Fuzzy Rule-Based System for Decision Making Support of Hybrid SVM-GMM Acoustic Event Detection

Ing-Jr Ding

Abstract

This paper presents a hybrid SVM-GMM mechanism based on a fuzzy rule-based system (FRBS), called FRBS-SVMGMM, for acoustic event detection (AED) applications. This method effectively combines the results of support vector machine (SVM) and Gaussian mixture model (GMM) calculations on acoustic data within the framework of FRBS operations. With the support of FRBS, which greatly increases AED recognition accuracy, AED can make reliable and correct detection decisions. This approach is essential to applications where low false recognition is a major concern. The proposed FRBS-SVMGMM mechanism is conceptually simple and computationally inexpensive. The comparative recognition performance experiments in this study demonstrate the effectiveness and superiority of the proposed FRBS-SVMGMM.

Keywords: Acoustic event detection, Fuzzy rule-based system, FRBS-SVMGMM, Gaussian mixture model, Support vector machine.

1. Introduction

Almost all living creatures are equipped with organs for both visual and aural perception; any security, surveillance or remote homecare system lacking acoustic information is effectively crippled. Acoustic event detection (AED), sometimes referred to as acoustic event classification, has received more attention in recent years [1-9], and plays an important role in the field of speech and audio information processing. Conventional security, surveillance, and remote homecare systems rely heavily, if not exclusively, on motion-tracking analysis of visual information (i.e., data captured by video camera) [10, 11]. Multimedia retrieval and indexing applications also focus on video information, but only recently have audio cues become an auxiliary method of detecting a specific shot in a video sequence [12, 13]. The image-acquiring process has inherent limitations that restrict the ability of visual data to capture status or situation development. Acoustic data can be a complementary source of information in this context.

The fundamental issues of AED include the following:

(1) Categorizing various kinds of sounds encountered in daily life [1, 2].
(2) Internal representation and modeling of a designated type of sound, to differentiate it from other sounds and background acoustics [3-7].
(3) Representation and modeling of background acoustics to allow the comparison required for acoustic event detection [8, 9].

This paper focuses on the second category of AED technical issues and addresses the problem of detecting “female screaming” in specific acoustic backgrounds. Early researches have explored detection of human activity, such as coughing, crying, talking, walking and running [3, 4]. Undoubtedly, among these detection activities, the detection of persons’ speaking is the major concern for practical applications on the real world, a smart speech conference system application, for example. However, it is interesting enough that AED researches on screaming detection are highly esteemed and rapidly increasing recently [5-7]. In fact, screaming detection is very helpful for those persons, security guards, family members and elder/younger care providers, and it is a great contribution to homecare and security applications where screaming usually represents serious or urgent cases. For closely meeting the real homecare and security applications, the detection of female screaming in this work is performed in three practical acoustic background environments, a living room, an indoor parking lot and an office space.

Acoustic event detection begins with a stream of acoustic frames entering the system at regular intervals. To decide whether the designated acoustic event has occurred, analysis is necessary each time a fixed number of frames is collected, or if a pre-determined time span, called the decision window, has elapsed. This study compares the input acoustic signals against two acoustic models (the singular and the normal). In acoustic or sound modeling, primary considerations are the type of representation and how to determine model parameters. The Gaussian mixture model (GMM) [14] and support

Corresponding Author: Ing-Jr Ding is with the Department of Electrical Engineering, National Formosa University, No.64, Wunhua Rd., Huwei Township, Yunlin County 632, Taiwan, R.O.C.
E-mail: ingjr@nfu.edu.tw

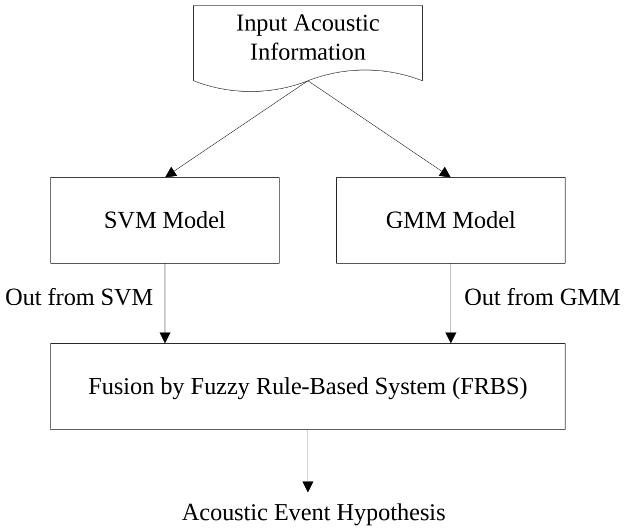

Figure 1. Model-based fusion of SVM and GMM by an FRBS for AED.

vector machine (SVM) [15, 16] appear the most popular acoustic modeling techniques for AED application, due to ease of approximation.

Model-based fusion (Fig. 1), also known as decision-level fusion, is an information fusion technique that combines outputs of different models (for example, the above-mentioned SVM and GMM classifiers) to yield a final classification score. The approach is frequently used in the field of pattern recognition, including acoustic event detection. The linear opinion pool (LOP) [17], also called weighted arithmetical mean (WAM) [18], is a fusion method that has become quite popular for its simplicity and speed in combining model outputs. Although the famous LOP approach simply utilizes the weighted sum of classifier outputs, it handles each output separately as an independent source of information. However, to ensure a reliable final decision, a method of considering the interactions among model outputs is necessary. For instance, instead of treating each model output separately as does the LOP scheme, approaches such as Sugeno's fuzzy integral (FI) and the associated fuzzy measure (FM) [19] can capture interactions among the various classifier outputs. These have been successfully employed to some pattern recognition applications [2, 20, 21, 22], including acoustic event detection application [2]. However, computation for these methods of increasing LOP estimate accuracy is complex and time-consuming, an adverse factor in on-line acoustic event detection.

This paper proposes a fuzzy rule-based system (FRBS) mechanism [23] for AED. This approach tackles inaccurate recognition due to using SVM modeling or GMM modeling alone, and avoids the unreliable fusion of SVM and GMM by linear opinion pool. It also forestalls the daunting cost of model fusion by fuzzy integral or fuzzy measure. The developed hybrid SVM-GMM scheme could be regulated by FRBS for rapid information fusion, reducing inaccurate determinations resulting from a poor fusion of SVM and GMM calculations. The popular Takagi-Sugeno (T-S) FRBS [24] has controlled a system as complicated as an electric power plant with success [25]. The author employs it in researching acoustic event detection.

Section 2 provides an overview of a general acoustic event detection framework, together with mathematical backgrounds for the two acoustic modeling techniques in popular use: SVM and GMM. The end of the section introduces the decision-window scheme for acoustic event detection. Section 3 describes the theoretical formulations of the conventional linear opinion pool approach for information fusion. It then explains the formulation and implementations of the proposed hybrid SVM-GMM scheme under T-S FRBS regulation for acoustic event detection. Section 4 presents experimental results that compare the effectiveness and performance of the proposed approach to conventional SVM-alone, GMM-alone and LOP-SVMGMM. Finally, Section 5 provides a conclusion.

