Description

Technical Field

This invention relates to determining whether or not speech contains a fundamental frequency which is commonly referred to as the unvoiced/voiced decision. More particularly, the unvoiced/voiced decision is made by a two stage voiced detector with the final threshold values being adaptively calculated for the speech environment utilizing statistical techniques.

Background of the Invention

In low bit rate voice coders, degradation of voice quality is often due to inaccurate voicing decisions. The difficulty in correctly making these voicing decisions lies in the fact that no single speech parameter or classifier can reliably distinguish voiced speech from unvoiced speech. In order to make the voice decision, it is known in the art to combine multiple speech classifiers in the form of a weighted sum. This method is commonly called discriminant analysis. Such a method is illustrated in D.P. Prezas, et. al., "Fast and Accurate Pitch Detection Using Pattern Recognition and Adaptive Time-Domain Analysis", Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc., Vol. 1, pp. 109-112, April 1986. As described in that article, a frame of speech is declared voice if a weighted sum of classifiers is greater than a specified threshold, and unvoiced otherwise. The weights and threshold are chosen to maximize performance on a training set of speech where the voicing of each frame is known.

A problem associated with the fixed weighted sum method is that it does not perform well when the speech environment changes. The reason is that the threshold is determined from the training set which is different from speech subject to background noise, non-linear distortion, and filtering.

B.S. Atal and L.R. Rabiner, in "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol ASSP-24, no.3, pages 201-212, June 1976, disclose a process for distinguishing between silence, voiced speech and unvoiced speech from a number of classifiers using a minimum-distance criterion in which the distance is defined by means of mean values and covariances of the classifiers for the three classes, determined from a manually classified set of training utterances.

P. de Souza, in "A Statistical Approach to the Design of an Adaptive Self-Normalizing Silence Detector", IEEE Trans. Acoust., Speech, Signal Processing, vol ASSP-31, no 3, pages 678-684, June 1983, discloses a process for detecting silence in which the first half-second of an input signal is assumed to be silent and means and covariances of the classifiers are calculated from this small sample. These are then used to detect, using a statistical test, a subsequent half-second of silence, which is then added to the original sample and the means and covariances updated, and so on until ten half-seconds of silence have been collected. The detector then returns to the start of the signal, using the means and covariances derived from the ten half-seconds of silence.

One method for adapting the threshold value to changing speech environment is disclosed in the paper of H. Hassanein, et al., "Implementation of the Gold-Rabiner Pitch Detector in a Real Time Environment Using an Improved Voicing Detector", IEEE Transactions on Acoustic, Speech and Signal Processing, 1986, Tokyo, Vol. ASSP-33, No. 1, pp.319-320. This paper discloses an empirical method which compares three different parameters against independent thresholds associated with these parameters and on the basis of each comparison either increments or decrements by one an adaptive threshold value. The three parameters utilized are energy of the signal, first reflection coefficient and zero-crossing count. For example, if the energy of the speech signal is less than a predefined energy level, the adaptive threshold is incremented. On the other hand, if the energy of the speech signal is greater than another predefined energy level, the adaptive threshold is decremented by one. After the adaptive threshold has been calculated, it is subtracted from a output of a elementary pitch detector. If the results of the subtraction yield a positive number, the speech frame is declared voice; otherwise, the speech frame is declared on unvoice. The problem with the disclosed method is that the parameters themselves are not used in the elementary pitch detector. Hence, the adjustment of the adaptive threshold is ad hoc and is not directly linked to the physical phenomena from which it is calculated. In addition, the threshold cannot adapt to rapidly changing speech environments.

Solution

The present invention accordingly provides a method and apparatus for making an adaptive

voiced/unvoiced determination for frames of speech as claimed in claims 1, 5 or 8.

The above described problem is solved and a technical advance is achieved by a voicing decision apparatus that adapts to a changing environment by utilizing adaptive statistical values to make the voicing decision. The statistical values are adapted to the changing environment by utilizing statistics based on an output of a voiced detector. The statistical parameters are calculated by the voiced detector generating a general value indicating the presence of a fundamental frequency in a speech frame in response to speech attributes of the frame. Second, the mean for unvoiced ones and voiced ones of speech frames is calculated in response to the generated value. The two means are then used to determine decision regions, and the determination of the presence of the fundamental frequency is done in response to the decision regions and the present speech frame.

Advantageously, in response to speech attributes of the present and past speech frames, the man for unvoiced frames is calculated by calculating the probability that the present speech frame is unvoiced, calculating the overall probability that any frame will be unvoiced, and calculating the probability that the present speech frame is voiced. The mean of the unvoiced speech frames is then calculated in response to the probability that the present speech frame is unvoiced and the overall probability. In addition, the mean of the voiced speech frame is calculated in response to the probability that the present speech frame is voiced and the overall probability. Advantageously, the calculations of probabilities are performed utilizing a maximum likelihood statistical operation.

