PATENT SPECIFICATION (11) 1 586 944

1 586 944

(21) Application No. 17821/78 (22) Filed 4 May 1978
(31) Convention Application No. 2720666 (32) Filed 7 May 1977 in
(33) Fed. Rep of Germany (DE)
(44) Complete Specification Published 25 Mar. 1981
(51) INT. CL[3] G10L 1/00
A61B 5/00
(52) Index at Acceptance
G4R 11A 11D 11X 12F 12G 1F 6B 7B
8F 9B
H4R PVA
(19)

(54) SOUND ANALYSIS

(71) We, N.V.PHILIPS' GLOEILAMPENFABRIEKEN, a limited liability Company, organised and established under the laws of the Kingdom of the Netherlands, of Emmasingel 29, Eindhoven, the Netherlands do hereby declare the invention, for which we pray that a patent may be granted to us, and the method by which it is to be performed, to be particularly described in and by the following statement:

The invention relates to a method of sound analysis of a kind which involves analysing an electrical signal representing a sound occurring during a limited period of time, and wherein respective short-term spectra formed of partial signals are produced from the electrical signal in consecutive short intervals said partial signals respectively indicating the energy content of the electrical signal representing the sound in adjacent spectral ranges and all short-term spectra being combined into a long-term spectrum. The invention also relates to an arrangement for performing this method.

Such long-term spectra have been used with success up to the present for sound analysis, particularly for speech and speaker identification. For instance, the identification of a speaker can be used in banking business for the situation in which a customer gives an instruction orally and verification of the customer is done by means of his speech, that is to say a recorded speech sample is compared in an analyser arrangement with a prior speech sample of the customer.

It is desirable, that a customer can also give such an oral instruction with associated verification over the telephone. In the transmission of speech over a telephone line the speech signal is, however, affected by the always different transfer functions of the transmission path. The speech signal can be changed to such an extent that a subsequent identification by means of automatic systems is greatly affected or even impossible. This also applies to other sounds which are, for example, transmitted for monitoring purposes to a monitoring central station and compared there with standard sounds.

It is an object of the invention to provide a method of sound analysis of the kind referred to, which is suitable for speech and speaker identification, respectively, in which the disturbing influence of transfer functions of the transmission path between the sound source and an analyser arrangement are eliminated to a very large extent. According to the invention a method of the kind referred to is characterised in that the difference between each short-term spectrum and the preceding short-term spectrum, or the long-term spectrum is formed, that an intermediate value is derived from all the differences so formed and this intermediate value is divided by the long-term spectrum to form a quotient and that quotient is applied to an arrangement for the identification of the sound. Consequently, the changes in the long-term spectrum are used as a characteristic feature for the analysis instead of the long-term spectrum itself. As these changes relate to a long-term spectrum, the influence of a linear transfer function which can be represented as a vector of the coefficients in the individual spectral ranges is completely eliminated in the ideal case, as can be easily deduced mathematically. If the difference to the preceding average value is formed, the differences can immediately be processed in parallel with the speech signal, while the formation of the difference to the long-term spectrum can only be performed at the end of the period and the long-term spectrum must therefore be stored in an intermediate store.

Deriving the intermediate values can be done in various manners. Efficiently, the derived intermediate value is the square root of the sum of the squares of the differences divided by the number of differences. In this way the so-called standard deviation is determined. Another possibility consists in that the derived intermediate value is the sum of the amounts of the differences divided by the number of differences. This possibility is somewhat easier and somewhat cheaper.

Efficiently, the differences are formed from average-term spectra representing the average value of a predetermined number of consecutive short-term spectra. In this case only a small number of signals need be stored and the required processing rate of the whole arrangement becomes lower. With a proper choice of the number of consecutive short-term spectra which are combined, that is to say with a proper choice of the periods for each average value spectrum, the identification possibility is, in addition, improved.