2. Acoustic Event Detection (AED) System

The acoustic event detection system is designed to identify a designated acoustic phenomenon when it appears in a certain acoustic background. Operations compare input acoustic signals against two acoustic models (the singular and the normal) and decide whether an acoustic event has occurred or not. Fig. 2 shows the architecture of a typical acoustic event detection system

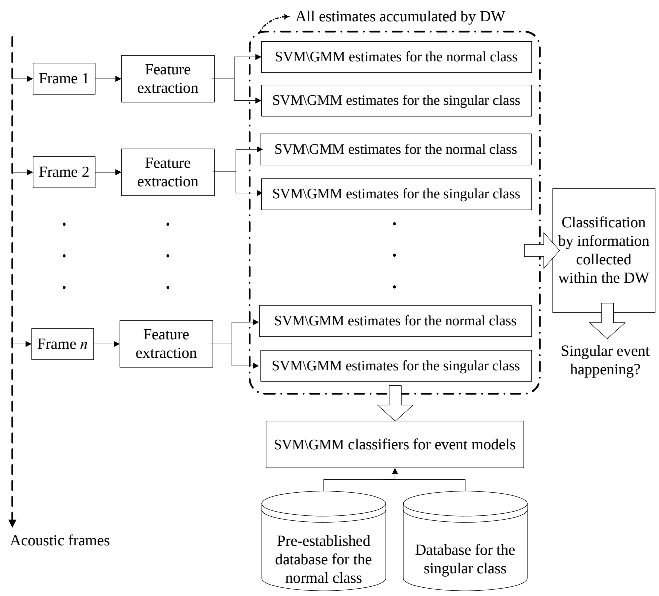

Figure 2. Acoustic event detection (AED) system.

associated with two sound models. It segments the input acoustic stream into the frame sequence, and extracts acoustic features to estimate the scores of both the normal and the singular situation via the classifier operation. After making a decision based on score estimates, the classifier makes its call.

To construct such a system, several essential issues such as feature extraction, acoustic model representation for classification, and decision-making criteria must be resolved. This paper considers popular LPC, LPCC, and MFCC acoustic features; SVM and GMM modeling techniques are frequently employed in the field of speaker recognition and therefore are extremely appropriate for acoustic event detection application in the study. Score estimate calculations provide decision-making criteria.

A. *SVM Classification Model*

SVM is often used as a data classifier. SVM is based on the theory of structural risk minimization of statistics [15, 16]. SVM classifies new input data by using a separating hyperplane. If the SVM model attempted to determine whether an input datum was *A*, it would first try to find the SVM model of *A* in the SVM database. Next, the separating hyperplane of the SVM model of *A* would classify the input datum as *A* or not-*A*.

Suppose a set of labeled training points is $\{(x_1, y_1)$, (x_2, y_2),…, $(x_n, y_n)\}$. Each training point x_i belongs to either of two classes and is given a label, $y_i \in \{-1, 1\}$ for $i = 1,2,...,n$. From these training data, the hyperplane is

$$w \cdot x + b = 0\,, \tag{1}$$

as defined by the pair (w, b), such that point x_i can be separated according to function [15]:

$$f(x_i) = sign(w \cdot x_i + b) = \begin{cases} 1, & if\ y_i = 1 \\ -1, & if\ y_i = -1 \end{cases}. \tag{2}$$

The set *S* is linearly separable if there exists a pair (w, b) such that the inequalities

$$\begin{cases} (w \cdot x_i + b) \geq 1, & if\ y_i = 1, \\ (w \cdot x_i + b) \leq -1, & if\ y_i = -1, \end{cases} \quad i = 1,2,...,n \tag{3}$$

are valid for all elements of set *S*. If the set *S* is linearly separable, a unique optimal hyperplane exists, and for this hyperplane, the margin between the projections of the training points of two different classes is maximized. If set *S* is not linearly separable, classification violations must be allowed in the SVM formulation [15].

The abovementioned SVM classifier verifies whether input acoustic data belongs to the class of the singular acoustic event.

B. *GMM Classifier*

Mathematically, a GMM is a weighted sum of *M* Gaussians, denoted as

$$\lambda = \{w_i, \mu_i, \Sigma_i\},\ i = 1, 2, ..., M\,,\ \sum_{i=1}^{M} w_i = 1\,, \tag{4}$$

where w_i is the weight, μ_i is the mean and Σ_i is the covariance [14]. To determine GMM model parameters for a certain sound class, the E-M algorithm [26] is readily applicable.

After GMM model training is complete, the recognition procedure can be executed based on these models. The GMM classifier consists of two separate GMM models, one for background sound, and the other for singular sound. Consider the classifier operating with a decision window (or equivalently, over a time interval) covering n acoustic feature vectors of D dimensions, $\mathrm{X} = \{x_i \mid i = 1, 2, ..., n\}$, together with two sound models, λ_1 for normal events and λ_2 for singular events.

During the recognition phase, the class of X is determined by maximizing a *posteriori* probability $P(\lambda_s \mid X)$,

$$\hat{s} = \max_{s=\{1,2\}} P(\lambda_s \mid X) = \max_{s=\{1,2\}} \frac{f(X \mid \lambda_s)}{f(X)} \cdot P(\lambda_s)\,. \tag{5}$$

Note that

$$f(x_i \mid \lambda_s) = \sum_{j=1}^{M} w_j \cdot b_j(x_i)\,, \tag{6}$$

and

$$b_j(x_i) = \frac{1}{(2\pi)^{D/2} \cdot |\Sigma_s|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x_i - \mu_s)^T (\Sigma_s)^{-1} (x_i - \mu_s)\right\}. \tag{7}$$

However, for simplicity in real implementation,

$$\hat{s} = \max_{s=\{1,2\}} \sum_{i=1}^{n} \log f(x_i \mid \lambda_s) \tag{8}$$

replaces (5). At the end of the recognition procedure, the signal X is then classified as one of the two sound classes indicated by $\hat{s}$.

C. *Decisions of the Classifier Made by DW*

The decision window (DW) is a time period covering a predetermined number of acoustic frames, within which analysis determines whether an acoustic event has occurred. For event detection by GMM, two likelihood scores are computed for each acoustic frame, the normal and the singular, using (6) based on the two GMM models. Within the decision window, all normal and singular estimates are respectively taken in log-values and accumulated, and whichever is greater determines whether the DW class is normal or singular, (8). For event detection by SVM, the label (normal or singular) of each acoustic frame is determined by (2). Within the decision

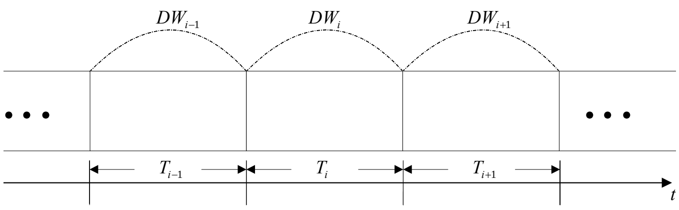

Figure 3. Decision window (DW) for event detection by SVM\GMM models.

window, all normal and singular acoustic frames are added separately, and the class of the DW is the one that has more acoustic frames.

Fig. 3 depicts a stream of decision windows, each of which covers the same number of acoustic frames and thus the same time span. The recognition performance of an acoustic event detection system using the DW criteria is evaluated according to the formula as follows,

$$\text{Recognition rates} = \frac{\text{Numbers of DW with correct detection}}{\text{Numbers of all DW}} \times 100\,(\%) \,. \quad (9)$$

3. Hybrid SVM-GMM Using Fuzzy Rule-Based System for AED

Fusing information sources from SVM and GMM models can promote recognition performance of AED in decision-making. As mentioned, LOP is simple and direct, but its effectiveness is doubtful. This section first presents the concept and formulation of LOP-SVMGMM (using LOP to consider the hybrid fusion of SVM and GMM).

