International Journal of Innovative Computing, Information and Control
Volume **9**, Number **2**, February **2013**
 ISSN 1349-4198
pp. 555–572

SVM-EMBEDDED FLCMAP SPEAKER ADAPTATION USING A SUPPORT VECTOR MACHINE TO IMPROVE FUZZY CONTROLLERS OF FCMAP

Ing-Jr Ding

Department of Electrical Engineering
National Formosa University
No. 64, Wunhua Rd., Huwei Township, Yunlin County 632, Taiwan
ingjr@nfu.edu.tw

Received November 2011; revised March 2012

Abstract. *Maximum a posteriori (MAP) adaptation is currently one of the most widely used speaker adaptation techniques. However, the effectiveness of MAP relies significantly on the quality and quantity of adaptation data. Although an FCMAP approach that incorporates a fuzzy logic controller (FLC) into MAP could avoid performance degradation that is caused by a poor MAP estimate resulting from relatively small amounts of adaptation data, the performance of FCMAP cannot be maintained in a situation where the quality of the acquired adaptation data is low. To address the issue of FCMAP, this research proposes a support vector machine (SVM)-embedded FLCMAP method that improves the fuzzy controller in the conventional FCMAP. The developed SVM-embedded FLCMAP determines the quality of adaptation data by using an SVM. The SVM-estimated data quality information, together with the data quantity, can then be used to drive the fuzzy inference operations of the FLC. The proposed SVM-embedded FLCMAP provides a method to ensure the robustness of FCMAP against data inferiority. The experimental results show that an SVM-embedded FLCMAP outperforms FCMAP, especially when encountering substandard data. The proposed SVM-embedded FLCMAP performs more efficiently than hybrid SVM-FLC MAP and is more flexible for adaptive data utilization.*
Keywords: Speaker adaptation, Maximum a *posteriori*, Adaptation data, SVM, FLC

1. **Introduction.** Automatic speech recognition (ASR) processing has recently received increasing attention [1-6]. A speech recognition system that can always maintain a satisfactory recognition performance during adverse conditions has been a challenging issue in the field of ASR. During the process of speech recognition, the mismatch between the pre-established reference templates and the testing template from a new speaker (or from a speaker known to the system but in poor "vocal form") disrupts the recognition system and significantly compromises recognition performance. Speaker adaptation (SA), also referred to as model-based adaptation, is the countermeasure that addresses this issue. SA is a process that adapts a complete speaker-independent (SI) model into a speaker-dependent (SD) one. This achieves an SD-like performance, although it requires a fraction of the speaker-specific training data.

There are currently three primary categories of SA techniques [3]: 1) maximum a *posteriori* (MAP) adaptation [7,8], which represents Bayesian-based adaptation; 2) maximum likelihood linear regression (MLLR) adaptation [9,10], which represents transformation-based adaptation; and 3) eigenvoice-based adaptation [11-15]. The eigenvoice-based adaptation category is a relatively novel SA technique that first appeared around 2000. It is also known as a speaker-clustering-based adaptation process where an SD speech model is established for every member in a group of speakers. Feature vectors (eigenvoices) are extracted through principal component analysis (PCA) to build the eigenvoice speech model. Prior to the advent of the eigenvoice approach, MAP and MLLR adaptation were the most common techniques for speaker adaptation. They are still used in the majority of speech recognition systems. MAP adaptation is provided with local but specific effects of adaptation, in contrast to MLLR adaptation, which brings about an overall and yet somewhat coarser speech model adaptation with the same adaptation samples.

This research focuses on MAP adaptation, which updates only the speech model associated with the adaptation samples, whereas the remaining models are unchanged. The principle of "direct model adaptation" supporting MAP adaptation is shown in Figure 1. The mean vectors that are not adapted (μ_4 and μ_5) have no adaptation samples acquired from the speaker's adaptation utterances, whereas the mean vectors (μ_1, μ_2 and μ_3) with the respective adaptation samples N_1, N_2 and N_3 available are adapted. The quality of MAP adaptation is determined significantly by the quantity and characteristics of the utterances acquired from adapted speakers. Insufficient or inadequate adaptation data would probably result in an unreliable speech model adaptation that inevitably jeopardizes recognition performance.

To address the issues caused by the scarcity of adaption data, the FCMAP variant of MAP [16] adheres to the principle that adaptation should not deviate significantly from prior speech model means when adaptation data are limited. In FCMAP, the framework of a fuzzy logic control (FLC) mechanism is incorporated into MAP to resolve the unreliable adaptation resulting from insufficient training samples. FCMAP has proven its reliability by achieving superior recognition rates while addressing adverse conditions of adaptation data scarcity. However, although FCMAP enhances MAP and is effective when the available adaptation data are limited, FCMAP performance cannot be maintained when the quality of the acquired adaptation data is low. Previous research [1] has explored a validation scheme of FCMAP adaptation data that enhances the quality of FCMAP adaptation, despite the inferiority of adaptation samples. A hybrid support vector machine (SVM)-FLC MAP adaptation approach was developed to use an SVM classifier [17,18] that verifies the appropriateness of adaptation data before performing FCMAP adaptation. The hybrid SVM-FLC MAP method may provide an immediate solution to the problem of unreliable FCMAP adaptation caused by inadequate adaptation data. However, hybrid SVM-FLC MAP encounters significant difficulty when the majority of adaptation data are unacceptable. The SVM classifier design of hybrid SVM-FLC MAP adopts a direct "hard classification" scheme that classifies each adaptation utterance from the speaker as valid or invalid. Only utterances verified as valid are re-sent to the following adaptation procedure. This adaptation process is questionable: only a slight adaptation occurs, or possibly, no adaptation at all. Therefore, speech recognition performance would not be improved.

