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

(11) EP 3 584 573 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
04.01.2023 Bulletin 2023/01

(21) Application number: **17896482.1**

(22) Date of filing: **14.09.2017**

(51) International Patent Classification (IPC):
G01N 29/52 (2006.01) ***G01N 29/46*** (2006.01)
G01N 29/44 (2006.01) ***G06N 3/08*** (2006.01)
G01N 29/14 (2006.01) ***G01H 17/00*** (2006.01)
G06N 3/04 (2006.01)

(52) Cooperative Patent Classification (CPC):
G01N 29/4427; G01H 17/00; G01N 29/14; G01N 29/4445; G01N 29/4472; G01N 29/46; G06N 3/0454; G06N 3/0472; G06N 3/088;
G01N 2291/0258

(86) International application number:
PCT/JP2017/033275

(87) International publication number:
WO 2018/150616 (23.08.2018 Gazette 2018/34)

(54) ABNORMAL SOUND DETECTION TRAINING DEVICE AND METHOD AND PROGRAM THEREFOR

TRAININGSVORRICHTUNG ZUR ERKENNUNG EINES ANOMALEN GERÄUSCHES SOWIE VERFAHREN UND PROGRAMM DAFÜR

DISPOSITIF D'APPRENTISSAGE DE DÉTECTION DE SON ANORMAL ET PROCÉDÉ ET PROGRAMME ASSOCIÉS

(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

(30) Priority: **15.02.2017 JP 2017025865**

(43) Date of publication of application:
25.12.2019 Bulletin 2019/52

(73) Proprietor: **Nippon Telegraph and Telephone Corporation**
Tokyo 100-8116 (JP)

(72) Inventors:
- **KOIZUMI, Yuma**
  Musashino-shi,
  Tokyo 180-8585 (JP)
- **SAITO, Shoichiro**
  Musashino-shi,
  Tokyo 180-8585 (JP)
- **UEMATSU, Hisashi**
  Musashino-shi,
  Tokyo 180-8585 (JP)

(74) Representative: **MERH-IP Matias Erny Reichl Hoffmann**
Patentanwälte PartG mbB
Paul-Heyse-Strasse 29
80336 München (DE)

(56) References cited:
WO-A1-2016/132468 JP-A- 2005 345 154
US-A1- 2011 040 496 US-A1- 2015 219 530

- **KOIZUMI ET AL: "2-7-5 Automatic design of acoustic features for detecting abnormal sounds in machine operation sounds", PROCEEDINGS OF THE ACOUSTICAL SOCIETY OF JAPAN MEETING; AUTUMN MEETING ACOUSTICAL SOCIETY OF JAPAN ; (TOYAMA) : 2016.09.14-16, THE ACOUSTICAL SOCIETY OF JAPAN, 1880-7658, vol. 2016, 31 August 2016 (2016-08-31), pages 365-368, XP009515527,**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

- MARCHI ERIK ET AL: "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks", 2015 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), IEEE, 19 April 2015 (2015-04-19), pages 1996-2000, XP033187110, DOI: 10.1109/ICASSP.2015.7178320 [retrieved on 2015-08-04]
- Sangwan Abhijeet ET AL: "Environmentally Aware Voice Activity Detector", , 27 August 2007 (2007-08-27), pages 2929-2932, XP055779570, Retrieved from the Internet: URL:https://www.isca-speech.org/archive/archive_papers/interspeech_2007/i07_2929.pdf [retrieved on 2021-02-24]
- KOIZUMI, YUUMA et al.: "Automatic design of acoustic features for detecting abnormal sounds in machine operation sounds", Proceedings of the 2016 Autumn Meeting of the Acoustical Society of Japan, vol. 2016, 31 August 2016 (2016-08-31), pages 365-368, XP009515527,

Description

[TECHNICAL FIELD]

[0001] The present invention relates to training techniques for detecting anomalous waves of anomalous sound or the like from signals such as a sound signal.

[BACKGROUND ART]

[0002] In a factory or the like, even a shutdown of an industrial device installed therein, such as a large-sized manufacturing or molding machine, due to a failure significantly hinders the factory's operation. It is thus necessary to routinely monitor the operational status of the device and immediately take measures in the event of an anomaly. One solution is to regularly dispatch a maintenance person to the field from an industrial device management service to check machine components for wear and the like. Since this requires enormous personnel and/or traveling costs and labor, however, it is difficult to implement this on all industrial devices and/or factories.

[0003] A solution to this is to install a microphone inside a machine so as to routinely monitor the operational sound of the machine. Through analysis of the operational sound, any occurrence of sound that is likely to be an anomaly (i.e., anomalous sound) is detected and an alert is raised, thereby solving the anomaly. However, setting the types of anomalous sound and/or methods for detecting them for every machine type or individual unit is even more expensive than manual monitoring. Thus, there is a need for automated design of rules for automatically detecting anomalous sound.

[0004] As a way to address this problem, anomalous sound detection based on statistical approaches is well known (see Non-patent Literature 1, for instance). Anomalous sound detection based on statistical approaches is generally classified into supervised anomalous sound detection and unsupervised anomalous sound detection. In supervised anomalous sound detection, a classifier is trained from training data for normal sound and anomalous sound; whereas in unsupervised anomalous sound detection, a classifier is trained only from training data for normal sound. For an industrial application, unsupervised anomalous sound detection is often used because training data for anomalous sound is difficult to collect.

[0005] A training/detection flow for unsupervised anomalous sound detection is as shown in Fig. 7. During training, an acoustic feature obtained from sound data (training data) at the time of normal operation is extracted. Thereafter, a normal sound model (a probability density function) is trained from the acoustic feature. Then, during determination, an acoustic feature is extracted in relation to a newly obtained observation, and a negative logarithmic likelihood (i.e., degree of anomaly) is evaluated against an already trained normal sound model. If the value is smaller than a threshold, it is determined to be normal; and if the value is larger than the threshold, it is determined to be anomalous. In other words, this is evaluation of how well an observed sound fits the normal sound model. It is based on the idea that sound "resembling" training data for normal sound should be produced if the observation represents normal sound and sound "not resembling" the training data for normal sound should be produced if the observation represents anomalous.

[0006] Description based on formulas is provided so as to further embody Fig. 7. The issue of anomalous sound detection is an issue for determining whether an observation signal $X_{\omega,\tau} \in C^{\Omega \times T}$ is normal or anomalous. Here, $\omega \in \{1, ..., \Omega\}$ and $\tau \in \{1, ..., T\}$ are indices of frequency and time respectively.

[0007] An acoustic feature $f_\tau \in R^D$ is first extracted from an observation signal.

$$f_\tau = F(x_\tau) \qquad (1)$$

[0008] Here, F represents a feature amount extraction function. Further, x_τ represents a vector in which a plurality of $X_{\omega,\tau}$ required for extracting an acoustic feature are lined up and is set as the following, for example.

$$x_\tau = \left(X_{\tau-P_b}, X_{\tau-P_b+1}, \ldots, X_{\tau+P_f}\right)^{\mathrm{T}} \qquad (2)$$

$$X_\tau = \left(X_{1,\tau}, X_{2,\tau}, \ldots, X_{\Omega,\tau}\right) \qquad (3)$$

[0009] Here, T represents transposition. P_b and P_f respectively represent the number of past frames and the number of future frames included in x_τ. For example, approximately $P_b=P_f=5$ is set.

[0010] Next, a degree of anomaly $L(f_\tau)$ is calculated as follows.

$$L(f_\tau) = -\ln p(f_\tau | z = 0) \qquad (4)$$

[0011] Here, $p(f_\tau|z=0)$ represents a normal sound model. Further, z is an indicator which becomes z=0 when $X_{\omega,\tau}$ represents normal sound and becomes $z \neq 0$ when $X_{\omega,\tau}$ represents anomalous sound. At the end, if a value of $L(f_\tau)$ is greater than a threshold φ, it is determined to be anomalous; and if the value of $L(f_\tau)$ is smaller than the threshold φ, it is determined to be normal.

$$\hat{R}_\tau = H(L(f_\tau), \varphi) = \begin{cases} 0\ (normal) & L(f_\tau) \le \varphi \\ 1\ (abnormal) & L(f_\tau) > \varphi \end{cases} \qquad (5)$$

[0012] Here, $H(L_\tau,\varphi)$ represents an anomalous determination function.

[PRIOR ART LITERATURE]

[NON-PATENT LITERATURE]

[0013]

Non-patent Literature 1: Tsuyoshi Ide and Masashi Sugiyama, "Anomaly Detection and Change Detection", Kodansha, pp. 6-7, 2015.

Non-patent Literature: KOIZUMI ET AL: "2-7-5 Automatic design of acoustic features for detecting abnormal sounds in machine operation sounds", ROCEEDINGS OF THE ACOUSTICAL SOCIETY OF JAPAN MEETING, THE ACOUSTICAL SOCIETY OF JAPAN, 1880-7658, vol. 2016, 31 August 2016 (2016-08-31), pages 365-368.

Non-patent Literature: ARCHI ERIK ET AL: "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks",2015 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), IEEE, 19 April 2015 (2015-04-19), pages 1996-2000.

Non-patent Literature: Sangwan Abhijeet ET AL: "Environmentally Aware Voice Activity Detector",, 27 August 2007 (2007-08-27), pages 2929-2932.

[PATENT LITERATURE]

[0014] US 2015/219530 A1 (LI WEICHANG [US] ET AL) 6 August 2015.

[SUMMARY OF THE INVENTION]

[PROBLEMS TO BE SOLVED BY THE INVENTION]

[0015] A challenge in unsupervised anomaly detection is design of a feature amount extraction function F(·). In supervised anomalous sound detection, an acoustic feature that can correctly identify a target of determination is manually designed. For example, if it is known that the normal sound is a sinusoidal wave of 1000 Hz and the anomalous sound is a sinusoidal wave of 2000 Hz, a log power of a Mel filter bank (log-MFBO) is extracted per frame because the two sounds have different tone timbres. If the normal sound is steady engine noise and the anomalous sound is a "bumping" sound produced by hitting of devices, a temporal difference of the power of a Mel filter bank (ΔMFBO) is extracted because the anomalous sound is a sporadic sound. "Deep training", a type of supervised training, is said to be able to automatically design an acoustic feature from training data.

[0016] In unsupervised anomaly detection, however, it cannot be known anomalous sound with what kind of sound characteristics will occur. It is accordingly hard to manually design the feature amount extraction function and use of deep training is also difficult. For example, if the anomalous sound is assumed to be a sinusoidal wave of 2000 Hz and log-MFBO is used as the acoustic feature due to the fact that the normal sound was a sinusoidal wave of 1000 Hz, anomalous sound like a "bumping" sound produced by hitting of devices cannot be detected. The opposite is also true. This has forced use of, for example, Mel filter bank cepstrum coefficient (MFCC), which is a general-purpose sound feature amount, resulting in low accuracy of detection compared to supervised training.

[0017] An object of the present invention is to provide an anomalous sound detection training apparatus that is capable of generating a feature amount extraction function for anomalous signal detection irrespective of whether training data for anomalous signals is available or not and a method and a program therefor.

[MEANS TO SOLVE THE PROBLEMS]

[0018] The present invention provides an anomalous sound detection training apparatus according to claim 1, an anomalous sound detection training method according to claim 3 and a program according to claim 4.

[EFFECTS OF THE INVENTION]

[0019] A feature amount extraction function for anomalous signal detection can be generated irrespective of whether training data for anomalous signals is available or not. Further, anomalous sound detection, degree-of-anomaly calculation, anomalous sound generation, anomalous sound detection training, anomalous signal detection, and anomalous signal detection training can be performed by using this feature amount extraction function.

