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

(11) EP 1 208 561 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
26.01.2005 Bulletin 2005/04

(21) Application number: **00925105.9**

(22) Date of filing: **16.05.2000**

(51) Int Cl.[7]: **G10L 11/00**

(86) International application number:
PCT/DK2000/000263

(87) International publication number:
WO 2000/072305 (30.11.2000 Gazette 2000/48)

(54) A METHOD AND APPARATUS FOR NOISE REDUCTION IN SPEECH SIGNALS

VERFAHREN UND VORRICHTUNG ZUR GERäUSCHREDUZIERUNG IN SPRACHSIGNALEN

PROCEDE ET DISPOSITIF DE REDUCTION DU BRUIT DANS DES SIGNAUX VOCAUX

(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: **19.05.1999 DK 99691**
08.02.2000 DK 200000201

(43) Date of publication of application:
29.05.2002 Bulletin 2002/22

(73) Proprietor: **Noisecom APS**
9220 Aalborg (DK)

(72) Inventor: **HERMANSEN, Kjeld**
DK-9260 Gistrup (DK)

(74) Representative: **Schmidt, Jens Joergen et al**
Patentgruppen ApS,
Arosgaarden,
Aaboulevarden 31
8000 Aarhus C (DK)

(56) References cited:
EP-A2- 0 537 948
WO-A1-96/16533
US-A- 5 479 560

- **20th International Conference on Industrial Electronics, Control and Instruments IECON'94., pages 1946-1951, vol. 3, 5-9 Sept. 1994, Witzke L.I. et al: "Speech synthesis based on feature extraction to enhance Noise-corrupted speech", abstract, page1948 - first paragraph page 1949, XP002901258**
- **1994 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1994., pages I/229-I/232 vol. 1, 19-22 April 1994 Gotoh Y. et al: "Using MAP estimated parameters to improve HMM speech recognition performance", abstract, summary XP002901259**

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).

EP 1 208 561 B1

Description

[0001] The present invention relates to noise reduction in speech signals, in particular noise reduction in speech signals employed in telecommunication, most particularly in telecommunication employing cellular phones.

[0002] Noise when added to a speech signal can impair the quality of the signal, reduce intelligibility, and increase listener fatigue. It is therefore of great importance to reduce noise in a speech signal, e.g. in relation to telecommunication, especially when employing cellular phones, or in relation to hearing aids.

[0003] Various methods of noise reduction in speech signals are known. These methods include spectral subtraction and other filtering methods. The noise reduction may e.g. be based on an estimate of the noise spectrum. Such methods depend on stationarity in the noise signal to perform optimally. As the noise in a speech signal is often non-stationary, the estimated noise spectrum used for spectral subtraction will be different from the actual noise spectrum during speech activity. This results in short duration random tones in the noise reduced signal, and such random tone noise tends to be very irritating to listen to due to psycho-acoustic effects.

[0004] WO 99/01942 disdoses a method of reducing the noise in a speech signal using spectral subtraction. According to this method a model based representation describing the quasi-stationary part of the speech signal is generated and manipulated, and the resulting speech signal is generated using the manipulated model and a second signal derived from the speech signal.

[0005] In 'Speech Synthesis Based on Feature Extraction to enhance noise-corrupted speech', L. I. Witzke et al., IECON' 1994, pages 1946-1951, there is disclosed a method for improving the intelligibility of noise-corrupted speech. This is done by extracting speech features (fundamental frequency, voiced/unvoiced decision, and formant frequencies) from the noise-corrupted speech signal. The speech signal can then be synthesized from the extracted features, whereby the intelligibility of the noise-corrupted speech signal is improved. The formant bandwidth is estimated from the extracted features, resulting in inaccuracy when regenerating the speech signal.

[0006] However, when employing known methods for noise reduction in speech signals, the actual speech signal as well as the noise is reduced. It is desirable to be able to eliminate noise in a signal without eliminating the actual signal.

[0007] The object of the present invention is to provide a method of noise reduction in speech signals which reduces the noise even more than known methods. It is further an object to provide a method of noise reduction in speech signals which reduces the noise without affecting the actual speech signal, i.e. a method which eliminates, or at least considerably reduces, unwanted components of a signal without, or at least only to a very limited extend, reducing wanted components.

[0008] The method according to the present invention employs a new model based spectral subtraction algorithm for noise suppression/noise reduction in speech. This new algorithm benefits from available knowledge of the speech dynamics. The method yields better results - especially for low Signal to Noise Ratios (SNR) - with less distortions and artefacts, like musical tones, than any other methods, e.g. the usual spectral subtraction.

[0009] The role of noise suppression in speech processing such as speech coding has gained an increased importance due to the advent of digital cellular telephones. With the low data rate speech coding algorithms the speech quality tends to degrade drastically in high noise. To prevent such quality loss, noise suppression must be achieved without introducing artefacts, speech distortion or significant loss of speech intelligibility.

[0010] Using an additive noise model the noisy signal is modelled as a sum of the speech signal and the noise assuming statistical independence. The spectral subtraction provides an estimate of the signal spectrum as the difference between the noisy spectrum and an estimate of the noise/background spectrum, the latter is obtained during periods of silence.

[0011] The main problem in spectral subtraction is artefacts caused by errors in the estimation of the difference spectrum. Using smoothed versions of the difference spectrum in order to avoid musical tone leads to audible signal distortions.

[0012] Transformation of the estimated speech signal to time domain requires knowledge of the phase of the signal. In most situations one uses the phase of the noisy signal. This works well for high Signal to Noise Ratios (SNR >10 dB). The problem is to use this phase for low SNR, which is a serious drawback of the classical spectral subtraction. A possibility to handle this problem is to use an alternative description of the signal.

[0013] One way of describing a signal in the frequency domain is by magnitude and phase. For speech this description does not relate directly to the articulation parameters. By using a model-based description of the signal it is possible to benefit from a priori knowledge of the speech and noise. These model parameters relate closely to the speech production. The speech signal is decomposed into two components: a generator signal (the residual signal) and a filter modelling of the vocal tract. This results in a separation of speech into a transient and a quasi-stationary part.

[0014] Determination of the noise free/reduced synthesis filter is done by a combination of classical spectral subtraction and model based characterisation of the difference spectrum of noisy speech and background noise.

[0015] The auto correlation function of the quasi stationary part of the speech is mapped into an LPC model spectrum of order 10 and the so-called f, b and g parameters (f formant frequency, b bandwidth and g gain) are determined from

this spectrum. This is a pseudo decomposition of the spectrum into second order sections.

[0016] One of the main problems in spectral subtraction is the "musical tones". Smoothing in time of concatenating spectra is often used with the result of reduced artefacts but also influencing the speech. The present method opens up for an efficient noise elimination in the quasi stationary component as well as in the transient part. Smoothing through lowpass filtering of the f, b and g parameters reduces noise in the quasi stationary part of the signal. The f, b and g parameters are organised in "fbg packets" called "fbg processes", and to each process is allocated a Kalman filter whose parameters correspond to the signal/speech model and noise. It is possible to benefit from the fact that the upper frequency of the f, b and g parameters is about 10 Hz.

[0017] Using the present method it is possible to further reduce noise and artefacts without influencing the speech signal. A noise robust pitch detector combined with a synthetic glottal pulse generator produces the new residual signal for voiced sounds). For unvoiced speech the residual signal is used as is. This residual signal is input to the noise free/reduced synthesis filter with noise free/reduced speech as output.

[0018] The dynamics of the synthesis filter are now constrained via the f, b and g parameters to the range 1 Hz to 10 Hz eliminating the main part of the usual musical tones and leaving the signal/speech component almost unchanged. The input to the synthesis filter depends on the SNR. For heavy noise a robust pitch detector determines the period of the synthetic glottal pulses used as input to the synthesis filter.

[0019] Thus, the present invention relates to a method of noise reduction in a speech signal as set forth in claim 1.

[0020] The frequency is preferably the formant frequency of the speech signal. The speech signal is preferably transmitted via a telecommunications means, most preferably via a cellular phone, but it may alternatively or additionally be transmitted via other means such as a hearing aid or other suitable microphone/speaker arrangements. Such microphone/speaker arrangements may be connected to telephones and/or video conference arrangements, thus allowing the person or persons using such arrangements to move freely within a certain distance from the telephone/video conference arrangement in the room in which the telephone/video conference arrangement is positioned. When using existing similar arrangements this is not possible due to the noise generated in the signals. Other suitable microphone/speaker arrangements may alternatively or additionally be employed. The parameters are preferably smoothed electronically.

[0021] The dynamic information regarding the f, b and g parameters may alternatively or additionally comprise information regarding the difference in frequency between the present speech signal and a previously measured speech signal. If such information is compared to knowledge concerning the human voice regarding the capability to change frequency within a certain time interval, it may be determined whether the present speech signal is in fact the speech signal that was previously measured, i.e. whether the present speech signal and the previously measured speech signal are in fact one and the same. If the difference in frequency exceeds a certain limit, the limit being determined on the basis of knowledge of the human voice and its capability of changing the frequency within a certain time interval, the two signals can not be the same. If the difference in frequency does not exceed such a limit, the two signal may be the same.

[0022] The dynamic information regarding the f, b and g parameters most preferably contain knowledge regarding the development of said parameters in time.

[0023] The a priori knowledge regarding the dynamics of the human voice may alternatively or additionally comprise knowledge regarding the maximum frequency span of a speech signal as described above.

[0024] The a priori knowledge regarding human speech production may be compared to measured parameters of the present speech signal as described above. The a priori knowledge may be obtained e.g. by knowledge regarding the anatomy of the mouth and throat region and/or of the vocal chord. The a priori knowledge may alternatively or additionally be based on a number of previous measurements of relevant parameters as described above. Such previous measurements, or alternatively or additionally a representative extract of such measurements, may be stored in look up tables. Such look up tables are preferably stored electronically in a computer or the like, but may alternatively or additionally be stored in a printed medium such as a book or a sheet of paper.

[0025] Preferably, the method comprises a step in which the speech signal is deemed to belong to a process, the process being a signal which may extend over one or more measurement frames. The process is preferably a formant process. It may, e.g., correspond to the pronunciation of a word.

[0026] The process is an active process at a certain time if it extends over one or more preceding measurement frames. Thus, the process is active if there is a detectable signal. A process may also be regarded as active if there is presently no detectable signal, but such a signal has been present for a predefined number of measurement frames preceding the present measurement frame. Thereby a process may be kept artificially alive even though the signal disappears for a short time interval. This is very useful in telecommunication employing cellular phones, since the signal may temporarily disappear during such communication, e.g. due to noise or fall out due to an uneven geographical distribution of sending masts resulting in an uneven network coverage. It is not desirable to deem a process "inactive" in such a case.

[0027] The smoothing step may comprise the step of determining whether a new formant frequency belongs to an

active process. This may be based on a comparison between the a priori knowledge regarding human speech production and the obtained dynamic information.

[0028] The method may in this case further comprise the step of defining a new process in case the new formant frequency does not belong to an active process, and the new formant frequency is then deemed to belong to said new process.

[0029] The process may be deemed to be inactive in case no new formant frequency is deemed to belong to said process. Thus, in case the signal is permanently terminated, the process is deemed to be inactive.

[0030] The method may further comprise the step of artificially maintaining the speech signal for a predetermined number of measurement frames in case the corresponding process is abruptly deemed to be inactive. This makes it possible to keep a process alive in case the signal is temporarily interrupted as described above.

[0031] The predetermined number of measurement frames may correspond to the maximum duration of the speech signal. Thus, a process may be artificially maintained for a time interval corresponding to the time interval it normally takes to produce such a sound.

[0032] The maximum duration of the speech signal is preferably between 40 ms and 80 ms, such as between 50 ms and 70 ms, such as approximately 60 ms.

[0033] The new formant may be deemed to belong to an active process if the difference in frequency between said formant and said process does not exceed a predetermined level as described above.

[0034] The predetermined level is preferably between 200 Hz and 600 Hz, such as between 300 Hz and 500 Hz, such as approximately 400 Hz.

[0035] The smoothing step preferably comprises the step of filtering the f, b and g parameters.

[0036] The filtering step is most preferably performed using a first order Infinite Impulse Response (IIR) filter, but it may alternatively or additionally be performed using any other suitable kind of filter.

[0037] The first order IIR filter is preferably a feedback filter of the form:

$$y[n]=b\cdot x[n]+a\cdot y[n-1],$$

wherein x designates the speech signal, y designates the filter output, and wherein a and b are parameters to be determined, and wherein the parameters a and b are preferably determined by using model knowledge of the speech process.

[0038] The method may further comprise the steps of

- determining whether voiced speech is present,
- using a noise eliminated pitch period to create a synthetic glottal pulse in case voiced speech is present, and
- replacing at least part of the original speech signal by said synthetic glottal pulse in case voiced speech is present.

[0039] The noise eliminated pitch period may be noise eliminated by using known methods or in the manner described above. Most preferably, it is noise eliminated using known methods as well as in the manner described above.

[0040] The determining step may comprise the steps of comparing the variance of the speech signal to an upper threshold value and to a lower threshold value. Voiced speech is considered to be present in case the variance of the speech signal exceeds the lower threshold value.

[0041] In case the variance of the speech signal also exceeds the upper threshold value, the speech signal is considered to contain purely voiced speech, i.e. no unvoiced component is present. In this case the original speech signal is completely replaced by the synthetic glottal pulse.

[0042] If the variance of the speech signal is between the two threshold values voiced as well as unvoiced components are considered to be present. In this case the original speech signal is replaced by a new pulse which is an appropriate combination of the synthetic glottal pulse and the original speech signal.

