted to the second party. After the pair of speaker-dependent code books 24 has been received, it can be used for the future synthesizing of wideband speech signals 25. Since these code books 24 are generated for the actual speaker's communication environment and the individual speech characteristics, the quality of the synthesized speech signals should be improved significantly as compared to the speech signals generated on the basis of the speaker-independent code books 23.

[0075] Some operative elements of the system for bandwidth extension of bandlimited speech signals according to an embodiment of the present invention are illustrated in Figure 3. Speech data 30 is input in the system as bandlimited speech signals X_{Lim} 31. The input speech signal is analyzed by an analyzing means 32. The analyzing means comprises means for extracting the bandlimited spectral envelope and for determining the power of the bandlimited excitation signal.

[0076] The analysis data are transmitted to a control unit 33. The analyzed bandlimited parameters are used to generate at least one characteristic vector that may be a cepstral vector. The characteristic vector is assigned to the vector of the bandlimited code book with the smallest distance to this characteristic vector. As a distance measure, e.g., the Itakuro-Saito distance measure, may be used. The vector determined in the bandlimited code book is mapped to the corresponding characterizing vector of the wideband code book. The bandlimited and the wideband code book constitute the pair of code books 34.

[0077] According to the inventive system, not only

speech data 30 are transmitted from one party to another but also for one or both of the communication partners speaker-dependent code books are generated before and/or during the communication. After the code books are completely generated by one party, they are transmitted to the other party. Thus, in addition to speech data 30 speaker-dependent data 35 comprising a pair of speaker-dependent code books 34 are transmitted via a further data channel.

[0078] A means for generating wideband excitation signals 36 is also controlled by the control unit 33 and provided to generate the wideband excitation signals corresponding to the respective lowband excitation signals that are obtained by the analyzing means 32. A wideband synthesizing means 37 eventually generates the wideband signals x_{WB} 38 on the basis of the wideband excitation signals and the wideband spectral envelopes.

[0079] The wideband signals x_{WB} 38 comprise lowband and highband speech portions that are missing in the detected bandlimited signals 31. If, e.g., the bandlimited signal shows a frequency range from 300 Hz to 3,4 kHz, the lowband and the highband signals may show frequency ranges from 50 - 300 Hz and from 3,4 kHz to a predefined upper frequency limit with a maximum of half of the sampling rate, respectively.

Claims

1. Method for generating wideband speech signals from bandlimited speech signals transmitted and received by a first party and by a second party, comprising
 - generating and transmitting first speaker-dependent data by the first party;
 - receiving the first speaker-dependent data by the second party;
 - analyzing the bandlimited speech signals received by the second party to determine at least one first analyzed bandlimited speech parameter; and
 - generating first wideband speech signals from the at least one first analyzed bandlimited speech parameter on the basis of the first speaker-dependent data by the second party.
2. Method according to claim 1, further comprising
 - generating and transmitting second speaker-dependent data by the second party;
 - receiving the second speaker-dependent data by the first party; and
 - generating second wideband speech signals on the basis of the second speaker-dependent data by the first party.
3. Method according to claim 1, further
 - providing a database for the second party that is not transmitted by the first party; and
 - wherein the first wideband speech signals are gen-