A. *LOP-SVMGMM*

Though there are numerous methods of combining the output information of different classifiers, the most direct is the linear opinion pool technique [17, 18] which calculates weighted sums of the classifier outputs. In this study, LOP consists of a weighted sum of SVM classifier output and GMM classifier output:

$$P_{SVM-GMM}(x) = \alpha \cdot P_{SVM}(x) + (1-\alpha) \cdot P_{GMM}(x),\ 0 \le \alpha \le 1, \quad (10)$$

where $P_{SVM-GMM}(x)$, $P_{SVM}(x)$, and $P_{GMM}(x)$ are the probability outputs of the singular event occurrence by the hybrid SVM-GMM system, the SVM classifier alone, and the GMM classifier alone, respectively, for x-th DW. α denotes the weight parameter controlling the balance between $P_{SVM}(x)$ and $P_{GMM}(x)$ (for the sake of simplicity, $P_{SVM-GMM}(x)$ is represented as P hereafter).

The LOP technique has been considered in several applications such as speaker recognition [27]. Although LOP is a common and simple fusion technique, its main weakness is independence of information sources [28]. It is also difficult to decide the value of α. A poorly estimated value for α would jeopardize the recognition performance of the AED system.

Combining the SVM classifier output and the GMM classifier output in an accurate way is conceptually plain: the more reliable the decisions of SVM and GMM classifiers are, the larger value for P will be. The following section formulates a solution within the framework of a fuzzy rule-based system.

B. *Hybrid SVM-GMM by T-S FRBS (FRBS-SVMGMM)*

The T-S procedure presents a systematic framework of fuzzy modeling design for a complex system. Overall system output is then a function of the subsystem outputs which could be as simple as a "linear" combination addressing fuzzy system behaviors in coefficient handling, or other, more elaborate forms. Under the T-S FRBS, a generic system can be formulated as a set of fuzzy implications (or rules) together with a system output determined by consequences in the set of implications. The system representation would have the form

Rule *i*: IF $x(1)$ is A_1^i and … and $x(n)$ is A_n^i,

$$\text{THEN} \quad y^i = a_0^i + a_1^i x(1) + \dots + a_n^i x(n),\ i = 1,2,\dots,l,$$

$$\text{System output:} \quad y = \frac{\sum_{i=1}^{l} w^i y^i}{\sum_{i=1}^{l} w^i}, \quad \text{given that}$$

$$w^i = \prod_{p=1}^{n} A_p^i(x(p)),$$

for a system of *n* inputs and *l* implications. Note that A_p^i, $p = 0,1,\dots,n,$ are fuzzy sets and $A_p^i(x(n))$ denotes the fuzzy values of the membership function associated with A_p^i for the input $x(n)$; a_p^i, $p = 0,1,\dots,n,$ are consequent parameters through which the *i*-th consequence y^i is expressed as a linear combination of *n* inputs.

1) Concepts of Indexes SVMVD and GMMLD

The inputs in a fuzzy rule-based system are usually signals or quantities of certain attribute in precise magnitudes. In acoustic event detection, two indexes, *SVMVD* and *GMMLD*, respectively derived from SVM and GMM classifiers, are devised as the inputs to the developed fuzzy rule-based system.

The index *SVMVD* (Support Vector Machine Vote Difference) that governs the probability output of the hybrid SVM-GMM system with the SVM separating hyperplane is designed in the form of (11):

$$SVMVD = N_{singular} - N_{normal}, \quad (11)$$

where $N_{singular}$ and N_{normal} denote the number of frames devoted to the singular sound and the normal sound classes, respectively. Note that $N_{singular} + N_{normal} = m$ if the decision window is set to cover m acoustic frames. The rationale for (11) is that a relatively large value of $SVMVD$ is estimated when the class inclination of the m frames in a decision window is obviously singular (the $N_{singular}$ value clearly larger than the N_{normal} value). In contrast, the value of $SVMVD$ would approach zero if resolution of the class inclination is difficult to make and be much smaller than zero if the class is normal.

The $GMMLD$ (Gaussian Mixture Model Likelihood Difference) index that governs the probability output of the hybrid SVM-GMM system in the case of two GMM sound models is devised as follows:

$$GMMLD = \sum_{i=1}^{m} \log f(x_i \mid \lambda_1) - \sum_{i=1}^{m} \log f(x_i \mid \lambda_2), \quad (12)$$

where λ_1 and λ_2 are the singular and normal sound models respectively in consideration, $f(x_i \mid \lambda_1)$ and $f(x_i \mid \lambda_2)$ are given by (6), representing the likelihood of λ_1 and λ_2 model classification, respectively, for frame x_i, and m represents the number of frames covered in a decision window. The rationale behind (12) is that at the decision stage covering m frames of a decision window, if the class inclination of the frames is clearly singular, the term $\sum_{i=1}^{m} \log f(x_i \mid \lambda_1)$ in (12) is substantially greater than the term $\sum_{i=1}^{m} \log f(x_i \mid \lambda_2)$. This yields a large $GMMLD$ value, indicating the occurrence of the singular acoustic event. If the class of the m frames cannot be resolved, both terms in (12) are competitive and lead to a 0-approaching value for $GMMLD$. A small $GMMLD$ value indicates that the m frames belong to the normal class.

2) Designs of SVMVD- and GMMLD-driven FRBS

The hybrid SVM-GMM approach that fuses SVM and GMM classifiers uses a fuzzy rule-based system. As already explained, two input variables (the $SVMVD$ and $GMMLD$ indexes) can be used to control the probability output (P) of the singular event occurrence of the hybrid SVM-GMM AED system. As a result, an FRBS dictated by six IF-THEN fuzzy rules can be designed accordingly:

Rule i: If $SVMVD$ is $A_1(SVMVD)$ and $GMMLD$ is $B_i(GMMLD)$,
then $P = f_i(SVMVD, GMMLD)$, $i = 1,2,3,$

Rule (j+3): If $SVMVD$ is $A_2(SVMVD)$ and $GMMLD$ is $B_j(GMMLD)$,
then $P = f_{j+3}(SVMVD, GMMLD)$, $j = 1,2,3,$

where

$$A_1(SVMVD) = \begin{cases} 1 & SVMVD < SVMVD_1, \\ \dfrac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} & SVMVD_1 \le SVMVD \le SVMVD_2, \\ 0 & SVMVD > SVMVD_2, \end{cases}$$

$$A_2(SVMVD) = \begin{cases} 0 & SVMVD < SVMVD_1, \\ \dfrac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} & SVMVD_1 \le SVMVD \le SVMVD_2, \\ 1 & SVMVD > SVMVD_2, \end{cases}$$

$$B_1(GMMLD) = \begin{cases} 1 & GMMLD \le GMMLD_1, \\ \dfrac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1} & GMMLD_1 \le GMMLD \le GMMLD_2, \\ 0 & GMMLD \ge GMMLD_2, \end{cases}$$

$$B_2(GMMLD) = \begin{cases} 0 & GMMLD \le GMMLD_1 \\ \dfrac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1} & GMMLD_1 < GMMLD \le GMMLD_2, \\ \dfrac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2} & GMMLD_2 \le GMMLD < GMMLD_3, \\ 0 & GMMLD \ge GMMLD_3, \end{cases}$$

$$B_3(GMMLD) = \begin{cases} 0 & GMMLD \le GMMLD_2, \\ \dfrac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2} & GMMLD_2 < GMMLD < GMMLD_3, \\ 1 & GMMLD \ge GMMLD_3, \end{cases} \quad (13)$$

along with the implication functions

$$f_i(SVMVD, GMMLD) = a_i \cdot SVMVD + b_i \cdot GMMLD + c_i, \quad i = 1,2,\ldots,6. \quad (14)$$

Note that $A_1(SVMVD)$ and $A_2(SVMVD)$ are membership functions associated respectively with small and large values of $SVMVD$, and $B_1(GMMLD)$, $B_2(GMMLD)$, and $B_3(GMMLD)$ are membership functions associated respectively with small, medium, and large values of $GMMLD$ (Figs. 4 and 5).

