Description

Technical Field

This invention relates to determining whether or not speech has a fundamental frequency present. This is also referred to as a voicing decision. More particularly, the invention is directed to selecting one of a plurality of voiced detectors which are concurrently processing speech samples for making the voicing decision with the selection being based on a distance measurement calculation.

Background and Problem

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 classifier can reliably distinguish voiced speech from unvoiced speech. The use of multiple voiced detectors and the selection of one of these detectors to make the determination of whether the speech is voiced or unvoiced is disclosed in the paper of J.P. Campbell, et al., "Voiced/Unvoiced Classification of Speech with Applications to the U.S Government LPC-10E Algorithm," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1986, Tokyo, Vol. 9. 11.4, pp. 473-476. This paper discloses the utilization of multiple linear discriminant voiced detectors each utilizing different weights and threshold values to process the same speech classifiers for each frame of speech. The weights and thresholds for each detector are determined by utilizing training data. For each detector, a different level of white noise is added to the training data. During the processing of actual speech, the detector to be utilized to make the voicing decision is determined by examining the signal-to-noise ratio, SNR. The range of possible values that the SNR can have is subdivided into subranges with each subrange being assigned to one of the detectors. For each frame, the SNR is calculated, the subrange is determined, and the detector associated with this subrange is selected to make the voicing decision.

In "A pattern recognition approach to voiced-un-voiced-silence classification with applications to speech recognition" (IEEE Transactions on Acoustics, Speech and Signal Processing, vol ASSP-24, No. 3, June 1976, pages 201 -212) Atal and Rabiner disclose a system in which various measures (zero crossings, log energy, autocorrelation, first linear prediction coefficient and normalised prediction error) are combined, using predetermined mean values and covariances for the three classes of signal (voiced, unvoiced and silence) to produce three distances, one for each class. The class with the smallest distance is then assigned to the signal.

A problem with the prior art approach is that it does not perform well with respect to a speech environment in which characteristics of the speech itself have been altered. In addition, the method used by Campbell is only adapted to white noise and cannot adjust for coloured noise. Therefore, there exists a need for a method of selecting between a plurality of voiced detectors that allows detection in a varying speech environment.

Summary of the Invention

According to the invention there is provided apparatus as claimed in claim 1 and a method as claimed in claim 17. Preferred forms of the invention are set out in the dependent claims.

Brief Description of the Drawing

The invention may be better understood from the following detailed description which when read with reference to the drawing in which:

FIG. 1 is a block diagram illustrating the present invention;

FIG. 2 illustrates, in block diagram form, statistical voice detector 103 of FIG. 1;

FIGS. 3 and 4 illustrate, in greater detail, the functions performed by statistical voiced detector 103 of FIG. 2; and

FIG. 5 illustrates, in greater detail, functions performed by block 340 of FIG. 4.

Detailed Description

FIG. 1 illustrates an apparatus for performing the unvoiced/voiced decision operation by selecting between one of two voiced detectors. It would obvious to one skilled in the art to use more than two voiced detectors in FIG. 1. The selection between detectors 102 and 103 is based on a distance measurement that is generated by each detector and transmitted to distance comparator 104. Each generated distance measurement represents a merit value indicating the correctness of the generating detector's voicing decision. Distance comparator 104 compares the two distance measurement values and controls a multiplexer 105 such that the detector generating the greatest distance measurement value is selected to make the unvoiced/voiced decision. However, for other types of measurements, the lowest merit value would indicate the detector making the most accurate voicing decision. Advantageously, the distance measurement may be the Mahalanobis distance. Advantageously, detector 102 is a discriminant detector, and detector 103 is a statistical detector. However, it would be obvious to one skilled in the art that the detectors could all be of the same type and that there could be more than two detectors present

in the system.

Consider now the overall operation of the apparatus illustrated in FIG. 1. Classifier generator 101 is responsive to each frame of speech 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 101 transmits the classifiers to detectors 102 and 103 via path 106.

Detectors 102 and 103 are responsive to the classifiers received via path 106 to make unvoiced/voiced decisions and transmit these decisions via paths 107 and 110, respectively, to multiplexer 105. In addition, the detectors determine a distance measure between voiced and unvoiced frames and transmit these distances via paths 108 and 109 to comparator 104. Advantageously, these distances may be Mahalanobis distances or other generalized distances. Comparator 104 is responsive to the distances received via paths 108 and 109 to control multiplexer 105 so that the latter multiplexer selects the output of the detector that is generating the largest distance.

FIG. 2 illustrates, in greater detail, statistical voiced detector 103. For each frame of speech, a set of classifiers also referred to as a vector of classifiers is received via path 106 from classifier generator 101. Silence detector 201 is responsive to these classifiers to determine whether or not speech is present in the present frame. If speech is present, detector 201 transmits a signal via path 210. If no speech (silence) is present in the frame, then only subtractor 207 and U/V determinator 205 are operational for that particular frame. Whether speech is present or not, the unvoiced/voiced decision is made for every frame by determinator 205.

In response to the signal from detector 201, classifier averager 202 maintains an average of the individual classifiers received via path 106 by averaging in the classifiers for the present frame with the classifiers for previous frames. If speech (non-silence) is present in the frame, silence detector 201 signals statistical calculator 203, generator 206, and averager 202 via path 210.

Statistical calculator 203 calculates statistical distributions for voiced and unvoiced frames. In particular, calculator 203 is responsive to the signal received via path 210 to calculate the overall probability that any frame is unvoiced and the probability that any frame is voiced. In addition, statistical calculator 203 calculates the statistical value that each classifier would have if the frame was unvoiced and the statistical value that each classifier would have if the frame was voiced. Further, calculator 203 calculates the covariance matrix of the classifiers. Advantageously, that statistical value may be the mean. The calculations performed by calculator 203 are not only based on the present frame but on previous frames as well. Statistical calculator 203 performs these calculations not only on the basis of the classifiers received for the present frame via path 106 and the average of the classifiers received path 211 but also on the basis of the weight for each classifiers and a threshold value defining whether a frame is unvoiced or voiced received via path 213 from weights calculator 204.

Weights calculator 204 is responsive to the probabilities, covariance matrix, and statistical values of the classifiers for the present frame as generated by calculator 203 and received via path 212 to recalculate the values used as weight vector a, for each of the classifiers and the threshold value b, for the present frame. Then, these new values of a and b are transmitted back to statistical calculator 203 via path 213.

Also, weights calculator 204 transmits the weights and the statistical values for the classifiers in both the unvoiced and voiced regions via path 214, determinator 205, and path 208 to generator 206. The latter generator is responsive to This information to calculate the distance measure which is subsequently transmitted via path 109 to comparator 104 as illustrated in FIG. 1.

U/V determinator 205 is responsive to the information transmitted via paths 214 and 215 to determine whether or not the frame is unvoiced or voiced and to transmit this decision via path 110 to multiplexer 105 of FIG. 1.