This study proposes a scheme for MAP adaptation (SVM-embedded FLCMAP adaptation) that follows the same principle as FCMAP adaptation by regulating the adaptation according to the quality and quantity of adaptation data under an FLC mechanism. This addresses the undesired effect of poor hybrid SVM-FLC MAP adaptation resulting from the inflexible design of accepting or rejecting adaptation data completely. Compared with FCMAP, the fuzzy controller of the proposed SVM-embedded FLCMAP is an enhanced version that considers one key issue on the quality of adaptation data in addition to the quantity of adaptation data; the fuzzy control mechanism in SVM-embedded FLCMAP would be regulated in a novel manner that allows speaker adaptation to be implemented entirely when adaptation data are both sufficient and highly qualified, and the adaptation speed would be adjusted to be as slow as possible if poor and substandard adaptation data are encountered. Furthermore, in contrast to the "hard classification" of hybrid SVM-FLC MAP for determining the quality of adaptation data to be only acceptable or unacceptable, SVM-embedded FLCMAP offers a "soft classification" scheme that is more flexible by using the SVM to estimate the degree of availability of adaptation data. Fuzzy theory has been used in the SVM classifier (Fuzzy SVM, FSVM) [19], where the method of employing the fuzzy mechanism was completely different to this research. The design of the FLC mechanism's fuzzy inference being driven by SVM-derived information has rarely been attempted and is therefore an interesting issue to be explored in this research. FLC is a popular modeling technique [20-22] and has been used in a wide range of successful applications [23-25], including speaker adaptation applications in previous studies [1,2,16]. In summary, the proposed SVM-embedded FLCMAP adaptation method in this research has several advantages compared with those without:

- robustness against both the scarcity and the inappropriateness of adaptation data,
- an improved adaptation method with an enhanced FLC as compared with previously developed FLC-based speaker adaptation methods that neglect the quality of adaptation data in FLC design (e.g., FCMAP [16] and FLC-MLED eigenvoice adaptation [2]), and
- improved flexibility and greater performances, compared with those with rigid verification schemes of adaptation data (e.g., hybrid SVM-FLC MAP adaptation [1]).

The remainder of this research is organized as follows. Section 2 details the theoretical formulation of FCMAP and hybrid SVM-FLC MAP adaptation methods. Section 3 introduces the concept that a confidence factor parameter derived from SVM is used for indicating the quality of adaptation utterances, followed by the formulation of the fuzzy control mechanism of the developed SVM-embedded FLCMAP for model adaptation in this research. Section 4 presents the experiment results where the effectiveness and performance of SVM-embedded FLCMAP are demonstrated, compared with the conventional MAP, FCMAP, and hybrid SVM-FLC MAP. Finally, Section 5 provides concluding remarks.

2. FCMAP and Hybrid SVM-FLC MAP.

As discussed, MAP adaptation is a type of direct model adaptation that attempts to re-estimate the model parameters directly [8]. However, MAP adaptation re-estimates only a portion of the model parameter units associated with the adaptation data. Therefore, MAP adaptation usually requires a significant amount of data. The recognition performance is improved as the adaptation data increase and the adaptation gets covering the model space. When a sufficient amount of data is available, the MAP estimation yields recognition performances equal to those obtained using maximum-likelihood estimation [7]. As shown in (1),

$$\hat{\mu}_k = \frac{N_k}{\tau + N_k}\bar{y}_k + \frac{\tau}{\tau + N_k}\mu_k \quad \text{(MAP adaptation formula)}, \tag{1}$$

the MAP estimate of the mean is essentially a weighted average of the prior mean (μ_k) and the sample mean ($\bar{y}_k$), and the weights are functions of the number of adaptation samples (N_k) if τ is fixed.

2.1. FCMAP.

The MAP adaptation formula in (1) shows that when N_k is equal to zero (i.e., no additional training data are available for adapting the kth Gaussian), the estimate is simply the prior mean of the kth Gaussian (e.g., μ_4 and μ_5 in Figure 1). Conversely, when a large number of training samples are used for the kth Gaussian ($N_k \to \infty$, as an extreme example), the MAP estimate in (1) converges asymptotically to the maximum likelihood estimate (i.e., the sample mean parameter with the kth Gaussian, $\bar{y}_k$). If N_k