[BRIEF DESCRIPTION OF THE DRAWINGS]

[0020]

Fig. 1 is a block diagram for describing an example of an anomalous sound detection training apparatus.
Fig. 2 is a flowchart for describing an example of an anomalous sound detection training method.
Fig. 3 is a block diagram for describing an example of an anomalous sound detection apparatus.
Fig. 4 is a flowchart for describing an example of an anomalous sound detection method.
Fig. 5 is a diagram for describing extraction of feature amounts and an image of distribution of the feature amounts.
Fig. 6 is a diagram for describing an intuitive image of a training procedure.
Fig. 7 is a diagram for describing a conventional technique.

[DETAILED DESCRIPTION OF THE EMBODIMENTS]

[Technical Background]

(Neyman-Pearson-type optimization index)

[0021] Unsupervised anomalous sound detection can be considered as a kind of hypothesis testing using null hypothesis and alternative hypothesis described below.

[0022] Null hypothesis: x_τ is a sample generated from $p(x|z=0)$.

[0023] Alternative hypothesis: x_τ is not a sample generated from $p(x|z=0)$.

[0024] Accordingly, it can be considered that an anomalous sound detection rate can be maximized by optimizing a feature amount extraction function in accordance with the theory of the hypothesis testing.

[0025] According to the Neyman Pearson theorem (see Reference Literature 1, for instance), it is known that the most powerful hypothesis testing function is a function which maximizes a true positive rate (TPR) when a false positive rate (FPR) is set as ρ. Here, a FPR and a TPR can be calculated by the following formulas. A false positive rate is a probability that normal sound is incorrectly detected as anomalous sound. On the other hand, a true positive rate is a probability that anomalous sound is detected as anomalous sound. A false positive rate and a true positive rate are also referred to as an error detection rate.

$$FPR(F,\phi)=\int H(L(F(x)),\phi)\,p(F(x),x|z=0)dx \qquad (6)$$

$$TPR(F,\phi)=\int H(L(F(x)),\phi)\,p(F(x),x|z\neq 0)dx \qquad (7)$$

[0026] [Reference Literature 1] J. Neyman, et al., "On the Problem of the Most Efficient Tests of Statistical Hypotheses", Phi. Trans. of the Royal Society, 1933.

[0027] When a threshold at which FPR=p is satisfied is represented as φ_ρ, an objective function to be maximized can be expressed as the following.

$$J=TPR(F,\phi_\rho)+\{\rho-FPR(F,\phi_\rho)\} \qquad (8)$$

[0028] Taking into account ρ as a constant having no relation to F when an issue of variation for maximizing this objective function for F is considered, the optimum feature amount extraction function F can be obtained by the following formula.

$$F\leftarrow \arg\min_F -TPR(F,\phi_\rho)+FPR(F,\phi_\rho) \qquad (9)$$

[0029] In other words, the feature amount extraction function F is set so that FPR(F,φ_ρ) is smaller and TPR(F,φ_ρ) is greater. Here, "FPR(F,φ_ρ) is smaller" corresponds to "a degree of anomaly calculated based on an acoustic feature of normal sound obtained by using the feature amount extraction function F is smaller than the threshold φ_ρ". Further, "TPR(F,φ_ρ) is greater" corresponds to "a degree of anomaly calculated based on an acoustic feature of input anomalous sound obtained by using the feature amount extraction function F is greater than the threshold φ_ρ". Accordingly, it can be said that the feature amount extraction function F is set so that a degree of anomaly calculated based on an acoustic feature of normal sound obtained by using the feature amount extraction function F is smaller than the threshold φ_ρ and a degree of anomaly calculated based on an acoustic feature of input anomalous sound obtained by using the feature amount extraction function F is greater than the threshold φ_ρ.

[0030] Hereinafter, an optimization index of Formula (9) is referred to as a "Neyman-Pearson-type optimization index". An implementing example for optimizing F by using this index is described below.

(Neyman Pearson variational autoencoder)

[0031] Formula (9) is deformed to have an optimizing form by using training data. First, expectation operations for a FPR and a TPR are replaced with arithmetic means of training data. Here, T represents the number of training data.

$$F\leftarrow \arg\min_F -\frac{1}{\mathrm{T}}\sum_{\tau=1}^{\mathrm{T}} H(L(F(x_\tau)),\phi_\rho) + \frac{1}{K}\sum_{k=1}^{K} H(L(F(x_k)),\phi_\rho) \qquad (10)$$

[0032] Here, x_τ and x_k respectively represent training data for normal sound and training data for anomalous sound. However, it is difficult to collect training data for anomalous sound (unsupervised training). Therefore, sampling is performed based on p(F(x),x|z≠0) in unsupervised training.

[0033] A probability distribution p(F(x),x|z≠0) followed by anomalous sound should be known so as to sample anomalous sound. However, information that what kind of anomalous sound is produced is often unknown, and it is accordingly difficult to directly estimate p(F(x),x|z≠0). It is therefore considered that estimation of a probability distribution p(F(x),x) followed by every kind of sound is easier than estimation of p(F(x),x|z≠0) and p(F(x),x) is accordingly estimated.

[0034] As to anomaly detection in mechanical sound of factories, for example, every kind of sound means every kind of mechanical sound recorded in various factories. In other words, every kind of sound is sound which can include normal sound and anomalous sound. More specifically, every kind of sound is sound which is produced in an environment, in which an anomalous sound detection apparatus is used, and which can include normal sound and anomalous sound. According to the Bayes' theorem, p(F(x),x) can be decomposed as follows. Here, "∝" represents proportion.

$$p(F(x),x)=\sum_{i=0}^{\infty} p(F(x),x|z=i)\,p(z=i) \qquad (11)$$

$$= p(F(x),x|z=0)\,p(z=0) + \sum_{i=0}^{\infty} p(F(x),x|z=i)\,p(z=i) \qquad (12)$$

$$\propto p(F(x),x|z=0) + p(F(x),x|z\neq 0) \qquad (13)$$

[0035] Formula (12) is deformed to Formula (13) in an assumption that class prior distribution p(z) is constant. That is, through estimation of p(F(x),x) and p(F(x),x|z=0), a probability distribution followed by sound other than normal sound, in other words, a probability distribution p(F(x),x|z≠0) followed by anomalous sound can be estimated by the following formula.

$$p(F(x),x|z\neq 0) \propto p(F(x),x) - p(F(x),x|z=0) \qquad (14)$$

[0036] A probability distribution followed by anomalous sound, a probability distribution followed by every kind of sound, and a probability distribution followed by normal sound are respectively expressed as p(F(x)|z≠0), p(F(x)), and p(F(x)|z=0) as well. A "probability distribution followed by sound" is also referred to as a "probability distribution modeling sound".

[0037] Thus, it can be said that a probability distribution p(F(x)|z≠0) modeling anomalous sound is a probability distribution obtained by removing a probability distribution p(F(x)|z=0) modeling normal sound from a probability distribution p(F(x)) modeling every kind of sound (sound which can include normal sound and anomalous sound).

[0038] Further, a feature amount extraction function is obtained based on this Formula (14) as described below, so that it can be said that a feature amount extraction function is based on the probability distribution p(F(x)) modeling every kind of sound (sound which can include normal sound and anomalous sound), the probability distribution p(F(x)|z=0) modeling normal sound, and the probability distribution p(F(x)|z≠0) modeling anomalous sound.

[0039] The above-described theory is intuitively illustrated in Fig. 5. Considering a space for feature amounts as the left in Fig. 5, every kind of sound should be widely distributed in the space for feature amounts and normal sound should be distributed in a part of the space. Therefore, anomalous sound corresponds to sound produced with high probability in a distribution of every kind of sound (e.g., mechanical sound that can be produced in real world) and with low probability in a distribution of normal sound (e.g., mechanical sound that is not similar to sound of devices which are monitoring objects).

[0040] A method for estimating p(F(x),x) with high precision is variational autoencoder, for example (see Reference Literature 2, for instance).

[0041] [Reference Literature 2] D. P. Kingma, and M. Welling, "Autoencoding variational Bayes", Proceedings of the International Conference on Learning Representations (ICLR), 2014.

[0042] Though details should be consulted with Reference Literature, variational autoencoder is a method in which a function for generating an observation signal from a latent variable (sound feature amount) f=F(x) (hereinafter, referred to as a "feature amount inverse transformation function"):

$$x = G(F(x)) \qquad (15)$$

is prepared and F and G are optimized so that the following objective function is minimized.

$$F,G \leftarrow \underset{F,G}{\arg\min}\, KL\left[q\left(F(x)|x\right)|p(F(x))\right] - \frac{1}{K}\sum_{k} \ln p\left(x_k \middle| G(f_k^s)\right) \qquad (16)$$

[0043] Here, KL[a|b] represents KL divergence of probability distributions a and b. In the present invention,

$$p(F(x)) = N(0, I_D) \qquad (17)$$

$$q(F(x)|x) = N(F(x), I_D) \qquad (18)$$

are defined for the sake of simplicity. Here, $N(\mu,\Sigma)$ represents a multidimensional normal distribution having a mean vector μ and a covariance matrix Σ, and I_D represents a D dimensional unit matrix. Further, f^s represents a value obtained through sampling from Formula (18) and a probability distribution of the second term of Formula (16) is expressed as:

$$p\left(x \middle| G(f^s)\right) = N\left(x \middle| G(f^s), I_H\right) \qquad (19).$$

[0044] From Formula (17) and Formula (19),

$$p(F(x), x) = p\left(x \middle| F(x)\, p\left(F(x)\right)\right) \qquad (20)$$

$$= p\left(x \middle| G(F(x))\, p\left(F(x)\right)\right) \qquad (21)$$

are obtained. Further, when it is assumed that F and G are deterministic information transformation, p(x|G(F(x)) is constantly a delta function based on Formula (1) and Formula (15). Therefore,

$$p(F(x), x) \propto \sum_{i=0}^{\infty} p\left(F(x)\middle|\, z = i\right) \qquad (22)$$

is obtained. Accordingly, K pieces of sound feature amounts f_K^s of anomalous sound are first generated by

$$f_k^s \sim p(F(x)) - p(F(x) | z = 0) \qquad (23)$$

so as to generate anomalous sound data. "~" in Formula (23) represents that f_k^s follows a probability distribution p(F(x))-p(F(x)|z=0). Then, anomalous sound data x_k may be generated by

$$x_k \leftarrow G(f_k^s) \qquad (24).$$

[0045] Thus, anomalous sound data x_k is generated by using at least the probability distribution p(F(x)) modeling sound which can include normal sound and anomalous sound, the probability distribution p(F(x)|z=0) modeling normal sound, and the feature amount inverse transformation function G.

[0046] Further, considering Formula (14), Formula (23), and Formula (24), it can be said that anomalous sound is generated by sampling an acoustic feature following the probability distribution p(F(x))-p(F(x)|z=0) modeling anomalous sound and using the sampled sound feature amount f_k and the feature amount inverse transformation function G.

[0047] Thus, optimization of a feature amount extraction function can be realized by optimizing a feature amount extraction function and a feature amount inverse transformation function while alternately using an optimization index of a variational autoencoder of Formula (16) and the Neyman-Pearson-type optimization index of Formula (10). However, anomalous sound data used for optimizing Formula (10) is generated based on Formula (23) and Formula (24).