Consider now in greater detail the operation of each block illustrated in FIG. 2 which is now given in terms of vector and matrix mathematics. Averager 202, statistical calculator 203, and weights calculator 204 implement 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, classifier averager 202 calculates the average for the classifiers 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 a vector representing the classifiers for the present frame, 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 classifiers over the present and past frames. Statistical calculator 203 is responsive to

receipt of the z, x_n and X_n information to calculate the covariance matrix, T, by first calculating the matrix of sums of squares and products, Q_n, as follows:

$$Q_n = (1-z)\,Q_{n-1} + z\,x_n\,x'_n. \quad (4)$$

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

$$T = Q_n - X_n\,X'_n. \quad (5)$$

The means are subtracted from the classifiers as follows:

$$x_n = x_n - X_n \quad (6)$$

Next, calculator 203 determines the probability that the frame represented by the present vector x_n is unvoiced by solving equation 7 shown below where, advantageously, the components of vector a 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 b initially equals -8.36454:

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

After solving equation 7, calculator 203 determines the probability that the classifiers represent a voiced frame by solving the following:

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

Next, calculator 203 determines the overall probability that any frame will be unvoiced by solving equation 9 for p_n:

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

After determining the probability that a frame will be unvoiced, calculator 203 then determines two vectors, u and v, which give the mean values of each classifier for both unvoiced and voiced type frames. Vectors u and v are the statistical averages for unvoiced and voiced frames, respectively. Vector u, statistical average unvoiced vector, contains the mean values of each classifier if a frame is unvoiced; and vector v, statistical average voiced vector, gives the mean value for each classifier if a frame is voiced. Vector u for the present frame is solved by calculating equation 10, and vector v is determined for the present frame by calculating equation 11 as follows:

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

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

Calculator 203 now communicates the u and v vectors, T matrix, and probability p to weights calculator 204 via path 212.

Weights calculator 204 is responsive to this information to calculate new values for vector a and scalar b. These new values are then transmitted back to statistical calculator 203 via path 213. This allows detector 103 to adapt rapidly to changing environments. Advantageously, if the new values for vector a and scalar b are not transmitted back to statistical calculator 203, detector 103 will continue to adapt to changing environments since vectors u and v are being updated. As will be seen, determinator 205 uses vectors u and v as well as vector a and scalar b to make the voicing decision. If n is greater than advantageously 99, vector a and scalar b are calculated as follows. Vector a is determined by solving the following equation:

$$a = \frac{T^{-1}(v_n-u_n)}{1-p_n(1-p_n)(u_n-v_n)'\,T^{-1}(u_n-v_n)}. \quad (12)$$

Scalar b is determined by solving the following equation:

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

After calculating equations 12 and 13, weights calculator 204 transmits vectors a, u, and v to block 205 via path 214. If the frame contained silence only equation 6 is calculated.

Determinator 205 is responsive to this transmitted information to decide whether the present frame is voiced or unvoiced. If the element of vector $(v_n - u_n)$ corresponding to power is positive, then, a frame is declared voiced if the following equation is true:

$$a'x_n - a'(u_n + v_n)/2 > 0; \quad (14)$$

or if the element of vector $(v_n - u_n)$ corresponding to power is negative, then, a frame is declared voiced if the following equation is true:

$$a'x_n - a'(u_n + v_n)/2 < 0. \quad (15)$$

Equation 14 can also be rewritten as:

$$a'\,x_n + b - \log[(1-p_n)/p_n] > 0.$$

Equation 15 can also be rewritten as:

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

If the previous conditions are not meet, determinator 205 declares the frame unvoiced. Equations 14 and 15 represent decision regions for making the voicing decision. The log term of the rewritten forms of equations 14 and 15 can be eliminated with some change of performance. Advantageously, in the present example, the element corresponding to power is the log of the speech energy.

Generator 206 is responsive to the information received via path 214 from calculator 204 to calculate the distance measure, A, as follows. First, the discriminant viable, d, is calculated by equation 16 as follows:

$$d = a'x_n + b - \log[(1-p_n)/p_n]. \quad (16)$$

Advantageously, it would be obvious to one skilled in the art to use different types of voicing detectors to generate a value similar to d for use in frame following equations. One such detector would be an auto-correlation detector. If the frame is voiced, the equations 17 through 20 are solved as follows:

$$m_1 = (1-z)m_1 + zd, \quad (17)$$

$$s_1 = (1-z)\,s_1 + zd^2, \text{ and} \quad (18)$$

$$k_1 = s_1 - m_1^2 \quad (19)$$

where m_1 is the mean for voiced frames and k_1 is the

variance for voiced frames.

The probability, P_d, that determinator 205 will declare a frame unvoiced is calculated by the following equation:

$$P_d = (1-z)P_d. \quad (20)$$

Advantageously, P_d is initially set to .5.

If the frame is unvoiced, equations 21 through 24 are solved as follows:

$$m_0 = (1-z)m_0 + zd, \quad (21)$$

$$s_0 = (1-z)s_0 + zd^2 \text{ and} \quad (22)$$

$$k_0 = s_0 - m_0^2. \quad (23)$$

The probability, P_d, that determinator 205 will declare a frame unvoiced is calculated by the following equation:

$$P_d = (1-z)\,P_d + z. \quad (24)$$

After calculating equation 16 through 22 the distance measure or merit value is calculated as follows:

$$A^2 = \frac{P_d(1-P_d)(m_1-m_0)^2}{(1-P_d)k_1 + P_d k_0}. \quad (25)$$

Equation 25 uses Hotelling's two-sample T^2 statistic to calculate the distance measure. For equation 25, the larger the merit value the greater the separation. However, other merit values exist where the smaller the merit value the greater the separation. Advantageously, the distance measure can also be the Mahalanobis distance which is given in the following equation:

$$A^2 = \frac{(m_1-m_0)^2}{(1-P_d)k_1 + P_d k_0}. \quad (26)$$

Advantageously, a third technique is given in the following equation:

$$A^2 = 2\frac{(m_1-m_0)^2}{(k_1+k_0)}. \quad (27)$$

Advantageously, a fourth technique for calculating the distance measure is illustrated in the following equation:

$$A^2 = a'(v_n - u_n) \quad (28)$$

Discriminant detector 102 makes the unvoiced/ voiced decision by transmitting information to multiplexer 105 via path 107 indicating a voiced frame if $a'x+b > 0$. If this condition is not true, then detector 102 indicates an unvoiced frame. The values for vector a and scalar b used by detector 102 are advantageously identical to the initial values of a and b for statistical voiced detector 103.

Detector 102 determines the distance measure in a manner similar to generator 206 by performing calculations similar to those given in equations 16 through 28.

In flow chart form, FIGS. 3 and 4 illustrate, in greater detail, the operations performed by statistical voiced detector 103 of FIG.2. Blocks 302 and 300 implement blocks 202 and 201 of FIG. 2, respectively. Blocks 304 through 318 implement statistical calculator 203. Blocks 320 and 322 implement weights calculator 204, and blocks 326 through 338 implement block 205 of FIG.2. Generator 206 of FIG. 2 is implemented by block 340. Subtractor 207 is implemented by block 308 or block 324.