[0043] The determining step may comprise the steps of comparing the first formant gain of the speech signal to an upper threshold value and to a lower threshold value. Voiced speech is present in case the first formant gain of the speech signal exceeds the lower threshold value.

[0044] The remarks above apply equally in this case.

[0045] The noise eliminated pitch period is preferably found from a residual signal of the speech signal.

[0046] The replacing step is preferably performed by fading out a residual signal and fading in the synthetic glottal pulse. Here "the synthetic glottal pulse" is to be understood as either a completely synthetic signal or an appropriate combination of the created synthetic pulse and the original speech signal as described above.

[0047] Preferably, at least the smoothing step is performed by a computer system, and the speech signal is most preferably generated in a cellular phone.

[0048] The present invention further relates to an apparatus for performing noise reduction in a speech signal as set

forth in claim 28.

[0049] The means for obtaining dynamic information regarding f, b and g parameters preferably comprises one or more suitable detectors, such as microphones, and/or one or more computers.

[0050] The smoothing means preferably comprises one or more computers and/or one or more suitable electronically filters, such as low pass filters and/or high pass filters and/or Infinite Impulse Response (IIR) filters and/or any other suitable kind of fillers.

[0051] The means for obtaining and storing a priori knowledge regarding human speech production preferably comprises one or more computers, most preferably comprising electronically storage means. It may further comprise one or more look up tables, the tables being created by using empirically obtained data (i.e. previous measurements of e. g. relevant parameters such as the dynamic information mentioned above) and/or by using theoretical calculation of relevant parameters. Such calculations may be based on knowledge regarding the anatomy of the mouth and throat region of humans as described above.

[0052] In an apparatus as described above, wherein the speech signal is deemed to belong to a process, the process being a signal which may extend over one or more measurement frames, and wherein the process is an active process at a certain time if it extends over one or more preceding measurement frames, the smoothing means may comprise means for determining whether a new formant frequency belongs to an active process. Such means may comprise means for comparing obtained relevant parameters of the speech signal (as described above) and theoretical and/or empirical data. It is then determined from a predefined set of criteria whether the new formant frequency belongs to the active process or not. This is further described above.

[0053] The smoothing means preferably comprises means for filtering the f, b and g parameters. It may comprise one or more electronically filters, such as low pass filters and/or high pass filters and/or IIR filters and/or any other suitable filters.

[0054] The apparatus may further comprise

- determining means for determining whether voiced speech is present,
- creating means for creating a synthetic glottal pulse by using a noise eliminated pitch period, and
- replacing means for replacing at least part of the original speech signal by said synthetic glottal pulse in case voiced speech is present.

[0055] The creating means may comprise one or more tone generators and/or it may comprise a computer.

[0056] The replacing means may comprise one or more faders, so that the synthetic pulse may be faded in as the original signal is faded out.

[0057] The determining means may comprise comparing means for comparing the variance of the speech signal to an upper threshold value and to a lower threshold value. Voiced speech is present in case the variance of the speech signal exceeds the lower threshold value. The comparison is most preferably performed by one or more computers having computer storage means for storing information regarding the threshold values. The threshold values may be obtained by theoretical calculations and/or by previous measurements, i.e. it may be empirically obtained.

[0058] The determining means may comprise comparing means for comparing the first formant gain of the speech signal to an upper threshold value and to a lower threshold value. Voiced speech is present in case the first formant gain of the speech signal exceeds the lower threshold value. The above is equally applicable here.

[0059] The apparatus may further comprise means for producing a noise eliminated pitch period from a residual signal of the speech signal.

[0060] A new model based spectral subtraction method has been presented, the method being applicable for noise elimination for high as well as low SNR avoiding linear and non-linear artefacts like musical tones. The method is attractive because of its flexibility, modularity and numerical robustness. Furthermore it is very suitable for real time implementation.

[0061] The invention will now be further described with reference to the accompanying drawings in which

Fig. 1 shows the frequency of a speech signal as a function of time indicating two different processes,

Fig. 2 is a flow diagram illustrating part of the smoothing step of the method according to the present invention,

Fig. 3 is a flow diagram illustrating the control of the maximum duration of a speech signal,

Fig. 4 is a flow diagram illustrating the allocation of a sign to a process,

Fig. 5 is a flow diagram illustrating the adjustment of the pitch of a speech signal,

Fig. 6 is a flow diagram illustrating the step of determining whether voiced speech is present and the step of replacing a noise eliminated pitch period by a synthetic glottal pulse,

Fig. 7a shows the intonation of a speech signal as a function of time,

Fig. 7b shows a synthetic internal glottal pulse,

Fig. 7c shows two pitch periods having different lengths,

Fig. 8 shows non-filtered f, b and g parameters of a speech signal ('three', 'four', 'five'),

Fig. 9 shows filtered f, b and g parameters corresponding to Fig. 8,

Fig. 10 shows an original noisy speech signal ('three', 'four', 'five'),

Fig. 11 shows the signal of Fig. 10, but with synthetic glottal pulses replacing the pitch periods where voiced speech is present.

Fig. 12 shows the speech signal of Figs. 10 and 11, but with the noise eliminated,

Fig. 13 shows a frequency spectrum corresponding to Fig. 10, and

Fig. 14 shows a frequency spectrum corresponding to Fig. 12.

[0062] Fig. 1 shows the frequency of a speech signal as a function of time. Two processes ('process 1' and 'process 2') are present.

[0063] When the signal of a process disappears it may be due to the signal "falling out", e.g. due to noise or a bad connection, or it may be due to the fact that the person stops speaking, i.e. the signal actually disappears. If the process has existed for a certain amount of time it is artificially kept alive for some time after the disappearance of the signal as indicated by "Δd" and the dotted lines of the figure.

[0064] At the time t_1 the signal of 'process 1' disappears and is artificially kept alive as described above. At time t_2 a new signal is detected, and this new signal is deemed to belong to 'process 1'. This is because the time elapsed (t_2-t_1) is smaller than Δd, and because the difference in frequency between the new signal and the signal of 'process 1' is below a predefined limit indicated by "Δf". The signal of'process 1', and the new signal are connected as indicated by the grey line.

[0065] The signal of 'process 2' also appears approximately at time t,. However, it is deemed not to belong to 'process 1', since the difference in frequency between the two signals exceeds the predefined limit (Δf). Since no other active processes are present a new process ('process 2') is defined, and the signal is deemed to belong to this process.

[0066] Eventually 'process 1' as well as 'process 2' expires since the signals disappear and no new signals appear within the predefined time limit, Δd.

[0067] Referring to Figs. 2-4 a filtering algorithm is described, the algorithm having the following four steps:

1. Determining whether new formant frequencies belong to active processes.
2. Elimination of non-updated processes.
3. Allocation of a sign to a certain process.
4. Filtering the processes and outputting the f, b and g parameters and the sign.

[0068] Fig. 2 is a flow diagram illustrating step one as defined above. New f, b and g parameters are used as input (FBG new), and the frequency nearest to a process is found. It is not possible to investigate the new formants one by one, since two or more of the new formants may fulfil the criteria for belonging to a certain process. Therefore, the formant having the frequency being nearest to the process is deemed to belong to said process.

[0069] In case the speech signal is found to belong to a process which already exists, that process is updated in accordance with the new signal. In case it is found that the speech signal does not belong to any of the existing processes, a new process is created, and the speech signal is deemed to belong to the new process. At this stage of the algorithm it is also tested whether the difference in frequency between the new formant and the existing process exceeds a certain predetermined level. The level may be set in the control system.

[0070] If additional new f, b and g parameters are present the formant which has thus been allocated to a process is "removed" from the "new parameters" and the steps described above are repeated until there are no more new f, b

and g parameters left.

[0071] Fig. 3 illustrates step 2 as defined above. It is investigated whether an active process has been updated. In case it has been updated nothing further is done during the present measurement frame.

[0072] In case the process has not been updated during the present measurement frame it is investigated how long the process in question has been present. In case it has only been present during one measurement frame the process is eliminated. In case it has been present for more than one measurement frame the age of the process is "counted down". Thus, the process will be regarded as having been present for one measurement frame less the next time the algorithm is performed. This is to ensure that an otherwise active process is not eliminated just because it is not updated during one measurement frame. Such processes will be kept artificially alive for a number of measurement frames depending on the maximum duration of the processes. This may be performed by maintaining the intemal values of the process during the measurement frame wherein the process was last updated, or it may be performed in a more sophisticated manner based on predictions made according to earlier measurements.

[0073] Fig. 4 is a flow diagram illustrating step 3 of the above algorithm. It is first investigated whether the sign of a given process (Process number X) is locked. If this is the case the energy in the process is compared to a lower threshold value ("energi_lav"). If the energy in the process is below said lower threshold value the sign is unlocked. Otherwise the sign is maintained in a locked mode. The energy in the process is then compared to an upper threshold value ("energi_hoej"). If the energy in the process is above said upper threshold value the process is allocated a sign in relation to locked processes. Otherwise it is investigated whether the process has a sign allocated. If this is not the case, the process is allocated a sign in relation to locked processes. Otherwise the mark is not updated, i.e. the previously allocated sign is maintained.

[0074] The formants are allocated mutually alternating signs in order to improve the quality of the sound. Processes having frequencies which are between the frequencies of two different other processes (the two different processes having different signs) are allocated a sign according to the process being closest in frequency to the formant, i.e. the sign allocated to this process is opposite to the sign of the "closest" process.

[0075] The allocation of signs to the processes is very important, since it makes it possible to "hold on" to a given process and thereby to maintain the control of the processes. This is very important for the execution of step 4 of the algorithm above. However; the allocation of signs per se is known and well described in the prior art.

[0076] In step 4 of the algorithm above the f, b and g parameters are individually filtered. Preferably, a first order Infinite Impulse Response (IIR) filter is used, the filter most preferably having the form:

$$y[n]=b\cdot x[n]+a\cdot y[n-1].$$

[0077] The coefficients a and b are calculated using a time constant in order to ensure that the filtering is independent of the frame shift. a is chosen in such a way that said time constant relates to the time constant of the speech signal, and b is chosen in such a way that the DC amplification is 1 (corresponding to the integral of the impulse response).

[0078] Fig. 5 is a flow diagram illustrating the adjustment of the pitch of a speech signal. This function determines the pitch period of a source sequence. The general idea of the function is based upon the assumption that the pitch is the most dominant periodical component in the speech signal. It is furthermore assumed that the pitch frequency, for physiological reasons, is limited to a certain frequency span. Thus, the main issue of determination of the pitch period is the calculation of the auto correlation and the determination of the pitch as the index of the maximum value of the auto correlation in a limited time interval.

[0079] The pitch sequence, which is used as an input to the function, is squared in order to avoid negative values and in order to enhance dynamical differences. The pitch sequence should be a frame of the speech signal or of the residual signal.

[0080] The squared pitch sequence is then rectified. This step emphasises the periodicity of the pitch by using knowledge regarding the structure of the pitch sequences. This is due to the fact that the pitch is much more powerful than other potential periodical components of the speech signal or the residual signal, and due to the fact that said other components are hidden by the rectification.

[0081] In case there are linear trends in the squared and rectified pitch sequence, such trends may now be removed. This may be done by creating a linear least squares approximation to the input sequence and then to subtract this approximation from the sequence. The linear least squares approximation may be done using the following equations:

$$y = a \cdot x + b$$

$$a = \frac{n\sum x_j y_j - \sum x_i \sum y_j}{n\sum x_j^2 - \left(\sum x_j\right)^2}$$

$$b = \bar{y} - a \cdot \bar{x}$$

wherein $\bar{x}$ and $\bar{y}$ designate the mean values of x and y, respectively.

[0082] Then the auto correlation is calculated for "allowed" pitch periods, i.e. for pitch periods having a duration which is between a lower threshold value and an upper threshold value, where said threshold values may be set initially. The calculated auto correlation is subsequently scaled using a linear weighting function. This is done in order to obtain a robust pitch detection.

[0083] In the next step the index of the maximum value of the weighted auto correlation function is used as an initial guess for the pitch period. In case the pitch period thus determined is large, and in case a shorter pitch period is more likely the pitch period is adjusted accordingly. A shorter pitch period may be more likely if e.g. the half pitch period is also an "allowed" pitch period. In this case it is possible that the sub-harmonic period of the pitch has been detected in stead of the actual pitch.

[0084] If the initial pitch period is not large or if a shorter pitch period is not more likely, the initial guess is maintained as the pitch period. Finally, the pitch period is used as an output.

[0085] Fig. 6 is a flow diagram illustrating the steps of the method according to the present invention in which it is determined whether voiced speech is present, and in which the noise eliminated pitch period is replaced by a synthetic glottal pulse.

[0086] The input parameter "x", which may be the variance or the gain of a signal, is compared to a lower threshold value ("taerskelnedre") as well as to an upper threshold value ("taerskeloevre"). If "x" is smaller than "taerskelnedre" the speech signal is considered to be completely unvoiced, and the original residual signal is maintained ("alfa1"=1 and "alfa2"=0). If "x" exceeds "taerskeloevre" the speech signal is considered to be completely voiced, and the original signal is replaced by a purely synthetic pitch pulse ("alfa1"=0 and "alfa2"=1). If "taerskelnedre" < "x" < "taerskeloevre" the speech signal is considered to be a mixture of voiced and unvoiced signals, and the original signal is replaced by a synthetic signal which is a suitable mixture of the original residual signal and the synthetic pitch pulse as described below.

$$\text{Synthetic residual} = \sqrt{\text{alfa1}} \cdot \text{residual} + \sqrt{\text{alfa2}} \cdot \text{synthetic pitch pulse}$$

wherein "Synthetic residual" is the resulting synthetic residual signal which replaces the original residual signal, "residual" is the original residual signal and "synthetic pitch pulse" is the purely synthetic pitch pulse which is created. "alfa1" and "alfa2" are determined as:

$$\text{alfa2} = \frac{\text{x - taerskelnedre}}{\text{taerskeloevre - taerskelnedre}}$$

$$\text{alfa1} = 1 - \text{alfa2}$$

[0087] Fig. 7a-c illustrate how the speech signal may be replaced by a synthetic glottal pulse in case voiced speech is present in the signal.