(Specific execution example)

[0048] Fig. 6 illustrates an intuitive image of an execution procedure according to the present embodiment. The present embodiment is realized by repeating a four-stage training procedure.

[0049] First, F and G are trained in accordance with an optimization index of a variational autoencoder. Here, F and G can be implemented by a fully-connected multilayer perceptron and a multilayer convolution neural network, for example. Further, in anomalous sound detection of mechanical sound, sound data recorded in various factories and sound data of human beings, for example, may be used as every kind of sound.

[0050] Then, an acoustic feature is extracted from training data $x_\tau (\tau \in \{1, ..., T\})$ of normal sound.

$$f_\tau = F(x_\tau) \qquad (25)$$

[0051] Subsequently, a normal sound model is trained based on the data. For this, a Gaussian mixture distribution:

$$p(F(x) | z = 0) = \sum_{c=1}^{C} w_c N(\mu_c, \Sigma_c) \qquad (26),$$

for example, can be used. Here, C represents the number of mixtures, and w_c, μ_c, and Σ_c respectively represent a mixture ratio, a mean vector, and a covariance matrix for the c-th distribution. This training can be realized by using an EM algorithm, for example (see Reference Literature 3, for instance).

[0052] [Reference Literature 3] Sadanori Konishi, "Introduction to Multivariate Analysis, Appendix C: EM algorithm", pp. 294-298, Iwanami Shoten, 2010.

[0053] At the end, a threshold φ_ρ is determined by using ρ which is a preset FPR. For this, the ρ_T-th degree of anomaly, which is obtained such that degrees of anomaly L(F(x)) are calculated by using all training data for normal sound and the calculated degrees of anomaly L(F(x)) are sorted in a descending order, may be used.

[0054] Then, anomalous sound data is generated based on Formula (23) and Formula (24). In order to more simply generate f_k^s, the following procedures 1. to 3. may be used. Based on these procedures 1. to 3., f_k^s following Formula (23) can be generated. Thus, f_k^s following Formula (23) may be generated by setting a value $\tilde{f}_k^s$ approximating to f_k^s which is generated through these procedures 1. to 3. and follows Formula (23) as f_k^s.

1. $\tilde{f}_k^s$ is generated from p(F(x)).
2. A degree of anomaly $L(\tilde{f}_k^s)$ is calculated.
3. If the degree of anomaly $L(\tilde{f}_k^s)$ is greater than φ_ρ, $f_k^s \leftarrow \tilde{f}_k^s$ is defined. If the degree of anomaly $L(\tilde{f}_k^s)$ is smaller than φ_ρ, $\tilde{f}_k^s$ is discarded, returning to 1.

[0055] Thus, anomalous sound may be generated by using the probability distribution p(F(x)) modeling sound which can include normal sound and anomalous sound, the feature amount inverse transformation function G which is an inverse function of a feature amount extraction function, and the threshold φ_ρ.
[0056] At the end, F is updated by using the Neyman-Pearson-type optimization index of Formula (10). When F is implemented by a multilayer perceptron or the like, the error back propagation method may be employed.

[Anomalous sound detection training apparatus and method]

[0057] As exemplarily shown in Fig. 1, an anomalous sound detection training apparatus includes a frequency domain conversion unit 1, an initialization unit 2, a first function updating unit 3, an acoustic feature extraction unit 4, a normal sound model updating unit 5, a threshold updating unit 6, an anomalous sound data sampling unit 7, and a second function updating unit 8. An anomalous sound detection training method is implemented by the units of the anomalous sound detection training apparatus executing the processing at step S1 to S9 described in Fig. 2 and below.
[0058] The anomalous sound data sampling unit 7 is an anomalous sound generation apparatus as well.
[0059] Normal sound data and every kind of sound data are input to the anomalous sound detection training apparatus. These sampling frequencies are appropriately set in accordance with a property of sound desired to be analyzed. The sampling frequency is set to approximately 16 kHz, for example.
[0060] Further, it is assumed that parameters of a feature amount extraction function, a feature amount inverse transformation function, and a normal sound model are set. As for a multilayer perceptron, for example, the number of layers of an intermediate layer and the number of hidden units are input. For a normal sound model, the number of mixtures is input in the case of a Gaussian mixture distribution. Further, the number of dimensions of a feature amount D=16 and p=0.05, approximately, may be set, for example.

<Frequency domain conversion unit 1>

[0061] The frequency domain conversion unit 1 converts each of input training data for normal sound and input every kind of sound data into frequency domains (step S1). Short-time Fourier transformation or the like may be employed for the conversion. At this time, the Fourier transformation length may be set at approximately 512 points and the shift length may be set at approximately 256 points, for example.
[0062] The training data for normal sound which is converted into the frequency domain is input to the acoustic feature extraction unit 4. The training data for normal sound which is converted into the frequency domain is input to the first function updating unit 3.

<Initialization unit 2>

[0063] The initialization unit 2 initializes a feature amount extraction function, a feature amount inverse transformation function, and a normal sound model in accordance with the input parameters (step S2).
[0064] The initialized feature amount extraction function is input to the acoustic feature extraction unit 4. The initialized feature amount extraction function and feature amount inverse transformation function are input to the first function updating unit 3. The initialized normal sound model is input to the normal sound model updating unit 5.

<First function updating unit 3>

[0065] The first function updating unit 3 updates the input feature amount extraction function and feature amount inverse transformation function based on an optimization index of a variational autoencoder of Formula (16), for example (step S3). In other words, the feature amount extraction function is subjected to the first update in the first function updating unit 3 based on the optimization index of the variational autoencoder.
[0066] A probabilistic gradient method, for example, may be used for this update. The batch size (the amount of data used for one update) in this case may be set at approximately 512, for example.
[0067] The updated feature amount extraction function and feature amount inverse transformation function are input to the second function updating unit 8.

<Sound feature amount extraction unit 4>

[0068] The acoustic feature extraction unit 4 extracts an acoustic feature of normal sound based on the input training data for normal sound by using the input feature amount extraction function (step S4).
[0069] The extracted sound feature amount of normal sound is output to the normal sound model updating unit 5 and the second function updating unit 8.
[0070] The first processing by the acoustic feature extraction unit 4 is performed by using the feature amount extraction function initialized by the initialization unit 2. The second and subsequent processing by the acoustic feature extraction unit 4 are performed by using a feature amount extraction function updated by the second function updating unit 8.

<Normal sound model updating unit 5>

[0071] The normal sound model updating unit 5 updates a normal sound model by using the acoustic feature extracted at the acoustic feature extraction unit 4 (step S5). The updated normal sound model is input to the second function updating unit 8.

<Threshold updating unit 6>

[0072] The threshold updating unit 6 obtains a threshold φ_ρ corresponding to the false positive rate p, which has a predetermined value, by using the input training data for normal sound and the input feature amount extraction function (step S6).
[0073] The obtained threshold φ_ρ is input to the anomalous sound data sampling unit 7 and the second function updating unit 8.
[0074] For example, the threshold updating unit 6 calculates degrees of anomaly L(F(x)) by using all training data for normal sound and uses the N-th degree of anomaly from the top among the degrees of anomaly L(F(x)) sorted in a descending order, as the threshold φ_ρ. Here, N represents a predetermined positive integer. N=round(pT) holds, for example. Here, round(·) represents round processing with respect to an integer. · is an arbitrary number.
[0075] The threshold φ_ρ is thus set by using degrees of anomaly obtained from normal sound, for example. More specifically, the threshold φ_ρ is set by using degrees of anomaly obtained from normal sound so that probability that input training data for normal sound is detected as anomalous sound is to have a preset false positive rate (error detection rate) ρ. However, the threshold φ_ρ may be set by using degrees of anomaly obtained from normal sound so that probability that input training data for anomalous sound is detected as anomalous sound is to have a preset true positive rate (error detection rate) ρ.
[0076] The first processing by the threshold updating unit 6 is performed by using the feature amount extraction function initialized by the initialization unit 2. The second and subsequent processing by the threshold updating unit 6 are performed by using a feature amount extraction function updated by the second function updating unit 8.

<Anomalous sound data sampling unit 7>

[0077] The anomalous sound data sampling unit 7 spuriously generates anomalous sound data and samples the anomalous sound data (step S7). The sampled anomalous sound data is input to the second function updating unit 8.
[0078] For example, the anomalous sound data sampling unit 7 spuriously generates anomalous sound data and samples the anomalous sound data by using a feature amount inverse transformation function and the threshold φ_ρ through the above-described procedures 1. to 3.
[0079] Specifically, the anomalous sound data sampling unit 7 generates a value $\tilde{f}_k^s$ obtained by approximating the acoustic feature f_k^s following the probability distribution P(F(x)) modeling sound which can include normal sound and anomalous sound through the procedure 1.
[0080] Then the anomalous sound data sampling unit 7 calculates a degree of anomaly $L(\tilde{f}_k^s)$ based on $\tilde{f}_k^s$ through the procedure 2.
[0081] Subsequently, the anomalous sound data sampling unit 7 compares the calculated degree of anomaly $L(\tilde{f}_k^s)$ with the threshold φ_ρ so as to determine whether or not $\tilde{f}_k^s$ can be accepted as the acoustic feature f_k^s through the procedure 3. If $L(\tilde{f}_k^s)>\varphi_\rho$, the anomalous sound data sampling unit 7 accepts $\tilde{f}_k^s$ as the acoustic feature f_k^s.
[0082] Then, the anomalous sound data sampling unit 7 calculates an output value, which is obtained when $\tilde{f}_k^s$ which is accepted as the acoustic feature f_k^s is input to the feature amount inverse transformation function G, based on Formula (24).
[0083] The anomalous sound data sampling unit 7 thus generates anomalous sound data, for example.
[0084] The anomalous sound data sampling unit 7 may generate anomalous sound data based on Formula (23) and Formula (24) so as to perform sampling for anomalous sound.

[0085] If training data for anomalous sound is available, namely in the case of supervised training, sampling is not performed. That is, the subsequent processing may be performed using the training data for anomalous sound as the sampling result. Needless to say, sampling may be performed in combination.

[0086] An acoustic feature of anomalous sound is input to the second function updating unit 8 as anomalous sound data. Therefore, the anomalous sound data sampling unit 7 may perform sound feature amount extraction processing for extracting an acoustic feature of sampled anomalous sound. The first processing of this sound feature amount extraction processing is performed by using the feature amount extraction function initialized by the initialization unit 2. The second and subsequent processing of this sound feature amount extraction processing are performed by using a feature amount extraction function updated by the second function updating unit 8.

<Second function updating unit 8>

[0087] The second function updating unit 8 updates the feature amount extraction function, which is updated in the first function updating unit 3, based on the Neyman-Pearson-type optimization index of Formula (10) defined by the threshold φ_ρ obtained in the threshold updating unit 6, by using the acoustic feature of normal sound extracted in the acoustic feature extraction unit 4 and the input sound feature amount of anomalous sound (step S8). In other words, the feature amount extraction function is subjected to the second update in the second function updating unit 8 based on an index defined based on an acoustic feature of normal sound, an acoustic feature of anomalous sound, and a threshold. The second function updating unit 8 may update a feature amount inverse transformation function in a similar manner in addition to the feature amount extraction function updated in the first function updating unit 3.

[0088] The updated feature amount extraction function is input to the first function updating unit 3, the acoustic feature extraction unit 4, the threshold updating unit 6, and the anomalous sound data sampling unit 7. When a feature amount inverse transformation function is updated, this updated feature amount inverse transformation function is input to the first function updating unit 3, the acoustic feature extraction unit 4, the threshold updating unit 6, and the anomalous sound data sampling unit 7.

