



(11) **EP 3 100 675 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
07.12.2016 Bulletin 2016/49

(21) Application number: **16166727.4**

(22) Date of filing: **22.04.2016**

(51) Int Cl.:

A61B 5/00 (2006.01)	A61B 5/08 (2006.01)
A61B 5/11 (2006.01)	A61B 5/02 (2006.01)
A61B 8/08 (2006.01)	G06F 19/00 (2011.01)
G06F 17/14 (2006.01)	A61B 5/0205 (2006.01)
A61N 1/39 (2006.01)	A61N 1/362 (2006.01)
A61B 5/0428 (2006.01)	A61B 5/113 (2006.01)

(84) Designated Contracting States:
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR**
Designated Extension States:
BA ME
Designated Validation States:
MA MD

(30) Priority: **03.06.2015 US 201514729423**

(71) Applicant: **IMEDI PLUS Inc.
Zhubei City 30261 (TW)**

(72) Inventors:
• **TSAI, Kun-Hsi
30261 Zhubei City (TW)**

- **TSAO, Yu
30261 Zhubei City (TW)**
- **KU, Shih-Hsuan
30261 Zhubei City (TW)**
- **LIANG, Tzu-Chen
30261 Zhubei City (TW)**
- **CHANG, Yun-Fan
30261 Zhubei City (TW)**
- **YANG, Shih-I
30261 Zhubei City (TW)**

(74) Representative: **Huang, Chongguang
Olswang LLP
90 High Holborn
London WC1V 6XX (GB)**

(54) **METHOD AND SYSTEM FOR RECOGNIZING PHYSIOLOGICAL SOUND**

(57) A system for recognizing physiological sound comprises a receiving module, a feature extracting module, a classifier, and a comparing module. A method for recognizing physiological sound comprises receiving a physiological sound by the receiving module; extracting at least one feature from the physiological sound by the feature extraction module; classifying the at least one

feature to identify at least one category by a classifier; and comparing the at least one category with a normal physiological sound and/or an abnormal physiological sound by the comparing module for evaluating a risk of disease. The method and system for recognizing physiological sound can precisely identify the specific physiological sound and exclude the noise.

EP 3 100 675 A1

Description

FIELD

[0001] The present invention relates to method for recognizing physiological sound, particularly to a method for extracting and classifying the feature of the physiological sound. The present invention also relates to a system for recognizing physiological sound.

BACKGROUND

[0002] Analysis of heart, lung, bowel and vascular disorders by means of noninvasive auscultation has long been a very useful tool for medical diagnosis of ailments. Conventional electronic stethoscopes were not invented until 1922. Modern electronic stethoscopes can improve sound quality and provide a visual indication of heart sounds, such as phonocardiogram (PCG). PCG and electrocardiogram (ECG) are applied in the basic examination for the heart. PCG could be obtained by recording the electric signals converted from mechanical vibration (collected via placing stethoscope in certain parts of the chest) by instrument. ECG could be obtained by placing electrodes on any two places of the heart and connecting two electrodes to the positive and negative poles of ECG machine to form leads and record the ECG voltage changes of the two places in the human body. ECG could be shown in ECG paper or monitor and further reflect the rhythms of beating heart as well as the weaker parts of myocardial muscles. The pitches and occurrence time of heart sounds follow certain regular patterns. The first heart sound (S1) and the second heart sound (S2) could be observed in a normal heart. The first heart sound occurs in the contraction period of the heart, which is caused by the blood flowing into great vessels during the contraction of ventricles (ventricle contracts and both mitral valve and tricuspid valve close). The first heart sound continues for relatively longer time and with low pitch. The second heart sound occurs in the relaxation period of heart, which is caused by the vibration of ventricular wall during ventricular relaxation (aortic and pulmonary valves close and atrioventricular valve opens to allow blood flowing from atrium to ventricle). The duration of the second heart sound is shorter. Clinically, abnormal third and fourth heart sounds sometimes would be detected. The third heart sound shows low frequency and amplitude, which is caused by the vibration of ventricular wall. The fourth heart sound is caused by the vibration of ventricular wall during atrial contraction owing to blood flow rapidly entering ventricle.

[0003] Many heart diseases could be effectively diagnosed through auscultation. In some deadly heart diseases (such as heart valve dysfunction, heart failure, etc.), cardiac auscultation has already become the most successful, reliable, and inexpensive method in early diagnosis. However, the correctness of cardiac auscultation is closely related to the experiences of doctors. Also, some diseases show obvious occurrence patterns (for example, during S1 and S2 or after S2, etc.). Therefore, how to automatically detect and preliminarily judge the occurring time of S1 and S2 has already become an important issue. This issue could effectively help doctors to conform the occurrences of diseases preliminarily. In normal situation, the time order of S1 and S2 could serve as the materials for making judgments. Nevertheless, time order is no longer reliable under the circumstances of arrhythmia. If voiceprint comparison for S1 and S2 is available, the judgment on the case of arrhythmia could be improved in quality. Researches about heart sound detection could be divided into two categories: ECG signal dependent and ECG signal independent. ECG signal dependent researches include ECG-based detections on instantaneous energy (Malarvili et al., 2003) and detection on QRS wave group and T wave (El-Segaier et al., 2005). Nonetheless, in low-quality ECG signals, it is not always possible to clearly detect T wave. Under such situation, S2 could be classified using unsupervised classifier (Carvalho et al., 2005), although such method should consider hardware equipment and the comfort of examinees. ECG-independent methods could be divided into unsupervised and supervised methods. Unsupervised methods include using normalized average Shannon energy (Liang et al., 1997) and high frequency-based methods (Kumar et al., 2006) for wavelet decomposition. Supervised methods include neural network classifier (Hebden et al., 1996) and decision making trees (Stasis et al., 1996) used for classification. In addition, the most advanced method used in recent years is to detect according to the features of the regular intervals between S1 and S2. Generally, average heart rate (Olmez et al., 2003, Kumar et al., 2006) would be assumed in research. However, such assumption is not applicable in the heart sound of arrhythmia patients.

[0004] It is relatively difficult to simultaneously and synchronously record and analyze ECG and PCG in actual clinical cases. Also, when pulseless electrical activity occurs, ECG cannot determine that the heart rate has stopped due to the maintenance of electrical activity. Thus, how to make diagnosis according to solely PCG became an important and mainstream research topic. Mainstream detection methods usually include the time interval features of S1 and S2. But this feature would become unreliable under the situation of arrhythmia and highly decrease the correctness of detection. Therefore, the disadvantages in prior arts should be resolved.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Many aspects of the present embodiments can be better understood with reference to the following drawings. The components in the drawings are not necessarily drawn to scale, the emphasis instead being placed upon clearly illustrating the principles of the present embodiments. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. 1 shows a block diagram of the system for recognizing physiological sound of the present invention;

FIG. 2 shows the classifier comprising the KNN module, the GMM module, the SVM module, and the DNN module of the present invention;

FIG. 3 is a DNN network model;

FIG. 4 shows a flow diagram illustrating the method for recognizing the physiological sound in FIG. 1; S201 to S206 represent Step 201 to Step 206 respectively;

FIG. 5 is a spectrum (upper portion) and wave pattern (lower portion) map of S 1 and S2 of the present invention;

FIG. 6 shows accuracy results of heart sound with classifiers KNN module, GMM module, SVM module, and DNN module of the present invention.

FIG. 7 shows segments (solid square lines) processed from energy differences of regular heart sounds by the VAD module.

FIG. 8 shows segments (solid square lines) processed from energy differences of irregular heart sounds by the VAD module.

FIG. 9 shows segments (solid square line) processed from energy differences of normal lung sounds by the VAD module.

FIG. 10 shows segments (solid square line) processed from energy differences of wheezing lung by the VAD module.

FIG. 11 shows segments (solid square lines) processed from energy differences of crackling lung sounds into energy differences by the VAD module.

DETAILED DESCRIPTION

[0006] The present invention provides a method and system for recognizing physiological sound. As shown in Figs. 1 and 2, the system for recognizing physiological sound 100 comprises a receiving module 110, a feature extraction module 120, a classifier 130, and a comparing module 140. The feature extraction module 120 comprises a voice activity detector (VAD) module 121, a Mel-frequency cepstral coefficient (MFCC) module 122, and a K-means algorithm module 123; the classifier 130 comprises a K-nearest neighbor (KNN) module 131, a Gaussian mixture model (GMM) module 132, a support vector machine (SVM) module 133, or a deep neural network (DNN) module 134.

Preparation 1 Feature extraction

(1) VAD module 121

[0007] The VAD module 121 is also called Speech Activity Detection or Speech Detection and is usually used to determine whether a voice signal is processed by vocal voice processing technology. The main application of the VAD module 121 is in speech coding and speech recognition. The VAD module 121 is also usually used in the pretreatment before classification to increase recognition correctness of later investigations. The purpose for this pretreatment is to find out the position of S1 and S2 and further detect whether this segment of voice is S1 or S2. Ideally, heart sound signals received through a stethoscope are usually stronger than noises, so we could determine whether this segment is the heart sound we want according to voice energy differences. The processing results after the VAD module 121 treatment are shown in FIG. 3. Energy difference serves as the accordance for determining where the heart sounds are from.