- erated on the basis of the first speaker-dependent data and the database that is not transmitted by the first party.
4. Method according to claim 2, further comprising providing a database for the second party that is not transmitted by the first party and/or providing another database for the first party that is not transmitted by the second party; and wherein the first wideband speech signals are generated on the basis of the first speaker-dependent data and the database that is not transmitted by the first party and/or the second wideband speech signals are generated on the basis of the second speaker-dependent data and the other database that is not transmitted by the second party.
 5. Method according to claim 3 or 4, wherein in generating the wideband speech signals priority is given to the speaker-dependent data.
 6. Method according to one of the preceding claims, wherein the transmitting of speaker-dependent data starts after the generation of the entire speaker-dependent data is completed.
 7. Method according to one of the preceding claims, wherein the speaker-dependent data comprise speaker-dependent code books and/or weights for artificial neural networks.
 8. Method according to one of the preceding claims, wherein the speaker-dependent data are generated using data sampled before the first and/or the second parties transmit and receive the bandlimited speech signals.
 9. Method according to claim 8, wherein the data sampled before the first and/or the second party transmit and receive the bandlimited speech signals have been generated by a speech recognizing means.
 10. Computer program product, comprising one or more computer readable media having computer-executable instructions performing the steps of the method according to one of the preceding claims. when said computer program product is run on a computer.
 11. System for bandwidth extension of bandlimited speech signals transmitted and received by a first party and by a second party, comprising a first database generated by the first party comprising first speaker-dependent data; a first transmitting means for transmitting the first database to the second party; a first analyzing means configured to analyze the bandlimited speech signals received by the second party to determine at least one first analyzed band-limited speech parameter; a first mapping means configured to determine at least one first wideband speech parameter from the at least one first analyzed bandlimited speech parameter on the basis of the first database; and a first wideband synthesizing means for synthesizing first wideband speech signals on the basis of the at least one first wideband speech parameter.
 12. System according to claim 11, further comprising a second database generated by the second party comprising second speaker-dependent data; a second transmitting means for transmitting the second database to the first party; a second analyzing means configured to analyze the bandlimited speech signals received by the first party to determine at least one second analyzed bandlimited speech parameter; a second mapping means configured to determine at least one second wideband speech parameter from the at least one second analyzed bandlimited speech parameter on the basis of the second database; a second wideband synthesizing means for synthesizing second wideband speech signals on the basis of the at least one second wideband speech parameter.
 13. System according to claim 11, further comprising a third database provided for the second party and not transmitted by the first party; and wherein the first mapping means is configured to determine the at least one first wideband speech parameter on the basis of the first database and on the basis of the third database.
 14. System according to claim 12, further comprising a third database provided for the second party and not transmitted by the first party; and wherein the first mapping means is configured to determine the at least one first wideband speech parameter on the basis of the first database and on the basis of the third database and/or a fourth database provided for the first party and not transmitted by the second party; and wherein the second mapping means is configured to determine the at least one second wideband speech parameter on the basis of the second database and on the basis of the fourth database.
 15. System according to claims 13 or 14, wherein the first and/or second mapping means are configured to give priority to the speaker-dependent data when determining the at least one wideband speech parameter.
 16. System according to one of the claims 11 - 15, where-

in the first and/or second transmitting means are configured to start the transmission of speaker-dependent data after the generation of the entire speaker-dependent data is completed.

17. System according to one of the claims 11 - 16, wherein the first and/or second databases comprise speaker-dependent code books and/or weights for artificial neural networks. 5
18. System according to one of the claims 11 - 17, wherein the first and/or second databases comprise speaker-dependent data sampled before the first and/or the second party transmit and receive the bandlimited speech signals 10
19. System according to claim 18, further comprising a speech recognizing means for generating speaker-dependent data. 15
20. System according to one of the claims 11 - 19, further comprising a control unit for controlling the determining of the at least one wideband speech parameter and the synthesizing of the wideband speech signals, and wherein the control unit controls the synthesizing means to adapt parameters that are needed to generate highband and/or lowband speech signals. 20
21. Hands-free set comprising a system according to one of the claims 11 - 20. 25
22. Mobile phone comprising a system according to one of the claims 11 - 20. 30

Patentansprüche

1. Verfahren zum Erzeugen von Breitband - Sprachsignalen aus bandbeschränkten Sprachsignalen, die von einer ersten Partei und einer zweiten Partei übermittelt und empfangen werden, umfassend Erzeugen und Übermitteln von ersten sprecherabhängigen Daten durch die erste Partei; Empfangen der ersten sprecherabhängigen Daten durch die zweite Partei; Analysieren der von der zweiten Partei empfangenen bandbeschränkten Sprachsignale, um zumindest einen ersten analysierten bandbeschränkten Sprachparameter zu bestimmen, und Erzeugen von ersten Breitband - Sprachsignalen aus dem zumindest einen ersten analysierten bandbeschränkten Sprachparameter auf der Grundlage der ersten sprecherabhängigen Daten durch die zweite Partei. 40
2. Verfahren gemäß Anspruch 1, weiterhin umfassend Erzeugen und Übermitteln von zweiten sprecherab-

hängigen Daten durch die zweite Partei; Empfangen der zweiten sprecherabhängigen Daten durch die erste Partei; und Erzeugen zweiter Breitband - Sprachsignale auf der Grundlage der zweiten sprecherabhängigen Daten durch die erste Partei. 45