[0089] Also, the feature amount extraction function and the normal sound model as finally updated after repeated control by the control unit 9 are output as a final result of training performed by the anomalous sound detection training apparatus and method.

<Control unit 9>

[0090] The control unit 9 repeatedly performs processing of the first function updating unit 3, the acoustic feature extraction unit 4, the normal sound model updating unit 5, and the second function updating unit 8, and processing of the threshold updating unit 6 and the anomalous sound data sampling unit 7, with a feature amount extraction function updated by the second function updating unit 8 and used as an input. If a feature amount inverse transformation function is further updated by the second function updating unit 8, the control unit 9 repeatedly performs processing of the first function updating unit 3, the acoustic feature extraction unit 4, the normal sound model updating unit 5, and the second function updating unit 8, and processing of the threshold updating unit 6 and the anomalous sound data sampling unit 7, with the feature amount extraction function and the feature amount inverse transformation function which are updated by the second function updating unit 8 and are used as inputs. These repeated processing are performed until the feature amount extraction function and the normal sound model converge (step S9).

[0091] Setting a first convergence condition as that the number of executions of repeated processing reaches a certain number of times (for instance, 1000 times), for example, the control unit 9 performs control so as to repeatedly perform the above-mentioned repeated processing until the first convergence condition is satisfied. The first convergence determination condition may also be a different condition.

[Anomalous sound detection apparatus and method]

[0092] As exemplarily shown in Fig. 3, an anomalous sound detection apparatus includes a spectrum calculation unit 11, an acoustic feature extraction unit 12, a degree-of-anomaly calculation unit 13, and a determination unit 14. An anomalous sound detection method is implemented by the units of the anomalous sound detection apparatus executing the processing at step S11 to S14 described in Fig. 4 and below.

[0093] The degree-of-anomaly calculation unit 13 is also a degree-of-anomaly calculation apparatus.

<Spectrum calculation unit 11>

[0094] Operational sound of a machine for which anomalous sound is to be detected is collected through a microphone. The sampling rate used for this collection is similar to the one used for training. The collected sound signal is input to the spectrum calculation unit 11.

[0095] The spectrum calculation unit 11 obtains an acoustic feature based on the collected sound signal in a similar manner to the frequency domain conversion unit 1 (step S11). The obtained sound feature amount is output to the acoustic feature extraction unit 12.

<Sound feature amount extraction unit 12>

[0096] The acoustic feature extraction unit 12 uses the obtained sound feature amount to extract an acoustic feature of the collected sound signal based on the feature amount extraction function that has been output by the anomalous sound detection training apparatus and method as the final result of training (step S12). In other words, the acoustic feature extraction unit 12 extracts an acoustic feature of input sound by using the feature amount extraction function.

[0097] The extracted sound feature amount is output to the degree-of-anomaly calculation unit 13.

<Degree-of-anomaly calculation unit 13>

[0098] The degree-of-anomaly calculation unit 13 calculates a degree of anomaly $L(F(x))$ which is a negative logarithmic likelihood by using the extracted sound feature amount and a normal sound model that has been output by the anomalous sound detection training apparatus and method as the final result of training (step S13). In other words, the degree-of-anomaly calculation unit 13 calculates a degree of anomaly of input sound by using the extracted sound feature amount.

[0099] The calculated degree of anomaly is output to the determination unit 14.

<Determination unit 14>

[0100] The determination unit 14 outputs "anomalous" if the degree of anomaly of a current frame is equal to or greater than threshold ϕ (step S14). In other words, the determination unit 14 determines whether or not input sound is anomalous sound based on the obtained degree of anomaly and the threshold.

[0101] Although the threshold should be adjusted in accordance with the machine and/or environment in question, it may be set at about 1500, for example.

[0102] As with voice activity detection, "hangover", which suppresses detection errors with heuristic rules, may be used. While the hangover processing to be applied may be any of various types, the hangover processing should be set in accordance with the types of false detection of anomalous sound.

[0103] As an example, musical noise that occurs during noise suppression could be determined as a sporadic anomalous sound. Since sound like a sporadic hitting sound often exhibits change in spectrum shape for 100 ms or more, the degree of anomaly would remain equal to or greater than the threshold for $\lceil$(100/the frame shift width for STFT$\rceil$ frames continuously. However, because in musical noise an anomalous amplitude spectrum value only occurs in the relevant frames, there would be several frames at most where the degree of anomaly remains equal to or greater than the threshold continuously. Thus, a rule for anomaly determination may be set as "outputting "anomalous" if the degree of anomaly remains equal to or greater than the threshold for F_1 frames or more continuously", for example.

[0104] As another example, it is also conceivable that anomalous sound continues for a long period of time with the degree of anomaly slightly below the threshold due to a low volume of the anomalous sound. In such a case, a rule like "detecting a sound as anomalous if a total sum of its degree of anomaly over the last F_2 frames is equal to or greater than ϕ_1" may be added as a determination rule for continuous anomalous sound. Although ϕ_1 should be determined by tuning, it may be set at about $\phi_1=F_2\times(\phi-250)$, for example.

[0105] By detecting anomalous sound from a large-sized manufacturing or molding machine installed in a factory or the like using the anomalous sound detection apparatus and method as described above, faster handling of a failure and/or failure prediction become possible. This can contribute to increased efficiency of an industry, for example, manufacturing industry in particular.

[Program and recording medium]

[0106] In the case of implementing the processing in the anomalous sound detection training apparatus, the anomalous sound detection apparatus, the degree-of-anomaly calculation apparatus, or the anomalous sound generation apparatus with a computer, the processing details of the functions to be provided by the anomalous sound detection training apparatus or the anomalous sound detection apparatus are described by a program. By the computer then executing the program, the processing is implemented on the computer.

[0107] The program describing the processing details may be recorded in a computer-readable recording medium. The computer-readable recording medium may be any kind of medium, for example, a magnetic recording device, an optical disk, a magneto-optical recording medium, and a semiconductor memory.

[0108] In addition, individual processing means may be embodied by execution of a predetermined program on a computer or at least some of their processing details may be implemented in hardware.

[Modifications]

[0109] In addition to being executed chronologically in the order described, the processing described for the anomalous sound detection training apparatus, the anomalous sound detection apparatus, the degree-of-anomaly calculation apparatus, or the anomalous sound generation apparatus may also be executed in parallel or individually depending on the processing ability of the apparatus executing the processing or on any necessity.

Claims

1. An anomalous sound detection training apparatus comprising:

a frequency domain conversion unit (1) adapted to convert each of input training data for normal sound and every kind of sound input data into frequency domains;
a first function updating unit (3) adapted to receive the input training data for normal sound and to update a feature amount extraction function and a feature amount inverse transformation function, the feature amount extraction function and the feature amount inverse transformation function being updated, based on an optimization index of a variational autoencoder;
an acoustic feature extraction unit (4) adapted to extract an acoustic feature of normal sound based on the input training data for normal sound by using the updated feature amount extraction function;
a normal sound model updating unit (5) adapted to update a normal sound model by using the acoustic feature that is extracted;
a threshold updating unit (6) adapted to obtain a threshold φ_ρ corresponding to a false positive rate p, the false positive rate ρ having a predetermined value, by using the input training data for normal sound and the updated feature amount extraction function;
an anomalous sound data sampling unit (7) adapted to generate an acoustic feature of anomalous sound, whereby the acoustic feature of anomalous sound is generated by using probability distribution followed by anomalous sound being a probability distribution obtained by removing a probability distribution modeling normal sound from a probability distribution modeling every kind of mechanical sound and by using the updated inverse transformation function;
a second function updating unit (8) adapted to update the feature amount extraction function that is updated, based on a Neyman-Pearson-type optimization index defined by the threshold φ_ρ that is obtained, by using the acoustic feature of normal sound that is extracted and the acoustic feature of anomalous sound,
the anomalous sound detection training apparatus being adapted to repeatedly perform processing of the first function updating unit (3), processing of the acoustic feature extraction unit (4), processing of the normal sound model updating unit (5),
processing of the second function updating unit (8), processing of the threshold updating unit (6) and processing of the anomalous sound data sampling unit (7) by using the feature amount extraction function, the feature amount extraction function being updated by the second function updating unit (8), as an input.

2. The anomalous sound detection training apparatus according to Claim 1, wherein

the anomalous sound sampling unit (7) generates the acoustic feature of anomalous sound by generating an acoustic feature f_k^s of anomalous sound following Formula (23) below:

$$f_k^s \sim p(F(x)) - p(F(x) | z = 0) \qquad (23)$$

when p(F(x)|z≠0) is set as a probability distribution followed by anomalous sound, p(F(x)) is set as a probability distribution followed by sound which includes normal sound and anomalous sound, and p(F(x)|z=0) is set as a probability distribution followed by normal sound.

3. An anomalous sound detection training method comprising:

a frequency domain conversion step in which a frequency domain conversion unit (1) converts each of input training data for normal sound and every kind of sound input data into frequency domains;
a first function updating step in which a first function updating unit (3) receives the input training data for normal sound and updates a feature amount extraction function and a feature amount inverse transformation function, the feature amount extraction function and the feature amount inverse transformation function being updated, based on an optimization index of a variational autoencoder;
an acoustic feature extraction step in which an acoustic feature extraction unit (4) extracts an acoustic feature of normal sound based on the input training data for normal sound by using the updated feature amount extraction function;
a normal sound model updating step in which a normal sound model updating unit (5) updates a normal sound model by using the acoustic feature that is extracted;
a threshold updating step in which a threshold updating unit (6) obtains a threshold φ_ρ corresponding to a false positive rate ρ, the false positive rate ρ having a predetermined value, by using the input training data for normal sound and the updated feature amount extraction functior
an anomalous sound data generation step in which an anomalous sound data sampling unit generates an acoustic feature of anomalous sound, whereby the acoustic feature of anomalous sound is generated by using probability distribution followed by anomalous sound being a probability distribution obtained by removing a probability distribution modeling normal sound from a probability distribution modeling every kind of mechanical sound and by using the updated inverse transformation function,
a second function updating step in which a second function updating unit (8) updates at least one of the feature amount extraction function that is updated and the feature amount inverse transformation function that is updated, based on a Neyman-Pearson-type optimization index defined by the threshold φ_ρ that is obtained, by using the acoustic feature of normal sound that is extracted and the acoustic feature of anomalous sound,
the anomalous sound detection training method repeatedly performs processing of the first function updating step, processing of the acoustic feature extraction step, processing of the normal sound model updating step, processing of the second function updating, processing of the threshold updating step and processing of the anomalous sound data generation step by using the feature amount extraction function and the feature amount inverse transformation function, the feature amount extraction function and the feature amount inverse transformation function being updated by the first function updating step, as inputs.

4. A program comprising instructions which, when the program is executed by a computer, cause the computer to carry out the method of Claim 3.