[0008] First of all, the maximum standard deviation of each voice signal was calculated according to Formula (1).

$$\tau = \max_m \sqrt{\frac{\sum_{i=1}^n (x[n,m])^2 - \mu^2}{n}}$$

Formula (1)

[0009] In Formula (1), m is sound frame and n is frequency domain. μ is average. Then is the calculation for confirming whether each sound frame has heart sound in each segment of voice, as shown in Formula (2).

$$f_i = \begin{cases} 1, & \text{if } std_i > \tau - \alpha \text{ and } std_i > \beta \\ 0, & \text{else} \end{cases}$$

Formula (2)

f_i is i th sound frame. 1 means heart sound and 0 means still sound. std_i is the deviation for i th sound frame. α and β are parameters.

(2) MFCC module 122

[0010] In general, humans show different perception sensitivity in different intervals of frequency domain. Under normal circumstance, low frequency shows higher resolution, which means smaller frequency difference could be detected in low frequency situation. Critical band phenomenon should also be considered. The critical band width is about 100 Hz below 1 kHz. When the frequency is higher than 1 kHz, the critical band width would show exponential increase. Thus, we could divide frequency bands in frequency domain according to the characteristics of human auditory. The frequency compositions in one frequency band are integrated into energy intensity. These frequency band intensities are then converted into cepstrum after the calculation of Discrete Cosine Transform (DCT), which is exactly Mel-Frequency Cepstrum (MFC). Since MFC is the parameter produced from the model for human auditory, MFC is successfully applied in speech recognition and classification. There are six sequential operations for the MFCC module 122: pre-emphasis, windowing, fast fourier transform (FFT), mel filtering, non-linear transformation and discrete cosine transform (DCT).

[0011] Generally, thirteen dimensions of voice features could be obtained after the 6 steps described above, including one logarithmic energy parameter and twelve cepstrum parameters. Furthermore, differential cepstrum parameters are added to reveal how cepstrum parameters change with time. The meaning of differential cepstrum parameter is the slope of cepstrum parameter versus time, which means the dynamic change of cepstrum parameter. Hence, thirty-nine dimensions of voice features could be obtained after adding the discussion about velocity and acceleration. The calculation methods are shown below.

$$vel[i, t] = \frac{\sum_{m=1}^4 m \cdot [c[i, t+m] - c[i, t-m]]}{2 \sum_{m=1}^4 m^2}$$

Formula (3)

$$acc[i, t] = \frac{\sum_{m=1}^4 m \cdot [vel[i, t+m] - vel[i, t-m]]}{2 \sum_{m=1}^4 m^2}$$

Formula (4)

$c[i]$ is the i th dimension of cepstrum parameter and t is the time indicator for sound frame.

(3) K-means algorithm module 123

[0012] The main goal of the K-means algorithm module 123 is to find out representative data points from huge amount of high dimensional data points. These data points could be called population centers. Data compression (using small number of data points to represent large amount of data to compress data) and classification (using small number of representative points to represent specific categories to lower data amount and calculation and avoid adverse effects brought by noises) are conducted on population centers. The calculation steps of algorithm are shown below.

A. Initialization: divide training materials randomly into K groups and arbitrarily select K values as the initial population

center y_k , $k=1,2,\dots,K$

B. Recursive calculation:

- a. Let each x find the nearest population center and make it belong to said population center. Calculate the distance between x and the population center to which x belongs.

$$k^* = \arg_k \min d(x, y_k), x \in C_k$$

Formula (5)

- b. All of x belonging to C_k forms a group. Calculate the population center y_k again.

- c. If the new groups of population centers are the same as the original population center set, the training is completed. Otherwise, new groups of populations replace the original population center groups. Step a is repeated to continue recursive calculations.

Preparation 2 Classifier

(1) KNN module 131

[0013] The basic concept behind the KNN module 131 is that "like attracts like". In other words, objects of the same type should gather together. In mathematical language, if objects of the same type could be represented by point in high dimension of space, the distances between these points should become closer. Therefore, while faced with a set of data in an unknown category, we only need to find out the nearest point in training data. By doing so, the categories of these data could be regarded as the same as the nearest point. The classification steps of the KNN module 131 is to categorize measurement data x to one of the C categories. The steps are described in detail as follows:

- A. Determine which data x is used and find the nearest point number K . Use appropriate distance formula to calculate the distance.

- B. When testing data x shows relatively higher similarities with certain category (among K category), x will be regarded as that category.

[0014] The pre-needed information for classifier is the nearest K number of point, selected distance calculation formula, and training data.

[0015] Assuming that our training set is paired as (x_i, z_i) , $i = 1, \dots, n$, x_i is the i th vector for training data and z_i is the corresponding classification index (For example, $z_i = j$ represents that i th training data vector is the sample for ω_j in j th category). The distance between testing data vector x and training data vector y is expressed as $d(x, y)$. In the disclosure, Euclidean metric was used as the distance calculation method, which is shown in Formula (6).

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$$

Formula (6)

(2) GMM module 132

[0016] Gaussian Classifier is a common classifier, which applies Bayes' Theorem as fundamental concept. That is why Gaussian Classifier is also called Bayes' Classifier. The concepts of Gaussian Classifier is to utilize known categorized data to calculate probability distribution and find out the possibilities of unknown categories data in each possibility distribution of known category. The category shows the highest possibility would be chosen as the category of the unknown data. The average value and standard deviation serve as the two parameters that determine the whole model.

[0017] Single Gaussian distribution is to describe the distribution shape according to one average and one standard deviation. However, signals are usually distributed in a more complex manner. So, one Gaussian distribution is usually insufficient to approximate said signal. Gaussian mixture model is to involve multiple Gaussian distributions to approximate signals. As such, the distribution of the signals could be better approximated. Nevertheless, as the number of mixed Gaussian curves increases, the distribution shapes would become similar and the complexity would relatively increase.

[0018] Gaussian mixture model is composed of 3 items: mean vector μ_i , covariance matrix Σ_i , and weighed value ω_i . To show Gaussian model in a simplified way λ , the expression formula is shown as Formula (7).

$$\lambda = \{\omega_i, \mu_i, \Sigma_i\}, i = 1, \dots, K \quad \text{Formula (7)}$$

[0019] The mixed probability density function of feature vector z could be expressed as Formula (8):

$$p(z|z)|\lambda = \sum_{i=1}^K \omega_i p_i(z[n]) \quad \text{Formula (8)}$$

$p_i(z)$ is the i th Gaussian probability density function of feature vector z .

$$p_i(z[n]) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (z[n] - \mu_i)^T \Sigma_i^{-1} (z[n] - \mu_i)\right) \quad \text{Formula (9)}$$

[0020] Since the maximum probability value is 1, the weighed value ω_i of mixed probability density function in Gaussian mixture model must conform to the condition described in Formula (10).

$$\sum_{i=1}^K \omega_i = 1, \omega_i \geq 0, 1 \leq i \leq K \quad \text{Formula (10)}$$

[0021] When there is/are N number of feature vector(s) z with dimension d , to train this training sample to be in line with Gaussian mixture model λ is also to obtain the proper 3 parameters μ_i , Σ_i , and ω_i ($i = 1, \dots, K$). The final goal is to allow trained model to fully represent the distribution of feature vector z , that is, to find a proper model parameter to allow feature vector z and Gaussian mixture model λ to show the maximal similarity $p(z|\lambda)$, as shown in Formula (11).

$$p(z|\lambda) = \prod_{n=1}^N p(z_n|\lambda) \quad \text{Formula (11)}$$

[0022] Since Formula (8) is a nonlinear equation, complex calculation is inevitable during solving. Generally, expectation-maximization (EM) algorithm is applied to find out the optimal parameter in Gaussian mixture model.

(3) SVM module 133

[0023] The SVM module 133 has been widely used in statistical classification and regression analysis. It has been proved by experiments that the SVM module 133 has strong classification capability. The central concept for the SVM module 133 is to reflect training data to high-dimensional feature plane and build up an optimal super plane (a plane in high dimension and with boundary at large intervals). Most of the SVMs mainly are applied in binary classification. It is also acceptable to combine multiple binary classifications to construct multi-category classification. The situations could be divided into linear data and non-linear data.

[0024] Assuming a set $\{x_i, i = 1, \dots, n\}$ and the set is assigned to either the category of ω_1 or ω_2 . The corresponding mark is noted as $y_i = \pm 1$. The goal is to find a super plane $g(x)$ ($g(x) = w^T x + w_0$) to allow all the points of $y_i = +1$ to fall in the side of $g(x) > 0$. By doing so, the sign of $g(x)$ could serve as the accordance for distinguishing in the hope to find the plane that shows the largest distances to the two side borders, called the best super plane. To make the distance of H_1 and H_2 become the largest, it is essential to solve Formula (12).