Advantageously, the generation of the general value is performed utilizing a discriminant analysis procedure, and the speech attributes are speech classifiers.

Advantageously, the decision regions are defined by the mean of the unvoiced and voiced speech frames and a weight and threshold value generated in response to the general values of past and present frames and the means of the voiced and unvoiced frames.

The method for detecting the presence of a fundamental frequency in speech frames comprises the steps of: generating a general value in response to a set of classifiers defining speech attributes of a present speech frame to indicate the presence of the fundamental frequency, calculating a set of statistical parameters in response to the general value, and determining the presence of the fundamental frequency in response to the general value and the calculated set of statistical parameters. The step of generating the general value is performed utilizing a discriminant analysis procedure. Further, the step of determining the fundamental frequency comprises the step of calculating a weight and a threshold value in response to the set of parameters.

Brief Description of the Drawing

FIG. 1 illustrates, in block diagram form, the present invention; and

FIGS. 2 and 3 illustrate, in greater detail, certain functions performed by the voiced detection apparatus of FIG. 1.

Detailed Description

FIG. 1 illustrates an apparatus for performing the unvoiced/voiced decision operation by first utilizing a discriminant voiced detector to process voice classifiers in order to generate a discriminant variable or general variable. The latter variable is statistically analyzed to make the voicing decision. The statistical analysis adapts the threshold utilized in making the unvoiced/voiced decision so as to give reliable performance in a variety of voice environments.

Consider now the overall operation of the apparatus illustrated in FIG. 1. Classifier generator 100 is responsive to each frame of voice to generate classifiers which advantageously may be the log of the speech energy, the log of the LPC gain, the log area ratio of the first reflection coefficient, and the squared correlation coefficient of two speech segments one frame long which are offset by one pitch period. The calculation of these classifiers involves digitally sampling analog speech, forming frames of the digital samples, and processing those frames and is well known in the art. Generator 100 transmits the classifiers to silence detector 101 and discriminant voiced detector 102 via path 106. Discriminant voiced detector 102 is responsive to the classifiers received via path 106 to calculate the discriminant value, x. Detector 102 performs that calculation by solving the equation: $x = c'y + d$. Advantageously, "c" is a vector comprising the weights, "y" is a vector comprising the classifiers, and "d" is a scalar representing a threshold value. Advantageously, the components of vector c are initialized as follows: component corresponding to log of the speech energy equals 0.3918606, component corresponding to log of the LPC gain equals -0.0520902, component corresponding to log area ratio of the first reflection coefficient equals 0.5637082, and

component corresponding to squared correlation coefficient equals 1.361249; and d initially equals -8.36454. After calculating the value of the discriminant variable x, the detector 102 transmits this value via path 111 to statistical calculator 103 and subtracter 107.

Silence detector 101 is responsive to the classifiers transmitted via path 106 to determine whether speech is actually present on the data being received on path 109 by classifier generator 100. The indication of the presence of speech is transmitted via path 110 to statistical calculator 103 by silence detector 101.

For each frame of speech, detector 102 generates and transmits the discriminant value x via path 111. Statistical calculator 103 maintains an average of the discriminant values received via path 111 by averaging in the discriminant value for the present, non-silence frame with the discriminant values for previous non-silence frames. Statistical calculator 103 is also responsive to the signal received via path 110 to calculate the overall probability that any frame is unvoiced and the probability that any frame is voiced. In addition, statistical calculator 103 calculates the statistical value that the discriminant value for the present frame would have if the frame was unvoiced and the statistical value that the discriminant value for the present frame would have if the frame was voiced. Advantageously, that statistical value may be the mean. The calculations performed by calculator 103 are not only based on the present frame but on previous frames as well. Statistical calculator 103 performs these calculations not only on the basis of the discriminant value received for the present frame via path 106 and the average of the classifiers but also on the basis of a weight and a threshold value defining whether a frame is unvoiced or voiced received via path 113 from threshold calculator 104.

Calculator 104 is responsive to the probabilities and statistical values of the classifiers for the present frame as generated by calculator 103 and received via path 112 to recalculate the values used as weight value a, and threshold value b for the present frame. Then, these new values of a and b are transmitted back to statistical calculator 103 via path 113.

Calculator 104 transmits the weight, threshold, and statistical values via path 114 to U/V determinator 105. The latter detector is responsive to the information transmitted via paths 114 and 115 to determine whether or not the frame is unvoiced or voiced and to transmit this decision via path 116.