Block 302 calculates the vector which represents the average of the classifiers for the present frame and all previous frames. Block 300 determines whether speech or silence is present in the present frame; and if silence is present in the present frame, the mean for each classfier is subtracted from each classifier by block 324 before control is transferred to decision block 326. However, if speech is present in the present frame, then the statistical and weights calculations are performed by blocks 304 through 322. First, the average vector is found in block 302. Second, the sums of the squares and products matrix is calculated in block 304. The latter matrix along with the vector X representing the mean of the classifiers for the present and past frames is then utilized to calculate the covariance matrix, T, in block 306. The mean X is then subtracted from the classifier vector x_n in block 308.

Block 310 then calculates the probability that the present frame is unvoiced by utilizing the present weight vector a, the present threshold value b, and the classifier vector 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 312. Then, the overall probability, p_n, that any frame will be unvoiced is calculated by block 314.

Blocks 316 and 318 calculate two vectors: u and v. The values contained in vector u represent the statistical average values that each classifier would have if the frame were unvoiced. Whereas, vector v contains values representing the statistical average values that each classifier would have if the frame were voiced. The actual vectors of classifiers for the present and previous frames are clustered around either vector u or vector v. The vectors representing the classifiers for the previous and present frames are clustered around vector u if these frames are found to be unvoiced; otherwise, the previous classifier vectors are clustered around vector v.

After execution of blocks 316 and 318, control is transferred to decision block 320. If N is greater than 99, control is transferred to block 322; otherwise, control is transferred to block 326. Upon receiving control, block 322 then calculates a new weight vector a and a new treshold value b. The vector a and value b are used in the next sequential frame by the preceding blocks in FIG. 3. Advantageously, if N is required to be greater than infinity, vector a and scalar b will never be changed, and detector 103 will adapt solely in response to vectors v and u as illustrated in blocks 326 through 338.

Blocks 326 through 338 implement u/v determinator 205 of FIG. 2. Block 326 determines whether the power term of vector v of the present frame is greater than or equal to the power term of

vector u. If this condition is true, then decision block 328 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 328, then the frame is so marked as voiced by block 330 otherwise the frame is marked as unvoiced by block 332. If the power term of vector v is less than the power term of vector u for the present frame, blocks 334 through 338 function are executed and function in a similar manner. Finally, block 340 calculates the distance measure.

In flow chart form, FIG. 5 illustrates, in greater detail the operations performed by block 340 of FIG. 4. Decision block 501 determines whether the frame has been indicated as unvoiced or voiced by examining the calculations 330, 332, 336, or 338. If the frame has been designated as voiced, path 507 is selected. Block 510 calculates probability P_d, and block 502 recalculates the mean, m_1, for the voiced frames and block 503 recalculates the variance, k_1, for voiced frames. If the frame was determined to be unvoiced, decision block 501 selects path 508. Block 509 recalculates probability P_d, and block 504 recalculates mean, m_0, for unvoiced frames, and block 505 recalculates the variance k_0 for unvoiced frames. Finally, block 506 calculates the distance measure by performing the calculations indicated.

Claims

1. An apparatus for determining the presence of a voiced sound in frames of speech,having a plurality of independent detecting means (102, 103) for detecting said voiced sound in a frame characterised in that

each of the detecting means comprises means (209) for calculating a merit value defining the separation between voiced and unvoiced ones of said frames of speech; and in that said apparatus comprises

means (104,105) for selecting one of said detecting means to indicate the presence of said voiced sound upon the selected one of said detecting means calculating a merit value better than any other one of said detecting means' calculated merit value.

2. The apparatus of claim 1 wherein said calculating means of each of said detecting means perform a statistical calculation to determine said merit value.

3. The apparatus of claim 2 wherein said statistical calculations are distance measurement calculations.

4. The apparatus of claim 2 wherein said one of said detecting means comprises means for indicating a frame is voiced upon detecting said voiced sound and indicating a frame is unvoiced upon said voiced sound being absent;

said calculating means for said one of said detecting means further comprises means for determining a discriminant variable for each ones of previous and present frames;

means for determining a mean value for voiced ones of said previous and present frames;

means for determining a variance value of said voiced ones of said previous and present frames;

means for determining a mean value of said unvoiced ones of said previous and present frames;

means for determining a variance value of said unvoiced ones of said previous and present frames; and

means for determining the merit value of said one of said detecting means from the determined voiced mean and variance values and the determined unvoiced mean and variance values.

5. The apparatus of claim 4 wherein said means for determining the merit value for said one of said detecting means comprises means for summing said variance values;

means for calculating a weighted sum of said variance values;

means for subtracting the mean value of said unvoiced frames from said mean value of said voiced frames;

means for squaring the subtracted value; and

means for dividing said weighted sum by the sum of said squared values, thereby generating said merit value for said one of said detecting means.

6. The apparatus of claim 5 wherein said means for calculating said weighted sum comprises means for calculating a first probability that said one of said detecting means indicates the presence of said voiced sound in said present frame;

means for calculating a second probability that said one of said detecting means indicates the absence of said voiced sound in said present frame;

means for multiplying said variance of said voiced ones of said previous and present frames by said first probability and said variance of said unvoiced ones of said previous and present frames by said second probability; and

means for forming said weighted sum from the results of said multiplications.

vector u. If this condition is true, then decision block 328 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 328, then the frame is so marked as voiced by block 330 otherwise the frame is marked as unvoiced by block 332. If the power term of vector v is less than the power term of vector u for the present frame, blocks 334 through 338 function are executed and function in a similar manner. Finally, block 340 calculates the distance measure.

In flow chart form, FIG. 5 illustrates, in greater detail the operations performed by block 340 of FIG. 4. Decision block 501 determines whether the frame has been indicated as unvoiced or voiced by examining the calculations 330, 332, 336, or 338. If the frame has been designated as voiced, path 507 is selected. Block 510 calculates probability P_d, and block 502 recalculates the mean, m_1, for the voiced frames and block 503 recalculates the variance, k_1, for voiced frames. If the frame was determined to be unvoiced, decision block 501 selects path 508. Block 509 recalculates probability P_d, and block 504 recalculates mean, m_0, for unvoiced frames, and block 505 recalculates the variance k_0 for unvoiced frames. Finally, block 506 calculates the distance measure by performing the calculations indicated.

Claims

1. An apparatus for determining the presence of a voiced sound in frames of speech,having a plurality of independent detecting means (102, 103) for detecting said voiced sound in a frame characterised in that

each of the detecting means comprises means (209) for calculating a merit value defining the separation between voiced and unvoiced ones of said frames of speech; and in that said apparatus comprises

means (104,105) for selecting one of said detecting means to indicate the presence of said voiced sound upon the selected one of said detecting means calculating a merit value better than any other one of said detecting means' calculated merit value.

2. The apparatus of claim 1 wherein said calculating means of each of said detecting means perform a statistical calculation to determine said merit value.