Patentansprüche

1. Trainingsvorrichtung zur Erkennung eines anomalen Geräuschs, die aufweist:

eine Frequenzdomäneumwandlungseinheit (1), die ausgebildet ist zum Umwandeln aller Eingangstrainingsdaten für normales Geräusch und jede Art von Geräuscheingangsdaten in Frequenzdomänen;
eine erste Funktionsaktualisierungseinheit (3), die ausgebildet ist zum Empfangen der Eingangstrainingsdaten für normales Geräusch und zum Aktualisieren einer Merkmalsmenge-Extraktionsfunktion und einer inversen Merkmalsmenge-Transformationsfunktion, wobei die Merkmalsmenge-Extraktionsfunktion und die inverse Merkmalsmenge-Transformationsfunktion basierend auf einem Optimierungsindex eines Variations-Autocodierers aktualisiert werden;
eine "akustisches Merkmal"-Extraktionseinheit (4), die ausgebildet ist zum Extrahieren eines akustischen Merkmals eines normalen Geräuschs basierend auf den Eingangstrainingsdaten für normales Geräusch unter Verwendung der aktualisierten Merkmalsmenge-Extraktionsfunktion;
eine "normales Geräusch"-Modellaktualisierungseinheit (5), die ausgebildet ist zum Aktualisieren eines "normales Geräusch"-Modells unter Verwendung des extrahierten akustischen Merkmals;
eine Schwellenwertaktualisierungseinheit (6), die ausgebildet ist zum Erhalten eines Schwellenwerts φ_ρ entsprechend einer Falsch-Positiv-Rate p, wobei die Falsch-Positiv-Rate ρ einen vorgegebenen Wert hat, unter Verwendung der Eingangstrainingsdaten für normales Geräusch und der aktualisierten Merkmalsmenge-Extraktionsfunktion;
eine "anomales Geräusch"-Datenabtasteinheit (7), die ausgebildet zum Erzeugen eines akustischen Merkmals eines anomalen Geräuschs, wobei das akustische Merkmal eines anomalen Geräuschs erzeugt wird unter Verwendung einer Wahrscheinlichkeitsverteilung gefolgt von einem anomalen Geräusch, die eine Wahrscheinlichkeitsverteilung ist, die durch Entfernen einer Wahrscheinlichkeitsverteilung modellierend ein normales Ge-

räusch von einer Wahrscheinlichkeitsverteilung modellierend jede Art von mechanischem Geräusch und unter Verwendung der aktualisierten inversen Transformationsfunktion erhalten wird;
eine zweite Funktionsaktualisierungseinheit (8), die ausgebildet ist zum Aktualisieren der Merkmalsmenge-Extraktionsfunktion, die aktualisiert ist, basierend auf einem Neyman-Pearson-Typ-Optimierungsindex, der durch den Schwellenwert φ_ρ definiert ist, der erhalten wird, unter Verwendung des akustischen Merkmals eines normalen Geräuschs, das extrahiert ist, und des akustischen Merkmals eines anomalen Geräuschs,
wobei die Trainingsvorrichtung zur Erkennung eines anomalen Geräuschs ausgebildet ist zum
wiederholten Durchführen einer Verarbeitung der ersten Funktionsaktualisierungseinheit (3), einer Verarbeitung der "akustisches Merkmal"-Extraktionseinheit (4), einer Verarbeitung der "normales Geräusch"-Modell-Aktualisierungseinheit (5), einer Verarbeitung der zweiten Funktionsaktualisierungseinheit (8), einer Verarbeitung der Schwellenwertaktualisierungseinheit (6) und einer Verarbeitung der "anomales Geräusch"-Datenabtasteinheit (7) unter Verwendung der Merkmalsmenge-Extraktionsfunktion, wobei die Merkmalsmenge-Extraktionsfunktion durch die zweite Funktionsaktualisierungseinheit (8) aktualisiert ist, als einen Eingang.

2. Die Trainingsvorrichtung zur Erkennung eines anomalen Geräuschs gemäß Anspruch 1, wobei

die "anomales Geräusch"-Abtasteinheit (7) das akustische Merkmal eines anomalen Geräuschs durch Erzeugen eines akustischen Merkmals f_k^s eines anomalen Geräuschs erzeugt gemäß folgender Formel (23):

$$f_k^s \sim p(F(x)) - p(F(x) \mid z = 0) \qquad (23)$$

wenn $p(F(x)|z \neq 0)$ als eine Wahrscheinlichkeitsverteilung gefolgt von einem anomalen Geräusch gesetzt ist, $p(F(x))$ als eine Wahrscheinlichkeitsverteilung gefolgt von einem Geräusch gesetzt ist, das ein normales Geräusch und ein anomales Geräusch enthält, und $p(F(x) \mid z=0)$ als eine Wahrscheinlichkeitsverteilung gefolgt von normalem Geräusch gesetzt ist.

3. Trainingsverfahren zur Erkennung eines anomalen Geräuschs, das aufweist:

einen Frequenzdomäneumwandlungsschritt, in dem eine Frequenzdomäneumwandlungseinheit (1) alle Eingangstrainingsdaten für normales Geräusch und jede Art von Geräuscheingangsdaten in Frequenzdomänen umwandelt;
einen ersten Funktionsaktualisierungsschritt, in dem eine erste Funktionsaktualisierungseinheit (3) die Eingangstrainingsdaten für normales Geräusch empfängt und eine Merkmalsmenge-Extraktionsfunktion und eine inverse Merkmalsmenge-Transformationsfunktion aktualisiert, wobei die Merkmalsmenge-Extraktionsfunktion und die inverse Merkmalsmenge-Transformationsfunktion basierend auf einem Optimierungsindex eines Variations-Autocodierers aktualisiert werden;
einen "akustisches Merkmal"-Extraktionsschritt, in dem eine "akustisches Merkmal"-Extraktionseinheit (4) ein akustisches Merkmal eines normalen Geräuschs basierend auf den Eingangstrainingsdaten für normales Geräusch extrahiert unter Verwendung der aktualisierten Merkmalsmenge-Extraktionsfunktion;
einen "normales Geräusch"-Modellaktualisierungsschritt, in dem eine "normales Geräusch"-Modellaktualisierungseinheit (5) ein "normales Geräusch"-Modell unter Verwendung des extrahierten akustischen Merkmals aktualisiert;
einen Schwellenwertaktualisierungsschritt, in dem eine Schwellenwertaktualisierungseinheit (6) einen Schwellenwert φ_ρ entsprechend einer Falsch-Positiv-Rate ρ erhält, wobei die Falsch-Positiv-Rate ρ einen vorgegebenen Wert hat, unter Verwendung der Eingangstrainingsdaten für normales Geräusch und der aktualisierten Merkmalsmenge-Extraktionsfunktion;
einen "anomales Geräusch"-Datenerzeugungsschritt, in dem eine "anomales Geräusch"-Datenabtasteinheit ein akustisches Merkmal eines anomalen Geräuschs erzeugt, wobei das akustische Merkmal eines anomalen Geräuschs erzeugt wird unter Verwendung einer Wahrscheinlichkeitsverteilung gefolgt von einem anomalen Geräusch, die eine Wahrscheinlichkeitsverteilung ist, die durch Entfernen einer Wahrscheinlichkeitsverteilung modellierend ein normales Geräusch von einer Wahrscheinlichkeitsverteilung modellierend jede Art von mechanischem Geräusch und unter Verwendung der aktualisierten inversen Transformationsfunktion erhalten wird;
einen zweiten Funktionsaktualisierungsschritt, in dem eine zweite Funktionsaktualisierungseinheit (8) zumindest eine der Merkmalsmenge-Extraktionsfunktion, die aktualisiert ist, und der inversen Merkmalsmengen-Transformationsfunktion, die aktualisiert ist, basierend auf einem Neyman-Pearson-Typ-Optimierungsindex aktualisiert, der durch den Schwellenwert φ_ρ definiert ist, der erhalten wird, unter Verwendung des akustischen Merk-

mals eines normalen Geräuschs, das extrahiert ist, und des akustischen Merkmals eines anomalen Geräuschs, wobei das Trainingsverfahren zur Erkennung eines anomalen Geräuschs wiederholt durchführt eine Verarbeitung des ersten Funktionsaktualisierungsschritts, eine Verarbeitung des "akustisches Merkmal"-Extraktionsschritts, eine Verarbeitung des "normales Geräusch"-Modell-Aktualisierungsschritts, eine Verarbeitung der zweiten Funktionsaktualisierung, eine Verarbeitung des Schwellenwertaktualisierungsschritts und eine Verarbeitung der "anomales Geräusch"-Datenerzeugungsschritts unter Verwendung der Merkmalsmenge-Extraktionsfunktion und der inversen Merkmalsmengen-Transformationsfunktion, wobei die Merkmalsmenge-Extraktionsfunktion und die inverse Merkmalsmengen-Transformationsfunktion durch den ersten Funktionsaktualisierungsschritt aktualisiert sind, als Eingänge.

4. Ein Programm, das Anweisungen aufweist, die bei Ausführung des Programms durch einen Computer den Computer veranlassen, das Verfahren gemäß Anspruch 3 auszuführen.

Revendications

1. Appareil d'apprentissage à la détection de sons anormaux comprenant :

une unité de conversion de domaine de fréquence (1) adaptée pour convertir chacune des données d'apprentissage d'entrée pour un son normal et chaque type de données d'entrée sonore en domaines de fréquence ;
une première unité de mise à jour de fonction (3) adaptée pour recevoir les données d'apprentissage d'entrée pour un son normal et pour mettre à jour une fonction d'extraction de quantité de fonctionnalité et une fonction de transformation inverse de quantité de fonctionnalité, la fonction d'extraction de quantité de fonctionnalité et la fonction de transformation inverse de quantité de fonctionnalité étant mises à jour, sur la base d'un indice d'optimisation d'un auto-encodeur variationnel ;
une unité d'extraction de fonctionnalité acoustique (4) conçue pour extraire une fonctionnalité acoustique du son normal basée sur les données d'apprentissage entrées pour un son normal en utilisant la fonction d'extraction de quantité de fonctionnalité mise à jour ;
une unité de mise à jour de modèle sonore normal (5) adaptée pour mettre à jour un modèle sonore normal en utilisant la fonctionnalité acoustique qui est extraite ;
une unité de mise à jour de seuil (6) adaptée pour obtenir un seuil φ_p correspondant à un taux de faux positifs p, le taux de faux positifs p ayant une valeur prédéterminée, en utilisant les données d'apprentissage entrées pour un son normal et la fonction d'extraction de quantité de fonctionnalité mise à jour ;
une unité d'échantillonnage de données sonores anormales (7) adaptée pour générer une fonctionnalité acoustique d'un son anormal, moyennant quoi la fonctionnalité acoustique d'un son anormal est générée en utilisant une distribution de probabilité suivie d'un son anormal étant une distribution de probabilité obtenue en supprimant une distribution de probabilité modélisant un son normal d'une distribution de probabilité modélisant chaque type de son mécanique et utilisant la fonction de transformation inverse mise à jour ;
une deuxième unité de mise à jour de fonction (8) adaptée pour mettre à jour la fonction d'extraction de quantité de fonctionnalité qui est mise à jour, sur la base d'un indice d'optimisation de type Neyman-Pearson défini par le seuil φ_p qui est obtenu, en utilisant la fonctionnalité acoustique de son normal qui est extraite et la fonctionnalité acoustique de son anormal,
l'appareil d'apprentissage à la détection de sons anormaux adapté afin d'effectuer de manière répétée le traitement de la première unité de mise à jour de fonction (3), le traitement de l'unité d'extraction de fonctionnalités acoustiques (4), le traitement de l'unité de mise à jour du modèle sonore normal (5), le traitement de la deuxième unité de mise à jour de fonction (8), le traitement de l'unité de mise à jour de seuil (6) et traitement de l'unité d'échantillonnage de données sonores anormales (7), en utilisant la fonction d'extraction de quantité de fonctionnalité,
la fonction d'extraction de quantité de fonctionnalité étant mise à jour par la deuxième unité de mise à jour de fonction (8), en tant qu'entrée.