[0088] Fig. 7a shows the intonation of a speech signal as a function of time. The intonation at time t2 is slightly larger than the intonation at time t1. The period where no signal is detected may be a period of silence or it may be a period where only completely unvoiced speech is present, i.e. a period during which no intonation may be detected, since an intonation may only be detected if the vocal chord is active, i.e. when voiced speech is present.

[0089] Fig. 7b shows a synthetic internal glottal pulse. In reality glottal pulses are very dependent upon the person speaking. The glottal pulse shown in Fig. 7b, however, is an "average" pulse which is constructed in such a way that it has a wide spectrum and at the same time has the maximum length.

[0090] Fig. 7c shows a signal with the synthetic glottal pulse of Fig. 7b phased in instead of a noisy signal. The synthetic glottal pulse has a certain length. The synthetic signal is artificially "extended" by a "zero-signal", so as to match the length of the original pitch period, 'ipitch'. In Fig. 7c the length of the second pulse is slightly larger than the length of the first pulse ('ipitch(t2)'>'ipitch(t1)'). This is due to the fact that the intonation at time t2 is slightly larger than

the intonation at time t1 as indicated in Fig. 7a. It is clear that the only difference between the two pulses is the length of the "zero-signal" following the synthetic glottal pulse.

[0091] Figs. 8 and 9 show non-filtered and filtered formant tracks (f, b and g parameters), respectively, for the words 'three', 'four' and 'five'. The parameters have been filtered using a so-called Kalman filter. When comparing Fig. 8 and Fig. 9, it is clear that the noise of the signal is considerably reduced during the filtering process, i.e. the fluctuations are reduced. It is also clear that the dynamics of the speech are at the same time left nearly unchanged. What is achieved is thus a speech signal, wherein the noise, which was initially present, is removed or at least considerably reduced in such a way that the original signal, i.e. the actual speech, is left nearly unchanged. It is thus possible to remove unwanted components of a signal (i.e. noise) without removing or changing wanted components of the signal (i.e. actual speech).

[0092] Figs. 10-12 also show speech signals representing the words 'three', 'four' and 'five'. Fig. 10 shows the original speech signal, including noise components. It is clear that this signal is very noisy, i.e. the signal to noise ratio (SNR) is very small.

[0093] In Fig. 11 part of the signal has been replaced by synthetic glottal pulses, and the figure thus represents the output of the flow diagram of Fig. 6. As is readily seen, the shifts between the regions in which the original signal has been replaced and the regions in which the original signal has been maintained are very abrupt. This is because the difference between the lower threshold value ("taerskelnedre") and the upper threshold value ("taerskeloevre") is relatively small. It is therefore very likely that the variance or gain (whichever parameter is chosen) is either below the lower threshold value or above the upper threshold value, rather than being between the two values. That is, the sound is most often considered to be either completely voiced or completely unvoiced, rather than being considered to contain voiced as well as unvoiced components.

[0094] Fig. 12 shows a signal in which the noise has been reduced. Conventional methods as well as the method according to the invention has been employed. It is very clear that the SNR has improved considerably as compared to Fig. 10.

[0095] Fig. 13 and Fig. 14 show frequency spectra corresponding to Fig. 10 and Fig. 12, respectively. When comparing Figs. 13 and 14 it is also clear that the SNR has improved considerably during the filtering process.

[0096] Thus, a method and apparatus for noise reduction in speech signals has been provided, the method and apparatus reducing the noise in the signal without, or at least only to a very limited extend, reducing the actual signal.

Claims

1. A method of reducing the amount of noise in a noisy speech signal, comprising the steps of

   - obtaining from a speech signal model based representations describing the quasi-stationary part of the speech,
   - obtaining, from said model based representation, dynamic information regarding frequency (f), bandwidth (b), and gain (g) parameters of said speech signal in relation to time,
   - defining processes as a function of time by letting said f, b, and g parameters belong to a process according to a priori knowledge regarding the dynamics of the human voice,
   - smoothing the f, b and g parameters with respect to time, the smoothing step being performed on said processes.

2. A method according to claim 1, wherein the a priori knowledge regarding the dynamics of the human voice comprises knowledge regarding the maximum frequency span of a speech signal.

3. A method according to claim 1 or 2, wherein the speech signal is deemed to belong to a process, the process being a signal which may extend over one or more measurement frames.

4. A method according to claim 3, wherein the process is an active process at a certain time if it extends over one or more preceding measurement frames.

5. A method according to claim 3 or 4, wherein the smoothing step comprises the step of determining whether a new formant frequency belongs to an active process.

6. A method according to claim 5, further comprising the step of defining a new process in case the new formant frequency does not belong to an active process, and wherein the new formant frequency is then deemed to belong to said new process.

7. A method according to any of claims 4-6, wherein a process is deemed to be inactive in case no new formant frequency is deemed to belong to said process.

8. A method according to claim 7, further comprising the step of artificially maintaining the speech signal for a predetermined number of measurement frames in case the corresponding process is abruptly deemed to be inactive.

9. A method according to claim 8, wherein the predetermined number of measurement frames corresponds to the maximum duration of the speech signal.

10. A method according to claim 9, wherein the maximum duration of the speech signal is between 40 ms and 80 ms.

11. A method according to claim 10, wherein the maximum duration of the speech signal is between 50 ms and 70 ms.

12. A method according to claim 11, wherein the maximum duration of the speech signal is approximately 60 ms.

13. A method according to any of claims 5-12, wherein the new formant is deemed to belong to an active process if the difference in frequency between said formant and said process does not exceed a predetermined level.

14. A method according to claim 13, wherein the predetermined level is between 200 Hz and 600 Hz.

15. A method according to claim 14, wherein the predetermined level is between 300 Hz and 500 Hz.

16. A method according to claim 15, wherein the predetermined level is approximately 400 Hz.

17. A method according to any of the preceding claims, wherein the smoothing step comprises the step of filtering the f, b and g parameters.

18. A method according to claim 17, wherein the filtering step is performed using a first order Infinite Impulse Response (IIR) filter.

19. A method according to claim 18, wherein the first order IIR filter is a feedback filter of the form:

$$y[n]=b\cdot x[n]+a\cdot y[n-1],$$

wherein x designates the speech signal, y designates the filter output, and wherein a and b are parameters to be determined.

20. A method according to claim 19, wherein the parameters a and b are determined by using model knowledge of the speech process.

21. A method according to any of the preceding claims, further comprising the steps of

- determining whether voiced speech is present,
- using a noise eliminated pitch period to create a synthetic glottal pulse in case voiced speech is present, and
- replacing at least part of the original speech signal by said synthetic glottal pulse in case voiced speech is present.

22. A method according to claim 21, wherein the determining step comprises the steps of comparing the variance of the speech signal to an upper threshold value and to a lower threshold value, and wherein voiced speech is present in case the variance of the speech signal exceeds the lower threshold value.

23. A method according to claim 21 or 22, wherein the determining step comprises the steps of comparing the first formant gain of the speech signal to an upper threshold value and to a lower threshold value, and wherein voiced speech is present in case the first formant gain of the speech signal exceeds the lower threshold value.

24. A method according to any of claims 21-23, wherein the noise eliminated pitch period is found from a residual signal of the speech signal.

25. A method according to any of claims 21-24, wherein the replacing step is performed by fading out a residual signal and fading in the synthetic glottal pulse.

26. A method according to any of the preceding claims, wherein at least the smoothing step is performed by a computer system.

27. A method according to any of the preceding claims, wherein the speech signal is generated in a cellular phone.

28. An apparatus for performing noise reduction in a speech signal, the apparatus comprising

- means for obtaining from a speech signal model based representations describing the quasi-stationary part of the speech,
- means for obtaining dynamic information regarding frequency (f), bandwidth (b) and gain (g) parameters of said speech signal in relation to time,
- means for defining processes as a function of time by letting said f, b, and g parameters belong to a process according to a priori knowledge regarding the dynamics of the human voice,
- smoothing means for smoothing the processes with respect to time.

29. An apparatus according to claim 28, wherein the a priori knowledge comprises the maximum frequency span of a speech signal.

30. An apparatus according to claim 28 or 29, wherein the speech signal is deemed to belong to a process, the process being a signal which may extend over one or more measurement frames, and wherein the process is an active process at a certain time if it extends over one or more preceding measurement frames, and wherein the smoothing means comprises means for determining whether a new formant frequency belongs to an active process.

31. An apparatus according to any of claims 28-30, wherein the smoothing means comprises means for filtering the f, b and g parameters.

32. An apparatus according to any of claims 28-30, further comprising

- determining means for determining whether voiced speech is present,
- creating means for creating a synthetic glottal pulse by using a noise eliminated pitch period, and
- replacing means for replacing at least part of the original speech signal by said synthetic glottal pulse in case voiced speech is present.

33. An apparatus according to claim 32, wherein the determining means comprises comparing means for comparing the variance of the speech signal to an upper threshold value and to a lower threshold value, and wherein voiced speech is present in case the variance of the speech signal exceeds the lower threshold value.

34. An apparatus according to claim 32 or 33, wherein the determining means comprises comparing means for comparing the first formant gain of the speech signal to an upper threshold value and to a lower threshold value, and wherein voiced speech is present in case the first formant gain of the speech signal exceeds the lower threshold value.

35. An apparatus according to any of claims 32-34, further comprising means for producing a noise eliminated pitch period from a residual signal of the speech signal.

Patentansprüche

1. Verfahren zum Verringern des Rauschpegels in einem gestörten Sprachsignal, umfassend folgende Schritte:

- Erlangen von auf einem Modell basierenden, den quasi-stationären Teil der Sprache beschreibenden Darstellungen aus einem Sprachsignal,

- Erlangen dynamischer Information betreffend Frequenz (f)-, Bandbreite (b)- und Verstärkung (g)-Parametern des Sprachsignals relativ zur Zeit aus der Modellbasierten Darstellung,

- Definieren von Prozessen als eine Funktion der Zeit, indem man die f-, b- und g-Parameter als zu einem Prozeß gemäß Apriori-Wissen betreffend die Dynamik der menschlichen Stimme zugehörend sein läßt,

- Glätten der f-, b- und g-Parameter im Hinblick auf die Zeit, wobei der Schritt des Glättens an den Prozessen ausgeführt wird.

2. Verfahren nach Anspruch 1, bei dem das Apriori-Wissen betreffend die Dynamik der menschlichen Stimme Kenntnisse über die maximale Frequenzspanne eines Sprachsignals umfaßt.

3. Verfahren nach Anspruch 1 oder 2, bei dem das Sprachsignal als zu einem Prozeß gehörend angenommen wird, wobei der Prozeß ein Signal ist, das sich über einen oder mehrere Messungsrahmen hinweg erstreckt.

4. Verfahren nach Anspruch 3, bei dem der Prozeß ein aktiver Prozeß zu einer bestimmten Zeit ist, wenn er sich über einen oder mehrere vorhergehende Messungsrahmen erstreckt.

5. Verfahren nach Anspruch 3 oder 4, bei dem der Schritt des Glättens den Schritt des Bestimmens umfaßt, ob eine neue Formantfrequenz zu einem aktiven Prozeß gehört.

6. Verfahren nach Anspruch 5, weiter umfassend den Schritt des Definierens eines neuen Prozesses für den Fall, daß die neue Formantfrequenz nicht zu einem aktiven Prozeß gehört, und wobei die neue Formantfrequenz dann als zu dem neuen Prozeß gehörend angenommen wird.

7. Verfahren nach einem der Ansprüche 4 bis 6, bei dem ein Prozeß als inaktiv angenommen wird für den Fall, daß keine neue Formantfrequenz als dem Prozeß zugehörend angenommen wird.

8. Verfahren nach Anspruch 7, weiter umfassend den Schritt des künstlichen Aufrechterhaltens des Sprachsignals für eine vorbestimmte Zahl Messungsrahmen in dem Fall, daß der entsprechende Prozeß abrupt als inaktiv angenommen wird.

9. Verfahren nach Anspruch 8, bei dem die vorbestimmte Zahl Messungsrahmen der Maximaldauer des Sprachsignals entspricht.

10. Verfahren nach Anspruch 9, bei dem die Maximaldauer des Sprachsignals zwischen 40 ms und 80 ms beträgt.

11. Verfahren nach Anspruch 10, bei dem die maximale Dauer des Sprachsignals zwischen 50 ms und 70 ms liegt.

12. Verfahren nach Anspruch 11, bei dem die maximale Dauer des Sprachsignals ungefähr 60 ms beträgt.

13. Verfahren nach einem der Ansprüche 5 bis 12, bei dem der neue Formant als zu einem aktiven Prozeß gehörend angenommen wird, wenn die Frequenzdifferenz zwischen dem Formanten und dem Prozeß ein vorbestimmtes Niveau nicht überschreitet.

14. Verfahren nach Anspruch 13, bei dem das vorbestimmte Niveau zwischen 200 Hz und 600 Hz liegt.

15. Verfahren nach Anspruch 14, bei dem das vorbestimmte Niveau zwischen 300 Hz und 500 Hz liegt.

16. Verfahren nach Anspruch 15, bei dem das vorbestimmte Niveau ungefähr 400 Hz beträgt.

17. Verfahren nach einem der vorhergehenden Ansprüche, bei dem der Schritt des Glättens den Schritt des Filterns der f-, b- und g-Parameter umfaßt.