3. The apparatus of claim 2 wherein said statistical calculations are distance measurement calculations.

4. The apparatus of claim 2 wherein said one of said detecting means comprises means for indicating a frame is voiced upon detecting said voiced sound and indicating a frame is unvoiced upon said voiced sound being absent;

said calculating means for said one of said detecting means further comprises means for determining a discriminant variable for each ones of previous and present frames;

means for determining a mean value for voiced ones of said previous and present frames;

means for determining a variance value of said voiced ones of said previous and present frames;

means for determining a mean value of said unvoiced ones of said previous and present frames;

means for determining a variance value of said unvoiced ones of said previous and present frames; and

means for determining the merit value of said one of said detecting means from the determined voiced mean and variance values and the determined unvoiced mean and variance values.

5. The apparatus of claim 4 wherein said means for determining the merit value for said one of said detecting means comprises means for summing said variance values;

means for calculating a weighted sum of said variance values;

means for subtracting the mean value of said unvoiced frames from said mean value of said voiced frames;

means for squaring the subtracted value; and

means for dividing said weighted sum by the sum of said squared values, thereby generating said merit value for said one of said detecting means.

6. The apparatus of claim 5 wherein said means for calculating said weighted sum comprises means for calculating a first probability that said one of said detecting means indicates the presence of said voiced sound in said present frame;

means for calculating a second probability that said one of said detecting means indicates the absence of said voiced sound in said present frame;

means for multiplying said variance of said voiced ones of said previous and present frames by said first probability and said variance of said unvoiced ones of said previous and present frames by said second probability; and

means for forming said weighted sum from the results of said multiplications.

* C L M S *

7. The apparatus of claim 6 wherein said means for dividing comprises means for multiplying the results of the division of said weighted sum by the sum of said squared values by said first and second probabilities to generate said merit value of said one of said detecting means.

8. The apparatus of claim 7 wherein said means for indicating said frame is voiced and unvoiced comprises a means responsive to a set of classifiers defining speech attributes of said present frame of speech for calculating a set of statistical parameters;

means responsive to the calculated set of parameters for calculating a set of weights each associated with one of said classifiers; and

means responsive to the calculated set of weights and classifiers and said set of parameters for determining the presence of said voiced sound in said frame of speech.

9. The apparatus of claim 8 wherein said means for calculating said set of weights comprises means for calculating a threshold value in response to said set of said parameters;

means for communicating said set of weights and said threshold value to said means for calculating said set of statistical parameters to be used for calculating another set of parameters for another one of said frames of speech; and

said means for calculating said set of statistical parameters further responsive to the communicated set of weights and another set of classifiers defining said speech attributes of said other frame for calculating another set of statistical parameters.

10. An apparatus for determining the presence of a voiced sound in frames of speech, comprising:

first means (102) for generating a first signal indicating the presence of said fundamental frequency in one of said frames of speech;

second means (103) for generating a second signal indicating the presence of said fundamental frequency in said one of said frames of speech; and characterised in that

said first means comprises means for calculating a first generalized distance value representing the degree of separation between voiced and unvoiced frames as determined by said first means;

said second means comprises means (206) for calculating a second generalized distance value representing the degree of separation between voiced and unvoiced frames as determined by said second means; and in that it includes

means (104,105) for selecting said first signal to indicate the presence of said voiced sound upon said first generalized value being better than said second generalized value and for selecting said second signal to indicate the presence of said voiced sound upon said second generalized value being better than said first generalized value.

11. The apparatus of claim 10 wherein said generalized distance values are the Mahalanobis distance values.

12. The apparatus of claim 11 wherein said first means further comprises a means responsive to a set of classifiers defining speech attributes of one frame of speech for calculating a set of statistical parameters;

means responsive to the calculated set of parameters for calculating a set of weights each associated with one of said classifiers; and

means responsive to the calculated set of weights and classifiers and said set of parameters for determining the presence of said voiced sound in said frame of speech.

13. The apparatus of claim 12 wherein said means for calculating said first generalized distant value comprises means responsive to said calculated set of parameters and said calculated set of weights for determining said first generalized distance value.

14. The apparatus of claim 13 wherein said second means is a discriminant voiced detector.

15. The apparatus of claim 14 wherein said means for calculating said second generalized distance value comprises means for determining a mean value for voiced ones of said previous and present frames;

means for determining a mean value of said unvoiced ones of said previous and present frames;

means for determining a variance value of said unvoiced ones of said previous and present frames; and

means for determining said second distance measurement value from the determined voiced mean and variance values and the determined unvoiced mean and variance values.

16. The apparatus of claim 15 wherein said means for determining said second distance measurement value comprises

means for calculating the weighted sum of said variance values;

means for subtracting the mean value of said unvoiced frames from said mean value of said voiced frames;

* C L M S *

means for squaring the subtracted value; and

means for dividing said weighted sum of said variance values by the sum of said squared values thereby generating said second distance measurement value.

17. A method for determining the presence of a voiced sound in frames of speech based on using a first and second voiced detection methods (102,103) for detecting said voiced sound in a frame, characterised by

calculating a first merit value defining the separation between voiced and unvoiced ones of said frames of speech by said first voiced detection method,

calculating a second merit value defining separation between said ones of said voiced and unvoiced frames of speech by said second voiced detection method; and

selecting (104,105) said first voiced detection method to indicate the presence of said voiced sound upon said first merit value being better than said second value and selecting said second voiced detection method to indicate the presence of said voiced sound upon said second merit value being better than said first value.

18. The method of claim 17 wherein said steps of calculating said first and second values each comprises the step of performing a statistical calculation to determine said first and second values, respectfully.

19. The method of claim 18 wherein said statistical calculations are distance measurement calculations.

20. The method of claim 18 wherein said method further comprises the steps of indicating a frame is voiced upon detecting said voiced sound and indicating a frame is unvoiced upon said voiced sound being detected as absent by said first voiced detection method

said step of calculating said first value further comprises the steps of determining a discriminant variable for each ones of previous and present frames;

determining a mean value for voiced ones of said previous and present frames;

determining a variance value of said voiced ones of said previous and present frames;

determining a mean value of said unvoiced ones of said previous and present frames;

determining a variance value of said unvoiced ones of said previous and present frames; and

determining said first value from the determined voiced mean and variance values and the determined unvoiced mean and variance values.

21. The method of claim 20 wherein said step of determining said first value comprises the steps of summing said variance values;

calculating the weighted sum of said variance values;

subtracting the mean value of said unvoiced frames from said mean value of said voiced frames;

squaring the subtracted values; and

dividing said weighted sum of variance values by the sum of said squared variance values thereby generating said statistical value.

22. The method of claim 21 wherein said step of calculating said weighted sum comprises the steps of calculating a first probability that said said step of determining said first value indicates the presence of said voiced sound in said present frame;

calculating a second probability that said step of determining said first value indicates the absence of said voiced sound in said present frame;

multiplying said variance of said voiced ones of said previous and present frames by said first probability and said variance of said unvoiced ones of said previous and present frames by said second probability; and

forming said weighted sum from the results of said multiplications.