Consider now in greater detail the operations of blocks 103, 104, 105, and 107 illustrated in FIG. 1. Statistical calculator 103 implements an improved EM algorithm similar to that suggested in the article by N. E. Day entitled "Estimating the Components of a Mixture of Normal Distributions", Biometrika, Vol. 56, No. 3, pp. 463-474, 1969. Utilizing the concept of a decaying average, calculator 103 calculates the average for the discriminant values for the present and previous frames by calculating following equations 1, 2, and 3:

$$n = n+1 \text{ if } n < 2000 \quad (1)$$

$$z = 1/n \quad (2)$$

$$X_n = (1-z)\,X_{n-1} + zx_n \quad (3)$$

x_n is the discriminant value for the present frame and is received from detector 102 via path 111, and n is the number of frames that have been processed up to 2000. z represents the decaying average coefficient, and X_n represents the average of the discriminant values for the present and past frames. Statistical calculator 103 is responsive to receipt of the z, x_n and X_n values to calculate the variance value, T, by first calculating the second moment of x_n, Q_n, as follows:

$$Q_n = (1-z)Q_{n-1} + zx_n^2 . \quad (4)$$

After Q_n has been calculated, T is calculated as follows:

$$T = Q_n - X_n^2 . \quad (5)$$

The mean is subtracted from the discriminant value of the present frame as follows:

$x_n = x_n - X_n$ (6)

Next, calculator 103 determines the probability that the frame represented by the present value x_n is unvoiced by solving equation 7 shown below:

$$P(u|x_n) = \frac{1}{1 + \exp(ax_n+b)} . \quad (7)$$

After solving equation 7, calculator 103 determines the probability that the discriminant value represents a voiced frame by solving the following:

$P(v|x_n) = 1-P(u|x_n)$. (8)

Next, calculator 103 determines the overall probability that any frame will be unvoiced by solving equation 9 for p_n:

$p_n = (1-z)\, p_{n-1} + z\, P(u|x_n)$. (9)

After determining the probability that a frame will be unvoiced, calculator 103 determines two values, u and v, which give the mean values of discriminant value for both unvoiced and voiced type frames. Value u, statistical average unvoiced value, contains the mean discriminant value if a frame is unvoiced, and value v, statistical average voiced value, gives the mean discriminant value if a frame is voiced. Value u for the present frame is solved by calculating equation 10, and value v is determined for the present frame by calculating equation 11 as follows:

$u_n = (1-z)\, u_{n-1} + z\, x_n\, P(u|x_n)/p_n - zx_n$ (10)

$v_n = (1-z)\, v_{n-1} + z\, x_n\, P(v|x_n)/(1-p_n) - zx_n$ (11)

Calculator 103 now communicates the u, v, and T values, and probability p_n to threshold calculator 104 via path 112.

Calculator 104 is responsive to this information to calculate new values for a and b. These new values are then transmitted back to statistical calculator 103 via path 113. This allows rapid adaptations to changing environments. If n is greater than advantageously 99, values a and b are calculated as follows. Value a is determined by solving the following equation:

$$a = T^{-1} \frac{(v_n-u_n)}{1-p_n(1-p_n)\, T^{-1}\, (u_n-v_n)^2} . \quad (12)$$

Value b is determined by solving the following equation:

$b = -\frac{1}{2}\, a(u_n + v_n) + \log[(1-p_n)/p_n]$. (13)

After calculating equations 12 and 13, calculator 104 transmits values a, u, and v to block 105 via path 114.

Determinator 105 is responsive to this transmitted information to decide whether the present frame is voiced or unvoiced. If the value a is positive, then, a frame is declared voiced if the following equation is true:

$ax_n - a(u_n + v_n)/2 > 0$; (14)

or if the value a is negative, then, a frame is declared voiced if the following equation is true:

$ax_n - a(u_n + v_n)/2 < 0$. (15)

Equation 14 can also be expressed as:

$$ax_n + b - \log[(1-p_n)/p_n] > 0 .$$

Equation 15 can also be expressed as:

$$ax_n + b - \log[(1-p_n)/p_n] < 0 .$$

If the previous conditions are not met, determinator 105 declares the frame unvoiced.

In flow chart form, FIGS. 2 and 3 illustrate, in greater detail, the operations performed by the apparatus of FIG. 1. Block 200 implements block 101 of FIG. 1. Blocks 202 through 218 implement statistical calculator 103. Block 222 implements threshold calculator 104, and blocks 226 through 239 implement block 105 of FIG. 1. Subtracter 107 is implemented by both block 208 and block 224. Block 202 calculates the value which represents the average of the discriminant value for the present frame and all previous frames. Block 200 determines whether speech is present in the present frame; and if speech is not present in the present frame, the mean for the discriminant value is subtracted from the present discriminant value by block 224 before control is transferred to decision block 226.

However, if speech is present in the present frame, then the statistical and weight calculations are performed by blocks 202 through 222. First, the average value is found in block 202. Second, the second moment value is calculated in block 206. The latter value along with the mean value X for the present and past frames is then utilized to calculate the variance, T, also in block 206. The mean X is then subtracted from the discriminant value x_n in block 208.