Known arrangements, used for performing the method of the kind referred to comprise a filter unit connected to receive the electrical signal representing the sound to be analysed and successively producing output signals corresponding to the consecutive short-term spectra, a first adder arrangement connected to the output of the filter unit for producing signals corresponding to the long-term spectrum and an identification arrangement. For performing the method according to the invention such a known arrangement is characterized in that a subtractor arrangement is connected to the output of the filter unit, the subtractor arrangement being also connected to the output of a store which is connected to the output of the filter unit onto the output of the first adder arrangement, the subtractor arrangement producing an output signal indicating the difference between the values of the signals applied to its two inputs, that a second adder arrangement is connected to the output of the subtractor arrangement for forming the average value of the differences produced by the subtractor arrangement that the outputs of the first adder arrangement and of the second adder arrangement are connected to the inputs of a divider arrangement which forms the output signals corresponding to the quotient of the values represented by the signals at its two inputs, and that the output of the divider arrangement is connected to the input of the identification arrangement. In this manner the method according to the invention is carried into effect with only a few additional elements added to the known arrangement.

For forming the differences from consecutive short-term spectra it is efficient that the input of the store is connected to the output of the filter unit and that the store supplies at its output the received signal at the instant the next value of the same spectral range appears at the input. Consequently this store needs only store the values of one short-term spectrum.

For forming the difference from the long-term spectrum it is, on the contrary, efficient that the input of the store is connected to the output of the first adder arrangement and that one input of the subtractor arrangement is connected to a second store connected to the output of the filter unit and storing all the signals produced during the limited period of time and successively supplying them to the subtractor arrangement at the end of this period. As the long-term spectrum is not available until the end of the period during which the sound is recorded, the differences cannot be calculated until after that instant so that all short-term spectra must be intermediately stored in the second store. The first store can be dispensed with when the first adder arrangement comprises already a store for each output value. A further elaboration for determining the standard variation is characterized in that the second adder arrangement is preceded by a squaring arrangement and followed by a square root extraction arrangement.

To combine each time several consecutive short-term spectra it is efficient to connect a third adder arrangement between the output of the filter unit and the input of the subtaction arrangement or the first or second store, respectively. This third adder arrangement can, in principle, be set-up in the same manner as the first and the second adder arrangements. The first adder arrangement can be connected directly to the output of the filter unit or to the output of the third adder arrangement.

It is also conceivable that the short-term spectra and the average term spectra as well as the long-term spectra each consist of one value for each spectral range. These individual values must be processed separately during the processing steps. As the processing steps are the same for each spectral value, the values of each spectral range can be consecutively processed in one processing step, so that only one arrangement is each time required, but possibly a corresponding number of intermediate stores. To yet increase the processing rate it is efficient that at least some of the adder arrangements and/or the subtractor arrangement and/or the divider arrangement consist of a number of parallel elements, which number does not exceed the number of adjacent spectral ranges, these elements producing separately and simultaneously the signals for several to all spectral ranges. However, this implementation requires additional cost and trouble and is therefore only necessary if particularly high requirements are imposed on the processing rate.

The arrangements for performing the individual processing steps can also be combined and realized by means of a correspondingly programmed multi-purpose computor, for example a micro-processor.

Embodiments of the invention will now be further explained with reference to the drawing in which:

Figure 1 shows the block diagram of an arrangement according to the invention,

Figure 2 shows an other implementation of the arrangement according to the invention.

In Figure 1 an electrical signal representing a sound to be analysed is applied to filter unit 1. This filter unit comprises a plurality of filters with adjacent passbands. The outputs of these filters can be led out directly. Often, however, the outputs of the filters lead to a multiplexing arrangement, which scans these output values successively in a cyclic manner and passes them to an output of the filter unit. The cycling time of the multiplexer is then, in general, in the range of from 1 to 20 ms. This means that the individual values of successive short-time spectra are supplied consecutively and that the distance between two successive short-time spectra is equal to the cycling time of the multiplexer. The example of Figure 1 shows a filter unit having one multiplex output only.