23. The method of claim 22 wherein said step of dividing comprises the step of multiplying the results of the division of said weighted sum by the sum of said squared values by said first and second probabilities to generate said first value.

Patentansprüche

1. Vorrichtung zur Bestimmung des Vorhandenseins eines stimmhaften Lautes in Rahmen von Sprache mit einer Vielzahl von unabhängigen Detektoreinrichtungen (102, 103) zur Feststellung des stimmhaften Lautes in einem Rahmen, dadurch gekennzeichnet,
daß jede der Detektoreinrichtungen eine Einrichtung (209) zur Berechnung eines Gütewertes umfaßt, der den Abstand zwischen stimmhaften und stimmlosen Sprachrahmen definiert, und
daß die Vorrichtung eine Einrichtung (104, 105) zur Auswahl einer der Detektoreinrichtungen für die Anzeige des Vorhandenseins des stimmhaften Lautes auswählt, wenn die gewählte Detek-

* C L M S *

toreinrichtung einen Gütewert berechnet, der besser ist als der von jeder anderen Detektoreinrichtung berechnete Gütewert.

2. Vorrichtung nach Anspruch 1, bei der die Berechnungseinrichtung jeder Detektoreinrichtung eine statistische Berechnung durchführt, um den Gütewert zu bestimmen.

3. Vorrichtung nach Anspruch 2, bei der die statistischen Berechnungen Abstandsmessungsberechnungen sind.

4. Vorrichtung nach Anspruch 2, bei der die eine Detektoreinrichtung eine Einrichtung umfaßt, die anzeigt, daß ein Rahmen stimmhaft ist, wenn der stimmhafte Laut festgestellt wird, und anzeigt, daß ein Rahmen stimmlos ist, wenn der stimmhafte Laut nicht vorhanden ist, die Berechnungseinrichtung für die eine Detektoreinrichtung ferner eine Einrichtung zur Bestimmung einer Diskriminanzvariablen für jeden vorhergehenden und augenblicklichen Rahmen umfaßt,
ferner eine Einrichtung zur Bestimmung eines Mittelwertes für stimmhafte Rahmen der vorhergehenden und augenblicklichen Rahmen,
eine Einrichtung zur Bestimmung eines Varianzwertes für stimmhafte Rahmen der vorhergehenden und augenblicklichen Rahmen,
eine Einrichtung zur Bestimmung eines Mittelwertes stimmloser Rahmen der vorhergehenden und augenblicklichen Rahmen,
eine Einrichtung zur Bestimmung eines Varianzwertes stimmloser Rahmen der vorhergehenden und augenblicklichen Rahmen und
eine Einrichtung zur Bestimmung des Gütewertes für die eine Detektoreinrichtung aus den bestimmten stimmhaften Mittel- und Varianzwerten und den bestimmten stimmlosen Mittel- und Varianzwerten.

5. Vorrichtung nach Anspruch 4, bei der die Einrichtung zur Bestimmung des Gütewertes für die eine Detektoreinrichtung eine Einrichtung zur Summierung der Varianzwerte umfaßt,
ferner eine Einrichtung zur Berechnung einer gewichteten Summe der Varianzwerte,
eine Einrichtung zum Subtrahieren des Mittelwertes der stimmlosen rahmen von dem Mittelwert der stimmhaften Rahmen,
eine Einrichtung zum quadrieren des subtrahierten Wertes und
eine Einrichtung zum Dividieren der gewichteten Summe durch die Summe der quadrierten Werte, wodurch der Gütewert für die eine Detektoreinrichtung erzeugt wird.

6. Vorrichtung nach Anspruch 5, bei der die Einrichtung zur Berechnung der gewichteten Summe eine Einrichtung zur Berechnung eines ersten Wahrscheinlichkeitswertes umfaßt dafür, daß die eine Detektoreinrichtung das Vorhandensein des stimmhaften Lautes im augenblicklichen Rahmen anzeigt,
ferner eine Einrichtung zur Berechnung eines zweiten Wahrscheinlichkeitswertes dafür, daß die eine Detektoreinrichtung das Fehlen des stimmhaften Lautes im augenblicklichen Rahmen anzeigt,
eine Einrichtung zur Multiplikation der Varianz der stimmhaften Rahmen aus den vorhergehenden und augenblicklichen Rahmen mit dem ersten Wahrscheinlichkeitswert und der Varianz der stimmlosen Rahmen aus den vorhergehenden und augenblicklichen Rahmen mit dem zweiten Wahrscheinlichkeitswert und
eine Einrichtung zur Bildung der gewichteten Summe aus dem Ergebnis der Multiplikationen.

7. Vorrichtung nach Anspruch 6, bei der die Divisionseinrichtung eine Einrichtung zur Multiplikation der Ergebnisse der Division der gewichteten Summe durch die Summe der quadrierten Werte mit dem ersten und dem zweiten Wahrscheinlichkeitswert umfaßt, um den Gütewert für die eine Detektoreinrichtung zu erzeugen.

8. Vorrichtung nach Anspruch 7, bei der die Einrichtung zur Anzeige dafür, daß der Rahmen stimmhaft und stimmlos ist, eine Einrichtung umfaßt, die unter Ansprechen auf einen Satz von Sprachattribute des augenblicklichen Sprachrahmens definierenden Klassifizierangaben einen Satz statistischer Parameter berechnet,
ferner eine Einrichtung, die unter Ansprechen auf den berechneten Satz von Parametern einen Satz von Gewichtungen berechnet, die je einer der Klassifizierangaben zugeordnet ist, und eine Einrichtung, die unter Ansprechen auf den berechneten Satz von Gewichtungen und Klassifizierangaben sowie den Satz von Parametern das Vorhandensein des stimmhaften Lautes in dem Sprachrahmen bestimmt.

9. Vorrichtung nach Anspruch 8, bei der die Einrichtung zur Berechnung des Satzes von Gewichtungen eine Einrichtung zur Berechnung eines Schwellenwertes unter Ansprechen auf den Satz von Parametern umfaßt,
ferner eine Einrichtung, die den Satz von Gewichtungen und den Schwellenwert an die Einrichtung zur Berechnung des Satzes von statistischen Parametern zur Verwendung für die Berechnung eines weiteren Satzes von Parametern für einen

* C L M S *

weiteren Sprachrahmen überträgt, und
wobei die Einrichtung zur Berechnung des Satzes statistischer Parameter ferner unter Ansprechen auf den übertragenen Satz von Gewichtungen und einen weiteren Satz von Klassifizierangaben, die die Sprachattribute des weiteren Rahmens definieren, einen weiteren Satz statistischer Parameter berechnet.