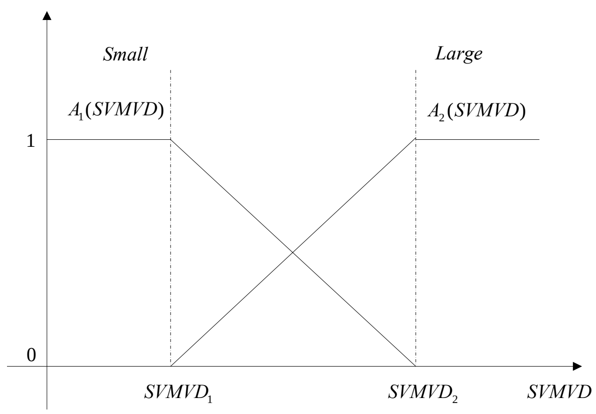

Figure 4. Membership functions of FRBS for the input $SVMVD$.

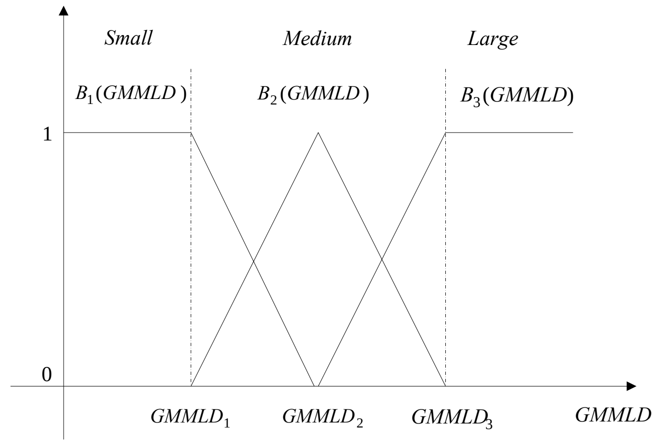

Figure 5. Membership functions of FRBS for the input $GMMLD$.

The final system output is as follows [24]

$$P = \frac{\sum_{i=1}^{6} w^i \cdot f_i(SVMVD, GMMLD)}{\sum_{i=1}^{6} w^i}, \quad (15)$$

where

$$w^1 = A_1(SVMVD) \cdot B_1(GMMLD),$$
$$w^2 = A_1(SVMVD) \cdot B_2(GMMLD),$$
$$w^3 = A_1(SVMVD) \cdot B_3(GMMLD),$$
$$w^4 = A_2(SVMVD) \cdot B_1(GMMLD),$$
$$w^5 = A_2(SVMVD) \cdot B_2(GMMLD),$$
$$w^6 = A_2(SVMVD) \cdot B_3(GMMLD). \quad (16)$$

The resulting system has 23 hyperparameters (a_1, a_2, a_3, a_4, a_5, a_6, b_1, b_2, b_3, b_4, b_5, b_6, c_1, c_2, c_3, c_4, c_5, c_6, $SVMVD_1$, $SVMVD_2$, $GMMLD_1$, $GMMLD_2$, $GMMLD_3$) that must be fixed. An iterative process sets these hyperparameters.

Step 1: Initialization of parameters. Let $GMMLD_1 : GMMLD_2 : GMMLD_3 = 1:2:3$ and $SVMVD_1 : SVMVD_2 = 1:3$. And initialize parameters $SVMVD_1$ and $GMMLD_1$.

Step 2: Estimate the parameters a_1, b_1 and c_1 under the conditions, $SVMVD \le SVMVD_1$ and $GMMLD \le GMMLD_1$,

where $A_1(SVMVD) = B_1(GMMLD) = 1$,
$A_2(SVMVD) = B_2(GMMLD) = B_3(GMMLD) = 0$,
$w^1 = 1$, $w^2 = w^3 = w^4 = w^5 = w^6 = 0$, and
$P = f_1(SVMVD, GMMLD) = a_1 \cdot SVMVD + b_1 \cdot GMMLD + c_1$.

In this case, i.e. the SVM vote difference $SVMVD \le SVMVD_1$ and the GMM likelihood difference $GMMLD \le GMMLD_1$, the appropriate values of a_1, b_1 and c_1 are determined by using the try-and-error experimental method that would maximize the recognition rate of acoustic event detection, R^i,

$$R^i = event\_detection(P = a_1 \cdot SVMVD + b_1 \cdot GMMLD + c_1, \; tunning\_database),$$

where the function $event\_detection(\cdot)$ is used to return the recognition rate of the proposed FRBS-SVMGMM detection with the probability output P controlled by selecting a_1, b_1 and c_1 for the testing data set *tunning_database*, and R^i denotes the returned recognition rate after performing the *i*-th iteration. Note that the try-and-error procedure for fixing the 3 hyperparameters of the fuzzy system would thus finally return an overall recognition rate that is better than the baseline R^0.

An algorithm for fixing a_1, b_1 and c_1 is developed as follows:

BEGIN
Input a_1, b_1 and c_1, of untrained hyperparameters.
Initialize $i = 0$, and the values of a_1, b_1 and c_1.
Increment a_1.
Increment i.
Calculate R^i using the function $event\_detection(\cdot)$.
IF ($R^i > R^{i-1}$) **THEN**
DO UNTIL ($R^i \le R^{i-1}$)
Increment a_1.
Increment i.
Determine R^i using the function $event\_detection(\cdot)$.
END DO UNTIL
ELSE
DO UNTIL ($R^i \le R^{i-1}$)
Decrement a_1.
Increment i.
Determine R^i using the function $event\_detection(\cdot)$.
END DO UNTIL
END IF
Increment b_1.
Increment i.
Calculate R^i using the function $event\_detection(\cdot)$.
IF ($R^i > R^{i-1}$) **THEN**
DO UNTIL ($R^i \le R^{i-1}$)
Increment b_1.
Increment i.
Determine R^i using the function $event\_detection(\cdot)$.

END DO UNTIL
ELSE
DO UNTIL ($R^i \le R^{i-1}$)
Decrement b_1.
Increment i.
Determine R^i using the function $event\_detection(\cdot)$.
END DO UNTIL
END IF
Increment c_1.
Increment i.
Calculate R^i using the function $event\_detection(\cdot)$.
IF ($R^i > R^{i-1}$) **THEN**
DO UNTIL ($R^i \le R^{i-1}$)
Increment c_1.
Increment i.
Determine R^i using the function $event\_detection(\cdot)$.
END DO UNTIL
ELSE
DO UNTIL ($R^i \le R^{i-1}$)
Decrement c_1.
Increment i.
Determine R^i using the function $event\_detection(\cdot)$.
END DO UNTIL
END IF
END

Step 3: Estimate the parameters a_3, b_3 and c_3 under the conditions, $SVMVD \le SVMVD_1$ and $GMMLD \ge GMMLD_3$,

where $A_1(SVMVD) = B_3(GMMLD) = 1$, $A_2(SVMVD) = B_1(GMMLD) = B_2(GMMLD) = 0$, $w^3 = 1$, $w^1 = w^2 = w^4 = w^5 = w^6 = 0$, and $P = f_3(SVMVD, GMMLD) = a_3 \cdot SVMVD + b_3 \cdot GMMLD + c_3$.