Connected to the output of the filter unit is an an adder arrangement 8 which each time adds a number of consecutive short-term spectra, that is to say each spectral component separately and successively. To this end adder 8 comprises preferably an adder arrangement as well as a shift register connected thereto, having a number of stages equal to the number of spectral ranges. Each stage of the shift register comprises a plurality of parallel storage positions equal to the maximum number of bits which can occur in the binary number indicating a spectral component in the average-term spectra. Each time the filter unit supplies a new signal for a spectral component, the associated value obtained sofar by means of addition has arrived at the shift register output, and the two values are added together and written into the first stage of the shift register, the oldest content which is then no longer required being removed from the shift register. After the predetermined number of short-term spectra the output of the shift register is switched over to the output of adder 8 and the next addition of the spectral components starts again at zero value. The division by the number of added short-term spectra for forming the average value is particularly simple if this number is an integral power of 2. In the other case a divider must be included between the output of the shift register and the output of adder 8. This basic implementation of adder 8 is known *per se* and is not an object of the invention.

The output of adder 8 is connected to the input of a further adder 2 which adds all average-termed spectra separately on the basis of the individual spectral values and which can be implemented in the same manner as adder 8, but for the fact that it has a correspondingly higher number of positions.

In addition, the output of adder 8 is connected to an input of a subtractor 4 and a store 5 whose output is connected to the other, subtracting input of subtractor 4. If store 5 delays the received signals for a number of steps equal to the number of spectral ranges, the spectral value available at the output of store 5 is the same as that at an input of subtractor 4, but for the fact that the former is spectral value of the preceding spectrum. To accomplish this, store 5 can be implemented as a shift register having a number of stages equal to the number of spectral values, each stage having a number of storage positions corresponding to the highest possible number of bits for a spectral value.

The subtractor 4 is arranged for processing two multibit binary numbers and has also a fundamentally known structure, for example subtraction by means of adding the complement values.

The output of subtractor 4 is connected to the input of adder 6 which can also be implemented in the same manner as adder 8. This adder 6, however, processes only the absolute values of the number supplied by subtractor 4, that is to say without taking the sign into account.

The outputs of adders 2 and 6 are connected to the inputs of a divider 7 which divides the value at the output of adder 6 by the value at the output of adder 2. This can, for example, be done in known manner by continuous subtraction, for which divider 7 is implemented in known manner.

In the described example the adders 2 and 6 need not supply the sum divided by the number of added values but the sums can be directly applied to divider 7 as the divisor which is substantially the same for the two sums is cancelled when these two values are divided. The slight deviation produced owing to the fact that adder 6 adds, owing to the difference formation of each time two consecutive values in subtractor 4, one value less than adder 2 can be neglected in most cases. The adders 2 and 6 supply the output values to the divider 7 consecutively and also separated on the basis of spectral values, so that the values for the consecutive spectral ranges appear successively at the output of this divider. These values are supplied to an analyser arrangement 3 and compared therein to a plurality of values obtained in the same manner. Analyser arrangement 3 can be similarly programmed multi-purpose computer and does not form part of the invention.

Figure 2 shows a similar arrangement in which the elements corresponding to the elements in Figure 1 have been given the same reference numerals. Here the output of filter unit 1 is connected in parallel to adders 8 and 2, which may be implemented in the same manner as described with reference to Figure 1. The output of adder 8 is connected to a store 9 which stores all the signals produced by adder 8 during the limited period of sound recording or speech transmission. In parallel therewith the long-term spectrum is again formed in adder 2, separated on the basis of spectral ranges, and entered into store 5 at the end of the period.