10. Vorrichtung zur Bestimmung des Vorhandenseins eines stimmhaften Lautes in Sprachrahmen mit
einer ersten Einrichtung (102) zur Erzeugung eines ersten Signals, das das Vorhandensein des stimmhaften Lautes in einem der Sprachrahmen anzeigt,
einer zweiten Einrichtung (103) zur Erzeugung eines zweiten Signals, das das Vorhandensein des stimmhaften Lautes in dem einen Sprachrahmen anzeigt,
dadurch gekennzeichnet,
daß die erste Einrichtung eine Einrichtung zur Berechnung eines ersten verallgemeinerten Distanzwertes umfaßt, der den Grad des Abstandes zwischen den stimmhaften und stimmlosen, durch die erste Einrichtung bestimmten Rahmen darstellt,
daß die zweite Einrichtung eine Einrichtung (206) zur Berechnung eines zweiten verallgemeinerten Distanzwertes aufweist, der den Grad des Abstandes zwischen stimmhaften und stimmlosen, durch die zweite Einrichtung bestimmten Rahmens darstellt, und
eine Einrichtung (104, 105) zur Auswahl des ersten Signals umfaßt, um das Vorhandensein des stimmhaften Lautes anzuzeigen, wenn der erste verallgemeinerte Distanzwert besser als der zweite verallgemeinerte Distanzwert ist, und zur Auswahl des zweiten Signals, um das Vorhandensein des stimmhaften Lautes anzuzeigen, wenn der zweite verallgemeinerte Distanzwert besser als der erste verallgemeinerte Distanzwert ist.

11. Vorrichtung nach Anspruch 10,
bei der die verallgemeinerten Distanzwerte die Mahalanobis-Distanzwerte sind.

12. Vorrichtung nach Anspruch 11,
bei der die erste Einrichtung ferner eine Einrichtung umfaßt, die unter Ansprechen auf einen Satz von Sprachattribute eines Sprachrahmens definierende Klassifizierangaben einen Satz von statistischen Parametern errechnet,
ferner eine Einrichtung, die unter Ansprechen auf den berechneten Satz von Parametern einen Satz von Gewichtungen berechnet, die je einer der Klassifizierangaben zugeordnet sind, und
eine Einrichtung, die unter Ansprechen auf den berechneten Satz von Gewichtungen und Klassifizierangaben sowie den Satz von Parametern das Vorhandensein des stimmhaften Lautes in dem Sprachrahmen bestimmt.

13. Vorrichtung nach Anspruch 12,
bei der die Einrichtung zur Berechnung des ersten verallgemeinerten Distanzwertes eine Einrichtung umfaßt, die unter Ansprechen auf den berechneten Satz von Parametern und den berechneten Satz von Gewichtungen den ersten verallgemeinerten Distanzwert bestimmt.

14. Vorrichtung nach Anspruch 13,
bei der die zweite Einrichtung ein Diskriminanten-Stimmhaft-Detektor ist.

15. Vorrichtung nach Anspruch 14,
bei der die Einrichtung zur Berechnung des zweiten verallgemeinerten Distanzwertes eine Einrichtung zur Bestimmung eines Mittelwertes für stimmhafte Rahmen der vorhergehenden und augenblicklichen Rahmen umfaßt, ferner eine Einrichtung zur Bestimmung eines Mittelwertes für stimmlose Rahmen der vorhergehenden und augenblicklichen Rahmen,
eine Einrichtung zur Bestimmung eines Varianzwertes der stimmlosen Rahmen von den vorhergehenden und augenblicklichen Rahmen und
eine Einrichtung zur Bestimmung des zweiten Distanzmeßwertes aus den bestimmten stimmhaften Mittel-und Varianzwerten sowie den bestimmten stimmlosen Mittel-und Varianzwerten.

16. Vorrichtung nach Anspruch 15,
bei der die Einrichtung zur Bestimmung des zweiten Distanzmeßwertes umfaßt:
eine Einrichtung zur Berechnung der gewichteten Summe der Varianzwerte,
eine Einrichtung zum Subtrahieren des Mittelwertes der stimmlosen Rahmen von dem Mittelwert der stimmhaften Rahmen,
eine Einrichtung zum quadrieren des subtrahierten Wertem und
eine Einrichtung zum Dividieren der gewichteten Summe der Varianzwerte durch die Summe der quadrierten Werte und dadurch Erzeugen des zweiten Distanzmeßwertes.

17. Verfahren zur Bestimmung des Vorhandenseins eines stimmhaften Lautes in Sprachrahmen unter Verwendung eines ersten und eines zweiten Stimmhaft-Detektorverfahrens (102, 103) zur Feststellung des stimmhaften Lautes in einem Rahmen,
gekennzeichnet durch:
Berechnen eines ersten Gütewertes, der den Ab-

* C L M S *

stand zwischen stimmhaften und stimmlosen Sprachrahmen durch das erste Stimmhaft-Detektorverfahren definiert, Berechnen eines zweiten Gütewertes, der den Abstand zwischen den stimmhaften und stimmlosen Sprachrahmen durch das zweite Stimmhaft-Detektorverfahren definiert, und
Auswählen (104, 105) des ersten Stimmhaft-Detektorverfahrens zur Anzeige des Vorhandenseins des stimmhaften Lautes, wenn der erste Gütewert besser als der zweite Gütewert ist, und Auswählen des zweiten Stimmhaft-Detektorverfahrens des Anzeige des Vorhandenseins des stimmhaften Lautes, wenn der zweite Gütewert besser als der erste Gütewert ist.

18. Verfahren nach Anspruch 17,
bei dem der Schritt zur Berechnung des ersten und zweiten Wertes je den Schritt der Durchführung einer statistischen Berechnung zur Bestimmung des ersten bzw. zweiten Wertes umfaßt.

19. Verfahren nach Anspruch 18, bei dem die statistischen Berechnungen Distanzmeßberechnungen sind.

20. Verfahren nach Anspruch 18,
bei dem ferner die Schritte vorgesehen sind, anzuzeigen, daß ein Rahmen stimmhaft ist, wenn der stimmhafte Laut festgestellt wird, und daß ein Rahmen stimmlos ist, wenn das erste Stimmhaft-Detektorverfahren feststellt, daß der stimmhafte Laut nicht vorhanden ist, und wobei der Schritt zur Berechnung des ersten Wertes ferner die Schritte umfaßt:
Bestimmen einer Diskriminanzvariablen für jeden vorhergehenden und augenblicklichen Rahmen,
Bestimmen eines Mittelwertes für stimmhafte Rahmen der vorhergehenden und augenblicklichen Rahmen,
Bestimmen eines Varianzwertes für die stimmhaften Rahmen der vorhergehenden und augenblicklichen Rahmen,
Bestimmen eines Mittelwertes für die stimmlosen Rahmen der vorhergehenden und augenblicklichen Rahmen,
Bestimmen eines Varianzwertes für die stimmlosen Rahmen der vorhergehenden und augenblicklichen Rahmen, und
Bestimmen des ersten Wertes aus den bestimmten stimmhaften Mittel- und Varianzwerten und den bestimmten stimmlosen Mittel- und Varianzwerten.