The values of a_3, b_3 and c_3 are fixed using the same process as for a_1, b_1 and c_1 with the initial condition $R^0 = R^i$ from step 2.

Step 4: Estimate the parameters a_4, b_4 and c_4 under the conditions, $SVMVD \ge SVMVD_2$ and $GMMLD \le GMMLD_1$,

where $A_2(SVMVD) = B_1(GMMLD) = 1$, $A_1(SVMVD) = B_2(GMMLD) = B_3(GMMLD) = 0$, $w^4 = 1$, $w^1 = w^2 = w^3 = w^5 = w^6 = 0$, and $P = f_4(SVMVD, GMMLD) = a_4 \cdot SVMVD + b_4 \cdot GMMLD + c_4$.

The values of a_4, b_4 and c_4 are fixed using the same process as for a_1, b_1 and c_1 with the initial condition $R^0 = R^i$ from step 3.

Step 5: Estimate the parameters a_6, b_6 and c_6 under the conditions, $SVMVD \ge SVMVD_2$ and $GMMLD \ge GMMLD_3$,

where $A_2(SVMVD) = B_3(GMMLD) = 1$, $A_1(SVMVD) = B_1(GMMLD) = B_2(GMMLD) = 0$, $w^6 = 1$, $w^1 = w^2 = w^3 = w^4 = w^5 = 0$, and $P = f_6(SVMVD, GMMLD) = a_6 \cdot SVMVD + b_6 \cdot GMMLD + c_6$.

The values of a_6, b_6 and c_6 are fixed using the same process as for a_1, b_1 and c_1 with the initial condition $R^0 = R^i$ from step 4.

Step 6: Estimate the parameters a_2, b_2 and c_2 under the conditions, $SVMVD \le SVMVD_1$ and $GMMLD_1 \le GMMLD \le GMMLD_2$,

where $A_1(SVMVD) = 1$,

$$A_2(SVMVD) = B_3(GMMLD) = 0,$$

$$B_1(GMMLD) = \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$B_2(GMMLD) = \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1},$$

$$w^1 = B_1(GMMLD) = \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$w^2 = B_2(GMMLD) = \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1},$$

$$w^3 = w^4 = w^5 = w^6 = 0, \text{ and}$$

$$P = \frac{w^1 \cdot f_1(SVMVD, GMMLD) + w^2 \cdot f_2(SVMVD, GMMLD)}{w^1 + w^2}$$
$$= \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1} \cdot (a_1 \cdot SVMVD + b_1 \cdot GMMLD + c_1) + \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1} \cdot (a_2 \cdot SVMVD + b_2 \cdot GMMLD + c_2).$$

With the initial condition $R^0 = R^i$ from step 5 and { a_1, b_1, c_1 } already obtained at step 2, the parameters a_2, b_2, c_2, could be determined through the same tuning process as in step 2 for best recognition rate too.

Step 7: Estimate the parameters a_5, b_5 and c_5 under the conditions, $SVMVD \ge SVMVD_2$ and $GMMLD_2 \le GMMLD \le GMMLD_3$,

where $A_2(SVMVD) = 1$,

$$A_1(SVMVD) = B_1(GMMLD) = 0,$$

$$B_2(GMMLD) = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2},$$

$$B_3(GMMLD) = \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2},$$
$$w^5 = B_2(GMMLD) = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2},$$
$$w^6 = B_3(GMMLD) = \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2},$$
$$w^1 = w^2 = w^3 = w^4 = 0, \text{ and}$$
$$P = \frac{w^5 \cdot f_5(SVMVD, GMMLD) + w^6 \cdot f_6(SVMVD, GMMLD)}{w^5 + w^6} = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2} \cdot (a_5 \cdot SVMVD + b_5 \cdot GMMLD + c_5) + \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2} \cdot (a_6 \cdot SVMVD + b_6 \cdot GMMLD + c_6).$$

With the initial condition $R^0 = R^i$ from step 6 and { a_6, b_6, c_6 } already obtained at step 5, the parameters a_5, b_5, c_5, could be estimated through the same tuning process as in step 2 for best recognition rate too.

Step 8: Re-estimate both the parameter sets { a_2, b_2, c_2 } and { a_3, b_3, c_3 } again under the conditions, $SVMVD \le SVMVD_1$ and $GMMLD_2 \le GMMLD \le GMMLD_3$, where

$$A_1(SVMVD) = 1, \quad A_2(SVMVD) = 0, \quad B_1(GMMLD) = 0,$$
$$B_2(GMMLD) = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2},$$
$$B_3(GMMLD) = \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2},$$
$$w^2 = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2}, \; w^3 = \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2},$$
$$w^1 = w^4 = w^5 = w^6 = 0, \text{ and}$$
$$P = \frac{w^2 \cdot f_2(SVMVD, GMMLD) + w^3 \cdot f_3(SVMVD, GMMLD)}{w^2 + w^3} = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2} \cdot (a_2 \cdot SVMVD + b_2 \cdot GMMLD + c_2) + \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2} \cdot (a_3 \cdot SVMVD + b_3 \cdot GMMLD + c_3).$$

With the initial condition $R^0 = R^i$ from step 7 and the initial values of { a_2, b_2, c_2 } and { a_3, b_3, c_3 } already separately calculated at steps 6 and 3 respectively, both the parameter sets { a_2, b_2, c_2 } and { a_3, b_3, c_3 } could be re-estimated with determined initial values through the same tuning process as in step 2 for best recognition rate too.

Step 9: Re-estimate the parameter set { a_6, b_6, c_6 } again under the conditions, $SVMVD_1 \le SVMVD \le SVMVD_2$ and $GMMLD \ge GMMLD_3$, where

$$A_1(SVMVD) = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1},$$
$$A_2(SVMVD) = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1},$$
$$B_1(GMMLD) = B_2(GMMLD) = 0, \quad B_3(GMMLD) = 1,$$
$$w^1 = w^2 = w^4 = w^5 = 0, \; w^3 = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1},$$
$$w^6 = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1}, \text{ and}$$
$$P = \frac{w^3 \cdot f_3(SVMVD, GMMLD) + w^6 \cdot f_6(SVMVD, GMMLD)}{w^3 + w^6} = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot (a_3 \cdot SVMVD + b_3 \cdot GMMLD + c_3) + \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot (a_6 \cdot SVMVD + b_6 \cdot GMMLD + c_6).$$

With the initial values of { a_6, b_6, c_6 } already individually determined at step 5 and the values of { a_3, b_3, c_3 } already re-estimated at step 8, new values for { a_6, b_6, c_6 } can now be acquired by tuning for a higher R^i value than in step 8.