[0025] Limitation is described in Formula (13)

$$y_i (w^T x_i + w_0) \geq 1, \quad i = 1, \dots, n \quad \text{Formula (13)}$$

[0026] When data is not linear and separable, kernel function is used as Formula (14), which projects data to higher dimension feature space.

$$k(x, y) = \langle \phi(x), \phi(y) \rangle \quad \text{Formula (14)}$$

[0027] Common kernel function includes linear polynomial Gaussian radial basis function. Different kernel functions could be selected according to different classification characteristics. In the disclosure, Gaussian radial basis function kernel is applied as kernel function, as shown in Formula (15).

$$K(x, y) = \exp(-||x - y||^2 / 2\sigma^2) \quad \text{Formula (15)}$$

(4) DNN module 134

[0028] Neural network (NN) is one mathematical model that mimics biological neural network structure and function to allow computer to undergo self-learning and deduce according to the rule of thumb, which makes neural network become more advantageous than logic inference calculus. Neural network algorithm has certain characteristics as follows: 1. parallel processing, 2. fault-tolerant, 3. combined memory, 4. solving optimal problems, 5. implementation of very-large-scale integration (VLSI), and 6. processing problems that are difficult for normal algorithm. So far, many scholars proposed different neural network models to solve different problems. Common neural network models include back-propagation network, Hopfield network and radial basis function network.

[0029] The operation of the DNN module 134 is to take output layer to serve as the input of the next hidden layer. The concept is to utilize the increase of hidden layer number to strengthen the system. FIG. 3 is the DNN module 134 having five layers. The relation between input layer and the output of 1st hidden layer is described as Formula (16).

$$a_2 = f(W_1 x) \quad \text{Formula (16)}$$

x is the input and W_1 is the weight. f is the activation function. Sigmoid function is used in the disclosure. a_2 is the output of the second layer.

[0030] After the output of the first hidden layer is obtained, the relation is shown as Formula (17). L is a layer number of the DNN module 134.

$$a_{i+1} = f(W_i a_i), \quad i = 2, \dots, L - 1 \quad \text{Formula (17)}$$

[0031] Besides, since the initial value of the parameter would influence the calculation results, the DNN module 134 usually uses restricted Boltzmann machines (RBM) to conduct the prediction for initial parameter and uses back-propagation to adjust parameters, as shown in Formula (18).

$$J(a_L, y) = \text{loss}(a_L, y) \quad \text{Formula (18)}$$

[0032] In Formula (18), y is label and a_L is the output of L th layer. In the disclosure, loss function applies softmax function. The detailed algorithm may refer to references (Bengio, 2009, Mohamed et al., 2013). Lastly, frequent use of dropout may avoid overtraining and obtain better efficiency.

Preparation 3 Evaluation methods

[0033] The evaluation methods are performed by the comparing module 140. The evaluation methods used in pattern recognition and information retrieval usually use precision, recall and F-measure to serve as the standards for evaluating the system is good or bad (Martin et al., 1997). The four situations considered are shown in Table 1, and each of the definitions are shown as in Formulae (19) to (21).

Table1. Objective assessment matrix

Predicted Class (Expectation)	Actual Class(Observation)	
	tp(true positive) Correct result	fp(false positive) False Alarm Unexpected result
	fn(false negative) Missed Detection Missing result	tn(true negative) Correct absence of result

$$\text{Precision} = \frac{tp}{tp+fp} \quad \text{Formula (19)}$$

$$\text{Recall} = \frac{tp}{tp+fn} \quad \text{Formula (20)}$$

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad \text{Formula (21)}$$

[0034] F-measure is also called F_1 measure, which represents the equal weights of precision and recall. Recall is usually called true positive rate or sensitivity, and precision is called positive predictive value. In classification research, accuracy is usually used to serve as evaluation standards, and the definition of accuracy is shown in Formula (22).

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn} \quad \text{Formula (22)}$$

Example 1 Experimental corpus and experimental steps for heart sounds

[0035] The flow chart of the method for recognizing physiological sound is shown in Figs. 1, 2 and 4. Step 201 is receiving a physiological sound by the receiving module 110. The place for receiving audios centralized on pulmonary valve auscultation area and second aortic valve auscultation area. The receiving module 110 is an electronic stethoscope, and the data used in this experiment is the actual voice data collected by the electronic stethoscope. The aim is to utilize heart sounds to find out the S1 and S2 audio clips. First of all, an analog microphone is used to record heart sounds and converted the recorded heart sounds into digital signals. Decoded by an audio codec, the digital audio signals are separated by two paths. One is to convert the filtered sounds into analog signals and deliver the analog signals through stethoscope ear tube. The other path is to store the non-processed digital sounds in the built-in memory. And these non-processed digital signals are used for analysis in the disclosure.

[0036] Step 202 is deciding at least one segment from the physiological sound by the VAD module 121 of the feature extraction module 120. Because heart sounds centralize in low frequency, the sampling rate is set as 5 kHz. Training data 111 was recorded by 17 healthy people (11 males and 5 females). After extracting heart sounds S1 and S2 by artificial selection, totally 322 S1 and 313 S2 were obtained. Testing data 112 was recorded by 4 healthy people (3 males and 1 female). After processing VAD module 121 of the feature extraction module 120, 122 heart sounds were cut, wherein 66 heart sounds were S1 and 56 heart sounds were S2.

[0037] Step 203 is transferring the at least one segment to the at least one MFCC feature vector by the MFCC module 122 of the feature extraction module 120. MFCC feature vectors extracted by the MFCC module 122 were extended from 13 dimensions to 39 dimensions.

[0038] Step 204 is finding at least one representative data point from the at least one MFCC feature vector by the K-means algorithm module 123 of the feature extraction module 120. In each clip of heart sound, the K-means algorithm module 123 used 2 central vectors to represent the heart sound and noise parts.

[0039] Step 205 is classifying at least one feature to identify at least one category by the classifier 130, wherein the classifier 130 comprises the KNN module 131, the GMM module 132, the SVM module 133, or the DNN module 134. Euclidean metric is the distance calculation method for the KNN module 131. In the SVM module 133, Gaussian radial basis function serves as the kernel function. S1 model and S2 model in the GMM module 132 each separately uses 8 mixed numbers. 3 hidden layers are set in the DNN module 134 and each layer has 100 neurons. Dropout was 70%.

[0040] Step 206 is comparing the at least one category with a normal physiological sound and/or an abnormal phys-

iological sound by the comparing module 140 for evaluating a risk of disease. The normal heart sounds or abnormal heart sounds are recorded by the receiving module 110. After extracting heart sounds S1 and S2 by artificial selection, heart sounds S 1 and S2 are trained to be training data.

Example 2 The accuracy result of heart sounds by different classifiers

[0041] In MFCC feature extraction of training data, the S1 and S2 spectrum and wave pattern map were observed and shown in FIG. 5. First, the frequency of heart sounds mainly concentrated in low frequency parts, which is highly different from voices concentrating below 8k Hz. Therefore, sampling was adjusted to 5k Hz. The baseband of heart sound is about 15 ms, so the frame size was set as 15 ms, with 10 ms overlapped. After adjusting the parameters of the VAD module 121 as the α and β shown in Formula (2), audio files cut manually served as goal to match the testing data processed by the VAD module 121 with training data.

[0042] In classifier part, the KNN module 131, the GMM module 132, the SVM module 133 and the DNN module 134 were used. However, the idea for using the KNN module 131 was relatively simpler, which merely used features to serve as the judgment standard for distances. The GMM module 132 was generation model. Each category was trained under Gaussian model separately. The possibility of testing data in individual model was also calculated. The SVM module 133 was to use linear or nonlinear (reflection) methods to classify training data to obtain the training model. Testing data then were introduced into model to obtain detection results. Finally, the DNN module 134 was the state-of-the-art identification method in recent years, which mimics the multiple layers of learning in human brains to obtain the training model. After introducing testing data, the detection results could be obtained.

Table 2. KNN experimental result

	Precision	Recall	F-measure	Accuracy
S1	85%	77.3%	81%	-
S2	75.8%	83.9%	79.7%	-
Average	-	-	-	80.3%

Table 3. GMM experimental result

	Precision	Recall	F-measure	Accuracy
S1	89.2%	87.9%	88.6%	-
S2	86%	87.5%	86.7%	-
Average	-	-	-	87.7%

Table 4. SVM experiment result

	Precision	Recall	F-measure	Accuracy
S1	96.7%	89.4%	92.9%	-
S2	88.5%	96.4%	92.3%	-
Average	-	-	-	92.6%

Table 5. DNN experiment result

	Precision	Recall	F-measure	Accuracy
S1	96.8%	90.9%	93.8%	-
S2	90%	96.4%	93.1%	-
Average	-	-	-	93.4%

[0043] From Tables 2 to 5 and FIG. 6 were the experiment results for the detections of S1 and S2 by using the proposed

systematic structure. From the experiment results, the SVM module 133 and the DNN module 134 both showed very high recognition rate. The DNN module 134 accuracy could reach 93.4%. S1 owned higher F-measures in four kinds of classifiers.