18. Verfahren nach Anspruch 17, bei dem der Filterschritt unter Verwendung eines IIR-Filters (Infinite Impulse Response = Unendlich-Impulsantwort-Filter) erster Ordnung durchgeführt wird.

19. Verfahren nach Anspruch 18, bei dem der IIR-Filter erster Ordnung ein Rückkopplungsfilter folgender Form ist:

$$y[n]=b\cdot x[n]+a\cdot y[n-1],$$

wobei x das Sprachsignal sowie y den Filterausgang bezeichnet und wobei a und b zu bestimmende Parameter sind.

20. Verfahren nach Anspruch 19, bei dem die Parameter a und b durch Verwendung von Modellkenntnis des Sprachprozesses bestimmt werden.

21. Verfahren nach einem der vorhergehenden Ansprüche, weiter umfassend folgende Schritte

- Bestimmen, ob stimmhafte Sprache vorhanden ist,
- Verwenden eines Teilungs- oder Schrittzeitraums, der rauschbereinigt ist, zum Schaffen eines synthetischen glottalen Impulses für den Fall, daß stimmhafte Sprache vorhanden ist, und
- Ersetzen wenigstens eines Teils des ursprünglichen Sprachsignals durch den synthetischen Glottalimpuls für den Fall, daß stimmhafte Sprache vorhanden ist.

22. Verfahren nach Anspruch 21, bei dem Schritt des Bestimmens die Schritte des Vergleichens der Varianz des Sprachsignals mit einem oberen Schwellenwert und einem unteren Schwellenwert umfaßt und bei dem stimmhafte Sprache für den Fall, daß die Varianz des Sprachsignals den unteren Schwellenwert überschreitet, vorhanden ist.

23. Verfahren nach Anspruch 21 oder 22, bei dem der Schritt des Bestimmens die Schritte des Vergleichens der ersten Formant-Verstärkung des Sprachsignals mit einem oberen Schwellenwert und einem unteren Schwellenwert umfaßt und bei dem stimmhafte Sprache für den Fall vorhanden ist, daß die erste Formant-Verstärkung des Sprachsignals den unteren Schwellenwert überschreitet.

24. Verfahren nach einem der Ansprüche 21 bis 23, bei dem der rauschbereinigte Schrittzeitraum aus einem Restsignal des Sprachsignals gefunden wird.

25. Verfahren nach einem der Ansprüche 21 bis 24, bei dem der Schritt des Ersetzens durch Abklingen eines Restsignals und Einblenden des synthetischen Glottalimpulses durchgeführt wird.

26. Verfahren nach einem der vorhergehenden Ansprüche, bei dem wenigstens der Schritt des Glättens von einem Computersystem durchgeführt wird.

27. Verfahren nach einem der vorhergehenden Ansprüche, bei dem das Sprachsignal in einem Zellulartelefon erzeugt wird.

28. Vorrichtung zum Durchführen einer Rauschverringerung in einem Sprachsignal, wobei die Vorrichtung umfaßt

- Mittel zum Erlangen auf einem Modell basierender, den quasi-stationären Teil der Sprache beschreibender Darstellungen aus einem Sprachsignal,
- Mittel zum Erlangen dynamischer Informationen betreffend Frequenz (f)-, Bandbreite (b)- und Verstärkung (g) -Parameter des Sprachsignals relativ zur Zeit,
- Mittel zum Bestimmen von Prozessen als Funktion der Zeit, indem man die f- b- und g-Parameter als zu einem Prozeß gemäß einer Apriori-Kenntnis betreffend die Dynamik der menschlichen Stimme zugehörend sein läßt,
- Glättungsmittel zum Glätten der Prozesse in bezug auf die Zeit.

29. Vorrichtung nach Anspruch 28, bei der die Apriori-Kenntnis die maximale Frequenzspanne eines Sprachsignals umfaßt.

30. Vorrichtung nach Anspruch 28 oder 29, bei der das Sprachsignal als zu einem Prozeß gehörend angesehen wird, wobei der Prozeß ein Signal ist, das sich über einen oder mehrere Messungsrahmen erstrecken kann, und bei der der Prozeß ein aktiver Prozeß zu einer bestimmten Zeit ist, wenn er sich über einen oder mehrere vorhergehende Messungsrahmen erstreckt, und bei der das Glättungsmittel Mittel zum Bestimmen umfaßt, ob eine neue Formantfrequenz zu einem aktiven Prozeß gehört.

31. Vorrichtung nach einem der Ansprüche 28 bis 30, bei der das Glättungsmittel Mittel zum Filtern der f-, b- und g-Parameter umfaßt.

32. Vorrichtung nach einem der Ansprüche 28 bis 30, weiter umfassend

- Bestimmungsmittel zum Bestimmen, ob stimmhafte Sprache vorhanden ist,
- Erzeugungsmittel zum Bilden eines synthetischen Glottalimpulses durch Verwenden einer rauschbereinigten Teilungs- oder Schrittperiode und
- Ersetzungsmittel zum Ersetzen wenigstens eines Teils des ursprünglichen Sprachsignals durch den synthetischen Glottalimpuls für den Fall, daß stimmhafte Sprache vorhanden ist.

33. Vorrichtung nach Anspruch 32, bei der das Bestimmungsmittel Vergleichsmittel zum Vergleichen der Varianz des Sprachsignals mit einem oberen Schwellenwert und einem unteren Schwellenwert umfaßt und bei der stimmhafte Sprache für den Fall vorhanden ist, daß die Varianz des Sprachsignals den unteren Schwellenwert überschreitet.

34. Vorrichtung nach Anspruch 32 oder 33, bei der das Bestimmungsmittel Vergleichsmittel zum Vergleichen der ersten Formantverstärkung des Sprachsignals mit einem oberen Schwellenwert und einem unteren Schwellenwert umfaßt und bei der stimmhafte Sprache für den Fall vorhanden ist, daß die erste Formantverstärkung des Sprachsignals den unteren Schwellenwert überschreitet.

35. Vorrichtung nach einem der Ansprüche 32 bis 34, weiter umfassend Mittel zum Erzeugen eines rauschbereinigten Teilungs- oder Schrittzeitraums aus einem Restsignal des Sprachsignals.

Revendications

1. Procédé de réduction de la quantité de bruit dans un signal vocal ayant du bruit, comprenant les étapes de :

- obtention à partir d'un signal vocal, des représentations sur la base d'un modèle décrivant la partie quasi stationnaire de la voix ;
- obtention, à partir de ladite représentation sur la base d'un modèle, d'informations dynamiques concernant des paramètres de fréquence (f), de bande passante (b), et de gain (g) dudit signal vocal par rapport au temps ;
- définition de traitements en fonction du temps en supposant que lesdits paramètres f, b et g appartiennent à un traitement selon une connaissance à priori concernant la dynamique de la voix humaine ;
- lissage des paramètres f, b et g par rapport au temps, l'étape de lissage étant réalisée sur lesdits traitements.

2. Procédé selon la revendication 1, dans lequel la connaissance à priori concernant la dynamique de la voix humaine comprend la connaissance concernant l'étendue de la fréquence maximum d'un signal vocal.

3. Procédé selon la revendication 1 ou 2, dans lequel le signal vocal est supposé appartenir à un traitement, le traitement étant un signal qui peut s'étendre sur une ou plusieurs trames de mesure.

4. Procédé selon la revendication 3, dans lequel le traitement est un traitement actif à un certain moment s'il s'étend sur une ou plusieurs trames de mesure précédentes.

5. Procédé selon la revendication 3 ou 4, dans lequel l'étape de lissage comprend l'étape de détermination si une nouvelle fréquence de formant appartient à un traitement actif.

6. Procédé selon la revendication 5, comprenant en outre l'étape de définition d'un nouveau traitement dans le cas où la nouvelle fréquence de formant n'appartient pas à un traitement actif, et dans lequel la nouvelle fréquence de formant est alors supposée appartenir audit nouveau traitement.

7. Procédé selon l'une quelconque des revendications 4 à 6, dans lequel un traitement est supposé être inactif dans le cas où aucune nouvelle fréquence de formant n'est supposée appartenir audit traitement.

8. Procédé selon la revendication 7, comprenant en outre l'étape de maintien artificiellement du signal vocal pendant

un nombre prédéterminé de trames de mesure dans le cas où le traitement correspondant est brusquement supposé être inactif.

9. Procédé selon la revendication 8, dans lequel le nombre prédéterminé de trames de mesure correspond à la durée maximum du signal vocal.

10. Procédé selon la revendication 9, dans lequel la durée maximum du signal vocal est entre 40 ms et 80 ms.

11. Procédé selon la revendication 10, dans lequel la durée maximum du signal vocal est entre 50 ms et 70 ms.

12. Procédé selon la revendication 11, dans lequel la durée maximum du signal vocal est approximativement 60 ms.

13. Procédé selon l'une quelconque des revendications 5 à 12, dans lequel le nouveau formant est supposé appartenir à un traitement actif si la différence de fréquence entre ledit formant et ledit traitement ne dépasse pas un niveau prédéterminé.

14. Procédé selon la revendication 13, dans lequel le niveau prédéterminé est entre 200 Hz et 600 Hz.

15. Procédé selon la revendication 14, dans lequel le niveau prédéterminé est entre 300 Hz et 500 Hz.

16. Procédé selon la revendication 15, dans lequel le niveau prédéterminé est approximativement 400 Hz.

17. Procédé selon l'une quelconque des revendications précédentes, dans lequel l'étape de lissage comprend l'étape de filtrage des paramètres f, b et g.

18. Procédé selon la revendication 17, dans lequel l'étape de filtrage est réalisée en utilisant un filtre à réponse impulsionnelle infinie (IIR) de premier ordre.

19. Procédé selon la revendication 18, dans lequel le filtre IIR de premier ordre est un filtre de rétroaction de la forme :

$$y[n] = b \cdot x[n] + a \cdot y[n - 1],$$

où x désigne le signal vocal, y désigne la sortie du filtre, et où a et b sont des paramètres à déterminer.

20. Procédé selon la revendication 19, dans lequel les paramètres a et b sont déterminés en utilisant une connaissance de modèle du traitement vocal.

21. Procédé selon l'une quelconque des revendications précédentes, comprenant en outre les étapes de :

- détermination si une voix exprimée est présente ;
- utilisation d'une période de pas du bruit éliminé pour créer une impulsion glottale synthétique dans le cas où la voix exprimée est présente ; et
- remplacement d'au moins une partie du signal vocal original par ladite impulsion glottale synthétique dans le cas où la voix exprimée est présente.

22. Procédé selon la revendication 21, dans lequel l'étape de détermination comprend les étapes de comparaison de la variance du signal vocal à une valeur de seuil supérieure et à une valeur de seuil inférieure, et dans lequel la voix exprimée est présente dans le cas où la variance du signal vocal dépasse la valeur de seuil inférieure.

23. Procédé selon la revendication 21 ou 22, dans lequel l'étape de détermination comprend les étapes de comparaison du premier gain de formant du signal vocal à une valeur de seuil supérieure et à une valeur de seuil inférieure, et dans lequel la voix exprimée est présente dans le cas où le premier gain de formant du signal vocal dépasse la valeur de seuil inférieure.

24. Procédé selon l'une quelconque des revendications 21 à 23, dans lequel la période de pas du bruit éliminé est trouvée à partir d'un signal résiduel du signal vocal.

25. Procédé selon l'une quelconque des revendications 21 à 24, dans lequel l'étape de remplacement est réalisée par l'affaiblissement d'un signal résiduel et l'affaiblissement dans l'impulsion glottale synthétique.

26. Précédé selon l'une quelconque des revendications précédentes, dans lequel au moins l'étape de lissage est réalisée par un système d'ordinateur.

27. Procédé selon l'une quelconque des revendications précédentes, dans lequel le signal vocal est généré dans un téléphone cellulaire.

28. Appareil pour réaliser une réduction du bruit dans un signal vocal, l'appareil comprenant :

- un moyen pour obtenir à partir d'un signal vocal des représentations sur la base d'un modèle décrivant la partie quasi stationnaire de la voix ;
- un moyen pour obtenir des informations dynamiques concernant des paramètres de fréquence (f), de bande passante (b) et de gain (g) dudit signal vocal par rapport au temps ;
- un moyen pour définir des traitements en fonction du temps en supposant que lesdits paramètres f, b et g appartiennent à un traitement selon une connaissance à priori concernant la dynamique de la voix humaine ;
- un moyen de lissage pour lisser les traitements par rapport au temps.

29. Appareil selon la revendication 28, dans lequel une connaissance à priori comprend l'étendue de la fréquence maximum d'un signal vocal.

30. Appareil selon la revendication 28 ou 29, dans lequel le signal vocal est supposé appartenir à un traitement, le traitement étant un signal qui peut s'étendre sur une ou plusieurs trames de mesure, et dans lequel le traitement est un traitement actif à un certain moment s'il s'étend sur une ou plusieurs trames de mesure précédentes, et dans lequel le moyen de lissage comprend un moyen de détermination si une nouvelle fréquence de formant appartient à un traitement actif.

31. Appareil selon l'une quelconque des revendications 28 à 30, dans lequel le moyen de lissage comprend un moyen pour filtrer les paramètres f, b et g.