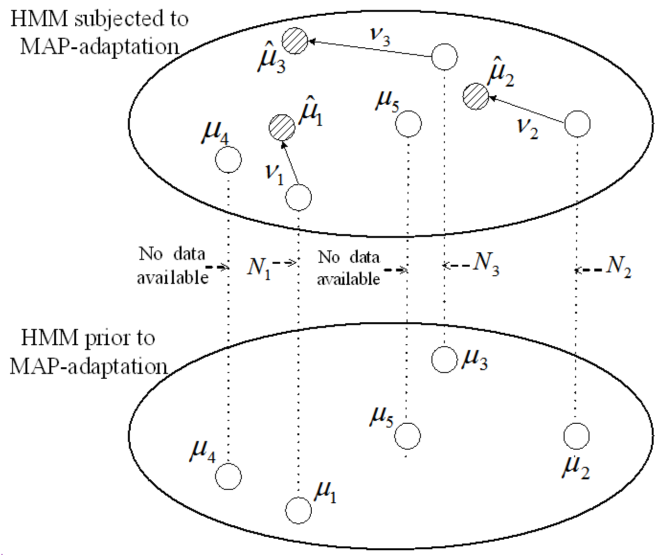

FIGURE 1. Rationale behind MAP adaptation

is fixed, the parameter τ controls the interpolation balance between the $\bar{y}_k$-term and the μ_k-term, (as N_k does). This is referred to as the "adaptation speed parameter" in [26,27] and allows the adaptation speed to be increased or decreased by choosing a small or a large value of τ, respectively. Parameter τ is also known as a "prior density parameter" because it determines which side of and the proximity to $\bar{y}_k$ or μ_k that the MAP-estimate of $\hat{\mu}_k$ would be.

The robustness of MAP adaptation against relatively small training samples N should not be overlooked either, and as yet in conventional schemes for MAP adaptation (e.g., [27-29]), a common value of τ was used for all the Gaussians of a given state, or for all states of an HMM, or even for all HMMs. To improve MAP, Juang et al. [16], in their proposed FCMAP adaptation, developed the concept that $\hat{\mu}_k$ should remain in proximity to μ_k when N is relatively small (as defined by a large τ value) to avoid performance degradation caused by a potentially poor estimate of $\bar{y}_k$. Conversely, when N is large enough, the adaptation should move quickly toward $\bar{y}_k$ (as defined by a small τ value). Stating this concept as simple rules leads to the following statements [16]:

Rule 1: If N is small, then τ should be large, so that $\hat{\mu}_k$ remains in proximity to μ_k.

Rule 2: If N is medium, then τ should be medium, so that $\hat{\mu}_k$ is between $\bar{y}_k$ and μ_k, accordingly.

Rule 3: If N is large, then τ should be small, so that $\hat{\mu}_k$ converges toward $\bar{y}_k$.

FCMAP incorporates an FLC into MAP to adjust the value of τ according to the value of adaptation samples N. The method for formulating statements of uncertain linguistic terms in quantized forms for subsequent computations is shown in [16]. However, the quality is more critical than the quantity of adaptation data for speaker adaptation techniques including MAP. The primary weakness of FCMAP is that the crucial issue of the quality of adaptation data is not considered in the FLC design of FCMAP.

2.2. **Hybrid SVM-FLC MAP.** Although FCMAP may be effective when limited adaptation data are available, FCMAP performance cannot be maintained in a situation where the quality of the acquired adaptation data is low. Previous studies have addressed this issue by developing a validation scheme of adaptation data for the FCMAP adaptation framework [1]. The issue of unreliable adaptation caused by inadequate training data was addressed, and a hybrid scheme of a support vector machine and fuzzy logic control for

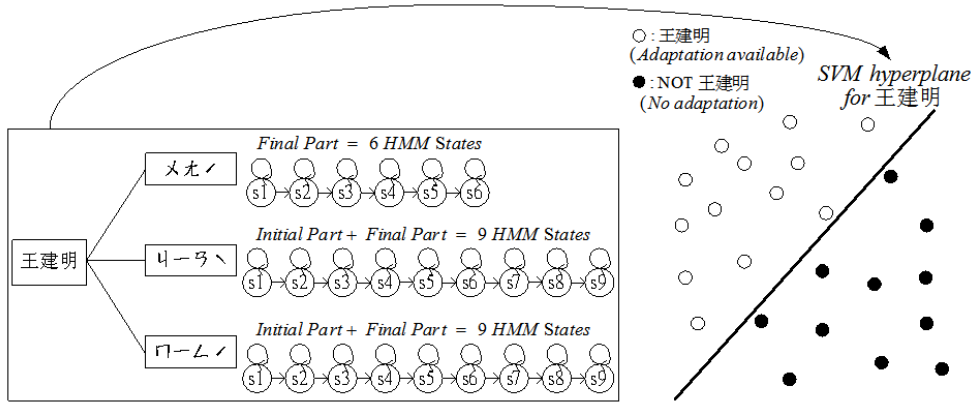

Figure 2. An example to show "hard classification" of adaptation data by hybrid SVM-FLC MAP

MAP adaptation (SVM-FLC MAP) was proposed. The hybrid SVM-FLC MAP approach incorporated an SVM classifier into the adaptation process to verify if the data are appropriate for adaptation before model parameters were adjusted by FCMAP adaptation [1]. If the adaptation data from the speaker were invalid, the data were rejected and the adaptation was not performed. Otherwise, the FCMAP adaptation procedure would start and the FLC of FCMAP would regulate the value of τ according to the amount of adaptation data.