3. Verfahren gemäß Anspruch 1, weiterhin eine Datenbank für die zweite Partei zur Verfügung stellend, die nicht von der ersten Partei übermittelt wird; und worin die ersten Breitband - Sprachsignale auf der Grundlage der ersten sprecherabhängigen Daten und der Datenbank, die nicht von der ersten Partei übermittelt wird, erzeugt werden. 50
4. Verfahren gemäß Anspruch 2, weiterhin umfassend Bereitstellen einer Datenbank für die zweite Partei, die nicht von der ersten Partei übermittelt wird und/oder Bereitstellen einer weiteren Datenbank für die erste Partei, die nicht von der zweiten Partei übermittelt wird; und worin die ersten Breitband - Sprachsignale auf der Grundlage der ersten sprecherabhängigen Daten und der Datenbank, die nicht von der ersten Partei übermittelt wird, erzeugt werden und/oder die zweiten Breitband - Sprachsignale auf der Grundlage der zweiten sprecherabhängigen Daten und der weiteren Datenbank, die nicht von der zweiten Partei übermittelt wird, erzeugt werden. 55
5. Verfahren gemäß Anspruch 3 oder 4, in dem bei dem Erzeugen der Breitband - Sprachsignale die Priorität den sprecherabhängigen Daten gegeben wird.
6. Verfahren gemäß einem der vorhergehenden Ansprüche, in dem das Übermitteln von sprecherabhängigen Daten beginnt, nachdem die Erzeugung der gesamten sprecherabhängigen Daten beendet worden ist.
7. Verfahren gemäß einem der vorhergehenden Ansprüche, in dem die sprecherabhängigen Daten sprecherabhängige Code - Bücher und/oder Gewichte für künstliche neuronale Netze umfassen.
8. Verfahren gemäß einem der vorhergehenden Ansprüche, in dem die sprecherabhängigen Daten unter Verwendung von Daten erzeugt werden, die gesammelt worden sind, bevor die erste und/oder zweite Partei die bandbeschränkten Sprachsignale übermittelt und empfängt.
9. Verfahren gemäß Anspruch 8, in dem die Daten, die gesammelt worden sind, bevor die erste und/oder zweite Partei die bandbeschränkten Sprachsignale übermittelt und empfängt, durch eine Spracherken-

nungseinrichtung erzeugt worden sind.