Step 10: Re-estimate the parameter set { a_5, b_5, c_5 } again under the conditions, $SVMVD_1 \le SVMVD \le SVMVD_2$ and $GMMLD_2 \le GMMLD \le GMMLD_3$, where $A_1(SVMVD) = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1}$,

$$A_2(SVMVD) = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1}, \; B_1(GMMLD) = 0,$$
$$B_2(GMMLD) = \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2},$$
$$B_3(GMMLD) = \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2}, \quad w^1 = w^4 = 0,$$
$$w^2 = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2},$$
$$w^3 = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2},$$
$$w^5 = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2},$$
$$w^6 = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2},$$
$$P = \frac{w^2 \cdot f_2(SVMVD, GMMLD)}{w^2 + w^3 + w^5 + w^6} + \frac{w^3 \cdot f_3(SVMVD, GMMLD)}{w^2 + w^3 + w^5 + w^6} + \frac{w^5 \cdot f_5(SVMVD, GMMLD)}{w^2 + w^3 + w^5 + w^6} + \frac{w^6 \cdot f_6(SVMVD, GMMLD)}{w^2 + w^3 + w^5 + w^6} =$$

$$\frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2} \cdot (a_2 \cdot SVMVD + b_2 \cdot GMMLD + c_2) + \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2} \cdot (a_3 \cdot SVMVD + b_3 \cdot GMMLD + c_3) + \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_3 - GMMLD}{GMMLD_3 - GMMLD_2} \cdot (a_5 \cdot SVMVD + b_5 \cdot GMMLD + c_5) + \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_2}{GMMLD_3 - GMMLD_2} \cdot (a_6 \cdot SVMVD + b_6 \cdot GMMLD + c_6).$$

With the initial values of $\{a_5, b_5, c_5\}$ already individually determined at step 7 and the values of $\{a_2, b_2, c_2\}$, $\{a_3, b_3, c_3\}$ and $\{a_6, b_6, c_6\}$ already re-calculated at step 9, new values for $\{a_5, b_5, c_5\}$ can now be obtained by tuning for a higher R^i value than in step 9.

Step 11: Re-estimate the parameter set $\{a_4, b_4, c_4\}$ again under the conditions, $SVMVD \geq SVMVD_2$ and $GMMLD_1 \leq GMMLD \leq GMMLD_2$, where

$$A_1(SVMVD) = 0, \quad A_2(SVMVD) = 1,$$

$$B_1(GMMLD) = \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$B_2(GMMLD) = \frac{GMMVD - GMMVD_1}{GMMVD_2 - GMMVD_1},$$

$$B_3(GMMLD) = 0, \quad w^1 = w^2 = w^3 = w^6 = 0,$$

$$w^4 = \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$w^5 = \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1}, \text{ and}$$

$$P = \frac{w^4 \cdot f_4(SVMVD, GMMLD) + w^5 \cdot f_5(SVMVD, GMMLD)}{w^4 + w^5} = \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1} \cdot (a_4 \cdot SVMVD + b_4 \cdot GMMLD + c_4) + \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1} \cdot (a_5 \cdot SVMVD + b_5 \cdot GMMLD + c_5).$$

Since initial values of $\{a_4, b_4, c_4\}$ have been already separately determined at step 7 and the re-estimated values of $\{a_5, b_5, c_5\}$ already calculated at step 10, new values for $\{a_4, b_4, c_4\}$ can now be acquired by tuning for a higher R^i value than in step 10.

Step 12: Re-estimate the parameter set $\{a_1, b_1, c_1\}$ again under the conditions, $SVMVD_1 \leq SVMVD \leq SVMVD_2$ and $GMMLD \leq GMMLD_1$, where

$$A_1(SVMVD) = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1},$$

$$A_2(SVMVD) = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1},$$

$$B_1(GMMLD) = 1, \quad B_2(GMMLD) = B_3(GMMLD) = 0,$$

$$w^2 = w^3 = w^5 = w^6 = 0, \quad w^1 = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1},$$

$$w^4 = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1}, \text{ and}$$

$$P = \frac{w^1 \cdot f_1(SVMVD, GMMLD) + w^4 \cdot f_4(SVMVD, GMMLD)}{w^1 + w^4} = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot (a_1 \cdot SVMVD + b_1 \cdot GMMLD + c_1) + \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot (a_4 \cdot SVMVD + b_4 \cdot GMMLD + c_4).$$

New values for $\{a_1, b_1, c_1\}$ can now be obtained by tuning for a higher R^i value than in step 11 since the initial values of $\{a_1, b_1, c_1\}$ have been already separately calculated at step 2 and the re-estimated values of $\{a_4, b_4, c_4\}$ already obtained at step 11.

Step 13: Re-estimate theses two parameters $SVMVD_2$ and $GMMLD_3$ under the conditions, $SVMVD_1 \leq SVMVD \leq SVMVD_2$ and $GMMLD_1 \leq GMMLD \leq GMMLD_2$, where

$$A_1(SVMVD) = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1},$$

$$A_2(SVMVD) = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1},$$

$$B_1(GMMLD) = \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$B_2(GMMLD) = \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1},$$

$$B_3(GMMLD) = 0, \; w^3 = w^6 = 0,$$

$$w^1 = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$w^2 = \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1},$$

$$w^4 = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1},$$

$$w^5 = \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1},$$

and

$$P = \frac{w^1 \cdot f_1(SVMVD, GMMLD)}{w^1 + w^2 + w^4 + w^5} + \frac{w^2 \cdot f_2(SVMVD, GMMLD)}{w^1 + w^2 + w^4 + w^5} + \frac{w^4 \cdot f_4(SVMVD, GMMLD)}{w^1 + w^2 + w^4 + w^5} + \frac{w^5 \cdot f_5(SVMVD, GMMLD)}{w^1 + w^2 + w^4 + w^5} =$$

$$\frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1} \cdot (a_1 \cdot SVMVD + b_1 \cdot GMMLD + c_1) + \frac{SVMVD_2 - SVMVD}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1} \cdot (a_2 \cdot SVMVD + b_2 \cdot GMMLD + c_2) + \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD_2 - GMMLD}{GMMLD_2 - GMMLD_1} \cdot (a_4 \cdot SVMVD + b_4 \cdot GMMLD + c_4) + \frac{SVMVD - SVMVD_1}{SVMVD_2 - SVMVD_1} \cdot \frac{GMMLD - GMMLD_1}{GMMLD_2 - GMMLD_1} \cdot (a_5 \cdot SVMVD + b_5 \cdot GMMLD + c_5).$$

With parameter sets { a_1, b_1, c_1 }, { a_2, b_2, c_2 }, { a_4, b_4, c_4 } and { a_5, b_5, c_5 } already re-estimated at steps 12, 8, 11 and 10 respectively, new values for $SVMVD_2$ and $GMMLD_3$ can now be acquired by tuning for a higher R^i value than in step 12.

Step 14: Update the values of $SVMVD_1$, $GMMLD_1$ and $GMMLD_2$ such that $SVMVD_1 : SVMVD_2 = 1:3$ and $GMMLD_1 : GMMLD_2 : GMMLD_3 = 1:2:3$,

$$\delta = \frac{\left| R^i - R^* \right|}{R^*}, \text{ /* } R^* \text{: desired recognition rate */}$$

$$R^0 = R^i.$$

Repeat from step 2 until δ is less than a predefined threshold.

When completing fixing hyperparameters by the above developed iterative process, the previous set of rules is expected to achieve the following IF-THEN implication representations:

Rule 1: If $SVMVD$ is small and $GMMLD$ is small, then P is extremely small,
Rule 2: If $SVMVD$ is small and $GMMLD$ is medium, then P is small,
Rule 3: If $SVMVD$ is small and $GMMLD$ is large, then P is slightly small,
Rule 4: If $SVMVD$ is large and $GMMLD$ is small, then P is slightly large,
Rule 5: If $SVMVD$ is large and $GMMLD$ is medium, then P is large,
Rule 6: If $SVMVD$ is large and $GMMLD$ is large, then P is extremely large,

where "extremely small," "small," "slightly small," "slightly large," "large" and "extremely large" P values can be determined by functions

$f_1(SVMVD, GMMLD)$, $f_2(SVMVD, GMMLD)$, $f_3(SVMVD, GMMLD)$, $f_4(SVMVD, GMMLD)$, $f_5(SVMVD, GMMLD)$ and $f_6(SVMVD, GMMLD)$, respectively.