Figure 2 shows an example of a Chinese name "王建明" (Wang, Jian-Ming) to detail how the hybrid SVM-FLC MAP approach decides to accept or reject adaptation data. For a human name database keyword-spotting speech recognition system, each keyword pattern has its own HMM speech model [30] and SVM determination model. A keyword-spotting procedure is used to extract a keyword pattern of the input speech utterance in this system. The HMM state sequence of the keyword pattern is then analyzed [30]. The SVM determination model of the keyword pattern then follows the analysis of the HMM speech model. Assume that the input speech utterance is recognized as the Chinese name "王建明" after the keyword-spotting procedure is completed. The pattern "王建明" has its own HMM state sequence corresponding to the speech utterance, and the SVM determination model of "王建明" is then selected. The SVM determination model of "王建明" has a separating hyperplane that has been trained in advance using training patterns. The separating hyperplane of the SVM determination model of "王建明" can be used to determine if the input speech signal matches "王建明". If the input speech signal is in the "王建明" side of the separating hyperplane, the input speech signal has the meaning of "王建明" and is then accepted as valid data for FCMAP adaptation. Otherwise, the FCMAP adaptation process does not occur until the valid adaptation data have been produced.

Nevertheless, one key issue has been neglected in the presented hybrid SVM-FLC MAP scheme from the perspective on the prohibition of the overall adaptation activity. In the extreme circumstance that all adaptation utterances from the speaker are declared as invalid data and prohibited from the following FCMAP adaptation, no adaptation is performed and the performance and FCMAP effectiveness is restricted.

3. **The Proposed SVM-Embedded FLCMAP.** To address the issue of an unreliable adaptation in FCMAP caused by the absence of a verification scheme for the quality of the acquired adaptation data and avoid the occurrence of the phenomenon of no any adaptation works conducted in hybrid SVM-FLC MAP, a more flexible adaptation framework,

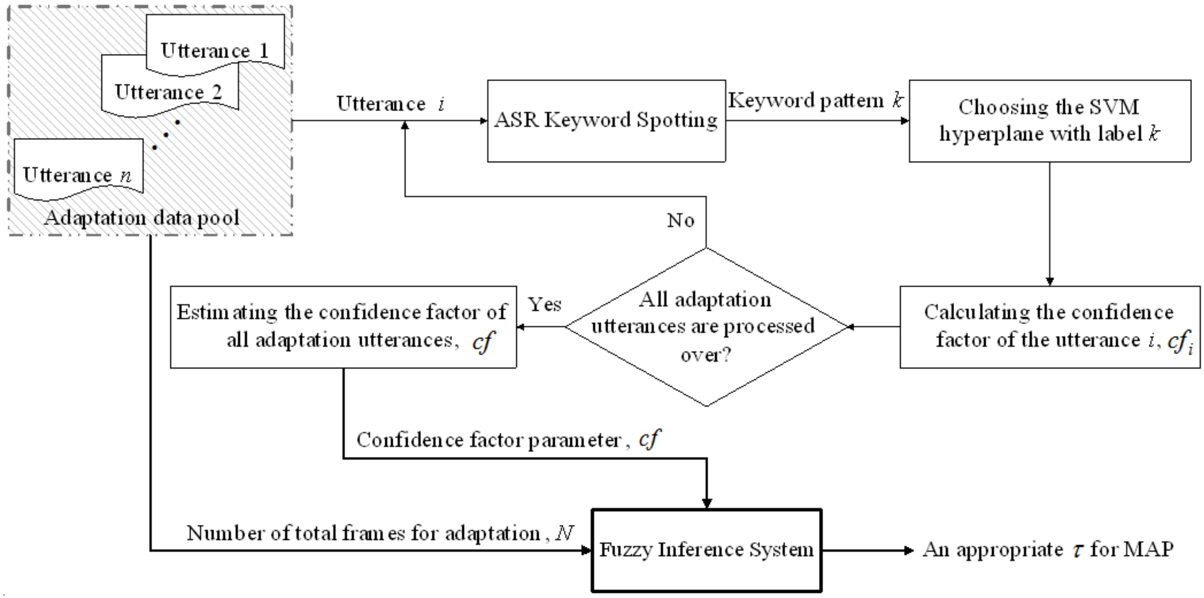

FIGURE 3. Adaptation frameworks of the proposed SVM-embedded FLCMAP

SVM-embedded FLCMAP, is developed and shown in Figure 3. The concept supporting the proposed SVM-embedded FLCMAP is that each adaptation utterance should be considered when performing MAP adaptation, regardless of the quality of the adaptation utterance. Compared with hybrid SVM-FLC MAP adaptation, SVM-embedded FLCMAP adaptation is a compromise for bringing in the impact of each adaptation utterance. To evaluate the attribute of total adaptation data from a new speaker, a confidence factor (*cf*) parameter is calculated to estimate the quality of each adaptation utterance according to its corresponding SVM model (Figure 3). By accumulating the values of each *cf* parameter, a representative value of total confidence factors is then derived for all adaptation utterances. Finally, the fuzzy inference engine of the FLC of SVM-embedded FLCMAP performs the reasoning of τ values for MAP adaptation based on two input signals: the calculated value of the *cf* parameter and the total amount of adaptation data (N).