10. Computerprogrammprodukt, das eines oder mehrere computerisierbare Medien umfasst, die computerisierbare Anweisungen aufweisen, die die Schritte des Verfahrens gemäß einem der vorhergehenden Ansprüche ausführen, wenn das genannte Computerprogrammprodukt auf einem Computer laufengelassen wird.
11. System für Bandbreitenerweiterung von bandbeschränkten Sprachsignalen, die von einer ersten Partei und von einer zweiten Partei übermittelt und empfangen werden, umfassend
 eine erste Datenbank, die durch die erste Partei erzeugt wird, welche erste sprecherabhängige Daten umfasst;
 eine erste Übermittlungseinrichtung zum Übermitteln der ersten Datenbank an die zweite Partei;
 eine erste Analyseinrichtung, die dazu konfiguriert ist, die bandbeschränkten Sprachsignale zu analysieren, die von der zweiten Partei empfangen werden, um zumindest einen ersten analysierten bandbeschränkten Sprachparameter zu bestimmen;
 eine erste Abbildungseinrichtung, die dazu konfiguriert ist, zumindest einen ersten Breitband - Sprachparameter aus dem zumindest einen ersten analysierten bandbeschränkten Sprachparameter auf der Grundlage der ersten Datenbank zu bestimmen; und
 eine erste Breitband - Synthetisierereinrichtung zum Synthetisieren von ersten Breitband - Sprachsignalen auf der Grundlage des zumindest einen ersten Breitband - Sprachparameters.
12. System entsprechend Anspruch 11, weiterhin umfassend
 eine zweite Datenbank, die von der zweiten Partei erzeugt wird, welche zweite sprecherabhängige Daten umfasst;
 eine zweite Übermittlungseinrichtung zum Übermitteln der zweiten Datenbank an die erste Partei;
 eine zweite Analyseinrichtung, die dazu konfiguriert ist, die bandbeschränkten Sprachsignale zu analysieren, die von der ersten Partei empfangen werden, um zumindest einen zweiten analysierten bandbeschränkten Sprachparameter zu bestimmen;
 eine zweite Abbildungseinrichtung, die dazu konfiguriert ist, zumindest einen zweiten Breitband - Sprachparameter aus dem zumindest einen zweiten analysierten bandbeschränkten Sprachparameter auf der Grundlage der zweiten Datenbank zu bestimmen;
 eine zweite Breitband - Synthetisierereinrichtung zum Synthetisieren von zweiten Breitband - Sprachsignalen auf der Grundlage des zumindest einen zweiten Breitband - Sprachparameters.
13. System gemäß Anspruch 11, weiterhin umfassend eine dritte Datenbank, die für die zweite Partei zur Verfügung gestellt wird und nicht von der ersten Partei übermittelt wird, und worin
 die erste Abbildungseinrichtung dazu konfiguriert ist, den zumindest einen ersten Breitband - Sprachparameter auf der Grundlage der ersten Datenbank und auf der Grundlage der dritten Datenbank zu bestimmen.
14. System gemäß Anspruch 12, weiterhin umfassend eine dritte Datenbank, die für die zweite Partei zur Verfügung gestellt wird und nicht von der ersten Partei übermittelt wird, und worin
 die erste Abbildungseinrichtung dazu konfiguriert ist, den zumindest einen ersten Breitband - Sprachparameter auf der Grundlage der ersten Datenbank und auf der Grundlage der dritten Datenbank zu bestimmen und/oder
 eine vierte Datenbank, die für die erste Partei zur Verfügung gestellt wird und nicht von der zweiten Partei übermittelt wird, und worin
 die zweite Abbildungseinrichtung dazu konfiguriert ist, den zumindest einen zweiten Breitband - Sprachparameter auf der Grundlage der zweiten Datenbank und auf der Grundlage der vierten Datenbank zu bestimmen.
15. System gemäß Anspruch 13 oder 14, in dem die erste und/oder zweite Abbildungseinrichtung so konfiguriert ist, dass sie die Priorität den sprecherabhängigen Daten gibt, wenn sie den zumindest-einen Breitband - Sprachparameter bestimmt.
16. System gemäß einem der Ansprüche 11 - 15, in dem die erste und/oder zweite Übermittlungseinrichtung konfiguriert ist, die Übertragung von sprecherabhängigen Daten zu beginnen, nachdem die Erzeugung der gesamten sprecherabhängigen Daten beendet worden ist.
17. System gemäß einem der Ansprüche 11 -16, in dem die erste und/oder zweite Datenbank sprecherabhängige Code - Bücher und/oder Gewichte für künstliche neuronale Netze umfasst.
18. System gemäß einem der Ansprüche 11 -17, in dem die erste und/oder zweite Datenbank sprecherabhängige Daten umfasst, die gesammelt worden sind, bevor die erste und/oder zweite Partei die bandbeschränkten Sprachsignale übermittelt und empfängt.
19. System gemäß Anspruch 18, weiterhin eine Spracherkennungseinrichtung zum Erzeugen von sprecherabhängigen Daten umfassend.
20. System gemäß einem der Ansprüche 11 - 19, wei-

terhin eine Steuereinheit zum Steuern des Bestimmen des zumindest einen Breitband - Sprachparameters and des Synthetisierens der Breitband - Sprachsignale umfassend, und worin die Steuereinheit die Synthetisierereinrichtung dahingehend steuert, Parameter anzupassen, die dazu benötigt werden, Hochband - und/oder Tiefband - Sprachsignale zu erzeugen.

21. Freisprecheinrichtung, die ein System gemäß einem der Ansprüche 11 - 20 umfasst.
22. Mobiltelefon, das ein System gemäß einem der Ansprüche 11 - 20 umfasst.