21. Verfahren nach Anspruch 20,
bei dem der Schritt zur Bestimmung des ersten Wertes die Schritte umfaßt:
Summieren der Varianzwerte,
Berechnen der gewichteten Summe der Varianzwerte,
Subtrahieren des Mittelwertes der stimmlosen Rahmen vom Mittelwert der stimmhaften Rahmen,
Quadrieren der subtrahierten Werte und
Dividieren der gewichteten Summe von Varianzwerten durch die Summe der quadrierten Varianzwerte und dadurch Erzeugen des statistischen Wertes.

22. Verfahren nach Anspruch 21,
bei dem der Schritt der Berechnung der gewichteten Summe die Schritte umfaßt:
Berechnen eines ersten Wahrscheinlichkeitswertes dafür, daß der Schritt zur Bestimmung des ersten Wertes das Vorhandensein des stimmhaften Lautes im augenblicklichen Rahmen anzeigt,
Berechnen eines zweiten Wahrscheinlichkeitswertes dafür, daß der Schritt der Bestimmung des ersten Wertes die Abwesenheit des stimmhaften Lautes im augenblicklichen Rahmen anzeigt,
Multiplizieren der Varianz der stimmhaften Rahmen von den vorhergehenden und augenblicklichen Rahmen mit dem ersten Wahrscheinlichkeitswert und der Varianz der stimmlosen Rahmen von den vorhergehenden und augenblicklichen Rahmen mit dem zweiten Wahrscheinlichkeitswert, und
Bilden der gewichteten Summe aus den Ergebnissen der Multiplikationen.

23. Verfahren nach Anspruch 22,
bei dem der Dividierschritt das Multiplizieren der Ergebnisse aus der Division der gewichteten Summe durch die Summe der quadrierten Werte mit dem ersten und dem zweiten Wahrscheinlichkeitswert zur Erzeugung des ersten Wertes umfaßt.

Revendications

1. Un dispositif destiné à déterminer la présence d'un son voisé dans des trames de parole, comportant un ensemble de moyens de détection indépendants (102, 103) qui sont destinés à détecter le son voisé dans une trame, caractérisé en ce que :
chacun des moyens de détection comprend des moyens (209) destinés à calculer une valeur de mérite définissant la séparation entre des trames voisées et des trames non voisées parmi les trames de parole; et en ce que ce dispositif comprend
des moyens (104, 105) destinés à sélectionner l'un des moyens de détection pour indiquer la présence du son voisé, lorsque le

* C L M S *

moyen sélectionné parmi les moyens de détection calcule une valeur de mérite meilleure que toute autre valeur de mérite calculée par les moyens de détection.

2. Le dispositif de la revendication 1, dans lequel les moyens de calcul de chacun des moyens de détection effectuent un calcul statistique pour déterminer la valeur de mérite.

3. Le dispositif de la revendication 2, dans lequel les calculs statistiques sont des calculs de mesure de distance.

4. Le dispositif de la revendication 2, dans lequel le moyen de détection sélectionné comprend des moyens qui indiquent qu'une trame est voisée en cas de détection du son voisé et qui indiquent qu'une trame est non voisée en cas d'absence du son voisé;

les moyens de calcul pour le moyen de détection sélectionné comprennent en outre des moyens qui sont destinés à déterminer une variable de discriminant pour chaque trame parmi les trames précédentes et la trame présente;

des moyens qui sont destinés à déterminer une valeur moyenne pour des trames voisées parmi les trames précédentes et la trame présente;

des moyens qui sont destinés à déterminer une valeur de variance des trames voisées parmi les trames précédentes et la trame présente;

des moyens qui sont destinés à déterminer une valeur moyenne des trames non voisées parmi les trames précédentes et la trame présente;

des moyens qui sont destinés à déterminer une valeur de variance des trames non voisées parmi les trames précédentes et la trame présente; et

des moyens qui sont destinés à déterminer la valeur de mérite du moyen de détection sélectionné, à partir de la valeur moyenne et de la valeur de variance déterminées pour les trames voisées, et à partir de la valeur moyenne et de la valeur de variance déterminées pour les trames non voisées.

5. Le dispositif de la revendication 4, dans lequel les moyens de détermination de la valeur de mérite pour le moyen de détection sélectionné comprennent des moyens qui sont destinés à faire la somme des valeurs de variance;

des moyens qui sont destinés à calculer une somme pondérée des valeurs de variance;

des moyens qui sont destinés à soustraire la valeur moyenne des trames non voisées, à partir de la valeur moyenne des trames voisées;

des moyens qui sont destinés à élever au carré la valeur qui résulte de la soustraction; et

des moyens qui sont destinés à diviser la somme pondérée par la somme des valeurs au carré, pour générer ainsi la valeur de mérite pour le moyen de détection sélectionné.

6. Le dispositif de la revendication 5, dans lequel les moyens destinés à calculer la somme pondérée comprennent des moyens qui sont destinés à calculer une première probabilité qui est la probabilité que le moyen de détection sélectionné indique la présence du son voisé dans la trame présente;

des moyens destinés à calculer une seconde probabilité qui est la probabilité que le moyen de détection sélectionné indique l'absence du son voisé dans la trame présente;

des moyens destinés à multiplier par la première probabilité la variance des trames voisées parmi les trames précédentes et la trame présente, et à multiplier par la seconde probabilité la variance des trames non voisées parmi les trames précédentes et la trame présente; et

des moyens qui sont destinés à former la somme pondérée précitée à partir des résultats de ces multiplications.

7. Le dispositif de la revendication 6, dans lequel les moyens de division comprennent des moyens qui sont destinés à multiplier par les première et seconde probabilités les résultats de la division de la somme pondérée par la somme des valeurs au carré, pour générer la valeur de mérite du moyen de détection sélectionné.

8. Le dispositif de la revendication 7, dans lequel les moyens qui sont destinés à indiquer que la trame est voisée ou non voisée comprennent des moyens qui réagissent à un jeu de classificateurs définissant des attributs de parole de la trame de parole présente, de façon à calculer un jeu de paramètres statistiques;

des moyens qui réagissent au jeu de paramètres calculé en calculant un jeu de poids qui sont respectivement associés à l'un des classificateurs; et

des moyens qui réagissent au jeu de poids calculé et aux classificateurs ainsi qu'au jeu de paramètres, de façon à déterminer la présence du son voisé dans la trame de parole.

9. Le dispositif de la revendication 8, dans lequel les moyens destinés à calculer le jeu de poids comprennent des moyens qui sont destinés à calculer une valeur de seuil sous la dépendance du jeu de paramètres;

des moyens qui sont destinés à transmettre

* C L M S *

le jeu de poids et la valeur de seuil aux moyens de calcul du jeu de paramètres statistiques, en vue de leur utilisation pour le calcul d'un autre jeu de paramètres pour une autre des trames de parole; et

les moyens destinés à calculer le jeu de paramètres statistiques réagissent en outre au jeu de poids transmis et à un autre jeu de classificateurs définissant les attributs de parole de cette autre trame, en calculant un autre jeu de paramètres statistiques.