Example 3 Identifying regular heart sounds and irregular heart sound

[0044] To evaluate the difference between regular and irregular heart sounds, as shown in FIG.7, three pairs of regular heart sounds S1 and three pairs of regular heart sounds S2 were recorded. Using five thousand (5K) sampling rate as the record factor, and 5K sampling rate was 0.2 milliseconds per sample. Each sample of S1 were respectively 1466, 6266, 10941, 15889, and 20564, and each sample of S2 were respectively 2995, 7796, 12608, 17421, and 22242. First heart rate of S1 was $60 / [(16266 - 14661) \times 0.2 \times 0.001] = 62.5$ beats per minute (BPM), and second heart rate of S1 was 64.2 BPM. The modulus difference between first and second heart rates was 1.7 (64.2 BPM - 62.5 BPM) which was smaller than 5% of the first heart rate 3.125 (62.5 BPM x 5%).

[0045] On the contrary, as shown in FIG.8, three pairs of irregular heart sounds S1 and three pairs of irregular heart sounds S2 were recorded. Each sample of S1 were respectively 1162, 8269, 11775, and 17555, and each sample of S2 were respectively 2661, 9630, 13344, and 18957. Second heart rate of S1 was 42.8 BPM, and third heart rate of S1 was 51.9 BPM. The modulus difference between second and third heart rates was 9.1, which was larger than 5% of the second heart rate 2.14.

[0046] As shown in FIG. 1, the regular and irregular heart sounds were received by the receiving module 110 and two segments of each pair of S 1 and S2 were detected by VAD module 121 of the feature extraction module 120, further calculating each heart rate and regularity. The modulus of differences between two heart rates is smaller than 5% of the prior heart rate, the heart rate is regular, and the modulus of differences between two heart rates is larger than 5% of the prior heart rate, the heart rate is irregular. The above method can identify the regular and irregular heart rate within two heart rate cycles.

[0047] As shown in Figs. 1 and 2, the method combines the DNN module 134 voiceprint analysis and the VAD module 121, the MFCC module 122 to detect heart sounds S1 and S2. Because this method and system do not need to rely on the time interval information of S1 and S2, Arrhythmia problems could be effectively solved. The K-means algorithm module 123 is then used to represent heart sounds and noises. Simultaneously comparing precision, recall, F-measure and accuracy usually used in classification, the experiment results showed that both S1 and S2 showed excellent detection results in each experiment and S1 showed higher correction rate.

[0048] In actual clinical use, the concept that S1 and S2 appear in pair or individually could be introduced in extracting heart sound with the VAD module 121. Then, the characteristics of S1 to have better recognition rate could be utilized to obtain S2 according to the order detections. Also, the third heart sound (S3) and the fourth heart sound (S4) are the same concept and could be recognized by using the present invention as well.

Example 4 Identifying regular and irregular lung sounds

[0049] As shown in Figs. 1 and 4, the same steps were processed to identify regular and irregular lung sounds. The place for receiving audios was on the sixth intercostal space of the left chest area. The key step to identify different physiological sounds is detecting the segment of the physiological sounds by the VAD module 121 of the feature extraction module 120 because the different physiological sounds have their own segments. As shown in FIG. 9, energy differences of normal lung sounds were processed by the VAD module 121 into segments. As shown in Figs. 1 and 10, energy differences of wheezing lung were processed by the VAD module 121 into segments. A wheeze (formally called "sibilant rhonchi" in medical terminology) is a continuous, coarse, whistling sound produced in the respiratory airways during breathing. Wheezing is commonly experienced by persons with asthma attacks. As shown in Figs. 1 and 11, energy difference of crackling lung sounds were processed by the VAD module 121 into segments. Crackles are caused by the "popping open" of small airways and alveoli collapsed by fluid, exudate, or lack of aeration during expiration. Crackles can be heard in patients with pneumonia, or pulmonary edema.

[0050] Even though numerous characteristics and advantages of the present invention have been set forth in the foregoing description, together with details of the structure and features of the invention, the disclosure is illustrative only. Changes may be made in the details, especially using the system or the method for recognizing physiological sound such as the bowel sound, vascular sound, tracheal breath sound, bronchioles breath sound or extremities sound also within the principles of the invention to the full extent indicated by the broad general meaning of the terms in which the appended claims are expressed.

Claims

1. A system for recognizing physiological sound, comprising: a receiving module configured to receive a physiological sound; a feature extracting module configured to extract at least one feature from the physiological sound; and a classifier configured to classify at least one feature to identify at least one category.
2. The system for recognizing physiological sound according to claim 1, wherein the receiving module is a physiological recording device converting an analog signal of the physiological sound into a digital signal of the physiological sound.
3. The system for recognizing physiological sound according to claim 2, wherein the physiological recording device is an electronic stethoscope.
4. The system for recognizing physiological sound according to claim 1, wherein the feature extracting module comprises: a voice activity detector (VAD) module configured to detect at least one segment from the physiological sound; and a Mel-frequency cepstrum coefficients (MFCC) module configured to transfer the at least one segment to at least one MFCC feature vector.
5. The system for recognizing physiological sound according to claim 4, wherein the system further comprises: a K-means algorithm module configured to find at least one representative data point from the at least one MFCC feature vector.
6. The system for recognizing physiological sound according to claim 1, wherein the classifier includes a supervised classifier.
7. The system for recognizing physiological sound according to claim 6, wherein the supervised classifier includes a K-nearest neighbor (KNN) module, a Gaussian mixture model (GMM) module, a support vector machine (SVM) module, or a deep neural network (DNN) module.
8. The system for recognizing physiological sound according to claim 1, wherein the physiological sound includes heart sound, lung sound, bowel sound, vascular sound, tracheal breath sound, bronchioles breath sound or extremities sound.
9. The system for recognizing physiological sound according to claim 8, wherein the heart sound comprises first heart sound (S1), second heart sound (S2) or a combination thereof.
10. The system for recognizing physiological sound according to claim 1, wherein the system further comprises: a comparing module configured to compare the at least one category with a normal physiological sound and/or an abnormal physiological sound for evaluating a risk of disease.
11. The system for recognizing physiological sound according to claim 1, wherein the system further comprises an automated external defibrillator, a Holter monitor, a Cardiopulmonary Resuscitation (CPR) machine, a pacemaker, an implantable cardioverter defibrillator (ICD), an electrocardiogram (EKG), or an ultrasonic wave device.
12. A method for recognizing physiological sound by the system as claimed in claim 1, comprising: receiving a physiological sound by the receiving module; extracting at least one feature from the physiological sound by the feature extraction module; and classifying the at least one feature to identify at least one category by the classifier.
13. The method for recognizing physiological sound according to claim 12, wherein extracting at least one feature from the physiological sound comprises: detecting at least one segment from the physiological sound by a VAD module; and transferring the at least one segment to at least one MFCC feature vector by an MFCC module.
14. The method for recognizing physiological sound according to claim 13, wherein the method further comprises: finding at least one representative data point from the at least one MFCC feature vector by a K-means algorithm module.
15. The method for recognizing physiological sound according to claim 12, wherein the classifier includes a supervised classifier.
16. The method for recognizing physiological sound according to claim 15, wherein the supervised classifier includes

a KNN module, a GMM module, an SVM module, or a DNN module.

5 17. The method for recognizing physiological sound according to claim 12, wherein the physiological sound includes heart sound, lung sound, bowel sound, vascular sound, tracheal breath sound, bronchioles breath sound or extremities sound.

18. The method for recognizing physiological sound according to claim 17, wherein the heart sound comprises first heart sound (S1), second heart sound (S2) or a combination thereof.

10 19. The method for recognizing physiological sound according to claim 12, wherein the method further comprises: comparing the at least one category with a normal physiological sound and/or an abnormal physiological sound by a comparing module for evaluating a risk of disease.