Block 210 calculates the probability that the present frame is unvoiced by utilizing the present weight value a, the present threshold value b, and the discriminant value for the present frame, x_n. After calculating the probability that the present frame is unvoiced, the probability that the present frame is voiced is calculated by block 212. Then, the overall probability, p_n, that any frame will be unvoiced is calculated by block 214.

Blocks 216 and 218 calculate two values: u and v. The value u represents the statistical average value that the discriminant value would have if the frame were unvoiced. Whereas, value v represents the statistical average value that the discriminant value would have if the frame were voiced. The actual discriminant values for the present and previous frames are clustered around either value u or value v. The discriminant values for the previous and present frames are clustered around value u if these frames had been found to be unvoiced; otherwise, the previous values are clustered around value v. Block 222 then calculates a new weight value a and a new threshold value b. The values a and b are used in the next sequential frame by the preceding blocks in FIG. 2.

Blocks 226 through 239 implement U/V determinator 105 of FIG. 1. Block 226 determines whether the value a for the present frame is greater than zero. If this condition is true, then decision block 228 is executed. The latter decision block determines whether the test for voiced or unvoiced is met. If the frame is found to be voiced in decision block 228, then the frame is so marked as voiced by block 230 otherwise the frame is marked as unvoiced by block 232. If the value a is less than zero for the present frame, blocks 234 through 238 are executed and function in a similar manner to blocks 228 through 232.

Claims

1. An apparatus for making a voiced/unvoiced determination for frames of speech of unknown voicing, comprising:

means (101) for detecting silence to select frames of speech of unknown voicing;

means (102) responsive to a set of classifiers from a classifier generator (100) defining speech attributes of one of said frames of speech of unknown voicing for generating a general value initially indicating voiced or unvoiced

means (103) responsive to said general value for calculating a set of statistical parameters; and

means (104) for calculating a threshold value in response to said set of said parameters;

means (104) for calculating a weight value in response to said set of said parameters;

means (105) responsive to said weight value and said threshold value and the calculated set of statistical parameters for determining voiced/unvoiced in said present one of said frames of speech of unknown voicing;

Characterized by means (113) for communicating said weight value and said threshold value to

* C L M S *

Equation 14 can also be expressed as:

$$ax_n + b - \log[(1-p_n)/p_n] > 0 .$$

Equation 15 can also be expressed as:

$$ax_n + b - \log[(1-p_n)/p_n] < 0 .$$

If the previous conditions are not met, determinator 105 declares the frame unvoiced.

In flow chart form, FIGS. 2 and 3 illustrate, in greater detail, the operations performed by the apparatus of FIG. 1. Block 200 implements block 101 of FIG. 1. Blocks 202 through 218 implement statistical calculator 103. Block 222 implements threshold calculator 104, and blocks 226 through 239 implement block 105 of FIG. 1. Subtracter 107 is implemented by both block 208 and block 224. Block 202 calculates the value which represents the average of the discriminant value for the present frame and all previous frames. Block 200 determines whether speech is present in the present frame; and if speech is not present in the present frame, the mean for the discriminant value is subtracted from the present discriminant value by block 224 before control is transferred to decision block 226.

However, if speech is present in the present frame, then the statistical and weight calculations are performed by blocks 202 through 222. First, the average value is found in block 202. Second, the second moment value is calculated in block 206. The latter value along with the mean value X for the present and past frames is then utilized to calculate the variance, T, also in block 206. The mean X is then subtracted from the discriminant value x_n in block 208.

Block 210 calculates the probability that the present frame is unvoiced by utilizing the present weight value a, the present threshold value b, and the discriminant value for the present frame, x_n. After calculating the probability that the present frame is unvoiced, the probability that the present frame is voiced is calculated by block 212. Then, the overall probability, p_n, that any frame will be unvoiced is calculated by block 214.

Blocks 216 and 218 calculate two values: u and v. The value u represents the statistical average value that the discriminant value would have if the frame were unvoiced. Whereas, value v represents the statistical average value that the discriminant value would have if the frame were voiced. The actual discriminant values for the present and previous frames are clustered around either value u or value v. The discriminant values for the previous and present frames are clustered around value u if these frames had been found to be unvoiced; otherwise, the previous values are clustered around value v. Block 222 then calculates a new weight value a and a new threshold value b. The values a and b are used in the next sequential frame by the preceding blocks in FIG. 2.

Blocks 226 through 239 implement U/V determinator 105 of FIG. 1. Block 226 determines whether the value a for the present frame is greater than zero. If this condition is true, then decision block 228 is executed. The latter decision block determines whether the test for voiced or unvoiced is met. If the frame is found to be voiced in decision block 228, then the frame is so marked as voiced by block 230 otherwise the frame is marked as unvoiced by block 232. If the value a is less than zero for the present frame, blocks 234 through 238 are executed and function in a similar manner to blocks 228 through 232.