32. Appareil selon l'une quelconque des revendications 28 à 30, comprenant en outre :

- un moyen de détermination pour déterminer si une voix exprimée est présente ;
- un moyen de création pour créer une impulsion glottale synthétique en utilisant une période de pas du bruit éliminé ; et
- un moyen de remplacement pour remplacer au moins une partie du signal vocal original par ladite impulsion glottale synthétique dans le cas où la voix exprimée est présente.

33. Appareil selon la revendication 32, dans lequel le moyen de détermination comprend un moyen de comparaison pour comparer la variance du signal vocal à une valeur de seuil supérieure et à une valeur de seuil inférieure, et dans lequel la voix exprimée est présente dans le cas où la variance du signal vocal dépasse la valeur de seuil inférieure.

34. Appareil selon la revendication 32 ou 33, dans lequel le moyen de détermination comprend un moyen de comparaison pour comparer le premier gain de formant du signal vocal à une valeur de seuil supérieure et à une valeur de seuil inférieure, et dans lequel la voix exprimée est présente dans le cas où le premier gain de formant du signal vocal dépasse la valeur de seuil inférieure.

35. Appareil selon l'une quelconque des revendications 32 à 34, comprenant en outre un moyen pour produire une période de pas du bruit éliminé à partir d'un signal résiduel du signal vocal.

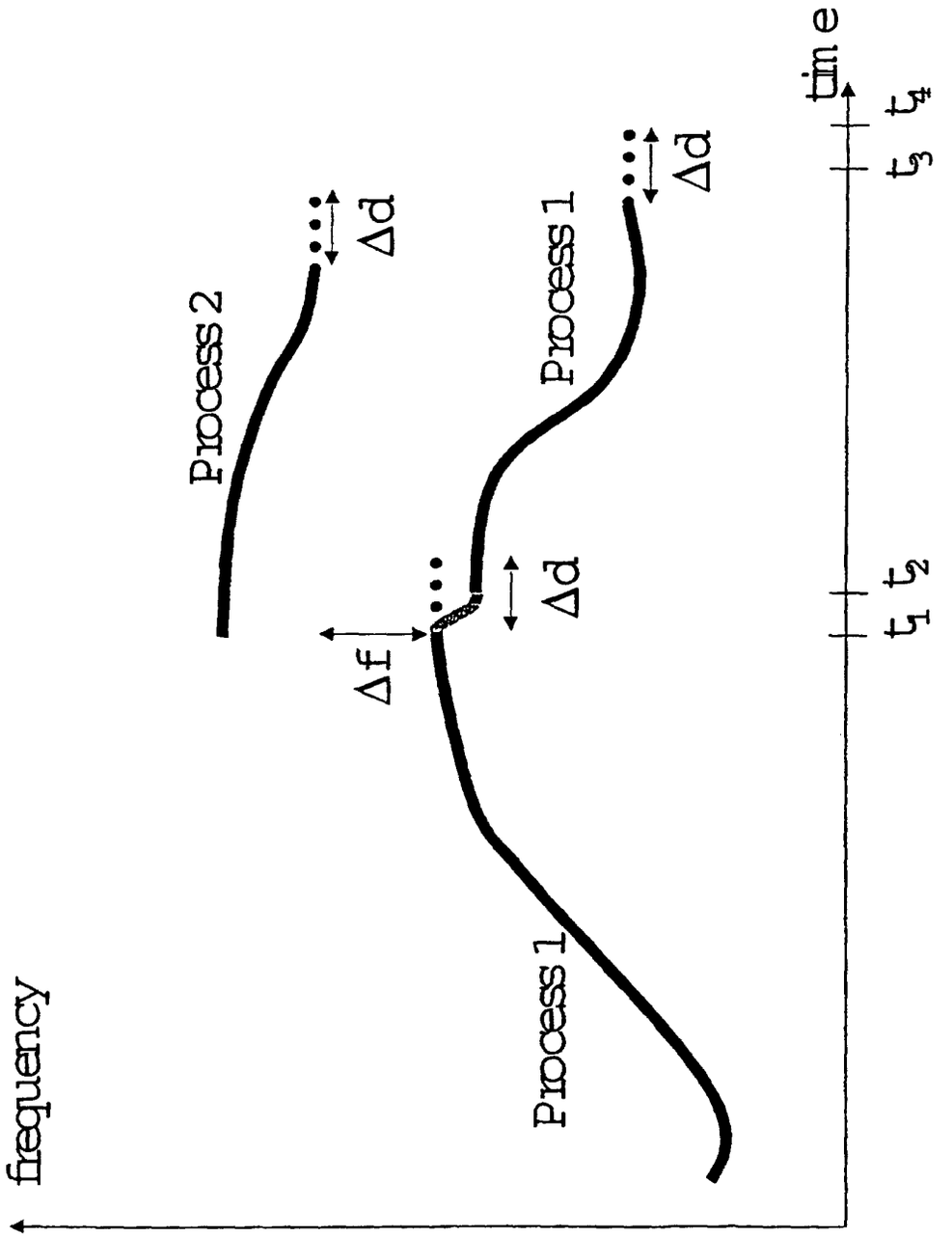

frequency
time
Process 1
Process 2
Process 1
Δf
Δd
Δd
Δd
t1 t2
t3 t4

Fig. 1

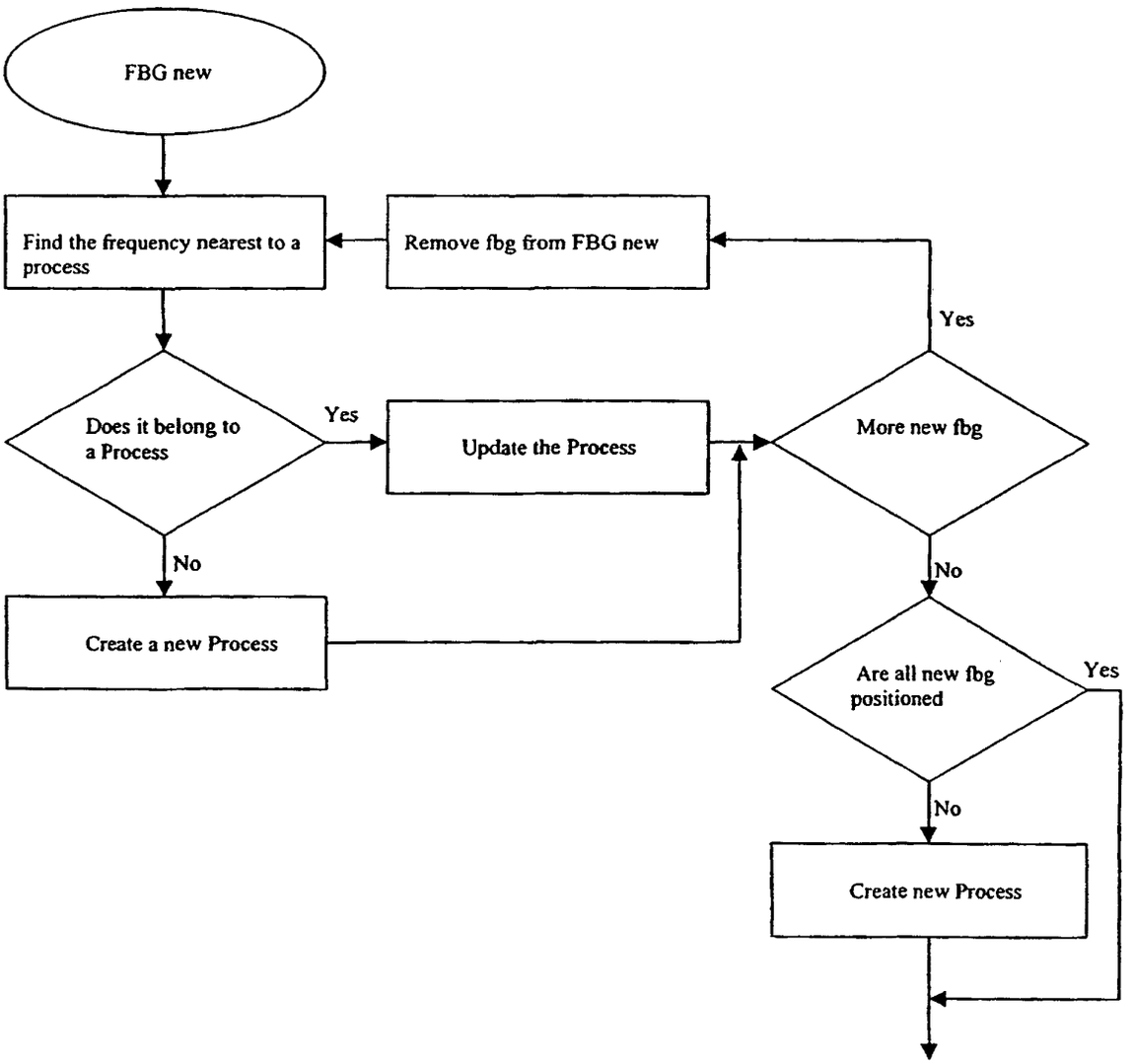

FBG new
Find the frequency nearest to a process
Remove fbg from FBG new
Does it belong to a Process
Yes
Update the Process
No
Create a new Process
Yes
More new fbg
No
Are all new fbg positioned
Yes
No
Create new Process

Fig. 2

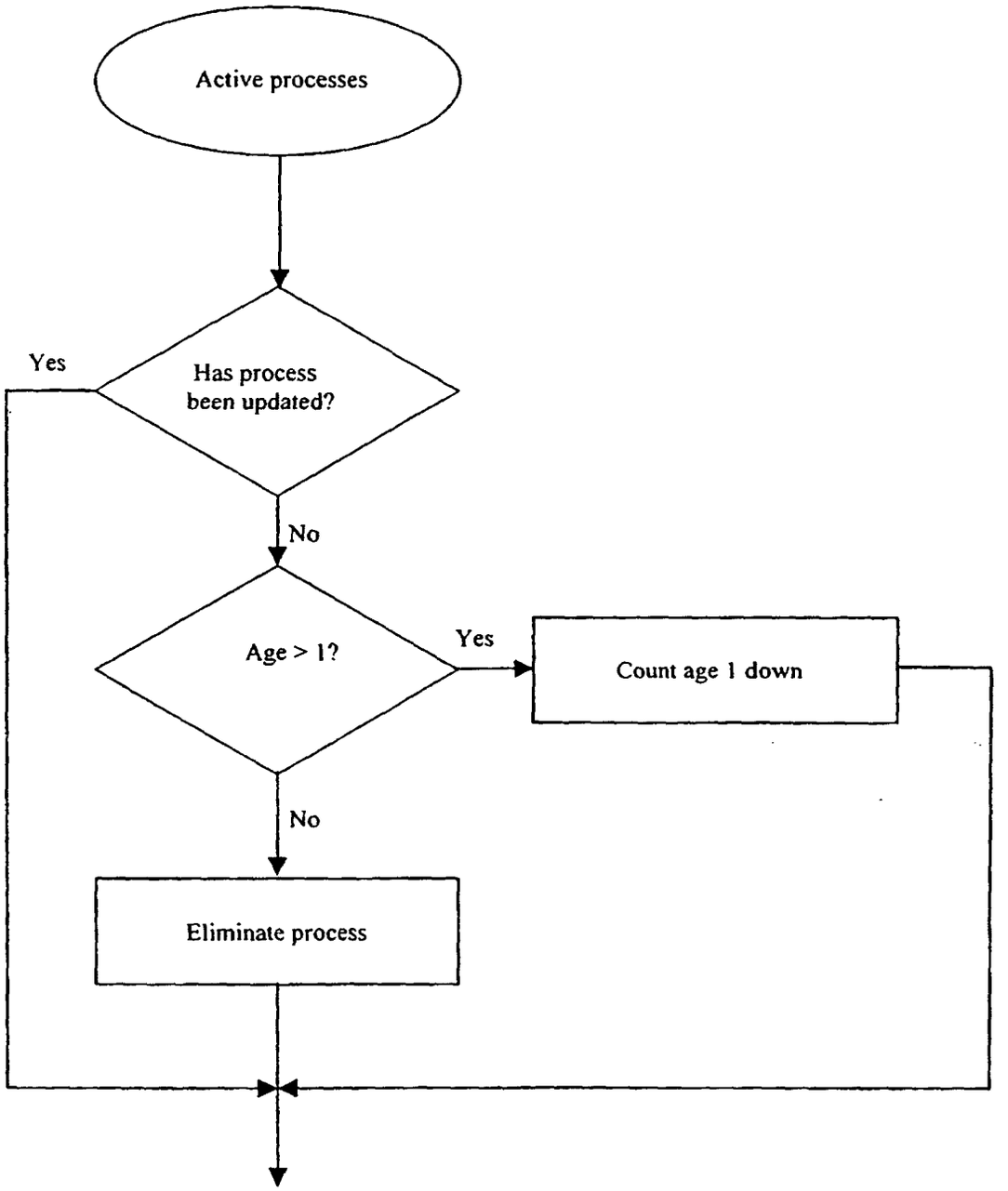

Active processes
Has process been updated?
Yes
No
Age > 1?
Yes
Count age 1 down
No
Eliminate process