2. Appareil d'apprentissage à la détection de sons anormaux selon la revendication 1, dans lequel

l'unité d'échantillonnage de sons anormaux (7) génère la fonctionnalité acoustique de son anormal en générant une fonctionnalité acoustique f_k^s de son anormal selon la formule (23) ci-dessous :

$$f_k^s \sim p(F(x)) - p(F(x) \mid z = 0) \qquad (23)$$

lorsque p(F(x)|z≠0) est défini comme une distribution de probabilité suivie d'un son anormal, p(F(x)) est défini comme une distribution de probabilité suivie d'un son qui comprend un son normal et un son anormal, et p(F(x)|z=0) est défini comme une distribution de probabilité suivie d'un son normal.

3. Méthode d'apprentissage à la détection de sons anormaux comprenant :

une étape de conversion de domaine de fréquence dans laquelle une unité de conversion dans le domaine de fréquence (1) convertit chacune des données d'apprentissage d'entrée pour un son normal et chaque type de données d'entrée sonore en domaines de fréquences ;
une première étape de mise à jour de fonction dans laquelle une première unité de mise à jour de fonction (3) reçoit les données d'apprentissage d'entrée pour le son normal et met à jour une fonction d'extraction de quantité de fonctionnalité et une fonction de transformation inverse de quantité de fonctionnalité, la fonction d'extraction de quantité de fonctionnalité et la fonction de transformation inverse de quantité de fonctionnalité étant mises à jour, sur la base d'un indice d'optimisation d'un auto-encodeur variationnel ;
une étape d'extraction de fonctionnalité acoustique dans laquelle une unité d'extraction de fonctionnalité acoustique (4) extrait une fonctionnalité acoustique du son normal sur la base des données d'apprentissage entrées pour un son normal en utilisant la fonction d'extraction de la quantité de fonctionnalité ;
une étape de mise à jour de modèle sonore normal dans laquelle une unité de mise à jour de modèle sonore normal (5) met à jour un modèle sonore normal en utilisant la fonctionnalité acoustique qui est extraite ;
une étape de mise à jour de seuil dans laquelle une unité de mise à jour de seuil (6) obtient un seuil φ_p correspondant à un taux de faux positif p, le taux de faux positif p ayant une valeur prédéterminée, en utilisant les données d'apprentissage entrées pour un son normal et la fonction d'extraction de quantité de fonctionnalité mise à jour ;
une étape de génération de données sonores anormales dans laquelle une unité d'échantillonnage de données sonores anormales génère une fonctionnalité acoustique d'un son anormal, moyennant quoi la fonctionnalité acoustique d'un son anormal est générée en utilisant une distribution de probabilité suivie d'un son anormal étant une distribution de probabilité obtenue en supprimant une distribution de probabilité modélisant le son normal à partir d'une distribution de probabilité modélisant chaque type de son mécanique et en utilisant la fonction de transformation inverse mise à jour,
une deuxième étape de mise à jour de fonction dans laquelle une deuxième unité de mise à jour de fonction (8) met à jour au moins une parmi la fonction d'extraction de quantité de fonctionnalité qui est mise à jour et la fonction de transformation inverse de quantité de fonctionnalité qui est mise à jour, sur la base d'un indice d'optimisation de type Neyman-Pearson défini par le seuil φ_p obtenu, en utilisant la fonctionnalité acoustique du son normal extrait et la fonctionnalité acoustique du son anormal,
le procédé d'apprentissage à la détection de sons anormaux exécute de manière répétée le traitement de la première étape de mise à jour de fonction, le traitement de l'étape d'extraction de fonctionnalité acoustique, le traitement de l'étape de mise à jour du modèle sonore normal, le traitement de la deuxième mise à jour de fonction, le traitement de l'étape de mise à jour de seuil et le traitement de l'étape de génération de données sonores anormales en utilisant la fonction d'extraction de quantité de fonctionnalité et la fonction de transformation inverse de quantité de fonctionnalité, la fonction d'extraction de quantité de fonctionnalité et la fonction de transformation inverse de quantité de fonctionnalité mise à jour par la première étape de mise à jour de fonction, en tant qu'entrées.

4. Programme comprenant des instructions qui, lorsque le programme est exécuté par un ordinateur, amènent l'ordinateur à mettre en œuvre le procédé de la revendication 3.

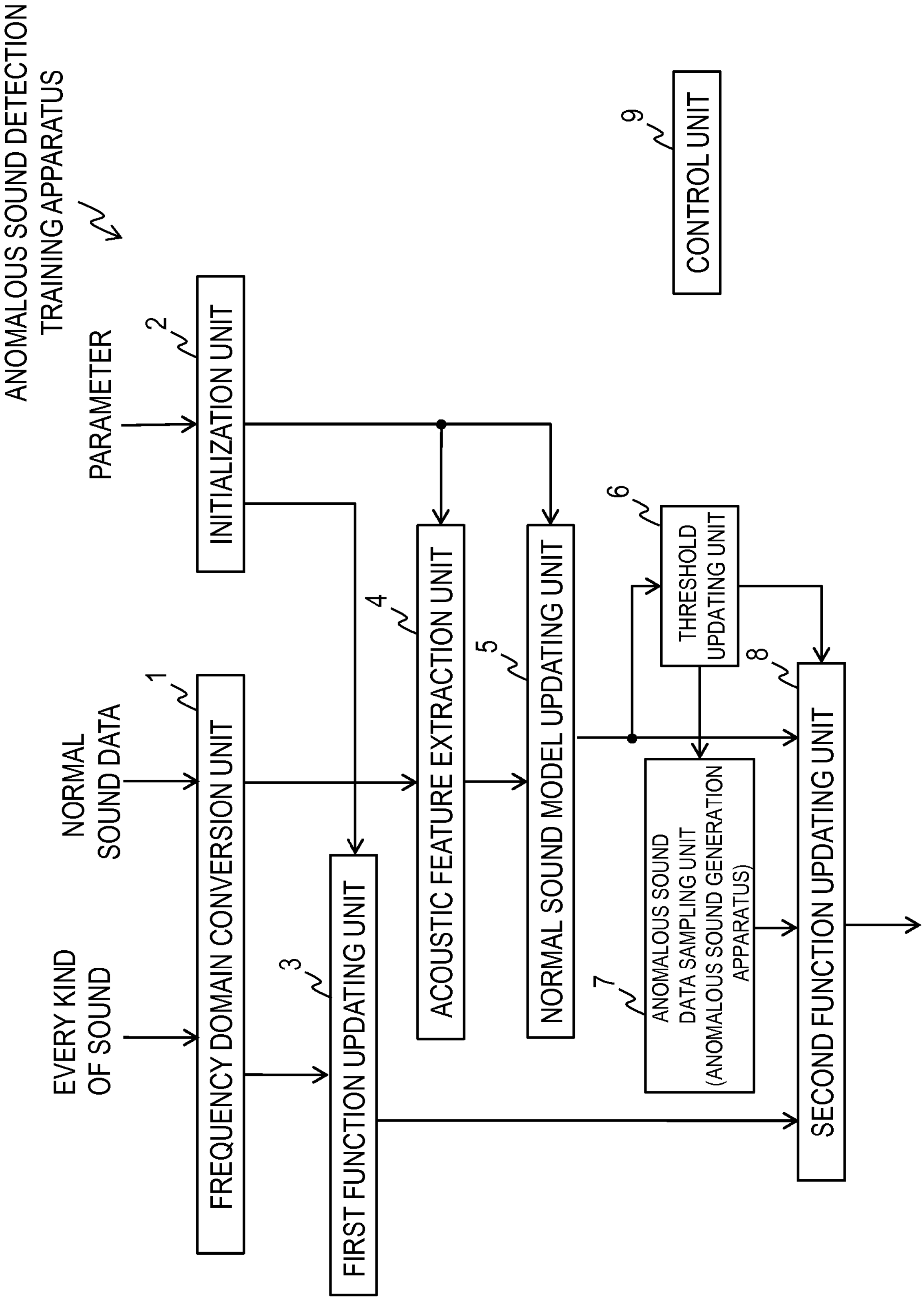
ANOMALOUS SOUND DETECTION TRAINING APPARATUS
PARAMETER
2
INITIALIZATION UNIT
NORMAL SOUND DATA
1
FREQUENCY DOMAIN CONVERSION UNIT
EVERY KIND OF SOUND
3
FIRST FUNCTION UPDATING UNIT
4
ACOUSTIC FEATURE EXTRACTION UNIT
5
NORMAL SOUND MODEL UPDATING UNIT
6
THRESHOLD UPDATING UNIT
7
ANOMALOUS SOUND DATA SAMPLING UNIT (ANOMALOUS SOUND GENERATION APPARATUS)
8
SECOND FUNCTION UPDATING UNIT
9
CONTROL UNIT

FIG. 1

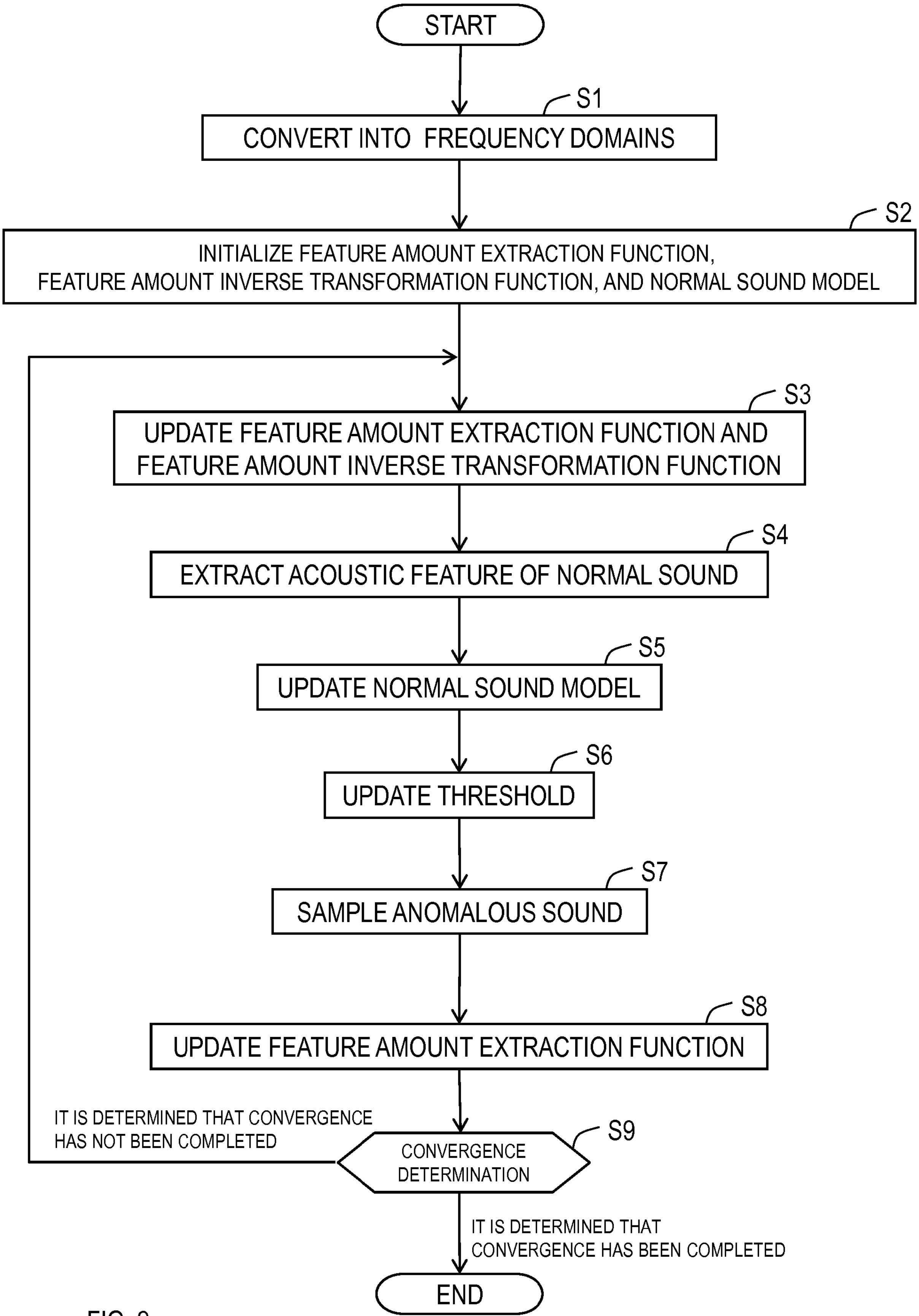
START
S1
CONVERT INTO FREQUENCY DOMAINS
S2
INITIALIZE FEATURE AMOUNT EXTRACTION FUNCTION,
FEATURE AMOUNT INVERSE TRANSFORMATION FUNCTION, AND NORMAL SOUND MODEL
S3
UPDATE FEATURE AMOUNT EXTRACTION FUNCTION AND
FEATURE AMOUNT INVERSE TRANSFORMATION FUNCTION
S4
EXTRACT ACOUSTIC FEATURE OF NORMAL SOUND
S5
UPDATE NORMAL SOUND MODEL
S6
UPDATE THRESHOLD
S7
SAMPLE ANOMALOUS SOUND
S8
UPDATE FEATURE AMOUNT EXTRACTION FUNCTION
IT IS DETERMINED THAT CONVERGENCE
HAS NOT BEEN COMPLETED
S9
CONVERGENCE
DETERMINATION
IT IS DETERMINED THAT
CONVERGENCE HAS BEEN COMPLETED
END