Claims

1. An apparatus for making a voiced/unvoiced determination for frames of speech of unknown voicing, comprising:

means (101) for detecting silence to select frames of speech of unknown voicing;

means (102) responsive to a set of classifiers from a classifier generator (100) defining speech attributes of one of said frames of speech of unknown voicing for generating a general value initially indicating voiced or unvoiced

means (103) responsive to said general value for calculating a set of statistical parameters; and

means (104) for calculating a threshold value in response to said set of said parameters;

means (104) for calculating a weight value in response to said set of said parameters;

means (105) responsive to said weight value and said threshold value and the calculated set of statistical parameters for determining voiced/unvoiced in said present one of said frames of speech of unknown voicing;

Characterized by means (113) for communicating said weight value and said threshold value to

said means (103) for calculating said set of parameters to be used for calculating another set of parameters for a subsequent one of said frames of speech of unknown voicing.

2. The apparatus of claim 1 wherein said generating means (102) comprises means for performing a discriminant analysis to generate said general value.

3. The apparatus of claim 2 wherein said means (104) for calculating said set of parameters is further responsive to the communicated weight and threshold values and another general value of said other one of said frames for calculating another set of statistical parameters.

4. The apparatus of claim 3 wherein said means (104) for calculating said set of parameters further comprises means for calculating the average of said general values over said present and previous ones of said speech frames; and

means responsive to said average of said general values for said present and previous ones of said speech frames and said communicated weight and threshold values and said other general value for determining said other set of statistical parameters.

5. An apparatus for making a voiced/unvoiced determination for frames of speech, comprising:

means (101) for detecting silence to select frames of speech;

means (102) responsive to a set of classifiers from a classifier generator (100) defining speech attributes of each of a present and past ones said frames of speech for generating a general value initially indicating voiced or unvoiced;

means for calculating (206) the variance of said general values over said present and previous ones of said speech frames;

means responsive to present and past ones of said frames for calculating (210) the probability that said present one of said frames is unvoiced;

means for calculating (212) the probability that said present one of said frames is voiced;

Characterized by means responsive to said present and past ones of said frames and said probability that said present one of frames is unvoiced for calculating (214) the overall probability that any frame will be unvoiced;

means responsive to said probability that said present one of said frames is unvoiced and said overall probability and said variance for calculating (216) a mean of said unvoiced ones of said frames;

means responsive to said probability that said present one of said frames is voiced and said overall probability and said variance for calculating (218) a mean of said voiced ones of said frames;

means responsive to said mean for unvoiced ones of said frames and said mean of voiced ones of said frames and said variance for determining (222) decision regions; and

means (105) for making the voiced/unvoiced determination in response to said decision regions for said present one of said frames.

6. The apparatus of claim 5 wherein said means for calculating said probability that said present one of said frames is unvoiced performs a maximum likelihood statistical operation.

7. The apparatus of claim 6 wherein said means for calculating said probability that said present one of said frames is unvoiced further responsive to a weight threshold value to perform said maximum likelihood statistical operation.

8. A method for making a voiced/unvoiced determination for frames of speech of unknown voicing comprising the steps of:

detecting silence (200) in order to select frames of speech of unknown voicing;

generating a general value in response to a set of classifiers from a classifier generator defining speech attributes of one of said frames of speech of unknown voicing to initially indicate voiced/unvoiced determination

calculating (103) a set of statistical parameters in response to said general value; and

calculating (104) a threshold value in response to said set of said parameters;

calculating (104) a weight value in response to said set of said parameters; and

determining (105) voice/unvoiced speech in said present one of said frames of speech of unknown voicing in response to said weight value and said threshold value and the calculated set of statistical parameters;

* C L M S *

Characterized by feeding back (113) said weight value and said threshold value for calculating another set of parameters for a subsequent one of said frames of speech.

9. The method of claim 8 wherein said step of generating comprises the step of performing a discriminant analysis to generate said general value.

10. The method of claim 9 wherein said step of calculating said set of parameters further responsive to the communicated weight and threshold values and another general value of said other one of said frames for calculating another set of statistical parameters.

11. The method of claim 10 wherein said step of calculating said set of parameters further comprises the steps of calculating the average of said general values over said present and previous ones of said speech frames; and

determining said other set of statistical parameters in response to said average of said general values for said present and previous ones of said speech frames and said communicated weight and threshold values and said other general values.