Fig. 3

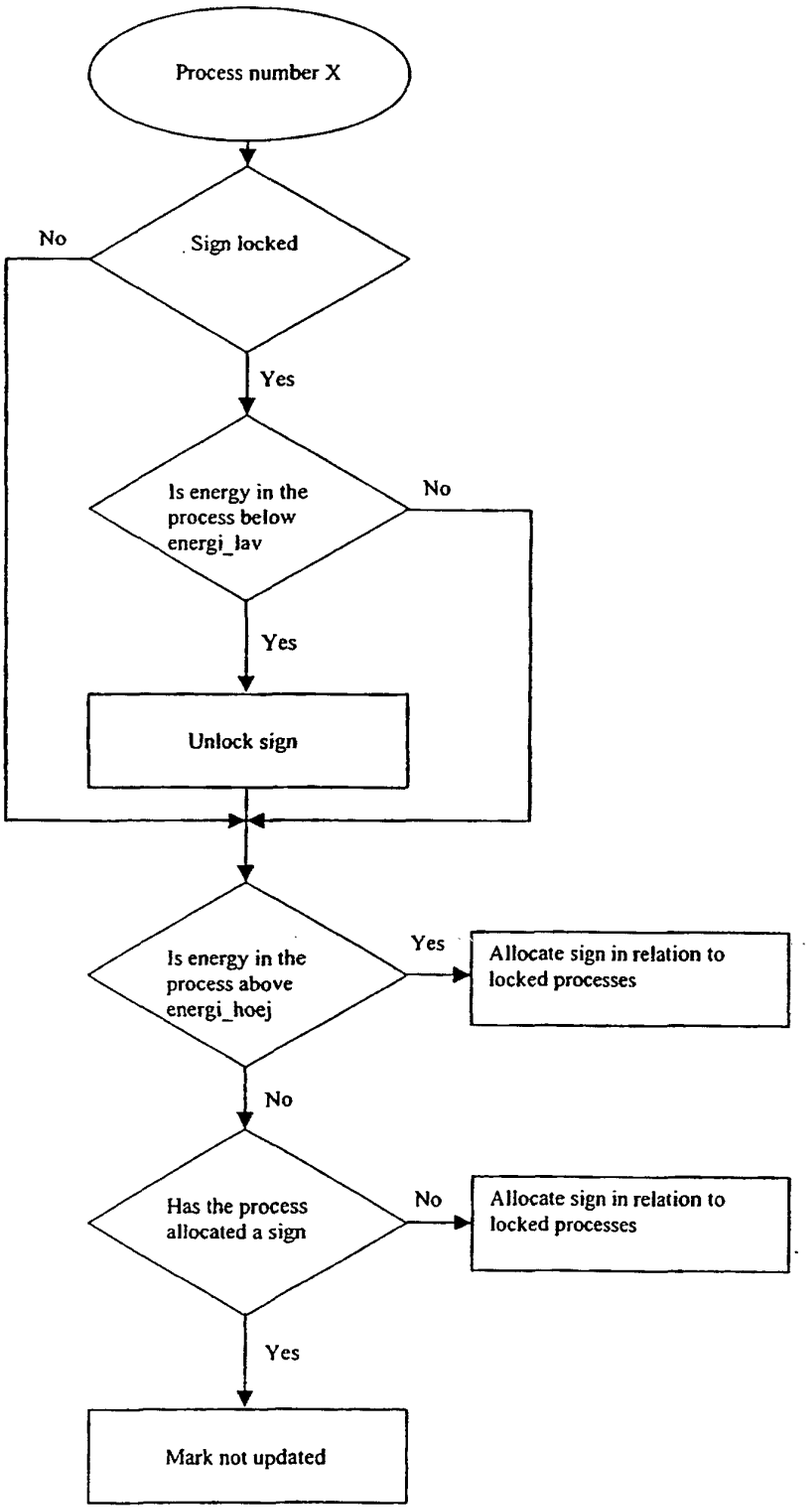

Process number X
Sign locked
No
Yes
Is energy in the process below energi_lav
No
Yes
Unlock sign
Is energy in the process above energi_hoej
Yes
Allocate sign in relation to locked processes
No
Has the process allocated a sign
No
Allocate sign in relation to locked processes
Yes
Mark not updated

Fig. 4

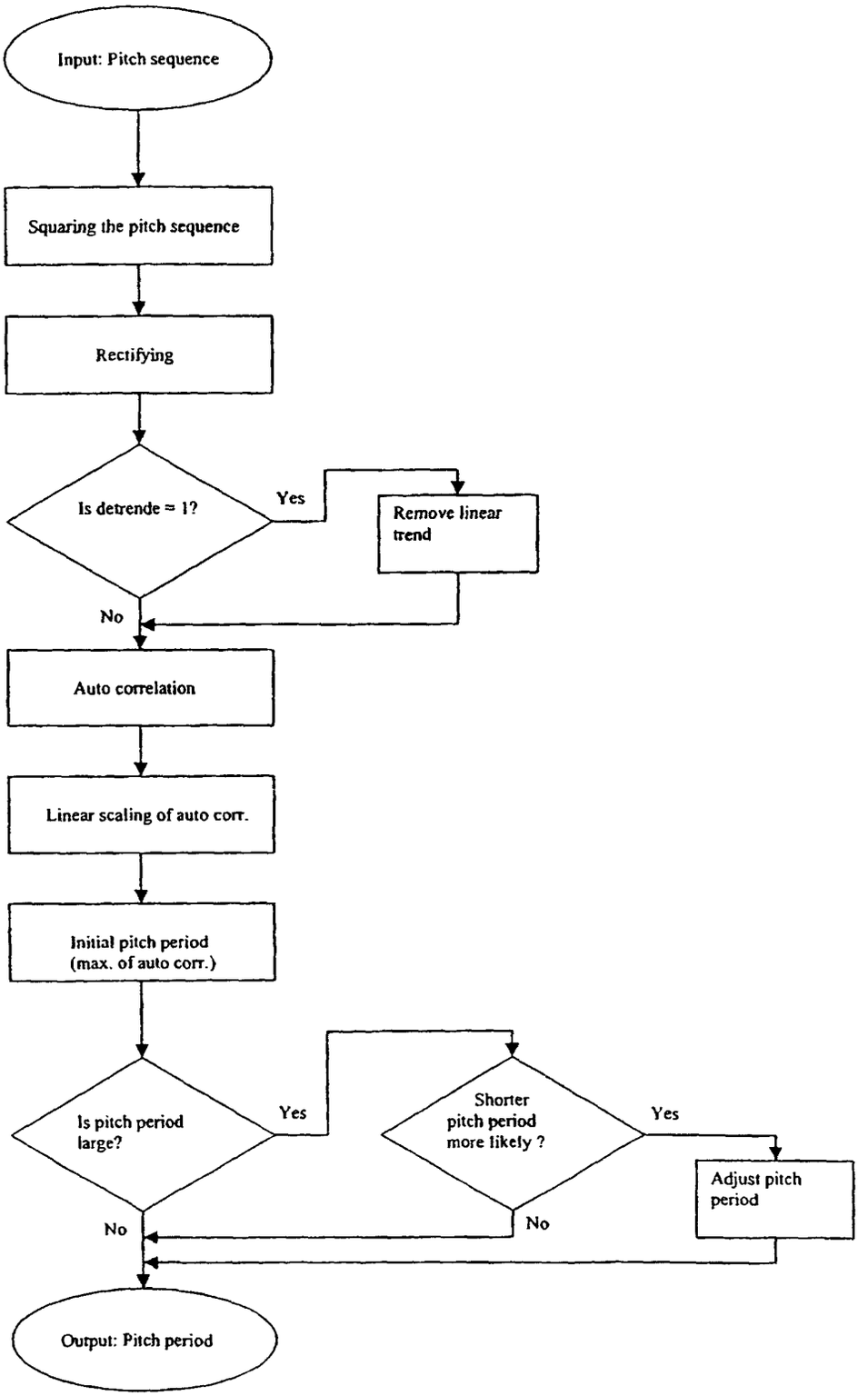

Input: Pitch sequence
Squaring the pitch sequence
Rectifying
Is detrende = 1?
Yes
Remove linear trend
No
Auto correlation
Linear scaling of auto corr.
Initial pitch period (max. of auto corr.)
Is pitch period large?
Yes
Shorter pitch period more likely ?
Yes
Adjust pitch period
No
No
Output: Pitch period

Fig. 5

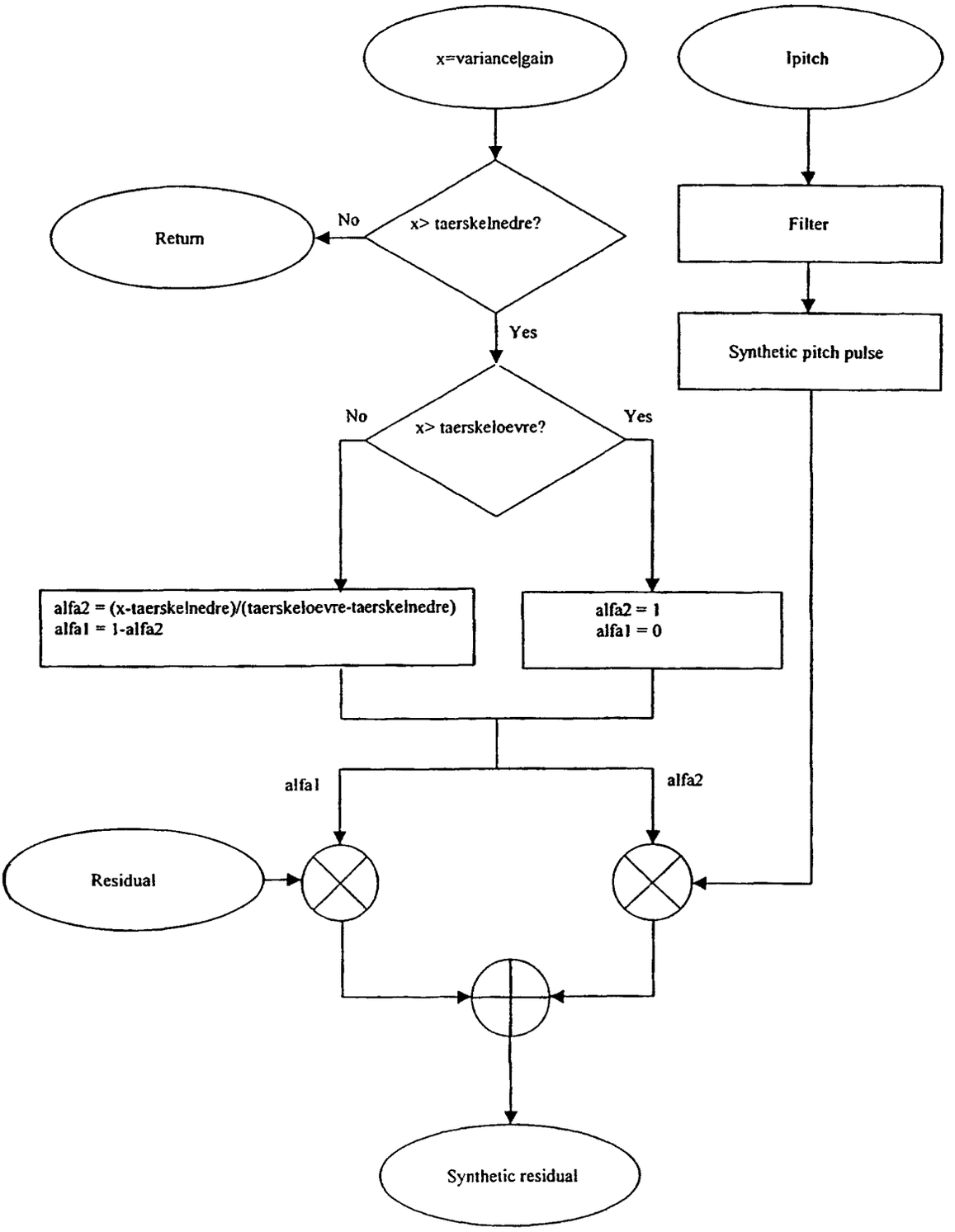

x=variance|gain
Ipitch
Return
No
x> taerskelnedre?
Filter
Yes
Synthetic pitch pulse
No
x> taerskeloevre?
Yes
alfa2 = (x-taerskelnedre)/(taerskeloevre-taerskelnedre)
alfa1 = 1-alfa2
alfa2 = 1
alfa1 = 0
alfa1
alfa2
Residual
Synthetic residual