The output of store 5 is again connected to the subtracting input of subtractor 4 one input of which is connected to the output of store 9. If the sound has been fully recorded, store 5 consequently contains the long-term spectrum whose spectral portions are successively and in cyclic manner applied to subtractor 4. Simultaneously the corresponding spectrum components of the average term spectra, stored in store 9, are applied to subtractor 4 whose output is connected to the input of a squaring device 11. By means of squaring all the numbers appearing at the output of squaring device 11 have a positive sign and are added in adder 6. In this example adder 6 and also adder 2 must comprise a divider arranged before the output, which each time divides the sums by the number of individual values. The output signal of adder 6 is furthermore applied to a root extraction device 12 which forms the root of the values supplied by adder 6 at the end of the processing operation.

The outputs of store 5 and root extractor 12 are connected to the input of divider 7 which supplies at the output the individual spectral values of the standard deviation and applies them to analyser arangement 3. This arrangement does not actually operate in real time, properly speaking, as the formation of the differences and their further processing can not start until the sound has been fully recorded at the end of the limited period of time and it is not possible until that instant to process successively all spectral values of all short-term spectra.

With a long cycle time of the multiplexer in the filter unit 1 the formation of average-term spectra in adder 8 markedly reduces the attainable accuracy. In this case it is possible to omit adder 8 and to connect the output of the filter unit 1 directly to the input of adder 2 and store 9 or to the input of the subtractor 4 and the store 5, respectively. This depends on the cycling time of the multiplexer and the highest oscillation frequency of the processed spectra. On the other hand it is advantageous, particularly in speech or speaker identification, to produce a given number of short-term spectra by means of adder 8 and to process them further to attain an optimum identification result.

If a high processing rate is required, that is to say the identification result must already be available immediately after the end of the period of the recorded sound it may be necessary, especially for the arrangement shown in Figure 2, to provide a plurality of elements arranged between the output of filter unit 1 and the inputs of analyser arrangement 3 so that each element successively processes only a portion of the total spectral values and several values of the same spectrum can be processed in parallel. This raises the processing rate with a corresponding factor and, in the limit case, a real time situation is in practice obtained again with one separate arrangement for each single spectral value.

The elements arranged between the output of filter unit 1 and the input of analyser arrangement 3 can also be combined in an arithmetic and logic unit, which is programmed correspondingly, for example in a micro-processor.

WHAT WE CLAIM IS:

1. A method of analysising an electrical signal representing a sound occurring during a limited period of time, and wherein respective short-term spectra formed of partial signals are produced from the electrical signal in consecutive short intervals, said partial signals respectively indicating the energy content of the electrical signal representing the sound in adjacent spectral ranges and all short-term spectra being combined into a long-term spectrum, characterized in that the difference between each short-term spectrum and the preceding short-term spectrum, or the long-term spectrum is formed, that an intermediate value is derived from all the differences so formed and this intermediate value is divided by the long-term spectrum to form a quotient and that quotient is applied to an arrangement for the identification of the sound.

2. A method as claimed in Claim 1, characterized in that the derived intermediate value is the square root of the sum of the squares of the differences divided by the number of differences.

3. A method as claimed in Claim 1, characterized in that the derived intermediate value is the sum of the amounts of the differences divided by the number of the differences.

4. A method as claimed in Claim 1, characterized in that the differences are formed from average-term spectra representing the average value of a predetermined number of consecutive short-term spectra.

5. An arrangement for performing the method as claimed in Claim 1, comprising a filter unit which is connected to receive the electrical signal representing the sound to be analysed and sucessively producing output signals corresponding to the consecutive short-term spectra, a first adder arrangement connected to the output of the filter unit for producing signals corresponding to the long-term spectrum, and an identification arrangement, characterized in that a subtractor arrangement is connected to the output of the subtractor arrangement being also connected to the output of a store which is connected to the output of the filter unit or to the output of the first adder arrangement, the subtractor arrangement producing an output signal indicating the difference between the values of the signals applied to its two inputs, that a second adder arrangement is connected to the output of the subtractor arrangement for forming the average value of the differences produced by the subtractor arrangement, that the outputs of the first adder arrangement and of the second adder arrangement are connected to the inputs of a divider arrangement which forms the output signals corresponding to the quotient of the values represented by the signals at its two inputs, and that the output of the divider arrangement is connected to the input of the identification arrangement.