3.1. **Derivation of confidence factor ($\boldsymbol{cf}$) parameter.** This section uses the SVM model to determine the *cf* value of each adaptation utterance. The SVM model is widely used because it is conceptually simple [17,18]. SVM classifies new input data by using a separating hyperplane. Assume that a set of labeled training points is $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. Each training point x_i belongs to either of two classes and is given a label y_i, $y_i \in \{-1, 1\}$, for $i = 1, 2, \ldots, n$. From these training data, the hyperplane is

$$w \cdot x + b = 0, \tag{2}$$

and is defined by the pair (w, b). Point x_i can be separated according to the following function:

$$f(x_i) = sign(w \cdot x_i + b) = \begin{cases} 1, & \text{if } y_i = 1 \\ -1, & \text{if } y_i = -1 \end{cases}. \tag{3}$$

For the ith adaptation utterance from a new speaker, a keyword-spotting ASR system extracts the keyword speech segment from an utterance and recognizes it as the keyword pattern k [30]. According to the recognized keyword pattern k, a trained SVM hyperplane with label k is selected to verify the ith adaptation utterance. An index cf_i (confidence factor of the ith adaptation utterance) that governs the degree of availability of the ith adaptation utterance is devised as follows:

$$cf_i = \begin{cases} 1, & \text{if } i\text{th utterance classified as pattern } k \text{ by } k\text{th SVM} \\ \exp\left(\frac{-d_{i,k}}{f}\right), & \text{if } i\text{th utterance classified as not pattern } k \text{ by } k\text{th SVM} \end{cases}, \tag{4}$$

where $d_{i,k}$ indicates the distance between the extracted keyword segment from the ith adaptation utterance and the kth SVM separating hyperplane corresponding to the recognized keyword pattern; f denotes the weight control parameter.

The index cf_i varies depending on the amount of confidence in the quality of the ith adaptation utterance. The rationale supporting (4) is that the adaptation utterance should be set as the larger *cf* value when its quality is superior, i.e., the adaptation utterance is more confident to perform adaptation; conversely, a small *cf* value would be suitable for a low-grade adaptation utterance to avoid endangering the system.

When checking the validity of adaptation data by SVM, the extracted keyword segment from the ith adaptation utterance has the meaning of the keyword pattern k. Its availability for adaptation is then verified by the SVM hyperplane with label k. The separating hyperplane of the SVM model of the keyword pattern k classifies the input extracted keyword segment as "keyword pattern k" or "not keyword pattern k". If the input speech signal falls on the side of keyword pattern k of the separating hyperplane, the input speech signal is identified as having the meaning of keyword pattern k. This means that the speech signal is of sufficient quality and qualifies for adaptation. Therefore, the *cf* value of the input speech signal is undoubtedly set to 1. Conversely, if the input speech signal falls on the side of "not keyword pattern k" of the separating hyperplane, the input speech signal does not match the meaning of keyword pattern k completely. This would be significantly problematic when performing a complete adaptation. There is a tendency to be more conservative when using imperfect training data to adapt the system, i.e., the effect of the adaptation should be restricted in this case, so that the degree of adaptation should be considered by the data relying on the similarity between the data and keyword pattern k.

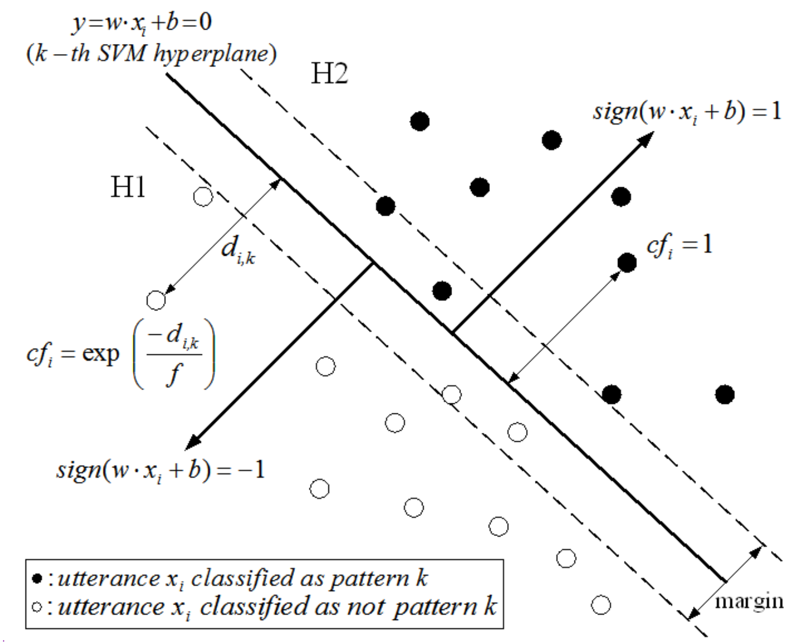

FIGURE 4. Rationale behind the devised confidence factor (*cf*) index

Nevertheless, one key issue has been neglected in the preceding hybrid SVM-FLC MAP adaptation scheme [1] from the perspective of the degree of availability of the faulty adaptation utterance. For the adaptation utterance that falls on the incorrect side of the SVM separating hyperplane, hybrid SVM-FLC MAP would reject it and not adopt it for adaptation. In this research, however, the *cf* value of the adaptation utterance could be viewed as a reference index for adaptation. As shown in (4), the *cf* value of the ith adaptation utterance could be determined purely geometrically by the distance $d_{ik} = \|\mu_i - \mu_k\|$, together with a tuning parameter f if the adaptation utterance is classified as "not keyword pattern k". When the quality of ith adaptation utterance is in doubt because of falling on the side of "not keyword pattern k" of the SVM hyperplane and the distance from the hyperplane is large (i.e., d_{ik} is large), then cf_i should be small. When the ith adaptation utterance is reliable, cf_i should be set as 1 in instances where the utterance is categorized as part of keyword pattern k, or cf_i should be set larger if the utterance still falls on the incorrect side of the SVM hyperplane but is in proximity to the hyperplane (i.e., d_{ik} is small). Figure 4 shows the "phenomenon" implicated by (4).