FIG. 2

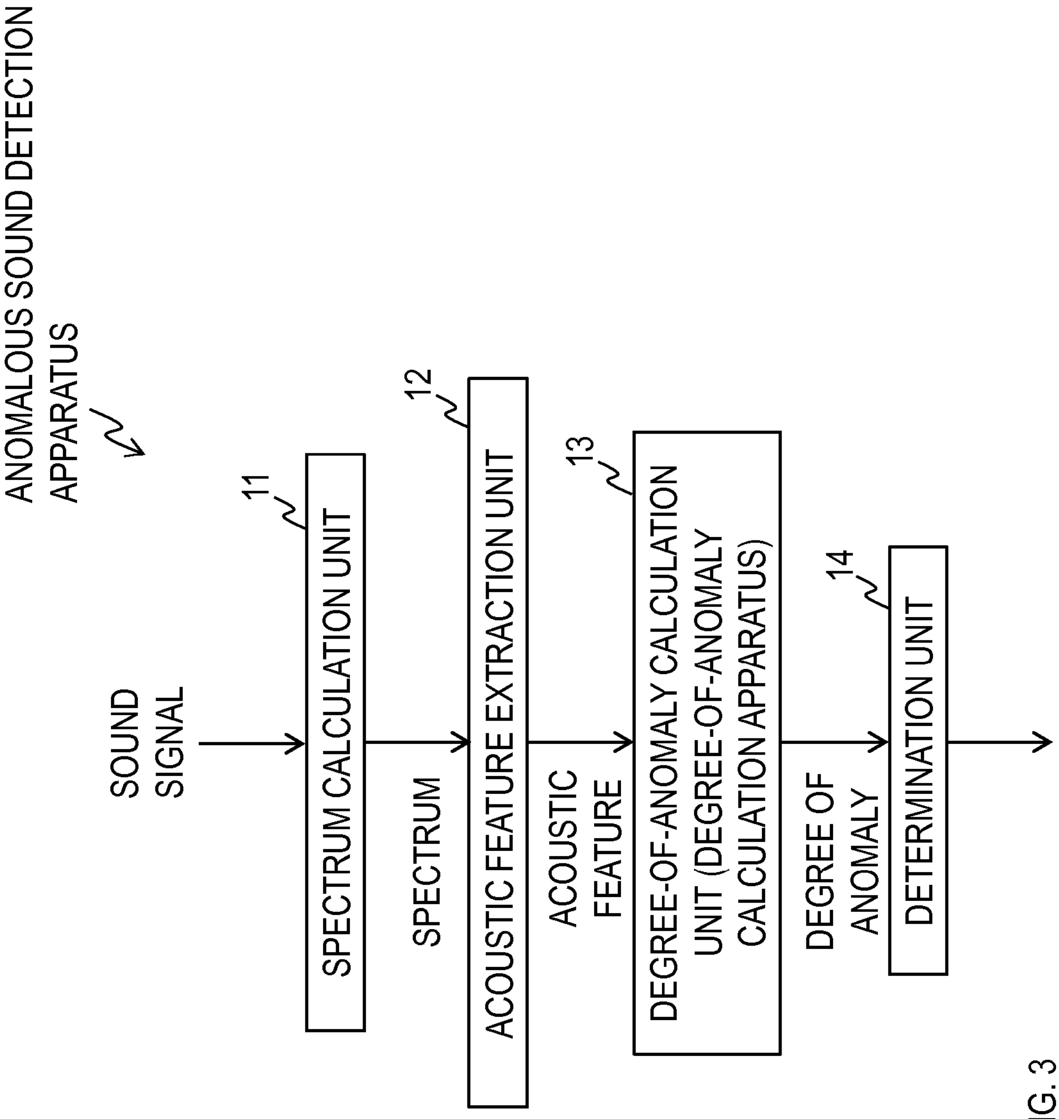
ANOMALOUS SOUND DETECTION APPARATUS
SOUND SIGNAL
11
SPECTRUM CALCULATION UNIT
SPECTRUM
12
ACOUSTIC FEATURE EXTRACTION UNIT
ACOUSTIC FEATURE
13
DEGREE-OF-ANOMALY CALCULATION UNIT (DEGREE-OF-ANOMALY CALCULATION APPARATUS)
DEGREE OF ANOMALY
14
DETERMINATION UNIT

FIG. 3

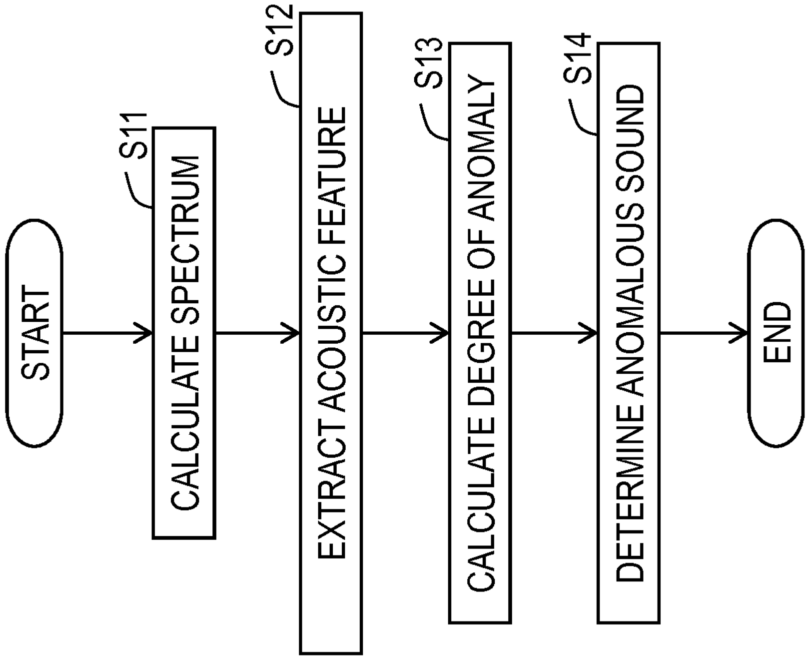
START
S11
CALCULATE SPECTRUM
S12
EXTRACT ACOUSTIC FEATURE
S13
CALCULATE DEGREE OF ANOMALY
S14
DETERMINE ANOMALOUS SOUND
END

FIG. 4

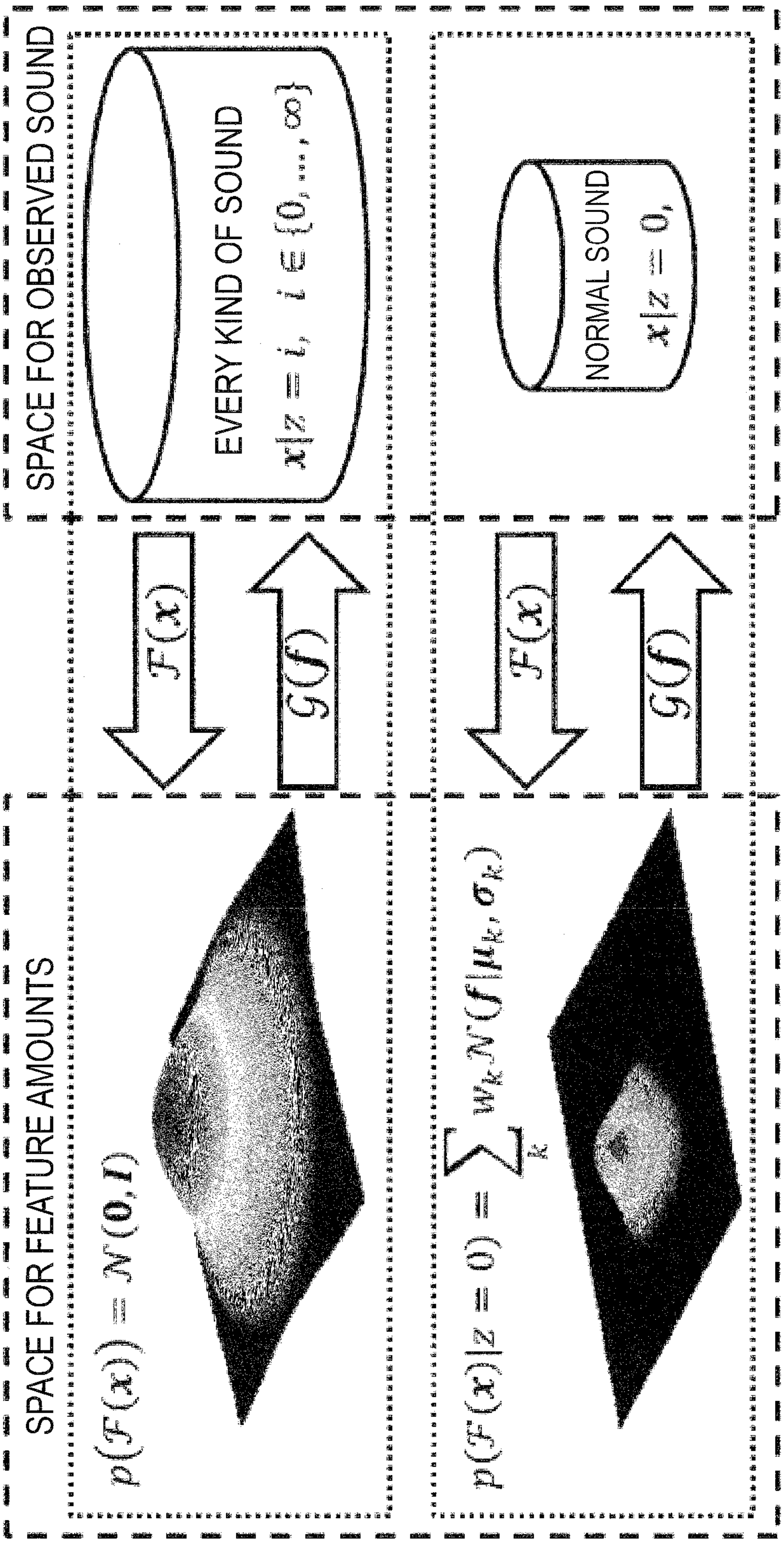
SPACE FOR FEATURE AMOUNTS
$p(\mathcal{F}(\boldsymbol{x})) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
$p(\mathcal{F}(\boldsymbol{x})|z=0) = \sum_k w_k \mathcal{N}(\boldsymbol{f}|\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)$
$\mathcal{F}(\boldsymbol{x})$
$\mathcal{G}(\boldsymbol{f})$
$\mathcal{F}(\boldsymbol{x})$
$\mathcal{G}(\boldsymbol{f})$
SPACE FOR OBSERVED SOUND
EVERY KIND OF SOUND
$\boldsymbol{x}|z = i, \ i \in \{0, \dots, \infty\}$
NORMAL SOUND
$\boldsymbol{x}|z = 0,$