Revendications

1. Procédé pour générer des signaux vocaux à large bande depuis des signaux vocaux à bande limitée transmis et reçus par un premier correspondant et par un second correspondant, comprenant générer et transmettre des premières données dépendantes du locuteur par le premier correspondant ; recevoir les premières données dépendantes du locuteur par le second correspondant ; analyser les signaux vocaux à bande limitée reçus par le second correspondant pour déterminer au moins un premier paramètre vocal à bande limitée analysée ; et générer des premiers signaux vocaux à large bande depuis au moins un premier paramètre vocal à bande limitée analysée sur la base des premières données dépendantes du locuteur par le second correspondant.
2. Procédé selon la revendication 1, comprenant en outre de générer et transmettre des secondes données dépendantes du locuteur par le second correspondant ; recevoir les secondes données du locuteur par le premier correspondant ; et générer des seconds signaux vocaux à large bande sur la base des secondes données dépendantes du locuteur par le premier correspondant.
3. Procédé selon la revendication 1, comprenant en outre fournir une base de données pour le second correspondant qui n'est pas transmise par le premier correspondant ; et dans lequel les premiers signaux vocaux à large bande sont générés sur la base des premières données dépendantes du locuteur et la base de données qui n'est pas transmise par le premier correspondant.
4. Procédé selon la revendication 2, comprenant en outre fournir une base de données pour le second correspondant qui n'est pas transmise par le premier correspondant et/ou fournir une autre base de données pour le premier correspondant qui n'est pas transmise par le second correspondant ; dans lequel les premiers signaux vocaux à large bande sont générés sur la base des premières données dépendantes du locuteur et la base de données qui n'est pas transmise par le premier correspondant et/ou les seconds signaux vocaux à large bande sont générés sur la base des secondes données dépendantes du locuteur et de l'autre base de données qui n'est pas transmise par le second correspondant.
5. Procédé selon la revendication 3 ou 4, dans lequel en générant les signaux vocaux à large bande la priorité est donnée aux données dépendantes du locuteur.
6. Procédé selon l'une quelconque des revendications précédentes, dans lequel la transmission des données dépendantes du locuteur commence après que la génération de la totalité des données dépendantes du locuteur soit achevée.
7. Procédé selon l'une quelconque des revendications précédentes, dans lequel les données dépendantes du locuteur comprennent des livres de codes dépendant du locuteur et/ou des poids pour réseaux neuronaux artificiels.
8. Procédé selon l'une quelconque des revendications précédentes, dans lequel les données dépendantes du locuteur sont générées en utilisant des données échantillonnées avant que les premier et/ou second correspondant(s) transmettent et reçoivent les signaux vocaux à bande limitée.
9. Procédé selon la revendication 8, dans lequel les données échantillonnées avant que les premier et/ou second correspondant(s) transmettent et reçoivent les signaux vocaux à bande limitée ont été générées par des moyens de reconnaissance vocale.
10. Produit de programme d'ordinateur, comprenant un ou plusieurs média(s) lisible(s) par ordinateur ayant des instructions exécutables par ordinateur réalisant les étapes de la méthode selon l'une des revendications précédentes quand ledit produit de programme d'ordinateur est exécuté par un ordinateur.
11. Système pour extension à large bande de signaux vocaux à bande limitée transmis et reçu par un pre-

mier correspondant et par un second correspondant, comprenant :

une première base de données générée par le premier correspondant comprenant des premières données dépendantes du locuteur ;
des premiers moyens de transmission pour transmettre la première base de données au second correspondant ;
des premiers moyens d'analyse configurés pour analyser les signaux vocaux à bande limitée reçus par le second correspondant pour déterminer au moins un premier paramètre vocal à bande limitée analysée ;
des premiers moyens de mise en correspondance configurés pour déterminer au moins un premier paramètre vocal à large bande depuis le au moins un premier paramètre vocal à bande limitée analysée sur la base de la première base de données ;
des premiers moyens de synthèse à large bande pour synthétiser des premiers signaux vocaux à large bande sur la base du au moins un premier paramètre vocal à large bande.

12. Système selon la revendication 11, comprenant en outre

une seconde base de données générée par le second correspondant comprenant des secondes données dépendantes du locuteur ;
des seconds moyens de transmission pour transmettre la seconde base de données au premier correspondant ;
des seconds moyens d'analyse configurés pour analyser les signaux vocaux à bande limitée reçus par le premier correspondant pour déterminer au moins un second paramètre vocal à bande limitée analysée ;
des seconds moyens de mise en correspondance configurés pour déterminer au moins un second paramètre vocal à large bande depuis le au moins un second paramètre vocal à bande limitée analysée sur la base de la seconde base de données ;
des seconds moyens de synthèse à large bande pour synthétiser des seconds signaux vocaux à large bande sur la base du au moins un second paramètre vocal à large bande.

13. Système selon la revendication 11, comprenant en outre

une troisième base de données prévue pour le second correspondant et non transmise par le premier correspondant ; et dans lequel
les premiers moyens de mise en correspondance sont configurés pour déterminer le au moins un premier paramètre vocal à large bande sur la base de la première base de données et sur la base de la troisième base de données.