Patentansprüche

1. Vorrichtung zum Treffen einer Stimmhaft/Stimmlos-Entscheidung für Sprachrahmen unbekannter Stimmhaftigkeit, mit:

Mitteln (101) zum Erkennen von Stille, um Sprachranmen unbekannter Stimmhaftigkeit auszuwählen;

Mitteln (102), die auf einen Satz von Klassifizierern von einem Klassifizierer-Generator (100) ansprechen, um Sprachattribute eines dieser Sprachrahmen unbekannter Stimmhaftigkeit zu definieren, um einen allgemeinen Wert zu erzeugen, der anfänglich "Stimmhaft" oder "stimmlos" angibt;

Mitteln (103), die auf den allgemeinen Wert ansprechen, um einen Satz von statistischen Parametern zu berechnen;

Mitteln (104) zum Berechnen eines Schwellenwertes abhängig von diesem Parametersatz;

Mitteln (104) zum Berechnen eines Gewichtungswertes abhängig von diesem Parametersatz; und

Mitteln (105), die auf den Gewichtungswert und den Schwellenwert sowie den berechneten Satz von statistischen Parametern ansprechen, um die Entscheidung stimmhaft/stimmlos für den vorliegenden der Sprachrahmen unbekannter Stimmhaftigkeit zu treffen;

gekennzeichnet durch Mittel (113) zum Mitteilen des Gewichtungswertes und des Schwellenwertes an die Mittel (103) zur Berechnung des Parametersatzes, der zur Berechnung eines weiteren Parametersatzes für einen folgenden der Sprachrahmen unbekannter Stimmhaftigkeit dient.

2. Vorrichtung nach Anspruch 1, bei welcher die Erzeugungsmittel (102) Mittel zum Durchführen einer Diskriminanzanalyse aufweisen, um den allgemeinen Wert zu erzeugen.

3. Vorrichtung nach Anspruch 2, bei welcher die Mittel (104) zum Berechnen des Parametersatzes ferner auf die mitgeteilten Gewichtungs- und Schwellenwerte sowie einen weiteren allgemeinen Wert jenes weiteren Rahmens ansprechen, um einen weiteren Satz von statistischen Parametern zu berechnen.

4. Vorrichtung nach Anspruch 3, bei welcher die Mittel (104) zum Berechnen des Parametersatzes ferner Mittel zum Berechnen des Durchschnittes der allgemeinen Werte des vorliegenden und vorhergehender Sprachrahmen aufweisen; und

Mittel, die auf den Durchschnitt der allgemeinen Werte für den vorliegenden und vorhergehende Sprachrahmen sowie die mitgeteilten Gewichtungs- und Schwellenwerte und jenen weiteren allgemeinen Wert ansprechen, um diesen weiteren Satz von statistischen Parametern zu berechnen.

5. Vorrichtung zum Treffen einer Stimmhaft/Stimmlos-Entscheidung für Sprachrahmen, mit:

Mitteln (101) zum Erkennen von Stille, um Sprachrahmen auszuwählen;

Mitteln (102), die auf einen Satz von Klassifizierern von einem Klassifizierer-Generator (100) ansprechen, um Sprachattribute für den vorliegenden und vorhergehende Sprachranmen zu definieren, um einen allgemeinen Wert zu erzeugen, der anfänglich "stimmhaft" oder "stimmlos" angibt;

Mittel zum Berechnen (206) der Varianz der allgemeinen Werte des vorliegenden und vorhergehender Sprachrahmen;

* C L M S *

Mittel, die auf den vorliegenden und vergangene Rahmen ansprechen, um die Wahrscheinlichkeit zu berechnen (210), daß der vorliegende der Rahmen stimmlos ist;

Mittel zum Berechnen (212) der Wahrscheinlichkeit, daß der vorliegende der Rahmen stimmhaft ist;

gekennzeichnet durch Mittel, die auf den vorliegenden und vergangene Rahmen sowie die Wahrscheinlichkeit ansprechen, daß der vorliegende Rahmen stimmlos ist, um die Gesamtwahrscheinlichkeit zu berechnen (214), daß irgendein Rahmen stimmlos ist;

Mittel, die auf diese Wahrscheinlichkeit ansprechen, daß der vorliegende Rahmen stimmlos ist, sowie auf die Gesamtwanrscheinlichkeit und die Varianz, um einen Mittelwert für diesen stimmlosen Rahmen zu berechnen (216);

Mittel, die auf diese Wahrscheinlichkeit ansprechen, daß der vorliegende Rahmen stimmhaft ist, sowie auf die Gesamtwahrscheinlichkeit und die Varianz, um einen Mittelwert für diesen stimmhaften Rahmen zu berechnen (218);

Mittel, die auf den Mittelwert der stimmlosen Rahmen und den Mittelwert der stimmhaften Rahmen sowie auf die Varianz ansprechen, um Entscheidungsbereiche zu bestimmen (222); und

Mittel (105), um die Stimmhaft/Stimmlos-Entscheidung abhängig von den Entscheidungsbereichen für den vorliegenden Rahmen zu treffen.

6. Vorrichtung nach Anspruch 6, bei welcher die Mittel zum Berechnen der Wahrscheinlichkeit, daß der vorliegende Rahmen stimmlos ist, eine statistische Maximalwahrscheinlichkeitsoperation durchführen.