With the *cf* of each adaptation utterance, what degree to the value of each confidence factor should be considered for all adaptation data? An average of all cf_i's as shown in (5), cf, is an obvious selection.

$$cf = \frac{\sum_{j=1}^{P} cf_j}{P}, \tag{5}$$

where cf is the averaged estimate of confidence factors of total P adaptation utterances. The value of cf denotes the quality of all adaptation data collected from a new speaker. A higher cf value indicates greater reliability of adaptation data.

3.2. **MAP under SVM-embedded FLC regulation.** As discussed, MAP adaptation performance is significantly dependent on the attributes of the adaptation data. As shown in (1), FCMAP enhances MAP by incorporating a fuzzy controller to adjust the value of τ according to the adaptation data size [16]. However, FCMAP does not consider the quality of adaptation data when developing the fuzzy controller. In this research, an SVM-embedded fuzzy logic control scheme with an index cf derived from SVM is proposed for MAP adaptation. The developed FLC considers the quantity and quality of data.

Consider the influence of the variation of both the total amount of adaptation data (N) and the quality of total adaptation data (cf) on the performance of MAP adaptation. When N and cf are large (i.e., available adaptation data are sufficient and appropriate for adaptation), adaptation should be fast by setting a small value for τ. Otherwise, a large value of τ should be set to prevent the adapted mean vector $\hat{\mu}_k$ from diverging significantly from the prior mean μ_k. A rule base with four fuzzy implications is given to govern the regulation of τ under the circumstance of two antecedents (N and cf), as follows:

Rule 1: If cf is small and N is small, then τ is set to be large,
Rule 2: If cf is small and N is large, then τ is set to be rather large,
Rule 3: If cf is large and N is small, then τ is set to be rather small,
Rule 4: If cf is large and N is large, then τ is set to be small.

Let $A_1(cf)$ and $A_2(cf)$ be the membership functions associated respectively with small and large values of *cf*, and $B_1(N)$ and $B_2(N)$ be the membership functions associated respectively with small and large amounts of adaptation data available, as shown in Figures 5 and 6.

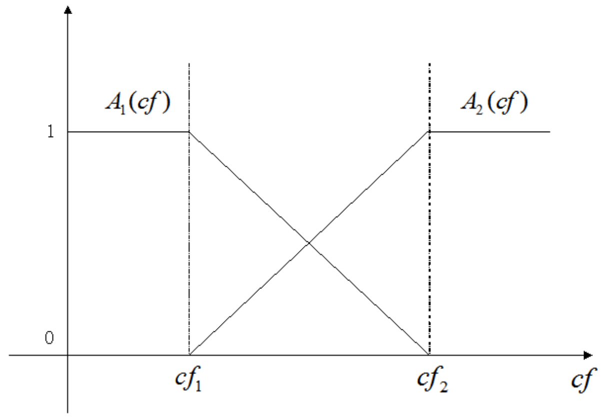

FIGURE 5. Membership functions associated with cf

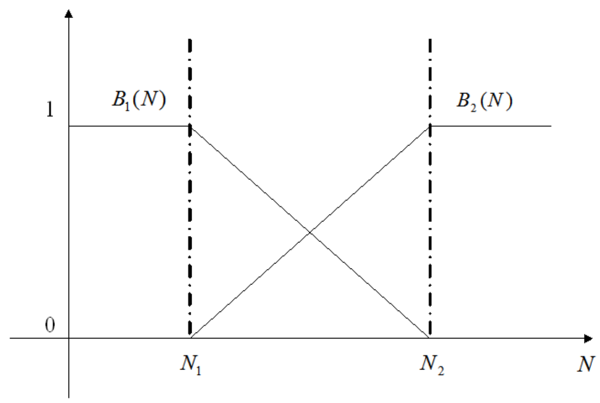

FIGURE 6. Membership functions associated with N

In addition, let the functions $f_1(cf, N)$, $f_2(cf, N)$, $f_3(cf, N)$ and $f_4(cf, N)$ set large, relatively large, relatively small, and small values of τ respectively in each of the four cases. The previous set of rules can then be further clarified as

Rule 1: If cf is $A_1(cf)$ and N is $B_1(N)$, then $\tau = f_1(cf, N)$,
Rule 2: If cf is $A_1(cf)$ and N is $B_2(N)$, then $\tau = f_2(cf, N)$,
Rule 3: If cf is $A_2(cf)$ and N is $B_1(N)$, then $\tau = f_3(cf, N)$,
Rule 4: If cf is $A_2(cf)$ and N is $B_2(N)$, then $\tau = f_4(cf, N)$,

where

$$A_1(cf) = \begin{cases} 1 & cf \le cf_1, \\ \frac{cf_2 - cf}{cf_2 - cf_1} & cf_1 < cf < cf_2, \\ 0 & cf \ge cf_2, \end{cases} \qquad A_2(cf) = \begin{cases} 0 & cf \le cf_1, \\ \frac{cf - cf_1}{cf_2 - cf_1} & cf_1 < cf < cf_2, \\ 1 & cf \ge cf_2, \end{cases}$$