14. Système selon la revendication 12, comprenant en outre

une troisième base de données prévue pour le second correspondant et non transmise par le premier correspondant ; et dans lequel
les premiers moyens de mise en correspondance sont configurés pour déterminer le au moins un premier paramètre vocal à large bande sur la base de la première base de données et sur la base de la troisième base de données et/ou
une quatrième base de données prévue pour le premier correspondant et non transmise par le second correspondant ; et dans lequel
les seconds moyens de mise en correspondance sont configurés pour déterminer le au moins un second paramètre vocal à large bande sur la base de la seconde base de données et sur la base de la quatrième base de données.

15. Système selon la revendication 13 ou 14, dans lequel les premier et/ou second moyens de mise en correspondance sont configurés pour donner la priorité aux données dépendantes du locuteur pendant la détermination du au moins un paramètre vocal à large bande.

16. Système selon l'une des revendications 11-15, dans lequel les premier et/ou second moyens de transmission sont configurés pour commencer la transmission des données dépendantes du locuteur après que la génération de la totalité des données dépendantes du locuteur soit achevée.

17. Système selon l'une des revendications 11-16, dans lequel les première et/ou seconde bases de données comprennent des livres de codes dépendantes du locuteur et/ou des poids pour des réseaux neuronaux artificiels.

18. Système selon l'une des revendications 11-17, dans lequel les première et/ou seconde bases de données comprennent des données dépendantes du locuteur échantillonnées avant que les premier et/ou second correspondant(s) transmettent et reçoivent les signaux vocaux à bande limitée.

19. Système selon la revendication 18, comprenant en outre des moyens de reconnaissance vocale pour générer des données dépendantes du locuteur.

20. Système selon l'une des revendications 11-19, comprenant en outre une unité de commande pour commander la détermination du au moins paramètre vocal à large bande et la synthèse des signaux vocaux à large bande, et dans lequel les unités de commande commandent les moyens de synthèse pour adapter les paramètres qui sont nécessaires pour générer des signaux vocaux à bande haute et/ou à bande

basse.