The $SVMVD$ and $GMMLD$ input variables in the fuzzy rule base are divided into a range of two states, "small" and "large," and a range of three states, "small," "medium," and "large," respectively. In most studies about speech and voice classification (including this study), the SVM approach achieves better performance than the GMM approach in recognition accuracy [1]. Compared with GMM, SVM has lower size requirements for training data. In addition, SVM is also less sensitive to the presence of irrelevant features than GMM [29]. In the proposed fuzzy rule base, the value of $SVMVD$ is the major concern for inferring an output of the rather large P value. As long as the $SVMVD$ value falls into the range of the "large" set, the output P value will be rather large (more than 0.5). The $GMMLD$ input can be used as a secondary variable to fine-tune the inferred output P value. Note that during the entire FRBS hyperparamters training process, a predefined threshold is used to be compared with the inferred output P of FRBS in such way that if the value of P is larger than the value of the threshold, then a decision of the singular acoustic event occurrence will be achieved, otherwise the normal condition decision will be made. The same threshold is used for decision making of online AED testing applications through the proposed FRBS-SVMGMM.

4. Experiments and Results

This study includes experiments to evaluate performance of the proposed FRBS-SVMGMM acoustic event detection approach and detect female screaming in three environments with different acoustic backgrounds: an office space, a parking lot, and a living room.

A. *Database and Experiment Design*

The experiments in this study established (1) SVM and GMM models for female screaming and acoustic backgrounds, (2) the training phase for fixing hyperparameters of the fuzzy rule-based system, and (3) the recognition phase for performance evaluation on the fusion of SVM and GMM decisions by FRBS (FRBS-SVMGMM) in Section 3.

The training phase of the GMM models constructs three models for "office space," "parking lot," and "living room" as background models using a ten-minute recording in each environment. The recordings were taken at 8 K Hz sampling rate, from which LPC, LPCC, and MFCC were extracted for each 20 ms frame (consisting of 160 samples). A 12-D LPC, a 12-D LPC/mel cepstrum, and a 12-D delta cepstrum were used. Three GMM models for "female screaming" in each of the three environments were built using two thirds of a 180-s recording (60 s for each environment) from each of 15 female participants to extract the same set of three acoustic features. The participants were requested to scream in every possible way they could during the recording. Three SVM models for "female screaming" in each of the three environments were also established using the same database as the GMM models training phase.

In the FRBS hyperparameter-tuning phase described above, the remaining one-third of the screaming data (20 s for each environment and 900 s in total for all 15 females in all three environments) was used as the training data.

In the performance evaluation phase, an entirely new group of 15 women was asked for screaming recordings of 60 s each (20 s for each of the three environments) as test data to compare recognition performance of the four acoustic event detection schemes: SVM alone, GMM alone, LOP-SVMGMM, and the proposed FRBS-SVMGMM.

B. *Experimental Results*

As mentioned, the 20 s recording segments of the acoustic event "female screaming" occurring in "office space," "parking lot" or "living room" backgrounds from 15 women formed an acoustic stream 15 min long. This stream was used to evaluate the performance of the proposed approach. The following observations were made during the experiment.

(1) Tables 1, 2, and 3 record the recognition performances of female screaming detection using the conventional LOP-SVMGMM with various α values by LPC, LPCC, and MFCC acoustic features, respectively. The best choices of α for LOP-SVMGMM by LPC, LPCC, and MFCC were 0.6, 0.7 and 0.7, respectively. Therefore, these values of α were selected in LOP-SVMGMM for comparison.

Table 1. Event detection by LOP-SVMGMM, using only LPC feature.

Weight α settings	Average recognition rates (%)		
	Selected backgrounds		
	Living room	Parking lot	Office space
0.1	82.33	86.67	93.67
0.2	82.33	87	94.33
0.3	82.67	87.33	94.67
0.4	82.67	87	94.67
0.5	83.33	88	95.33
0.6	**84**	**89**	**95.67**
0.7	83.67	88.33	95.67
0.8	83	87.67	95
0.9	82.67	87.33	94.33

Table 2. Event detection by LOP-SVMGMM, using only LPCC feature.

Weight α settings	Average recognition rates (%)		
	Selected backgrounds		
	Living room	Parking lot	Office space
0.1	89.67	89.67	92.33
0.2	89.67	90	92.67
0.3	90.67	90.67	93.33
0.4	90.33	91	93.33
0.5	90.67	91.67	94.67
0.6	90.67	92	95.67
0.7	**90.67**	**92.67**	**96.33**
0.8	90	92.67	96
0.9	89.67	92	95.67

Table 3. Event detection by LOP-SVMGMM, using only MFCC feature.

Weight α settings	Average recognition rates (%)		
	Selected backgrounds		
	Living room	Parking lot	Office space
0.1	89.67	90.67	94
0.2	90	91	94.33
0.3	89.67	91	94.33
0.4	90	91.67	95
0.5	90	92	95.67
0.6	90.67	92.33	96.33
0.7	**91.67**	**93**	**96.67**
0.8	91	92.67	96.67
0.9	90.33	92	96.33

(2) Figs. 6, 7, and 8 show the experimental records of female screaming detection by SVM alone, GMM alone, LOP-SVMGMM and FRBS-SVMGMM conducted in "living room," "parking lot," and "office space" backgrounds, respectively. The method exploiting hybrid SVM-GMM governed by the FRBS achieved an average of 85.33%, 91.67%, and 92.67% recognition rates for event detection using LPC, LPCC, and MFCC respectively in the testing context of living room (Fig. 6). In all cases, the recognition accuracy of SVM alone, GMM alone, and LOP-SVMGMM was inferior to the scores of FRBS-SVMGMM. The proposed FRBS-SVMGMM leads to the best recognition, followed by LOP-SVMGMM, then by SVM alone, and GMM alone yields the worst recognition. Furthermore, against the FRBS-regulated fusion of SVM and GMM, the LOP-fusion of SVM and GMM reaches competitive scores of 84%, 90.67%, and 91.67% on the use of LPC, LPCC, and MFCC features, respectively. Similar conclusions came from the test contexts of parking lot and office space (Figs. 7 and 8).

(3) For all testing in the three backgrounds, MFCC performed best in acoustic event detection, LPCC the second and LPC the third, regardless of which detection scheme being adopted (Figs. 6-8).

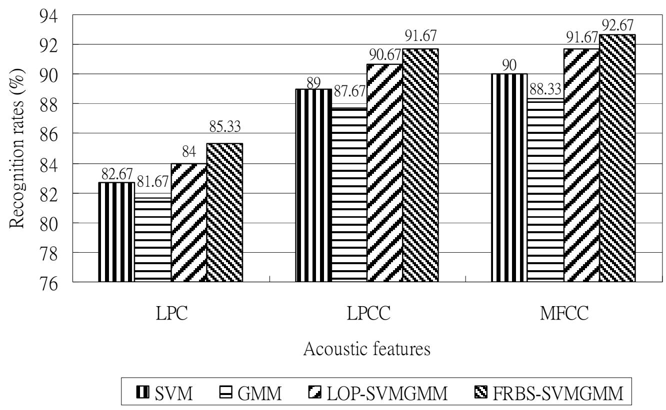

Figure 6. Living room acoustic event detection.