FIG. 5

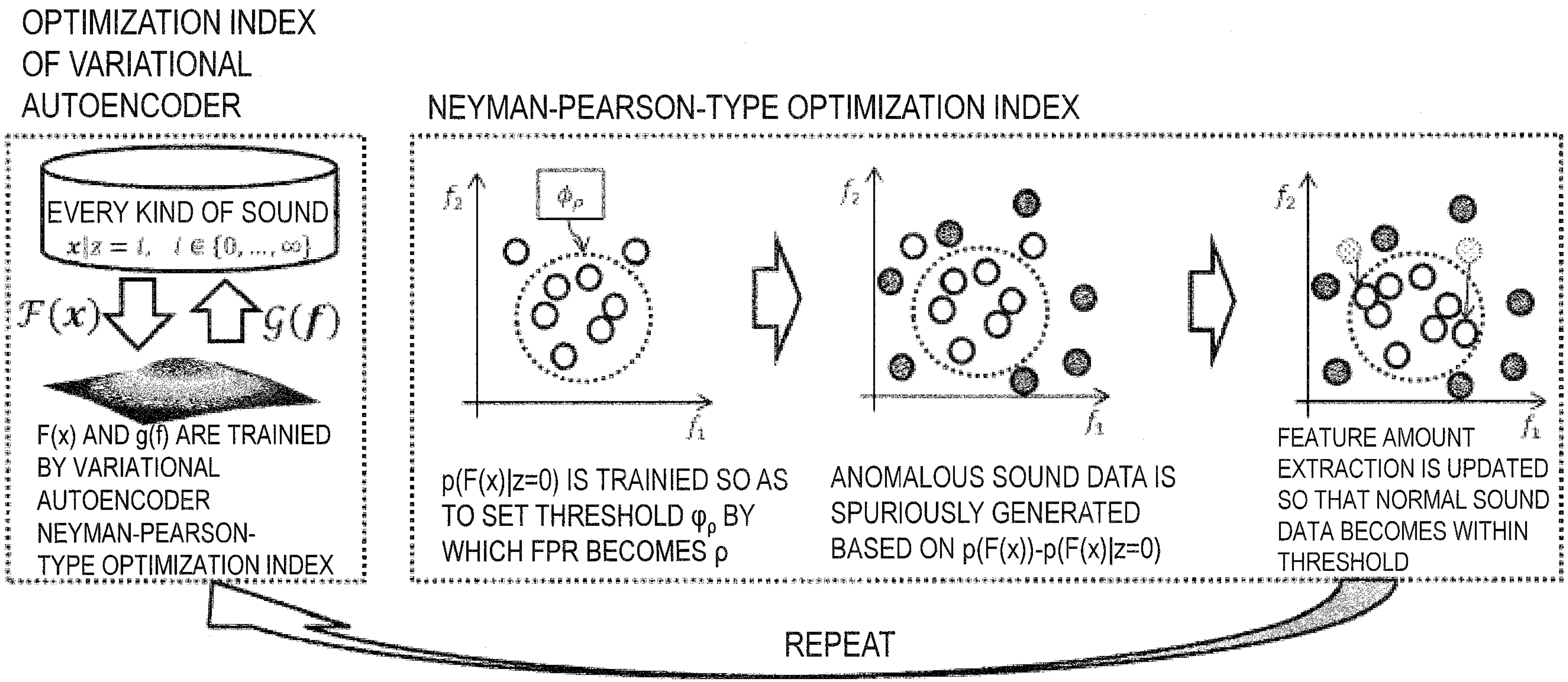
OPTIMIZATION INDEX OF VARIATIONAL AUTOENCODER
EVERY KIND OF SOUND
$x|z = i, \ i \in \{0, \ldots, \infty\}$
$\mathcal{F}(x)$
$\mathcal{G}(f)$
F(x) AND g(f) ARE TRAINIED BY VARIATIONAL AUTOENCODER NEYMAN-PEARSON-TYPE OPTIMIZATION INDEX
NEYMAN-PEARSON-TYPE OPTIMIZATION INDEX
f_2
f_1
ϕ_ρ
p(F(x)|z=0) IS TRAINIED SO AS TO SET THRESHOLD φ_ρ BY WHICH FPR BECOMES ρ
ANOMALOUS SOUND DATA IS SPURIOUSLY GENERATED BASED ON p(F(x))-p(F(x)|z=0)
FEATURE AMOUNT EXTRACTION IS UPDATED SO THAT NORMAL SOUND DATA BECOMES WITHIN THRESHOLD
REPEAT

FIG. 6

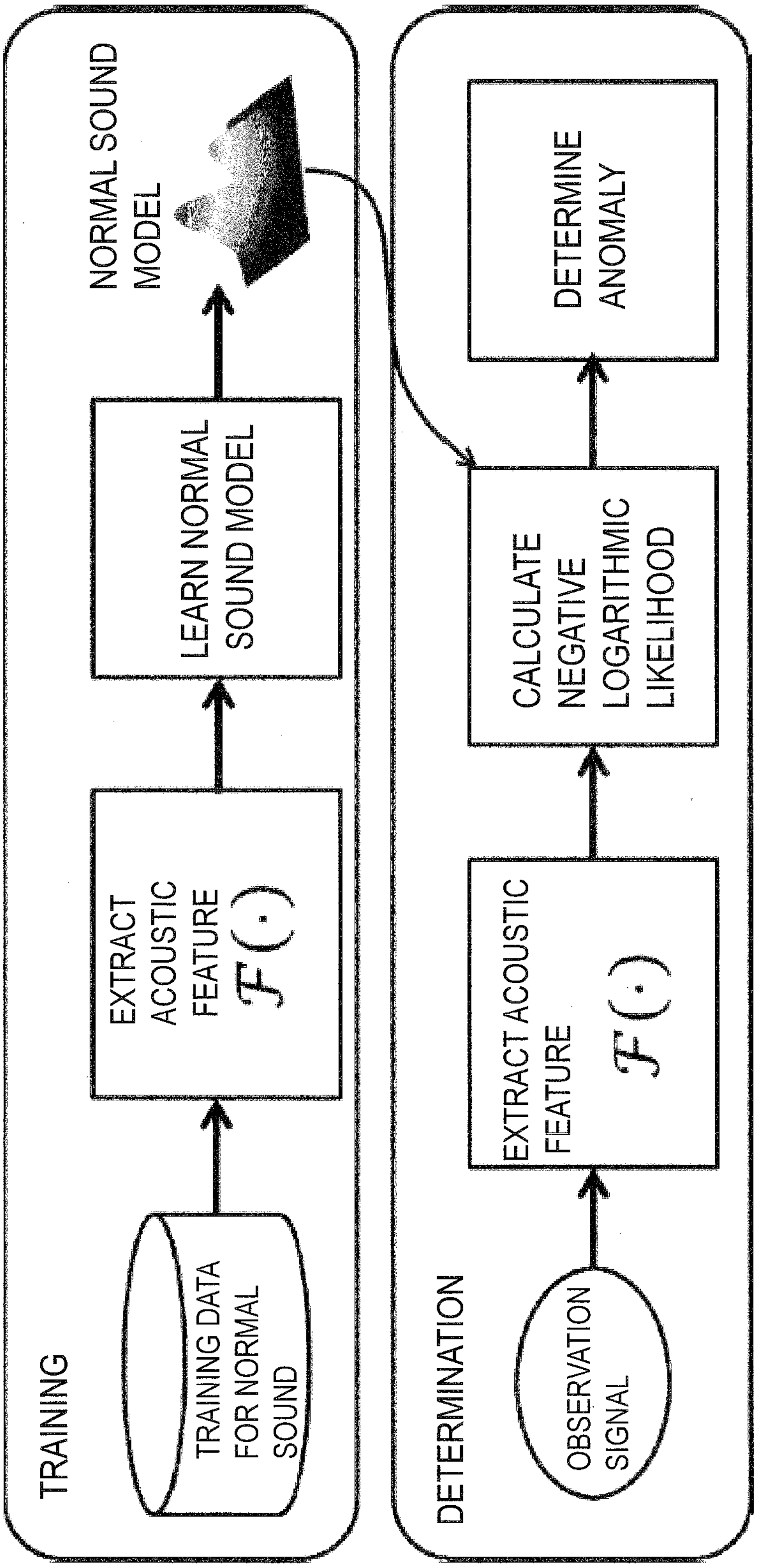
TRAINING
TRAINING DATA FOR NORMAL SOUND
EXTRACT ACOUSTIC FEATURE $\mathcal{F}(\cdot)$
LEARN NORMAL SOUND MODEL
NORMAL SOUND MODEL
DETERMINATION
OBSERVATION SIGNAL
EXTRACT ACOUSTIC FEATURE $\mathcal{F}(\cdot)$
CALCULATE NEGATIVE LOGARITHMIC LIKELIHOOD
DETERMINE ANOMALY

FIG. 7

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 2015219530 A1, LI WEICHANG **[0014]**

Non-patent literature cited in the description

- **TSUYOSHI IDE ; MASASHI SUGIYAMA.** Anomaly Detection and Change Detection. *Kodansha,* 2015, 6-7 **[0013]**
- 2-7-5 Automatic design of acoustic features for detecting abnormal sounds in machine operation sounds. **KOIZUMI et al.** ROCEEDINGS OF THE ACOUSTICAL SOCIETY OF JAPAN MEETING. THE ACOUSTICAL SOCIETY OF JAPAN, 31 August 2016, vol. 2016, 365-368 **[0013]**
- A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks. **ARCHI ERIK et al.** 2015 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP). IEEE, 19 April 2015, 1996-2000 **[0013]**
- **SANGWAN ABHIJEET et al.** *Environmentally Aware Voice Activity Detector,* 27 August 2007, 2929-2932 **[0013]**
- **J. NEYMAN et al.** On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Phi. Trans. of the Royal Society,* 1933 **[0026]**
- **D. P. KINGMA ; M. WELLING.** Autoencoding variational Bayes. *Proceedings of the International Conference on Learning Representations (ICLR),* 2014 **[0041]**
- **SADANORI KONISHI.** Introduction to Multivariate Analysis, Appendix C: EM algorithm. Iwanami Shoten, 2010, 294-298 **[0052]**