7. Vorrichtung nach Anspruch 6, bei welcher die Mittel zum Berechnen der Wahrscheinlichkeit, daß der vorliegende Rahmen stimmlos ist, ferner auf einen Gewichtungs-Schwellenwert ansprechen, um die statistische Maximalwahrscheinlichkeitsoperation durchzuführen.

8. Verfahren zum Treffen einer Stimmhaft/Stimmlos-Entscheidung für Sprachrahmen unbekannter Stimmhaftigkeit, mit den Schritten:

Erkennen von Stille (200), um Sprachrahmen unbekannter Stimmhaftigkeit auszuwählen;

Erzeugen eines allgemeinen Wertes abhängig von einem Satz von Klassifizierern aus einem Klassifizierer-Generator, welche Sprachattribute für einen Sprachrahmen unbekannter Stimmhaftigkeit definieren, um anfänglich eine Stimmhaft/Stimmlos-Entscheidung anzugeben;

Berechnen (103) eines Satzes von statistischen Parametern abhängig von diesem allgemeinen Wert; und

Berechnen (104) eines Schwellenwertes abhängig von diesem Parametersatz;

Berechnen (104) eines Gewichtungswertes abhängig von diesem Parametersatz; und

Bestimmen (105) von stimmhafter/stimmloser Sprache im vorliegenden Sprachrahmen unbekannter Stimmhaftigkeit abhängig vom Gewichtungswert, vom Schwellenwert und dem berechneten Satz statistischer Parameter;

gekennzeichnet durch Zurückführen (113) des Gewichtungswertes und des Schwellenwertes zum Berechnen eines weiteren Parametersatzes für einen folgenden Sprachrahmen.

9. Verfahren nach Anspruch 8, bei welchem der Schritt des Erzeugens den Schritt des Durchführens einer Diskriminanzanalyse umfaßt, um den allgemeinen Wert zu erzeugen.

10. Verfahren nach Anspruch 9, bei welchem der Schritt des Berechnens des Parametersatzes ferner auf die mitgeteilten Gewichtungs- und Schwellenwerte sowie einen weiteren allgemeinen Wert jenes weiteren Rahmens anspricht, um einen weiteren Satz statistischer Parameter zu berechnen.

11. Verfahren nach Anspruch 10, bei welchem der Schritt des Berechnens des Parametersatzes ferner die Schritte des Berechnens des Durchschnittes der allgemeinen Werte des vorliegenden und vorhergehender Sprachrahmen; und

des Bestimmens des weiteren Satzes statistischer Parameter abhängig vom Durchschnitt der allgemeinen Werte für den vorliegenden und vorhergehende Sprachrahmen sowie von den mitgeteilten Gewichtungswerten, Schwellenwerten und diesen weiteren allgemeinen Werten umfaßt.

Revendications

1. Un dispositif destiné à effectuer une détermination d'état voisé/non voisé pour des trames de parole dont le voisement est inconnu, comprenant :

des moyens (101) destinés à détecter une condition de silence pour sélectionner des trames de parole de voisement inconnu;

des moyens (102) qui réagissent à un ensemble de classificateurs provenant d'un générateur de classificateurs (100) définissant des attributs de parole de l'une des trames de parole de voisement inconnu, de façon à générer une valeur générale indiquant initialement un état voisé ou non voisé;

des moyens (103) oui réagissent à la valeur générale précitée en calculant un ensemble de paramètres statistiques; et

des moyens (104) qui sont destinés à calculer une valeur de seuil sous la dépendance de l'ensemble de paramètres;

des moyens (104) qui dont destinés à calculer une valeur de poids sous la dépendance de l'ensemble de paramètres;

des moyens (105) qui réagissent à la valeur de poids et à la valeur de seuil et à l'ensemble calculé de paramètres statistiques, en déterminant un état voisé/non voisé dans la trame présente parmi les trames de parole de voisement inconnu;

caractérisé par des moyens (113) qui sont destinés à communiquer la valeur de poids et la valeur de seuil aux moyens (103) destinés à calculer l'ensemble de paramètres, en vue de leur utilisation pour le calcul d'un autre ensemble de paramètres pour l'une suivante des trames de parole de voisement inconnu.

2. Le dispositif de la revendication 1, dans lequel les moyens de génération (102) comprennent des moyens qui sont destinés à effectuer une analyse par discriminant pour générer la valeur générale.

3. Le dispositif de la revendication 2, dans lequel les moyens (104) destinés à calculer l'ensemble de paramètres réagissent en outre aux valeurs de poids et de seuil communiquées, et à une autre valeur générale de l'autre trame parmi les trames précitées, en calculant un autre ensemble de paramètres statistiques.