15

20

25

30

35

40

45

50

55

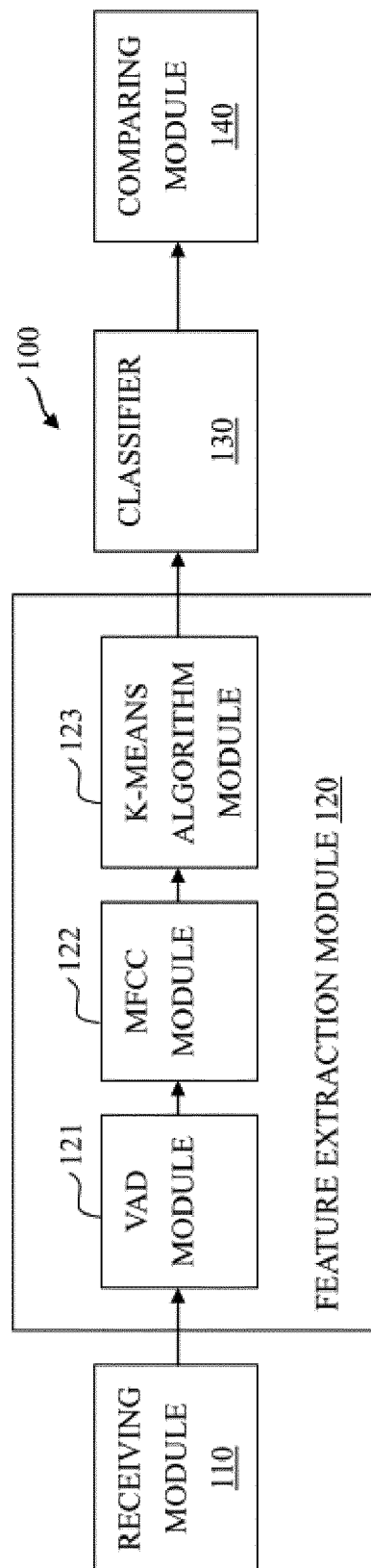


FIG.1

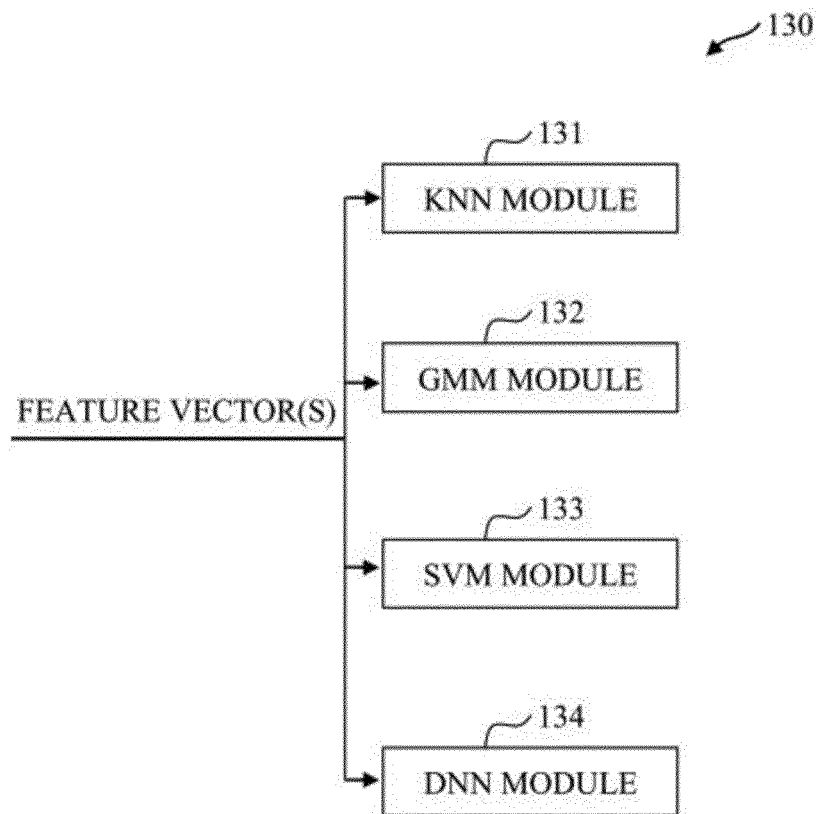


FIG.2

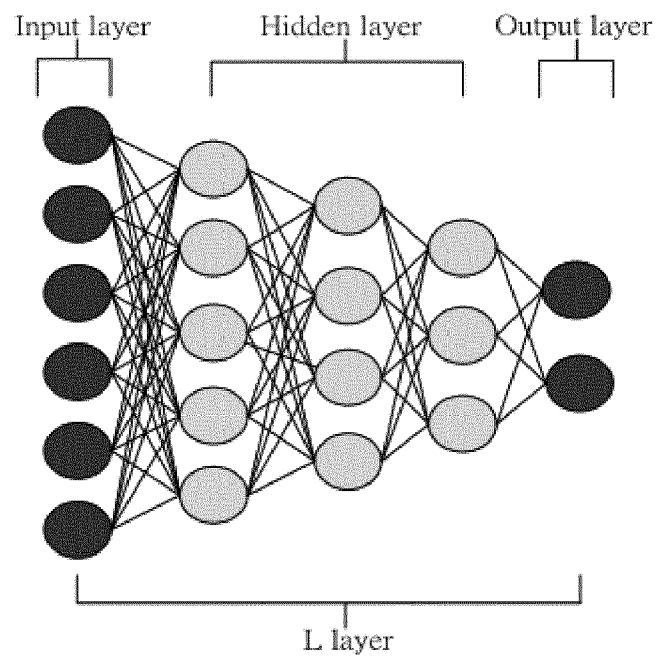


FIG.3