21. Kit mains libres comprenant un système selon l'une des revendications 11-20.

5

22. Téléphone mobile comprenant un système selon l'une des revendications 11-20.

10

15

20

25

30

35

40

45

50

55

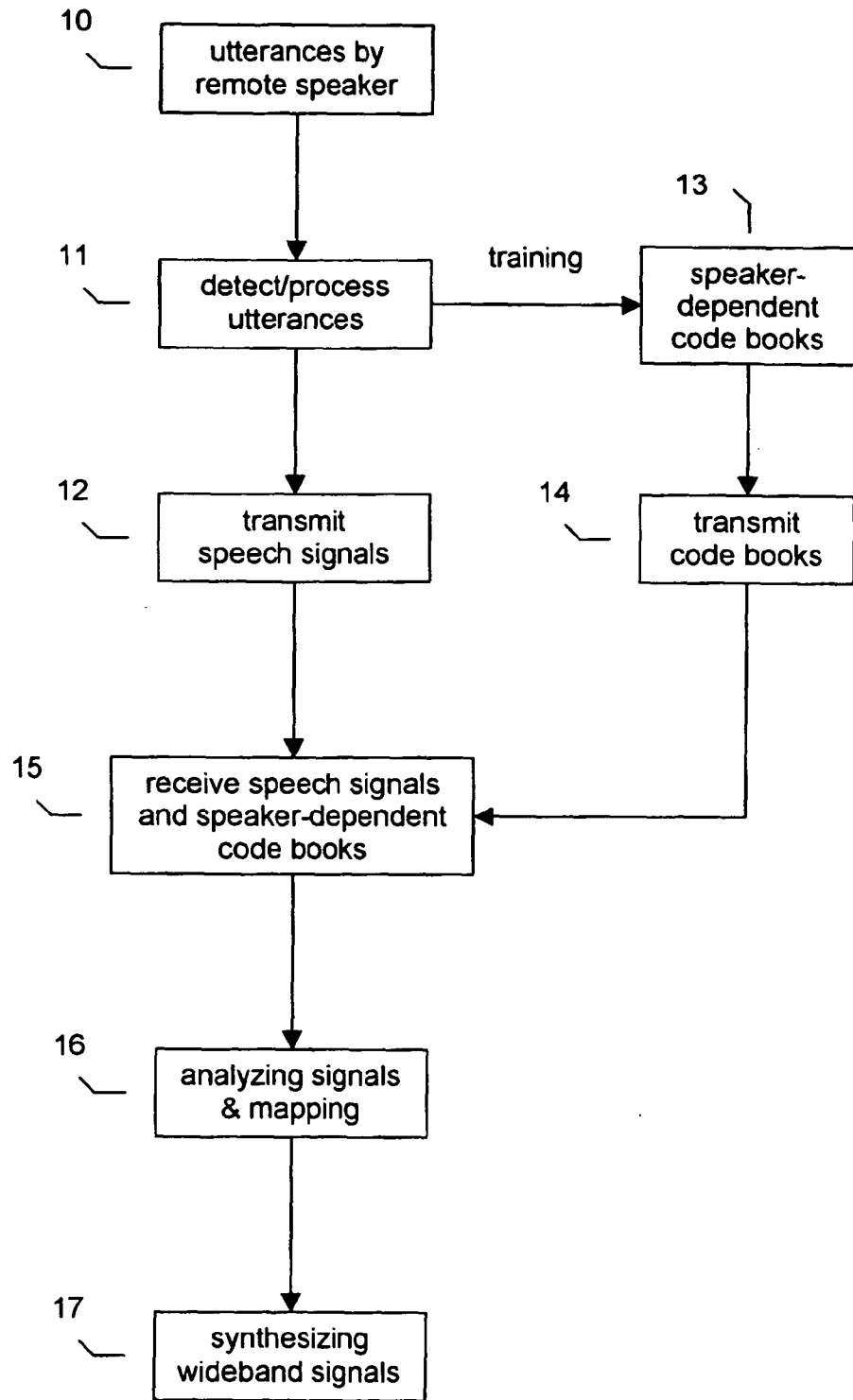


Fig. 1

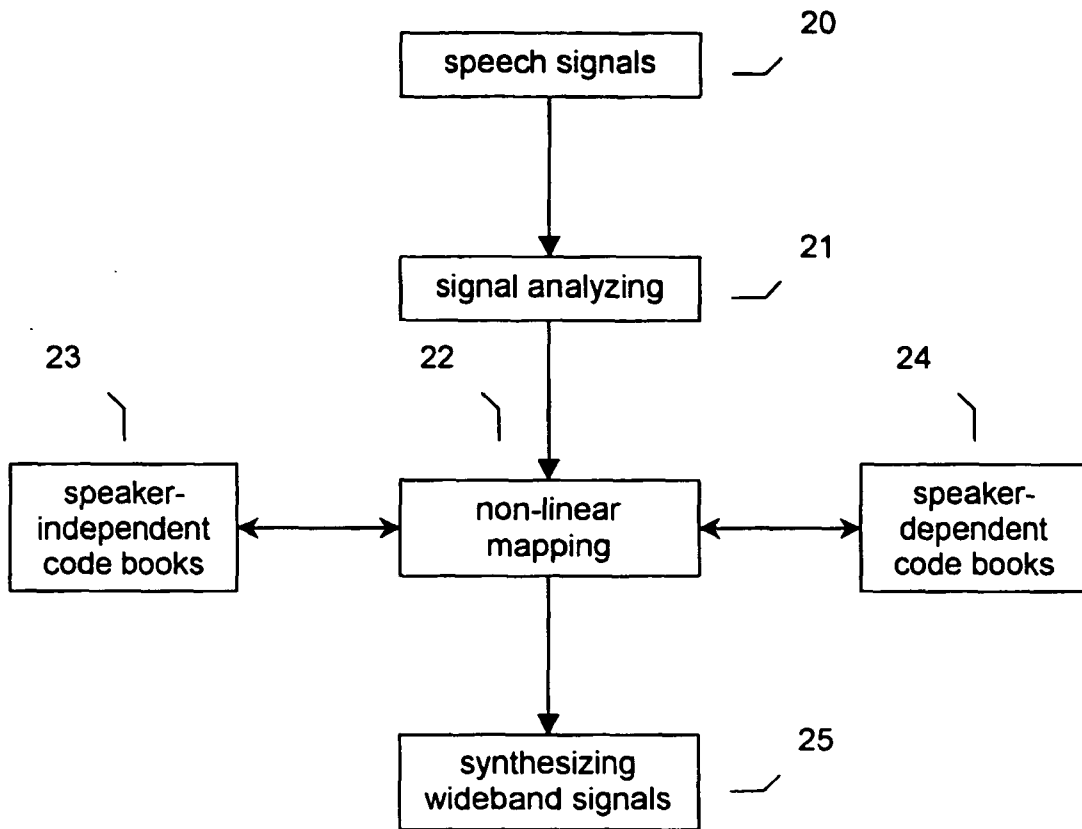


Fig. 2

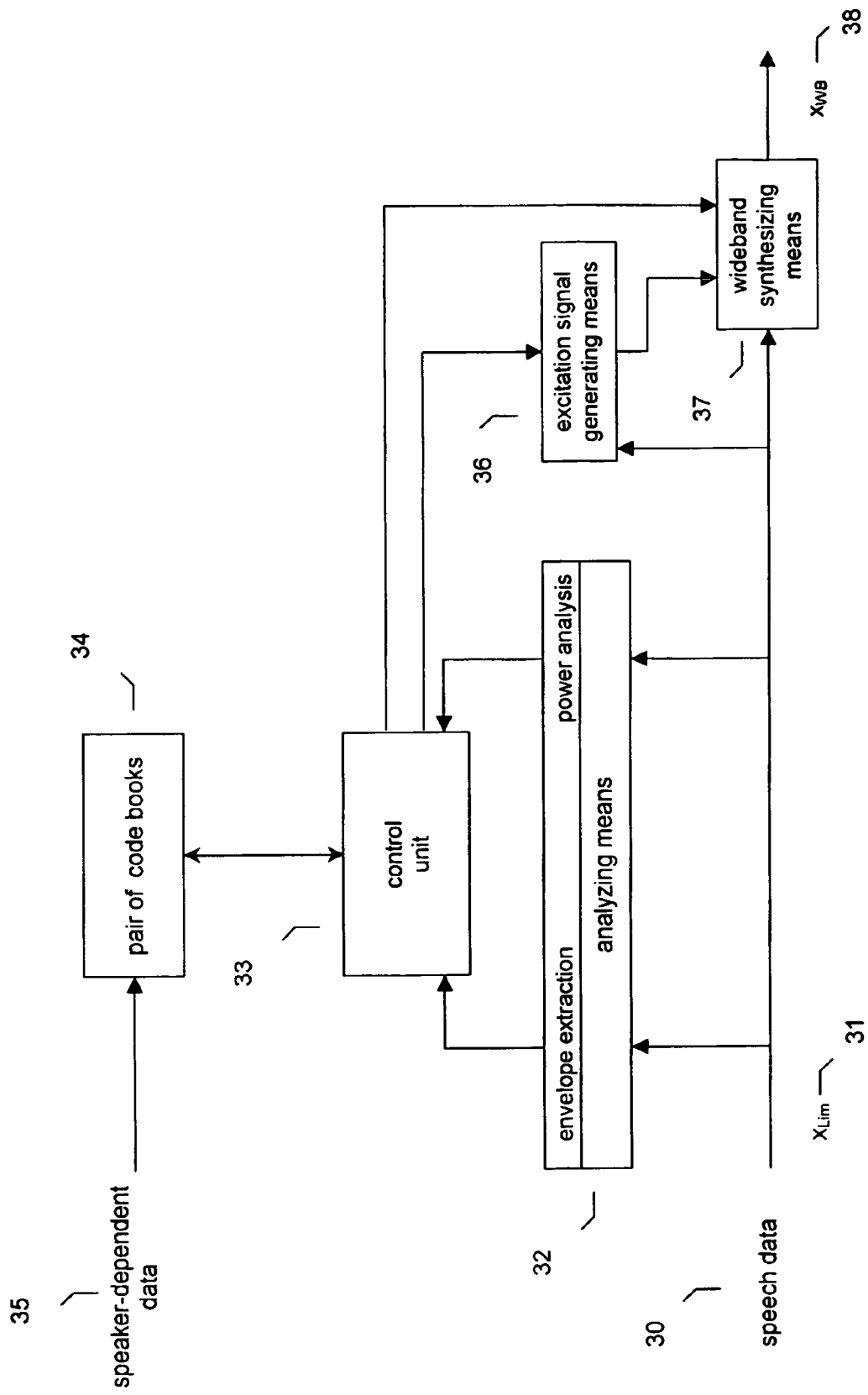


Fig. 3