Fig. 6

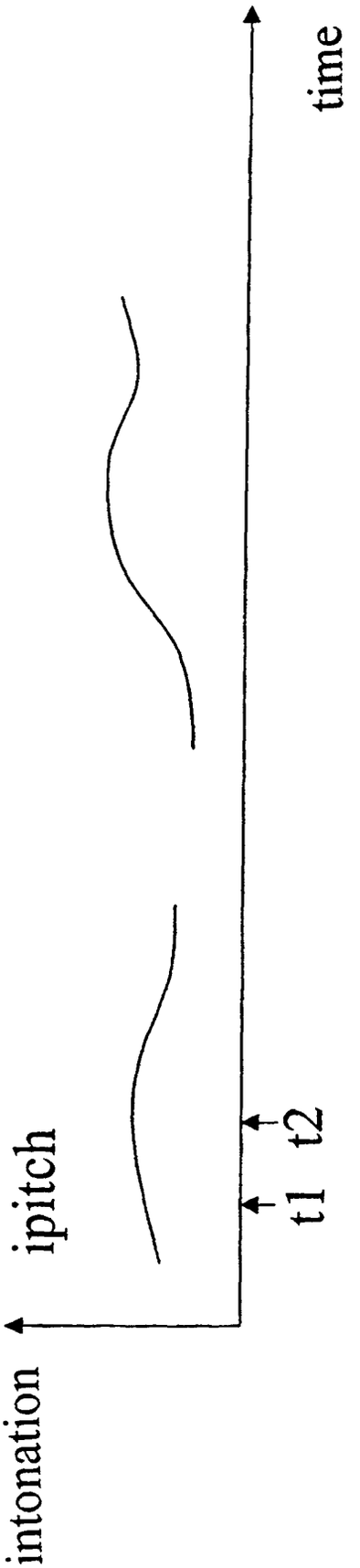


Fig. 7a

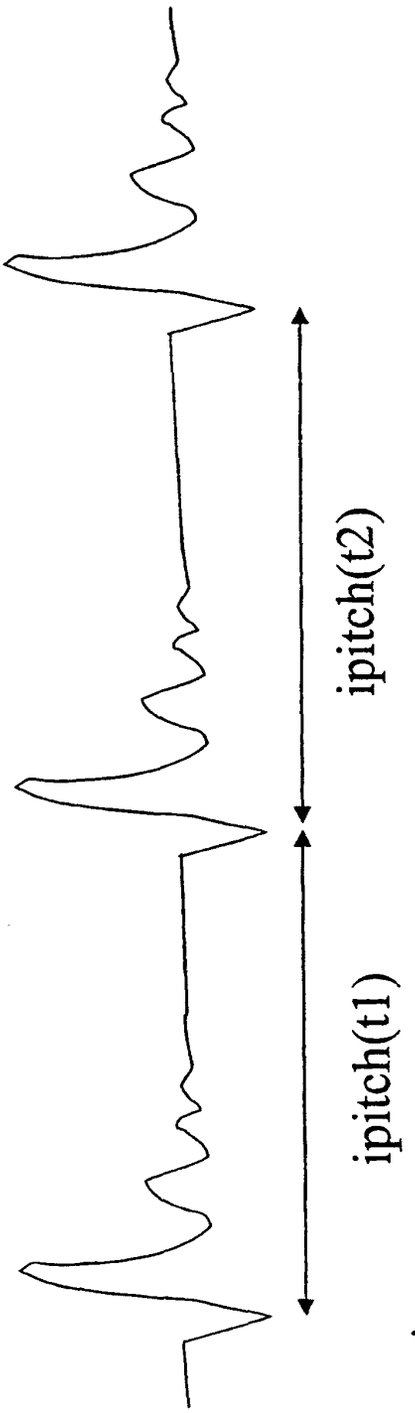


Fig. 7c



Fig. 7b

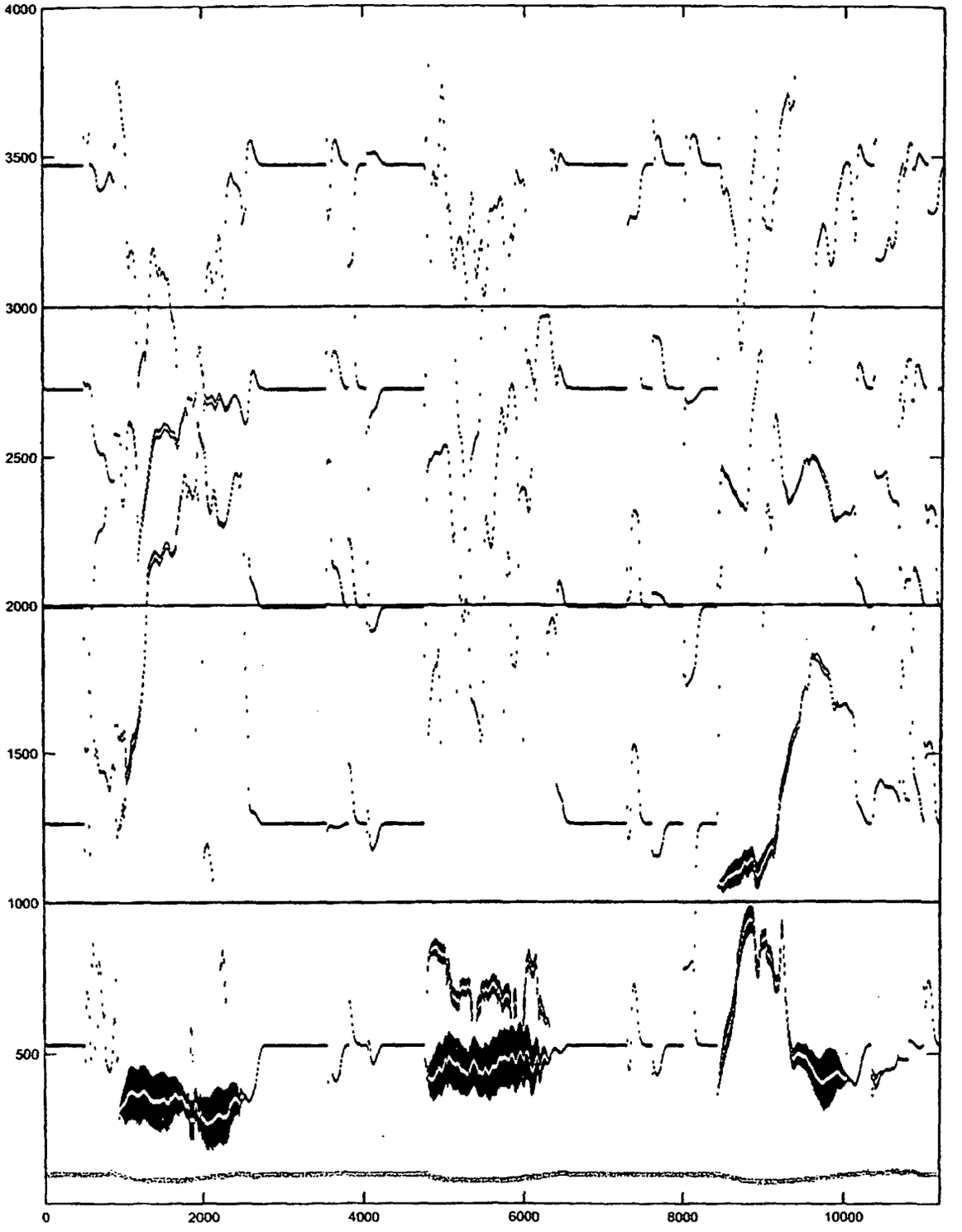

4000
3500
3000
2500
2000
1500
1000
500
0
2000
4000
6000
8000
10000

Fig. 8

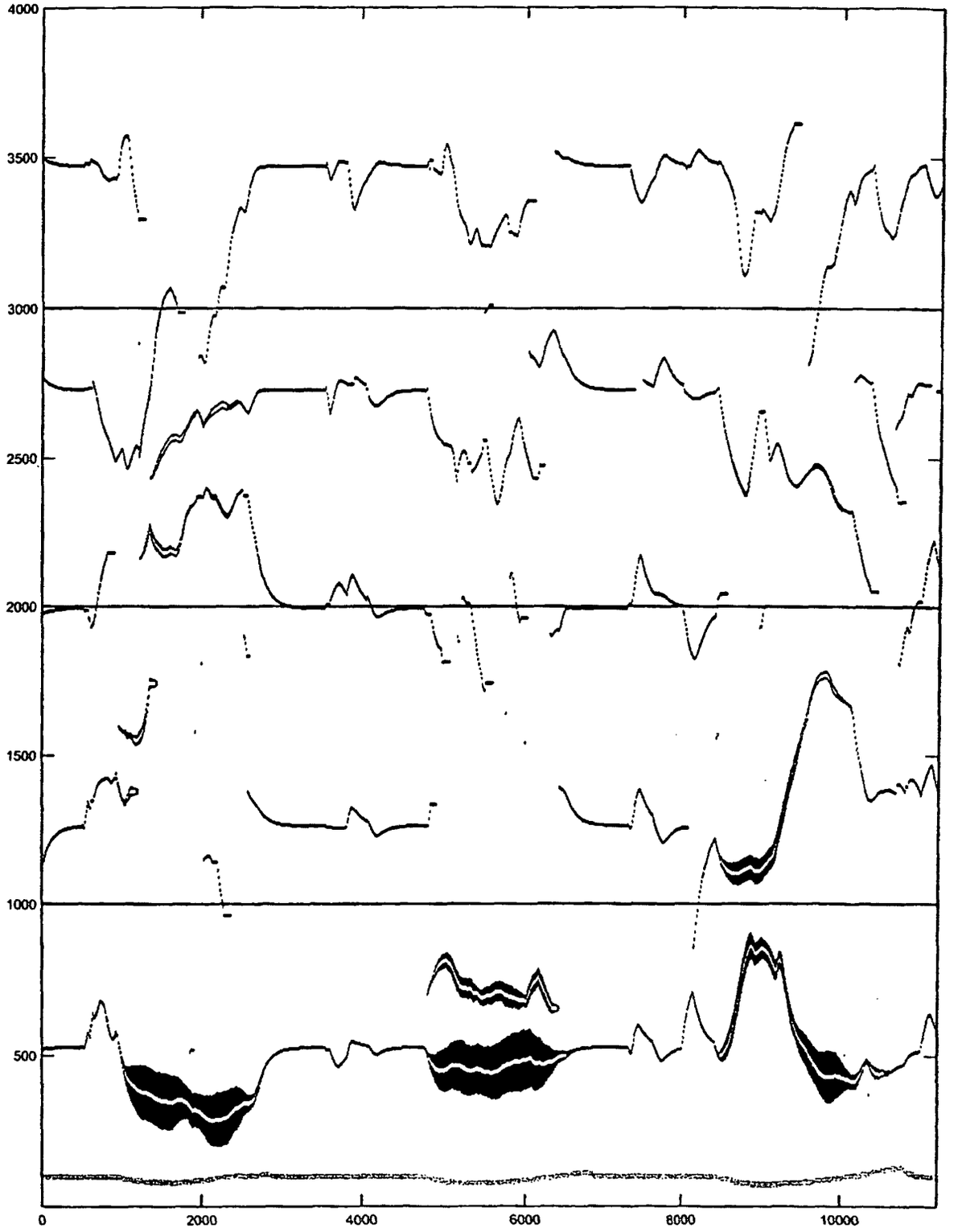

4000
3500
3000
2500
2000
1500
1000
500
0
2000
4000
6000
8000
10000

Fig. 9

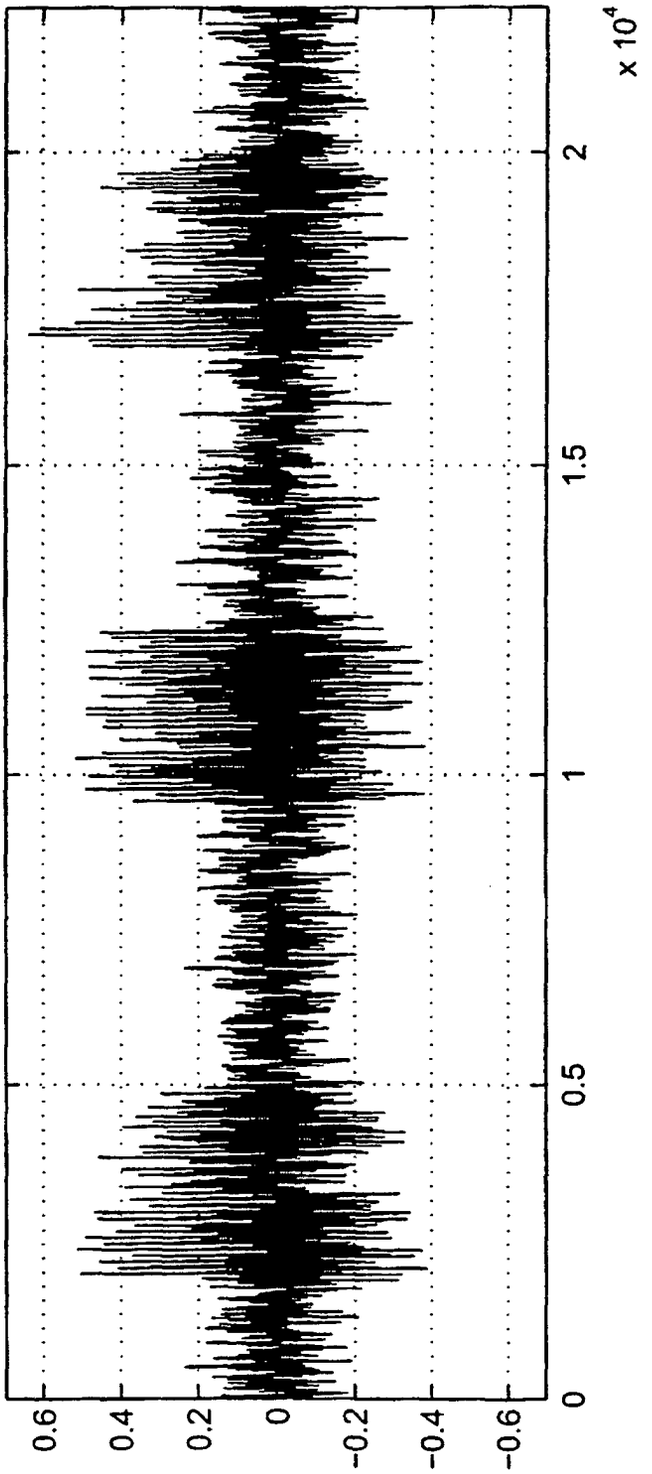


Fig. 10