4. Le dispositif de la revendication 3, dans lequel les moyens (104) destinés à calculer l'ensemble de paramètres, comprennent en outre des moyens qui sont destinés à calculer la moyenne des valeurs générales sur des trames présente et précédentes parmi les trames de parole; et

des moyens qui réagissent à la moyenne des valeurs générales pour la trame présente et des trames précédentes parmi les trames de parole, et aux valeurs de poids et de seuil communiquées et à l'autre valeur générale, de façon à déterminer l'autre ensemble précité de paramètres statistiques.

5. Un dispositif destiné à effectuer une détermination d'état voisé/non voisé pour des trames de parole, comprenant :

des moyens (101) qui sont destinés à détecter une condition de silence pour sélectionner des trames de parole;

des moyens (102) qui réagissent à un ensemble de classificateurs provenant d'un générateur de classificateurs (100) définissant des attributs de parole de chaque trame parmi des trames de parole présente et passées, de façon à générer une valeur générale indiquant initialement un état voisé ou non voisé;

des moyens (206) destinés à calculer la variance des valeurs généralessur les trames de parole présente et précédentes;

des moyens qui réagissent aux trames présente et passées en calculant (210) la probabilité que la trame présente parmi les trames précitées soit non voisée;

des moyens (212) destinés à calculer la probabilité que la trame présente parmi les trames précitées soit voisée;

caractérisé par des moyens qui réagissent à la trame présente et à des trames passées parmi les trames précitées, et à la probabilité que la trame présente soit non voisée, en calculant (214) la probabilité globale qu'une trame quelconque soit non voisée;

des moyens qui réagissent à la probabilité que la trame présente parmi les trames précitées soit non voisée, et à la probabilité globale et à la variance, en calculant (216) une moyenne des trames non voisées;

des moyens qui réagissent à la probabilité précitée que la trame présente parmi les trames précitées soit voisée, et à la probabilité globale et à la variance, en calculant (218) une moyenne des trames voisées parmi les trames précitées;

des moyens qui réagissent à la moyenne pour des trames non voisées parmi les trames précitées

* C L M S *

et à la moyenne des trames voisées parmi les trames précitées et à la variance, en déterminant (222) des régions de décision; et

des moyens (105) qui sont destinés à effectuer la détermination d'état voisé/non voisé sous la dépendance des régions de décision pour la trame présente parmi les trames précitées.

6. Le dispositif de la revendication 5, dans lequel les moyens destinés à calculer la probabilité que la trame présente parmi les trames précitées soit non voisée, effectuent une opération statistique du type à vraisemblance maximale.

7. Le dispositif de la revendication 6, dans lequel les moyens qui sont destinés à calculer la probabilité que la trame présente parmi les trames précitées soit non voisée, réagissent en outre à des valeurs de poids et de seuil en effectuant l'opération statistique du type à vraisemblance maximale.

8. Un procédé pour effectuer une détermination d'état voisé/non voisé pour des trames de parole dont le voisement est inconnu, comprenant les étapes suivantes :

on détecte une condition de silence (200) dans le but de sélectionner des trames de parole de voisement inconnu;

on génère une valeur générale sous la dépendance d'un ensemble de classificateurs, provenant d'un générateur de classificateurs, définissant des attributs de parole de l'une des trames de parole de voisement inconnu, pour indiquer initialement une détermination d'état voisé/non voisé;

on calcule (103) un ensemble de paramètres statistiques sous la dépendance de la valeur générale précitée;

on calcule (104) une valeur de seuil sous la dépendance de l'ensemble de paramètres;

on calcule (104) une valeur de poids sous la dependance de l'ensemble de paramètres; et

on détermine (105) l'état voisé/non voisé de la parole dans la trame présente parmi les trames de parole de voisement inconnu, sous la dépendance de la valeur de poids et de la valeur de seuil, ainsi que de l'ensemble calculé de paramètres statistiques;

caractérisé en ce qu'on renvoie (113) la valeur de poids et la valeur de seuil pour calculer un autre ensemble de paramètres pour l'une suivante des trames de parole.

9. Le procédé de la revendication 8, dans lequel l'étape de génération comprend l'étape qui consiste à effectuer une analyse par discriminant pour générer la valeur générale précitée.

10. Le procédé de la revendication 9, dans lequel l'étape de calcul de l'ensemble de paramètres réagit en outre aux valeurs de poids et de seuil communiquées, et à une autre valeur générale de l'autre des trames précitées, de façon à calculer un autre ensemble de paramètres statistiques.

11. Le procédé de la revendication 10, dans lequel l'étape de calcul de l'ensemble de paramètres comprend en outre les étapes qui consistent à calculer la valeur moyenne des valeurs générales précitées sur la trame présente et des trames précédentes parmi les trames de parole précitées; et

à déterminer l'autre ensemble précité de paramètres statistiques sous la dépendance de la moyenne des valeurs générales pour la trame présente et des trames précédentes parmi les trames de parole précitées, et sous la dépendance des valeurs de poids et de seuil communiquées et des autres valeurs générales précitées.