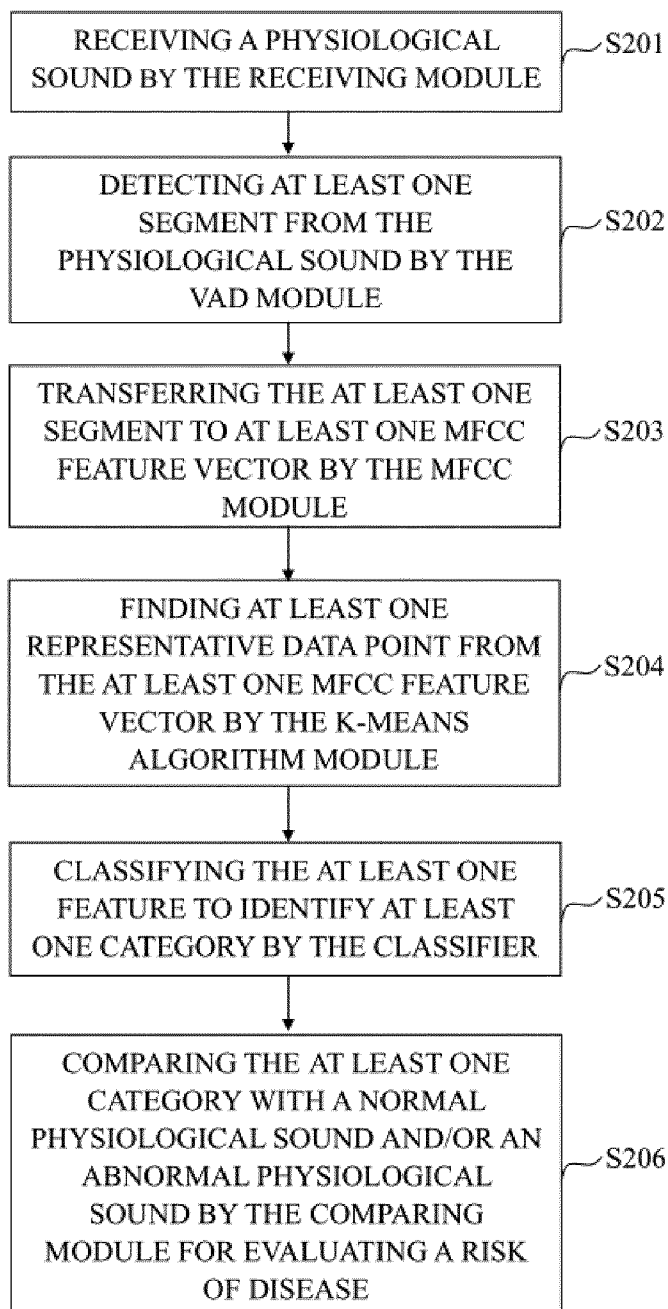


FIG.4

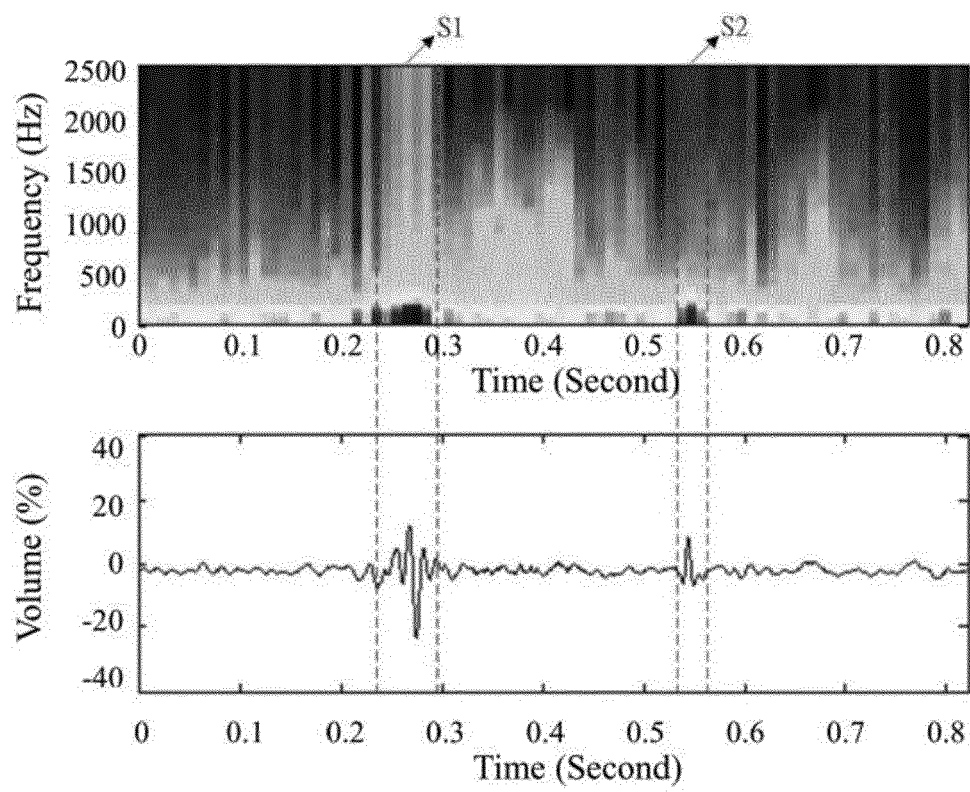


FIG.5

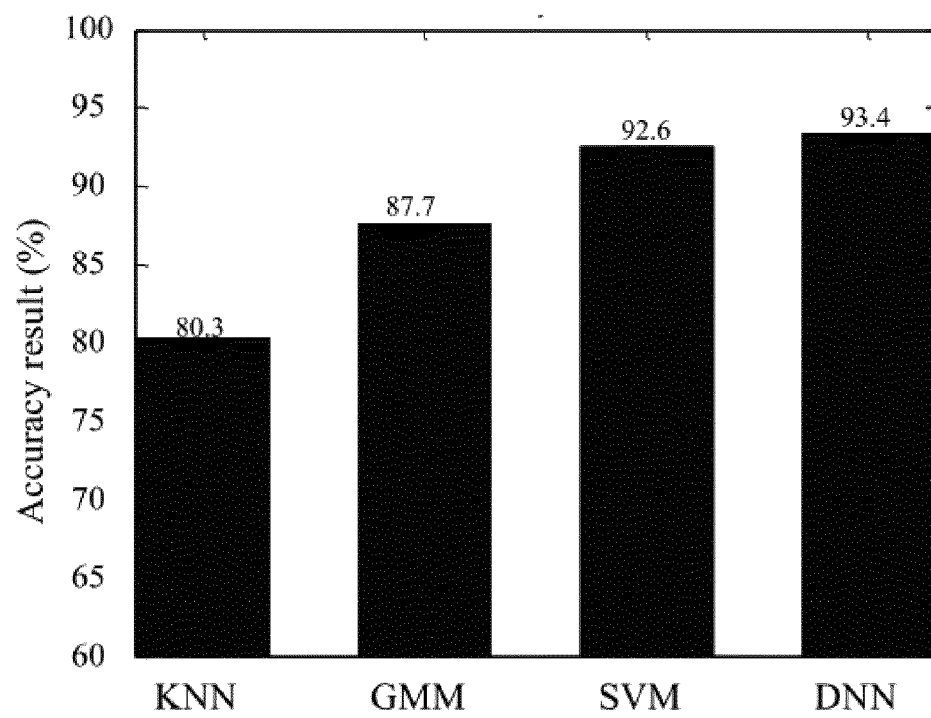


FIG.6

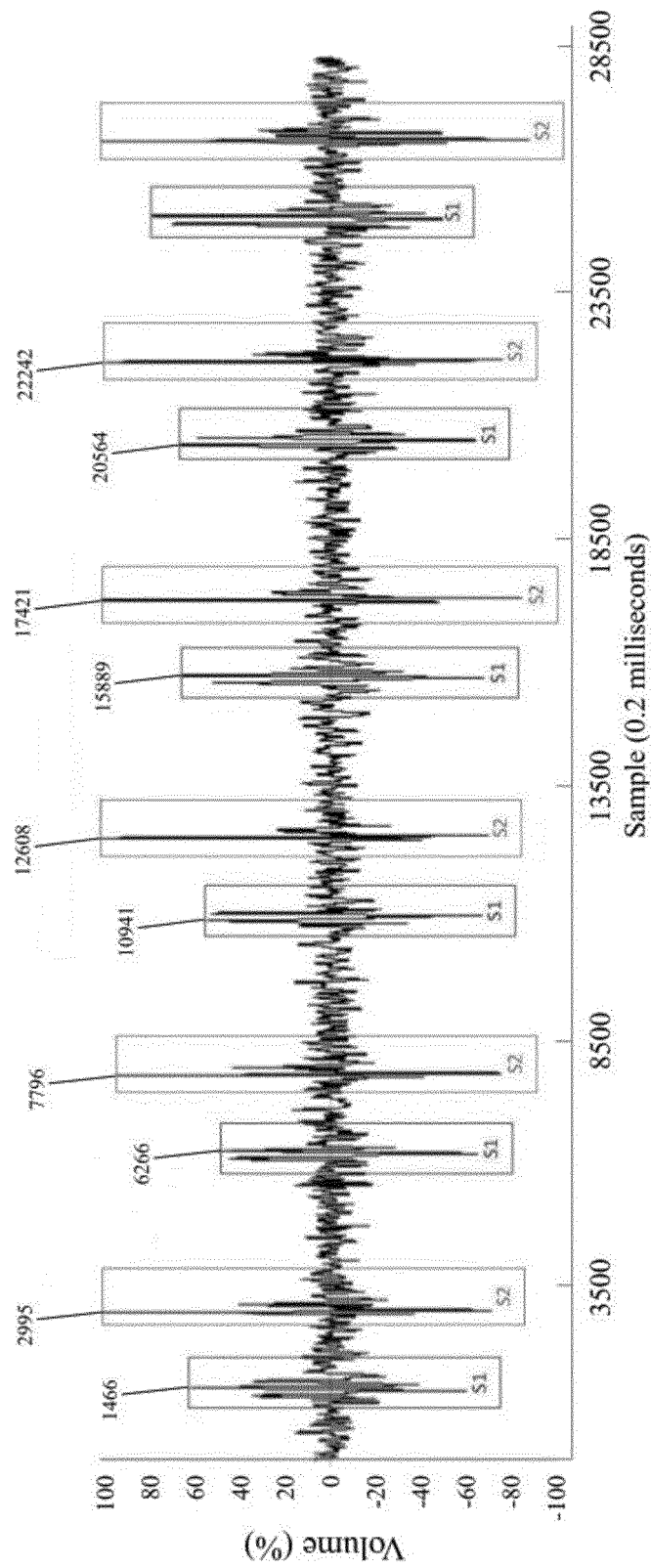