$$B_1(N) = \begin{cases} 1 & N \le N_1, \\ \frac{N_2 - N}{N_2 - N_1} & N_1 < N < N_2, \\ 0 & N \ge N_2, \end{cases} \qquad B_2(N) = \begin{cases} 0 & N \le N_1, \\ \frac{N - N_1}{N_2 - N_1} & N_1 < N < N_2, \\ 1 & N \ge N_2, \end{cases} \tag{6}$$

along with the implication functions

$$f_i(cf, N) = a_i \cdot cf + b_i \cdot N + c_i, \quad i = 1, 2, 3, 4, \tag{7}$$

and the final system output, as follows [20]:

$$\tau = \frac{\sum_{i=1}^{4} w^i \cdot f_i(cf, N)}{\sum_{i=1}^{4} w^i}, \tag{8}$$

where

$$\begin{aligned} &w^1 = A_1(cf) \cdot B_1(N), \quad w^2 = A_1(cf) \cdot B_2(N), \\ &w^3 = A_2(cf) \cdot B_1(N), \quad w^4 = A_2(cf) \cdot B_2(N). \end{aligned} \tag{9}$$

The system now has 16 hyperparameters (a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4, c_1, c_2, c_3, c_4, cf_1, cf_2 N_1, N_2) to be fixed. The following iterative process is developed to set these hyperparameters.

Step 1: An initialization setting for parameters is done in the first step here. Let $cf_1 : cf_2 = 1 : 3$ and $N_1 : N_2 = 1 : 3$. And initialize parameters cf_1 and N_1.

Step 2: Estimate the parameters a_1, b_1 and c_1 under the conditions, $cf \le cf_1$ and $N \le N_1$, where

$$\begin{aligned} &A_1(cf) = B_1(N) = 1, \quad A_2(cf) = B_2(N) = 0, \\ &w^1 = 1, \quad w^2 = w^3 = w^4 = 0 \text{ and} \\ &\tau = f_1(cf, N) = a_1 \cdot cf + b_1 \cdot N + c_1. \end{aligned}$$

In this case, i.e., the amount of adaptation data $N \le N_1$ and the confidence factor $cf \le cf_1$, 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 performance P^i,

$$P^i = \textit{speech\_recognition}(\tau = a_1 \cdot cf + b_1 \cdot N + c_1,\ \textit{tunning\_utterances}),$$

where the function *speech_recognition*$(\cdot)$ is used to return the recognition performance of the proposed SVM-embedded FLCMAP adaptation with the parameter τ controlled by selecting a_1, b_1 and c_1 for the testing data set *tunning_utterances*, and P^i denotes the returned recognition performance after performing the i-th iteration. Note that the try-and-error procedure for fixing the 3 hyperparameters of FLC would thus finally return an overall recognition rate that is better than the baseline P^0. The procedure for fixing a_1, b_1 and c_1 is explained in the following algorithm:

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 P^i using the function *speech_recognition*$(\cdot)$.
    IF $(P^i > P^{i-1})$ **THEN**
        DO UNTIL $(P^i \le P^{i-1})$
            Increment a_1.
            Increment i.
            Determine P^i using the function *speech_recognition*$(\cdot)$.
        END DO UNTIL
    ELSE
        DO UNTIL $(P^i \le P^{i-1})$
            Decrement a_1.
            Increment i.
            Determine P^i using the function *speech_recognition*$(\cdot)$.

```
        END DO UNTIL
    END IF
    Increment b1.
    Increment i.
    Calculate P^i using the function speech_recognition(·).
    IF (P^i > P^(i-1)) THEN
        DO UNTIL (P^i ≤ P^(i-1))
            Increment b1.
            Increment i.
            Determine P^i using the function speech_recognition(·).
        END DO UNTIL
    ELSE
        DO UNTIL (P^i ≤ P^(i-1))
            Decrement b1.
            Increment i.
            Determine P^i using the function speech_recognition(·).
        END DO UNTIL
    END IF
    Increment c1.
    Increment i.
    Calculate P^i using the function speech_recognition(·).
    IF (P^i > P^(i-1)) THEN
        DO UNTIL (P^i ≤ P^(i-1))
            Increment c1.
            Increment i.
            Determine P^i using the function speech_recognition(·).
        END DO UNTIL
    ELSE
        DO UNTIL (P^i ≤ P^(i-1))
            Decrement c1.
            Increment i.
            Determine P^i using the function speech_recognition(·).
        END DO UNTIL
    END IF
END
```

Step 3: Estimate the parameters a_2, b_2 and c_2 under the conditions, $cf \leq cf_1$ and $N \geq N_2$, where

$$\begin{aligned} &A_1(cf) = B_2(N) = 1, \quad A_2(cf) = B_1(N) = 0, \\ &w^2 = 1, \quad w^1 = w^3 = w^4 = 0 \quad \text{and} \\ &\tau = f_2(cf, N) = a_2 \cdot cf + b_2 \cdot N + c_2. \end{aligned}$$

The values of a_2, b_2 and c_2 are fixed using the same process as for a_1, b_1 and c_1 with the initial condition $P^0 = P^i$ from Step 2.

Step 4: Estimate the parameters a_3, b_3 and c_3 under the conditions, $cf \geq cf_2$ and $N \leq N_1$, where

$$\begin{aligned} &A_2(cf) = B_1(N) = 1, \quad A_1(cf) = B_2(N) = 0, \\ &w^3 = 1, \quad w^1 = w^2 = w^4 = 0 \quad \text{and} \\ &\tau = f_3(cf, N) = a_3 \cdot cf + b_3 \cdot N + c_3. \end{aligned}$$

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 $P^0 = P^i$ from Step 3.