6. An arrangement as claimed in Claim 5, characterized in that the input of the store is connected to the output of the filter unit and that the store supplies at its output the received signal at the instant the next value of the same spectral range appears at the input.

7. An arrangement as claimed in Claim 5, characterized in that the input of the store is connected to the output of the first adder arrangement and that one input of the subtractor arrangement is connected to a second store connected to the output of the filter unit and storing all the signals produced during the limited period of time and successively supplying them to the subtractor arrangement at the end of this period.

8. An arrangement as claimed in Claim 7, characterized in that the second adder arrangement is preceded by a squaring arrangement and followed by a square root extraction arrangement.

9. An arrangement as claimed in Claim 5, characterized in that a third adder arrangement is connected between the output of filter unit and the input of the subtractor arrangement or the first or second store, respectively.

10. An arrangement as claimed in Claim 5, characterized in that at least some of the adder arrangements and/or the subtractor arrangement and/or the divider arrangement (7) consist of a number of parallel elements which number does not exceed the number of adjacent spectral ranges, these elements producing separately and simultaneously the signals for several to all spectral ranges.

11. A method as claimed in Claim 1 substantially as herein described.

12. An arrangement as claimed in Claim 5 substantially as herein described.

R.J. BOXALL,
Chartered Patent Agents,
Mullard House,
Torrington Place,
London WC1E 7HD.
Agent for the Applicants.

Printed for Her Majesty's Stationery Office,
by Croydon Printing Company Limited, Croydon, Surrey, 1981.
Published by The Patent Office, 25 Southampton Buildings,
London, WC2A 1AY, from which copies may be obtained.

1586944 COMPLETE SPECIFICATION
1 SHEET
This drawing is a reproduction of the Original on a reduced scale

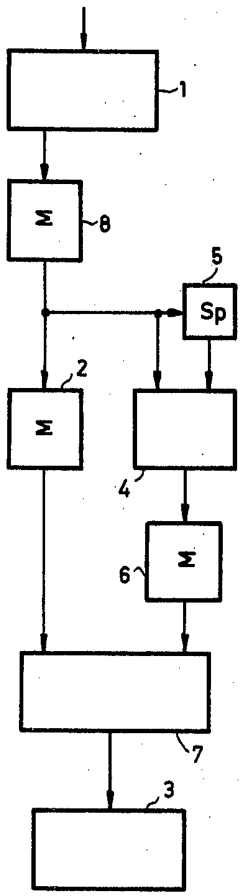

Fig.1

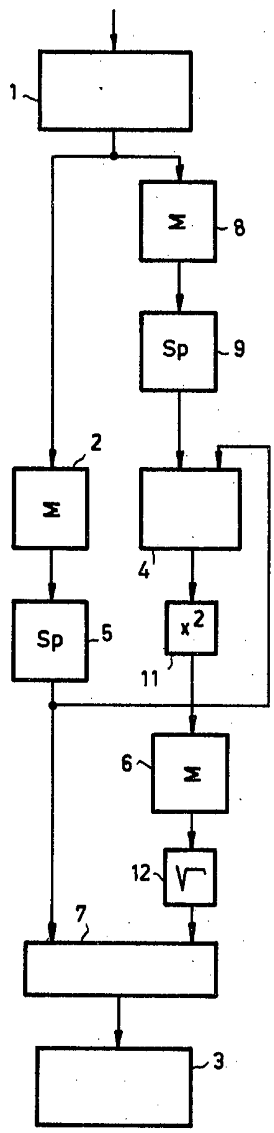

Fig.2