FIG. 7

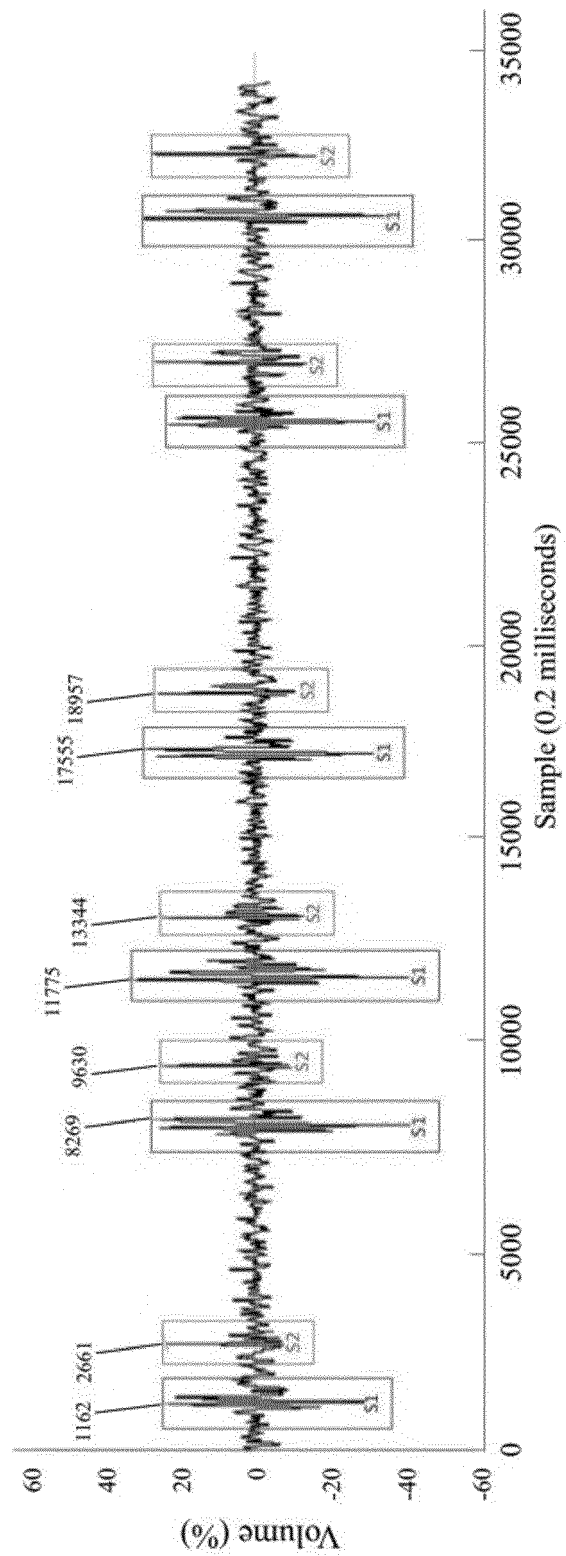


FIG.8

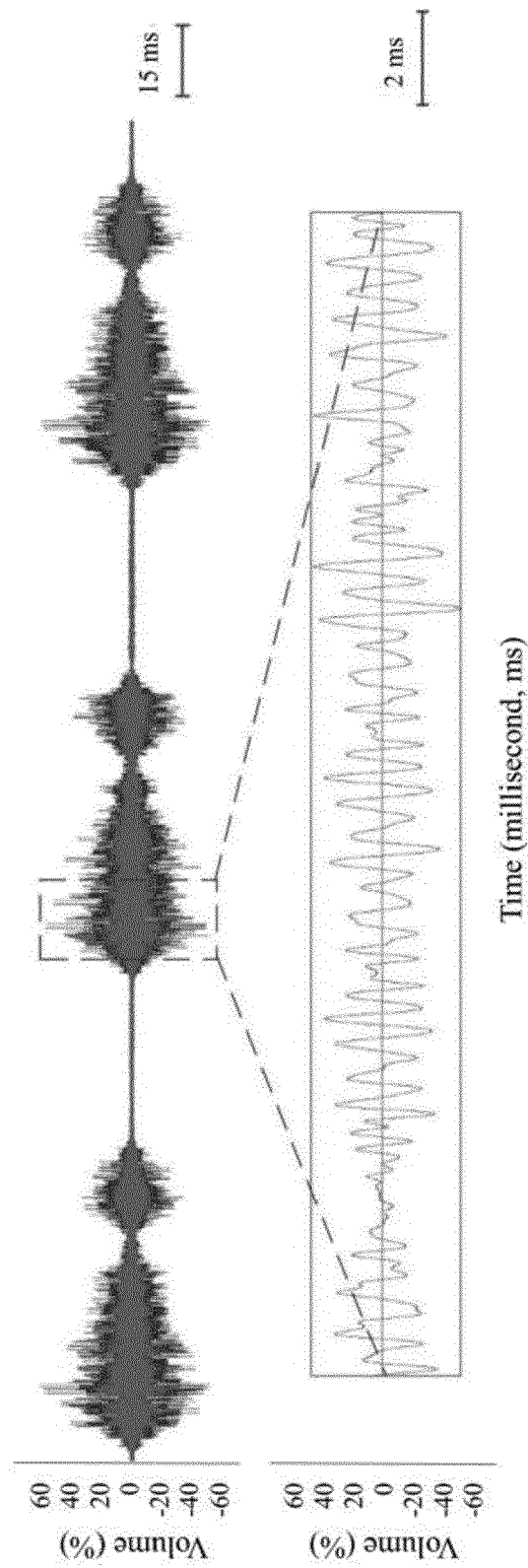


FIG.9

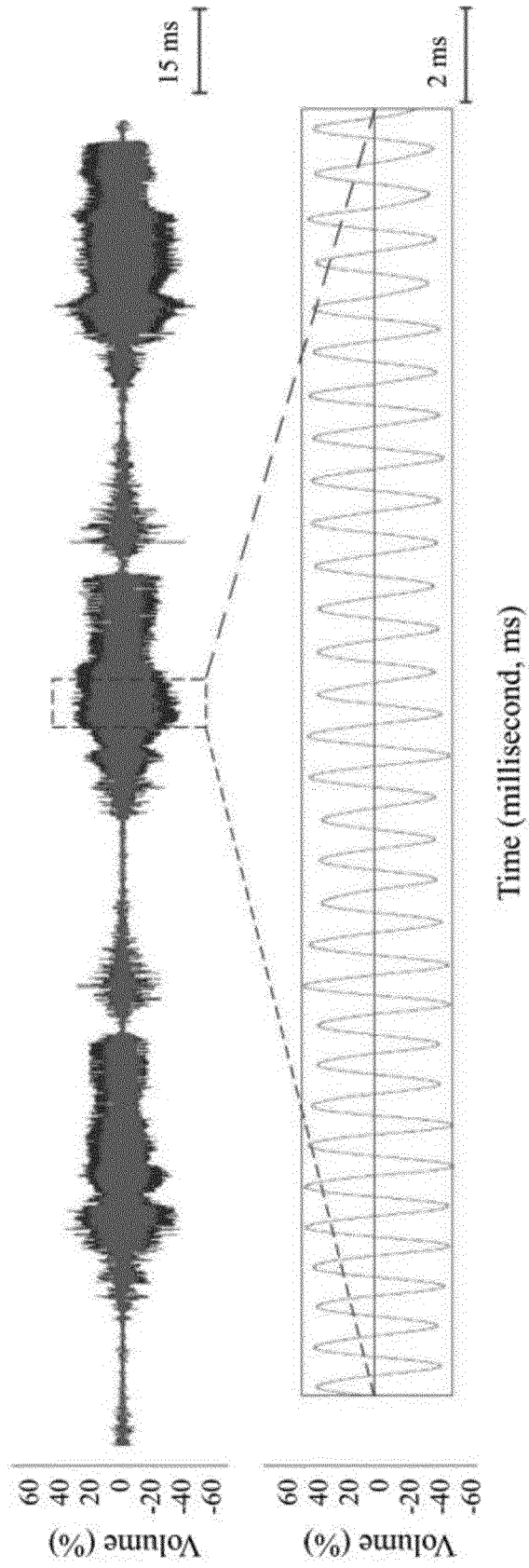
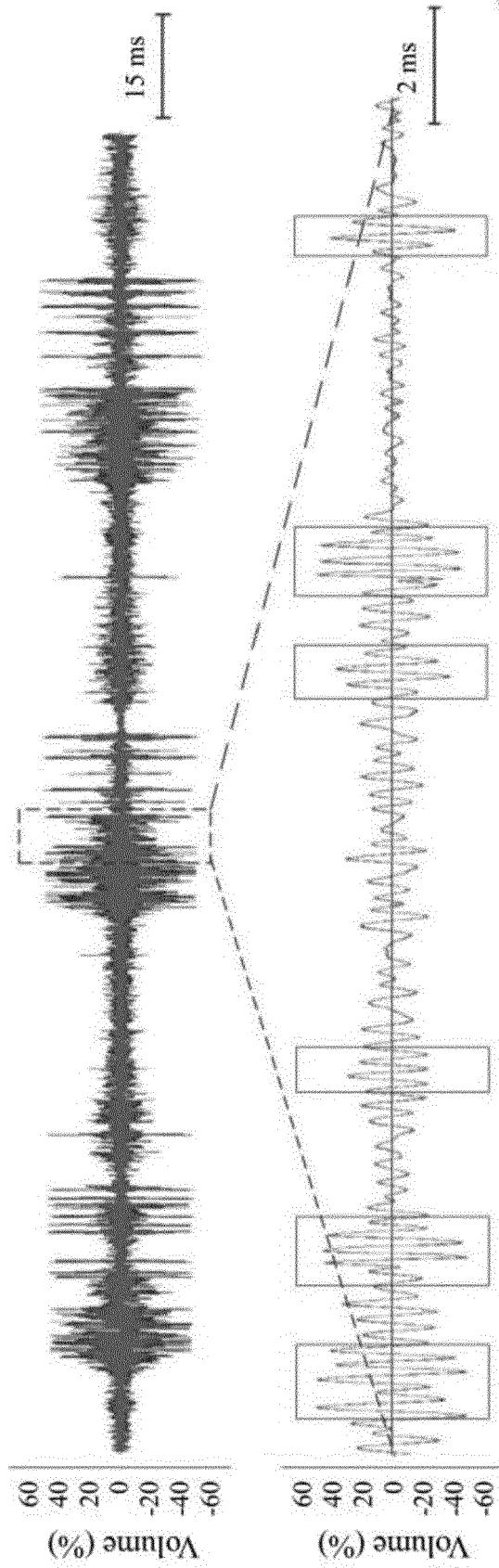