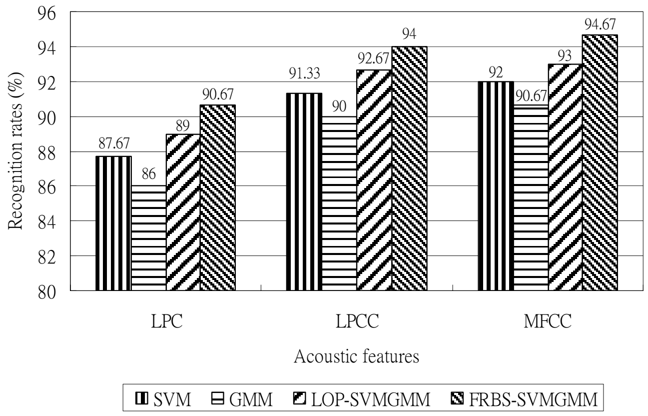

Figure 7. Parking lot acoustic event detection.

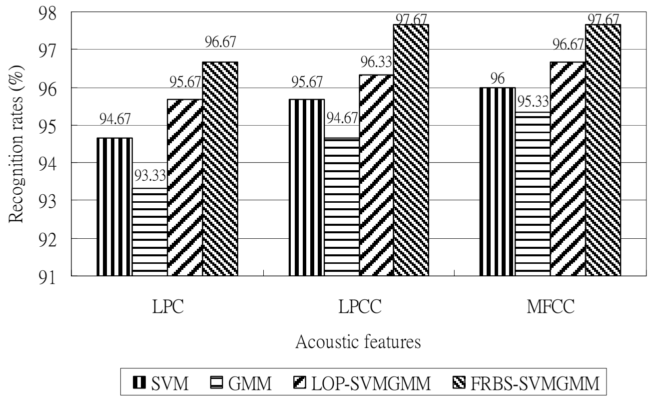

Figure 8. Office space acoustic event detection.

(4) In the experiment, the noisiest background was the living room (where family members exchanged conversation while children ran and played, with a TV set turned on aloud), followed by the parking lot, and then the office space. Such a phenomenon seems to reflect the recognition accuracy of event detection. To be specific, event detection had the best recognition performance in office space, followed by the parking lot, and then the living room, regardless of which of the three acoustic features was used.

5. Conclusions

This paper proposes an FRBS-SVMGMM mechanism that uses an FRBS to govern the decisions of SVM and GMM detection in acoustic event detection. This study examines the performance of female screaming detection in three operational backgrounds (office space, indoor parking lot, and living room) with three distinctive acoustic features. For all test cases, experimental records indicate that the proposed FRBS-SVMGMM scheme is superior to conventional acoustic event detection techniques: the SVM classifier alone, the GMM classifier alone, and the fusion of SVM and GMM classifiers by linear opinion pool (LOP-SVMGMM).

Acknowledgment

This research is partially supported by the National Science Council (NSC) in Taiwan under grant NSC 100-2221-E-150-092.

References

[1] A. Temko and C. Nadeu, "Classification of acoustic events using SVM-based clustering schemes," *Pattern Recognition*, vol. 39, no. 4, pp. 682-694, 2006.
[2] A. Temko, D. Macho, and C. Nadeu, "Fuzzy integral based information fusion for classification of highly confusable non-speech sounds," *Pattern Recognition*, vol. 41, no. 5, pp. 1831-1840, 2008.
[3] C. Clavel, T. Ehrette, and G. Richard, "Event detection for an audio-based surveillance system," In *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 1306-1309, 2005.
[4] P. K. Atrey, N. C. Maddage, and M. S. Kankanhalli, "Audio based event detection for multimedia surveillance," In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 813-816, 2006.
[5] W. Huang, T. K. Chiew, H. Li, T. S. Kok, and J. Biswas, "Scream detection for home applications," In *Proc. of IEEE Conference on Industrial Electronics and Applications*, pp. 2115-2120, 2010.
[6] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance," In *Proc. of IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 21-26, 2007.
[7] J. Montalvão, "Improved signal representation for event detection in remote health care through psychoanalytical masking," In *Proc. of XVIII Congresso Brasileiro de Automática (CBA)*, Bonito-MS, 2010.
[8] M. Cristani, M. Bicego, and V. Murino, "On-line adaptive background modeling for audio surveillance," In *Proc. of International Conference on Pattern Recognition*, vol. 2, pp. 399-402, 2004.
[9] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 246-252, 1999.
[10] R. J. Evans, E. L. Brassington, and C. Stennett, "Video motion processing for event detection and other applications," In *Proc. of International Conference on Visual Information Engineering*, pp.

93-96, 2003.
[11] A. Albiol, C. Sandoval, V. Naranjo, and J. M. Mossi, “Robust motion detector for video surveillance applications,” In *Proc. of International Conference on Image Processing*, vol. 3, pp. II–379-382, 2003.
[12] R. Cai, L. Lu, H.-J. Zhang, and L.-H. Cai, “Highlight sound effects detection in audio stream,” In *Proc. of IEEE International Conference on Multimedia and Expo*, vol. 3, pp. 37-40, 2003.
[13] O. Gillet and G. Richard, “Automatic transcription of drum sequences using audiovisual features,” In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. III–205-208, 2005.
[14] D. A. Reynolds and R. C. Rose, “Robust text-independent speaker identification using Gaussian mixture speaker models,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, 1995.
[15] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.
[16] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.
[17] J. A. Benediktsson and P. H. Swain, “Consensus theoretic classification methods,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 688-704, 1992.
[18] L. Kuncheva, “‘Fuzzy’ versus ‘Non-fuzzy’ in combining classifiers designed by boosting,” *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 729-741, 2003.
[19] M. Sugeno, “Fuzzy measures and fuzzy integrals – a survey,” *Fuzzy Automata and Decision Processes*, Amsterdam, North-Holland, pp. 89-102, 1977.
[20] T. Pham and M. Wagner, “Fuzzy-integration based normalization for speaker verification,” In *Proc. of International Conference on Spoken Language Processing*, pp. 3273-3276, 1998.
[21] S. Chang and S. Greenberg, “Application of fuzzy-integration-based multiple-information aggregation in automatic speech recognition,” In *Proc. of IEEE International Conference on Fuzzy Integration Processing*, Beijing, 2003.
[22] S. Chang and S. Greenberg, “Syllable-proximity evaluation in automatic speech recognition using fuzzy measures and a fuzzy integral,” In *Proc. of IEEE International Conference on Fuzzy Systems*, pp. 828-833 2003.
[23] H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications (3rd ed.)*, Kluwer Academic, 1996.
[24] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its application to modeling and control,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116-132, 1985.
[25] J. Yen, R. Langari, and L. A. Zadeh (Eds.), *Industrial Applications of Fuzzy Logic and Intelligent Systems*, IEEE Press, New York, 1995.
[26] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*, vol. 39, pp. 1-38, 1977.
[27] J. K. Shah, B. Y. Smolenski, and R. E. Yantorno, “Decision level fusion of usable speech measures using consensus theory,” In *Proc. of IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 245-250, 2003.
[28] J.-L. Marichal, “Behavioral analysis of aggregation in multicriteria decision aid,” in: J. Fodor, B. D. Baets and P. Perny (Eds.), *Preferences and Decisions under Incomplete Knowledge*, Physica-Verlag, pp. 153-178, 2000.
[29] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, “Feature selection for SVMs,” In *Proc. of Neural Information Processing Systems*, pp. 668-674, 2000.

Ing-Jr Ding was born in Taipei, Taiwan, in 1975. He received the B.S. degree from Chang-Gung University in 1999, M.S. degree from National Central University in 2001, and Ph.D. degree from National Chiao-Tung University in 2008. He joined the Graduate Institute of Automation and Control at National Taiwan University of Science and Technology as a project assistant professor from March 2009 to July 2009. Since August 2009, he has been an assistant professor in the Department of Electrical Engineering, National Formosa University. His research interests include speech recognition, artificial intelligence, and multimedia techniques. He is a member of IEICE.