10. Un dispositif destiné à déterminer la présence d'un son voisé dans des trames de parole, comprenant :

des premiers moyens (102) destinés à générer un premier signal indiquant la présence d'une fréquence fondamentale dans l'une des trames de parole;

des seconds moyens (103) destinés à générer un second signal indiquant la présence de la fréquence fondamentale dans la trame de parole précitée; et caractérisé en ce que :

les premiers moyens comprennent des moyens destinés à calculer une première valeur de distance généralisée représentant le degré de séparation entre des trames voisées et non voisées, qui est déterminé par les premiers moyens;

les seconds moyens comprennent des moyens (206) qui sont destinés à calculer une seconde valeur de distance généralisée représentant le degré de séparation entre des trames voisées et non voisées, qui est déterminé par les secondes moyens; et en ce qu'il comprend

des moyens (104, 105) destinés à sélectionner le premier signal pour indiquer la présence du son voisé lorsque la première valeur généralisée est meilleure que la seconde valeur généralisée, et à sélectionner le second signal pour indiquer la présence du son voisé lorsque la seconde valeur généralisée est meilleure que la première valeur généralisée.

11. Le dispositif de la revendication 10, dans lequel les valeurs de distance généralisées sont les valeurs de distance de Mahalanobis.

12. Le dispositif de la revendication 11, dans lequel les premiers moyens comprennent en outre des moyens qui réagissent à un jeu de classificateurs définissant des attributs de parole d'une trame de parole, en calculant un jeu de paramètres statistiques;

des moyens qui réagissent au jeu de paramètres calculé en calculant un jeu de poids dans lequel chaque poids est associé à l'un des classificateurs; et

des moyens qui réagissent au jeu de poids et aux classificateurs calculés, ainsi qu'au jeu de paramètres, de façon à déterminer la présence du son voisé dans la trame de parole.

13. Le dispositif de la revendication 12, dans lequel les moyens de calcul de la première valeur de distance généralisée comprennent des moyens qui réagissent au jeu de paramètres calculé et au jeu de poids calculé, de façon à déterminer la première valeur de distance généralisée.

14. Le dispositif de la revendication 13, dans lequel les seconds moyens sont un détecteur de parole voisée à discriminant.

15. Le dispositif de la revendication 14, dans lequel les moyens destinés à calculer la seconde valeur de distance généralisée comprennent des moyens qui sont destinés à déterminer une valeur moyenne pour des trames voisées parmi les trames précédentes et présente;

des moyens qui sont destinés à déterminer une valeur moyenne des trames non voisées parmi les trames précédentes et présente;

des moyens qui sont destinés à déterminer une valeur de variance des trames non voisées parmi les trames précédentes et présente; et

des moyens qui sont destinés à déterminer la seconde valeur de mesure de distance à partir de la valeur moyenne et de la valeur de variance qui sont déterminées pour les trames voisées, et à partir de la valeur moyenne et de la valeur de variance qui sont déterminées pour les trames non voisées.

16. Le dispositif de la revendication 15, dans lequel les moyens destinés à déterminer la seconde valeur de mesure de distance comprennent :

des moyens destinés à calculer la somme pondérée des valeurs de variance;

des moyens destinés à soustraire la valeur moyenne des trames non voisées, à partir de la valeur moyenne des trames voisées;

des moyens destinés à élever au carré la valeur qui résulte de la soustraction; et

des moyens destinés à diviser la somme pondérée des valeurs de variance par la somme des valeurs élevées au carré, pour générer ainsi la seconde valeur de mesure de distance.

17. Un procédé de détermination de la présence d'un son voisé dans des trames de parole, basé sur l'utilisation d'un premier et d'un second procédé de détection de parole voisée (102, 103), pour la détection du son voisé dans une trame, caractérisé en ce que :

on calcule une première valeur de mérite

* C L M S *

définissant la séparation entre des trames voisées et des trames non voisées parmi les trames de parole, en employant le premier procédé de détection de parole voisée;

on calcule une seconde valeur de mérite définissant la séparation entre les trames voisées et les trames non-voisées parmi les trames de parole, en employant le second procédé de détection de parole voisée; et

on sélectionne (104, 105) le premier procédé de détection de parole voisée pour indiquer la présence du son voisé lorsque la première valeur de mérite est meilleure que la seconde valeur, et on sélectionne le second procédé de détection de parole voisée pour indiquer la présence du son voisé lorsque la seconde valeur de mérite est meilleure que la première valeur.

18. Le procédé de la revendication 17, dans lequel chacune des étapes de calcul des première et seconde valeurs comprend l'étape qui consiste à effectuer un calcul statistique pour déterminer respectivement les première et seconde valeurs.

19. Le procédé de la revendication 18, dans lequel les calculs statistiques sont des calculs de mesure de distance.

20. Le procédé de la revendication 18, comprenant en outre les étapes qui consistent à indiquer qu'une trame est voisée sous l'effet de la détection du son voisé, et à indiquer qu'une trame est non voisée sous l'effet de l'absence de la détection du son voisé par le premier procédé de détection de parole voisée,

l'étape de calcul de la première valeur comprenant en outre les étapes suivantes :

on détermine une variable de discriminant pour chacune des trames précédentes et présente;

on détermine une valeur moyenne pour des trames voisées parmi les trames précédentes et présente;

on détermine une valeur de variance des trames voisées parmi les trames précédentes et présente;

on détermine une valeur moyenne des trames non voisées parmi les trames précédentes et présente;

on détermine une valeur de variance des trames non voisées parmi les trames précédentes et présente; et

on détermine la première valeur à partir de la valeur moyenne et de la valeur de variance déterminées pour les trames voisées, et à partir de la valeur moyenne et de la valeur de variance déterminées pour les trames non voisées.

21. Le procédé de la revendication 20, dans lequel l'étape de détermination de la première valeur comprend les étapes suivantes :

on fait la somme des valeurs de variance;

on calcule la somme pondérée des valeurs de variance;

on soustrait la valeur moyenne des trames non voisées, à partir de la valeur moyenne des trames voisées;

on élève au carré les valeurs qui résultent de la soustraction; et

on divise la somme pondérée des valeurs de variance par la somme des valeurs de variance au carré, pour générer ainsi la valeur statistique précitée.

22. Le procédé de la revendication 21, dans lequel l'étape de calcul de la somme pondérée comprend les étapes qui consistent à calculer une première probabilité qui est la probabilité que l'étape de détermination de la première valeur indique la présence du son voisé dans la trame présente;

on calcule une seconde probabilité qui est la probabilité que l'étape de détermination de la première valeur indique l'absence du son voisé dans la trame présente;

on multiplie par la première probabilité la variance des trames voisées parmi les trames précédentes et présente, et on multiplie par la seconde probabilité la variance des trames non voisées parmi les trames précédentes et présente; et

on forme la somme pondérée précitée à partir des résultats de ces multiplications.

23. Le procédé de la revendication 22, dans lequel l'étape de division comprend l'étape qui consiste à multiplier par les première et seconde probabilités les résultats de la division de la somme pondérée par la somme des valeurs élevées au carré, pour générer la première valeur.

* C L M S *