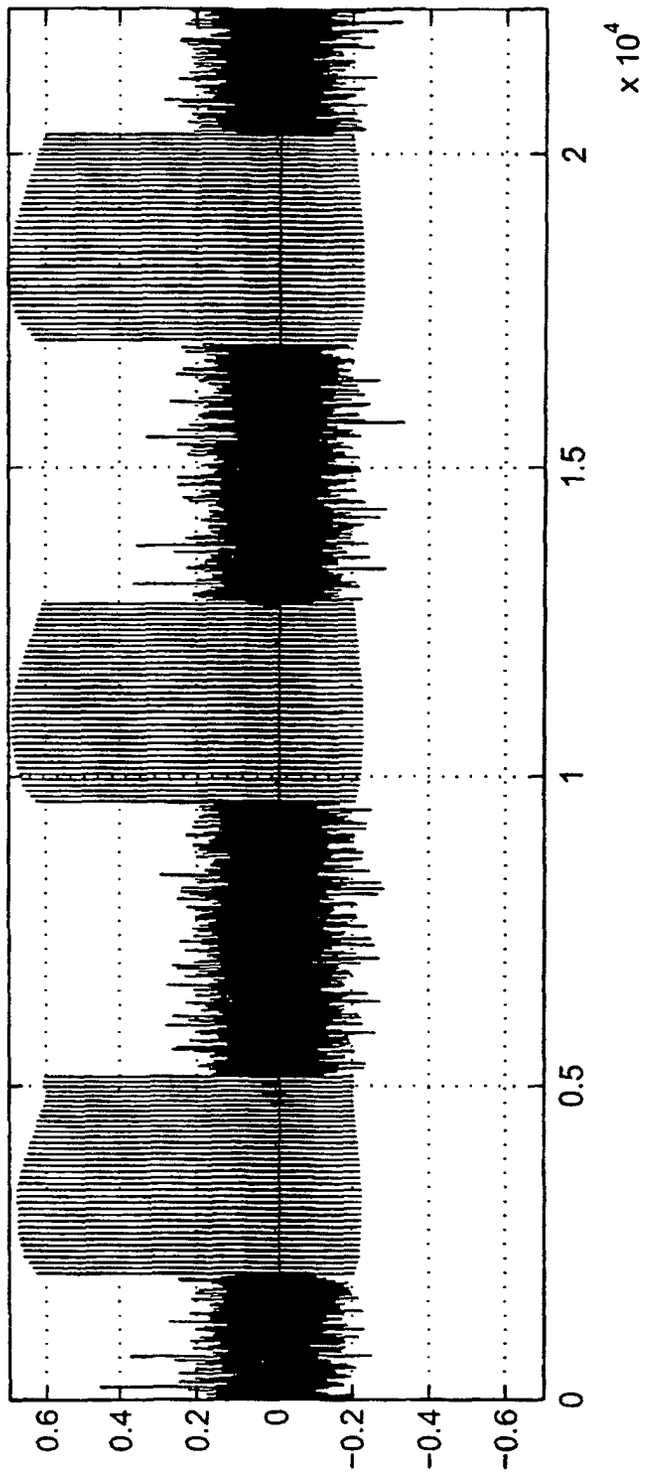


Fig. 11

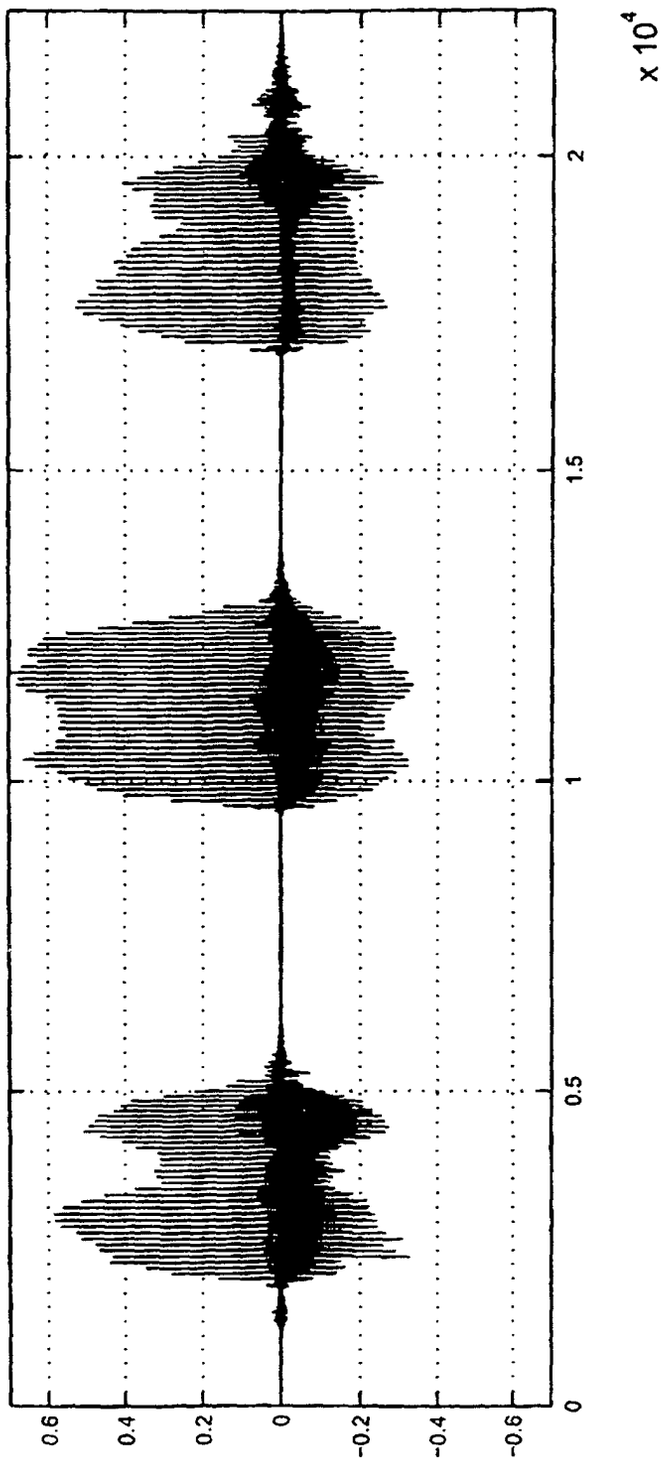


Fig. 12

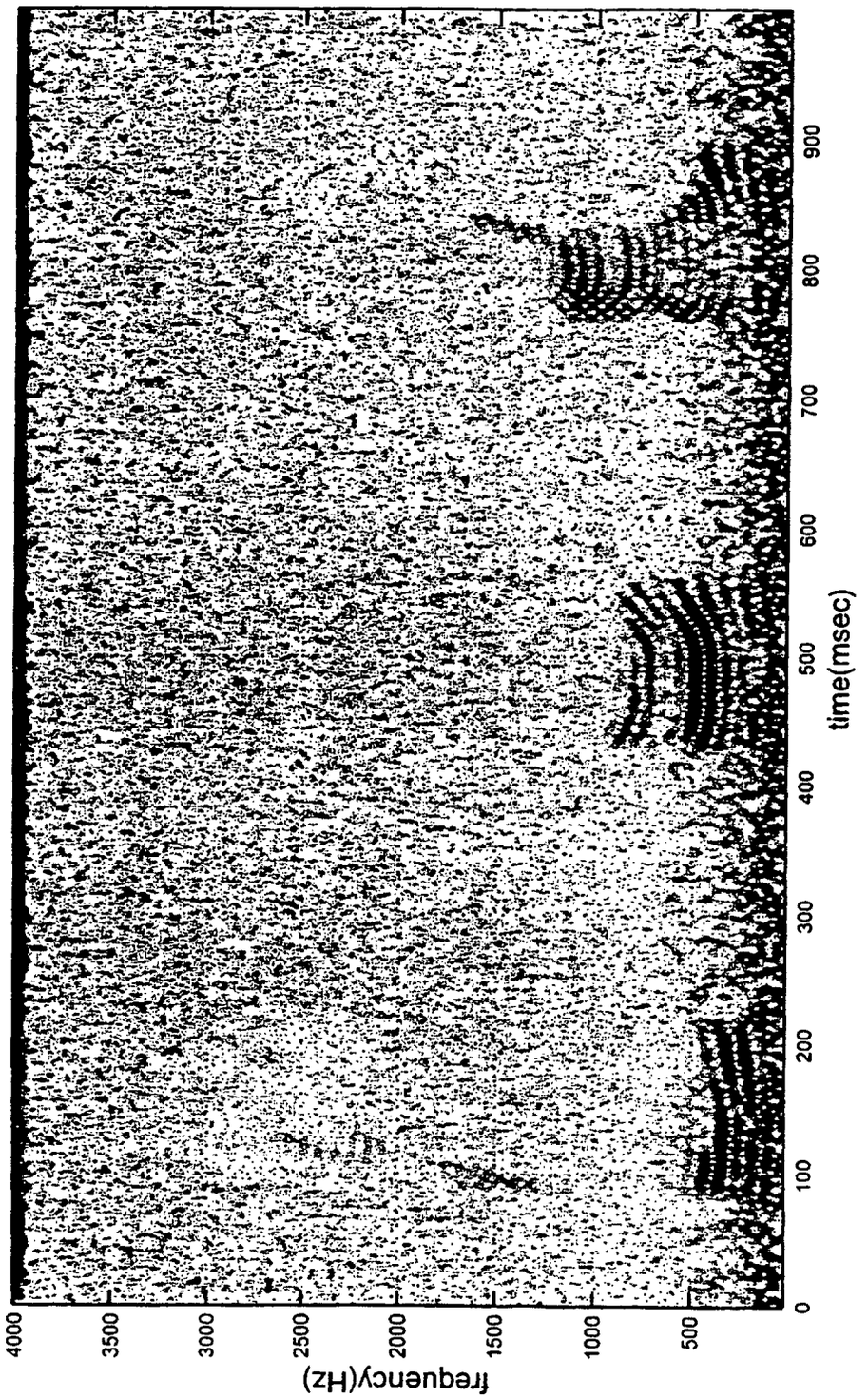

frequency(Hz)
4000
3500
3000
2500
2000
1500
1000
500
0
100
200
300
400
500
600
700
800
900
time(msec)

Fig. 13

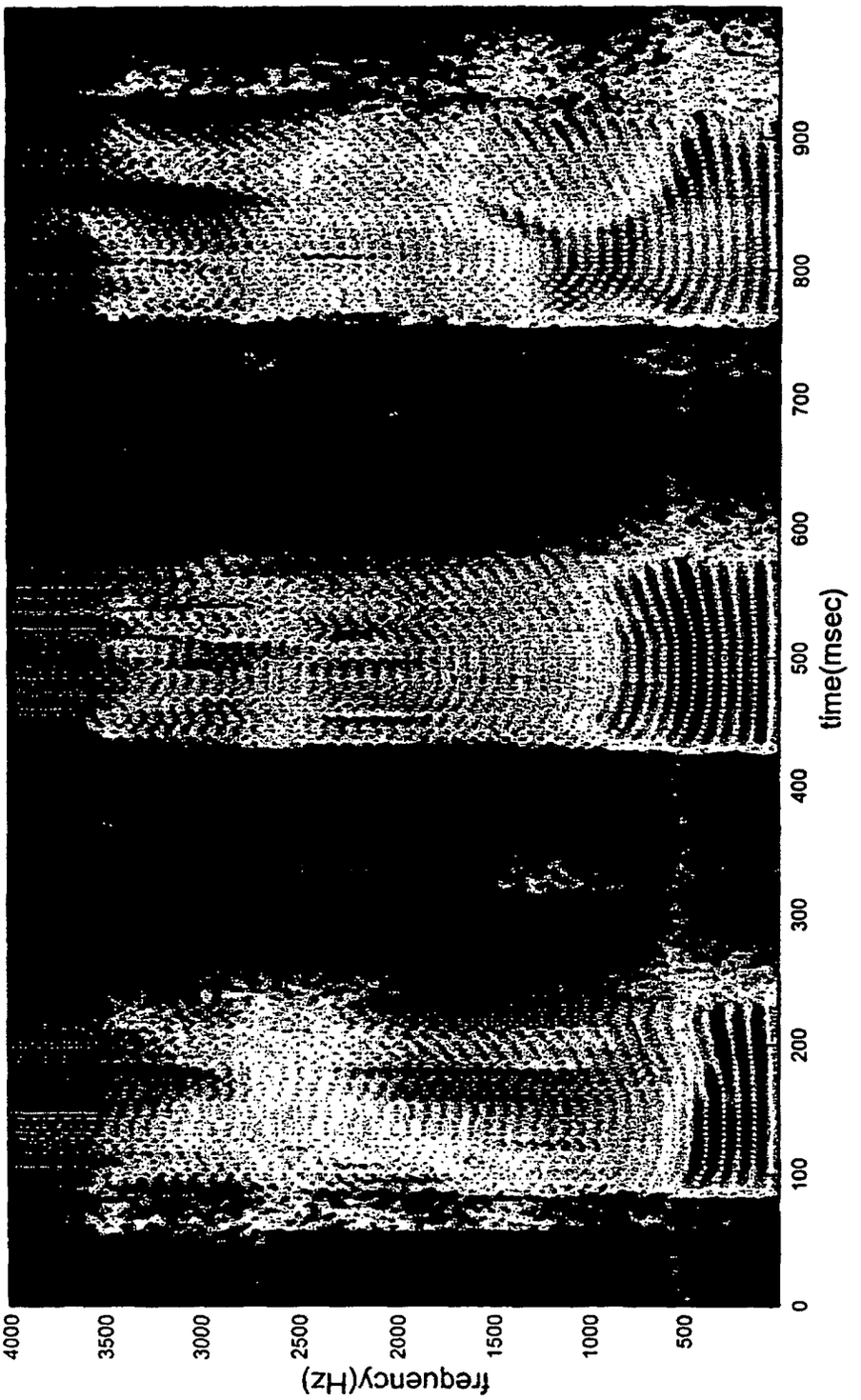

frequency(Hz)
4000
3500
3000
2500
2000
1500
1000
500
0
100
200
300
400
500
600
700
800
900
time(msec)

Fig. 14