FIG.10



Time (millisecond, ms)

FIG.11



EUROPEAN SEARCH REPORT

Application Number
EP 16 16 6727

5

10

15

20

25

30

35

40

45

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	WO 2008/024624 A2 (SAVIC RES LLC [US]; SAVIC MICHAEL [US]) 28 February 2008 (2008-02-28) * paragraphs [0017] - [0019], [0044], [0046], [0050] * * figure 1 *	1-3, 6-12, 15-19	INV. A61B5/00 A61B5/08 A61B5/11 A61B5/02 A61B8/08 G06F19/00
X	----- CN 102 697 520 B (TIANJIN WOKANG TECHNOLOGY CO LTD) 29 October 2014 (2014-10-29) * paragraphs [0002], [0005], [0018], [0022], [0026], [0031], [0063], [0072] *	1-19	ADD. G06F17/14 A61B5/0205 A61N1/39 A61N1/362 A61B5/0428 A61B5/113
A	----- US 2012/157864 A1 (THAKUR PRAMODSINGH HIRASINGH [US] ET AL) 21 June 2012 (2012-06-21) * paragraph [0033] *	9,11,18	
A	----- US 5 218 969 A (BREDESEN MARK S [US] ET AL) 15 June 1993 (1993-06-15) * the whole document *	1-19	TECHNICAL FIELDS SEARCHED (IPC)
A	----- CN 101 930 734 B (UNIV CHONGQING) 23 May 2012 (2012-05-23) * the whole document *	1-19	A61B G06F
A	----- KR 2011 0127785 A (NAT UNIV CHUNGBUK IND ACAD [KR]) 28 November 2011 (2011-11-28) * the whole document *	1-19	
A	----- WO 2014/172671 A1 (DIGIMARC CORP [US]) 23 October 2014 (2014-10-23) * the whole document *	1-19	
A	----- WO 2005/027720 A2 (BODYMEDIA INC [US]) 31 March 2005 (2005-03-31) * the whole document *	1-19	
	----- -/--		
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 25 October 2016	Examiner Meyer, Wolfgang
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	

EPO FORM 1503 03.82 (P04C01)

50

55



EUROPEAN SEARCH REPORT

 Application Number
 EP 16 16 6727

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	WO 97/36543 A1 (RUSH PRESBYTERIAN ST LUKE [US]) 9 October 1997 (1997-10-09) * the whole document *	1-19	
A	WO 2008/036911 A2 (UNIV NEW JERSEY MED [US]; SEMMLOW JOHN [US]) 27 March 2008 (2008-03-27) * the whole document *	1-19	
A	WO 2006/079062 A1 (3M INNOVATIVE PROPERTIES CO [US]; UNIV MINNESOTA [US]) 27 July 2006 (2006-07-27) * the whole document *	1-19	
A	TAUHIDUR RAHMAN ET AL: "BodyBeat", MOBILE SYSTEMS, APPLICATIONS, AND SERVICES, ACM, 2 PENN PLAZA, SUITE 701 NEW YORK NY 10121-0701 USA, 2 June 2014 (2014-06-02), pages 2-13, XP058049623, DOI: 10.1145/2594368.2594386 ISBN: 978-1-4503-2793-0 * the whole document *	1-19	
			TECHNICAL FIELDS SEARCHED (IPC)
The present search report has been drawn up for all claims			
Place of search		Date of completion of the search	Examiner
The Hague		25 October 2016	Meyer, Wolfgang
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03/02 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 16 16 6727

5

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

25-10-2016

10

15

20

25

30

35

40

45

50

55

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008024624 A2	28-02-2008	EP 2066233 A2	10-06-2009
		US 2008082017 A1	03-04-2008
		US 2009156950 A1	18-06-2009
		WO 2008024624 A2	28-02-2008

CN 102697520 B	29-10-2014	NONE	

US 2012157864 A1	21-06-2012	AU 2011344004 A1	11-07-2013
		CN 103370004 A	23-10-2013
		EP 2651291 A2	23-10-2013
		JP 2014502526 A	03-02-2014
		US 2012157864 A1	21-06-2012
		WO 2012082692 A2	21-06-2012

US 5218969 A	15-06-1993	NONE	

CN 101930734 B	23-05-2012	NONE	

KR 20110127785 A	28-11-2011	NONE	

WO 2014172671 A1	23-10-2014	CN 105393252 A	09-03-2016
		EP 2987106 A1	24-02-2016
		WO 2014172671 A1	23-10-2014

WO 2005027720 A2	31-03-2005	BR PI0414345 A	07-11-2006
		CA 2538710 A1	31-03-2005
		EP 1667579 A2	14-06-2006
		EP 2319410 A1	11-05-2011
		IL 174267 A	30-12-2010
		JP 5174348 B2	03-04-2013
		JP 2007504917 A	08-03-2007
		KR 20060129178 A	15-12-2006
		US 2005113703 A1	26-05-2005
		US 2008161707 A1	03-07-2008
		US 2008171943 A1	17-07-2008
		US 2008177193 A1	24-07-2008
		US 2008183082 A1	31-07-2008
		US 2008183090 A1	31-07-2008
		US 2010286532 A1	11-11-2010
		US 2014094707 A1	03-04-2014
		US 2014221849 A1	07-08-2014
		US 2014221850 A1	07-08-2014
		WO 2005027720 A2	31-03-2005

WO 9736543 A1	09-10-1997	AU 2555897 A	22-10-1997
		CA 2220334 A1	09-10-1997

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 16 16 6727

5 This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

25-10-2016

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		EP 0836415 A1	22-04-1998
		JP H11506970 A	22-06-1999
		US 6056703 A	02-05-2000
		US 6287266 B1	11-09-2001
		US 6425874 B1	30-07-2002
		WO 9736543 A1	09-10-1997

WO 2008036911 A2	27-03-2008	US 2010094152 A1	15-04-2010
		US 2014180153 A1	26-06-2014
		WO 2008036911 A2	27-03-2008

WO 2006079062 A1	27-07-2006	AU 2006206220 A1	27-07-2006
		CA 2595924 A1	27-07-2006
		EP 1855594 A1	21-11-2007
		JP 2008528124 A	31-07-2008
		US 2006167385 A1	27-07-2006
		WO 2006079062 A1	27-07-2006

专利名称(译)	用于识别生理声音的方法和系统		
公开(公告)号	EP3100675A1	公开(公告)日	2016-12-07
申请号	EP2016166727	申请日	2016-04-22
[标]申请(专利权)人(译)	创心医电股份有限公司		
申请(专利权)人(译)	伊梅季PLUS INC.		
当前申请(专利权)人(译)	伊梅季PLUS INC.		
[标]发明人	TSAI KUN HSI TSAO YU KU SHIH HSUAN LIANG TZU CHEN CHANG YUN FAN YANG SHIH I		
发明人	TSAI, KUN-HSI TSAO, YU KU, SHIH-HSUAN LIANG, TZU-CHEN CHANG, YUN-FAN YANG, SHIH-I		
IPC分类号	A61B5/00 A61B5/08 A61B5/11 A61B5/02 A61B8/08 G06F19/00 G06F17/14 A61B5/0205 A61N1/39 A61N1/362 A61B5/0428 A61B5/113		
CPC分类号	A61B7/003 A61B7/04 G10L15/02 G10L15/16 G10L25/18 G10L25/24 G10L25/27 G10L25/63 G10L25/66 G10L25/78 A61B5/02028 A61B5/0205 A61B5/0428 A61B5/0803 A61B5/1107 A61B5/113 A61B5/4255 A61B5/7246 A61B5/7253 A61B5/7264 A61B5/7267 A61B5/7275 A61B5/7282 A61B7/008 A61B8/0883 A61B8/0891 A61B8/5223 A61B2562/0204 A61N1/36578 A61N1/3904 G06F19/3481 G16H50/20 G16H50/30		
审查员(译)	MEYER , WOLFGANG		
优先权	14/729423 2015-06-03 US		
其他公开文献	EP3100675B1		
外部链接	Espacenet		

摘要(译)

一种识别生理声音的系统，包括接收模块，特征提取模块，分类器和比较模块。一种用于识别生理声音的方法包括：由接收模块接收生理声音；通过所述特征提取模块从所述生理声音中提取至少一个特征；分类所述至少一个特征以通过分类器识别至少一个类别；以及通过所述比较模块将所述至少一个类别与正常生理声音和/或异常生理声音进行比较，以评估疾病风险。用于识别生理声音的方法和系统可以精确地识别特定生理声音并排除噪声。

$$\tau = \max_m \sqrt{\frac{\sum_{i=1}^n (x_i(m))^2 - \bar{x}^2}{n}}$$

Formula (1)