Step 5: Estimate the parameters a_4, b_4 and c_4 under the conditions, $cf \geq cf_2$ and $N \geq N_2$, where

$$\begin{aligned}
&A_2(cf) = B_2(N) = 1, \quad A_1(cf) = B_1(N) = 0,\\
&w^4 = 1, \quad w^1 = w^2 = w^3 = 0 \quad \text{and}\\
&\tau = f_4(cf, N) = a_4 \cdot cf + b_4 \cdot N + c_4.
\end{aligned}$$

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 $P^0 = P^i$ from Step 4.

Step 6: Re-estimate both the parameter sets $\{a_1, b_1, c_1\}$ and $\{a_2, b_2, c_2\}$ again under the conditions, $cf \leq cf_1$ and $N_1 \leq N \leq N_2$, where

$$\begin{aligned}
&A_1(cf) = 1, \quad A_2(cf) = 0, \quad B_1(N) = \frac{N_2 - N}{N_2 - N_1}, \quad B_2(N) = \frac{N - N_1}{N_2 - N_1},\\
&w^1 = B_1(N), \quad w^2 = B_2(N), \quad w^3 = w^4 = 0 \quad \text{and}\\
&\tau = \frac{w^1 \cdot f_1(cf, N) + w^2 \cdot f_2(cf, N)}{w^1 + w^2}\\
& = \frac{N_2 - N}{N_2 - N_1} \cdot (a_1 \cdot cf + b_1 \cdot N + c_1) + \frac{N - N_1}{N_2 - N_1} \cdot (a_2 \cdot cf + b_2 \cdot N + c_2).
\end{aligned}$$

With the initial condition $P^0 = P^i$ from Step 5 and the initial values of $\{a_1, b_1, c_1\}$ and $\{a_2, b_2, c_2\}$ already separately determined at Steps 2 and 3 respectively, both the parameter sets $\{a_1, b_1, c_1\}$ and $\{a_2, b_2, c_2\}$ could be re-calculated with determined initial values through the same tuning process as in Step 2 for best recognition rate too.

Step 7: Re-estimate the parameter set $\{a_4, b_4, c_4\}$ again under the conditions, $cf_1 \leq cf \leq cf_2$ and $N \geq N_2$, where

$$\begin{aligned}
&A_1(cf) = \frac{cf_2 - cf}{cf_2 - cf_1}, \quad A_2(cf) = \frac{cf - cf_1}{cf_2 - cf_1}, \quad B_1(N) = 0, \quad B_2(N) = 1,\\
&w^1 = w^3 = 0, \quad w^2 = A_1(cf), \quad w^4 = A_2(cf) \quad \text{and}\\
&\tau = \frac{w^2 \cdot f_2(cf, N) + w^4 \cdot f_4(cf, N)}{w^2 + w^4}\\
& = \frac{cf_2 - cf}{cf_2 - cf_1} \cdot (a_2 \cdot cf + b_2 \cdot N + c_2) + \frac{cf - cf_1}{cf_2 - cf_1} \cdot (a_4 \cdot cf + b_4 \cdot N + c_4).
\end{aligned}$$

With the initial values of $\{a_4, b_4, c_4\}$ already individually determined at Step 5 and the values of $\{a_2, b_2, c_2\}$ already re-estimated at Step 6, new values for $\{a_4, b_4, c_4\}$ can now be acquired by tuning for a higher P^i value than in Step 6.

Step 8: Re-estimate the parameter set $\{a_3, b_3, c_3\}$ again under the conditions, $cf \geq cf_2$ and $N_1 \leq N \leq N_2$, where

$$\begin{aligned}
&A_1(cf) = 0, \quad A_2(cf) = 1, \quad B_1(N) = \frac{N_2 - N}{N_2 - N_1}, \quad B_2(N) = \frac{N - N_1}{N_2 - N_1},\\
&w^1 = w^2 = 0, \quad w^3 = B_1(N), \quad w^4 = B_2(N) \quad \text{and}\\
&\tau = \frac{w^3 \cdot f_3(cf, N) + w^4 \cdot f_4(cf, N)}{w^3 + w^4}\\
& = \frac{N_2 - N}{N_2 - N_1} \cdot (a_3 \cdot cf + b_3 \cdot N + c_3) + \frac{N - N_1}{N_2 - N_1} \cdot (a_4 \cdot cf + b_4 \cdot N + c_4).
\end{aligned}$$

Since initial values of $\{a_3,\ b_3,\ c_3\}$ already separately determined at Step 4 and the re-estimated values of $\{a_4,\ b_4,\ c_4\}$ already calculated at Step 7, new values for $\{a_3,\ b_3,\ c_3\}$ can now be obtained by tuning for a higher P^i value than in Step 7.

Step 9: Re-estimate the parameter cf_2 under the conditions, $cf_1 \le cf \le cf_2$ and $N \le N_1$, where

$$
\begin{aligned}
&A_1(cf) = \frac{cf_2 - cf}{cf_2 - cf_1}, \quad A_2(cf) = \frac{cf - cf_1}{cf_2 - cf_1}, \quad B_1(N) = 1, \quad B_2(N) = 0,\\
&w^1 = A_1(cf), \quad w^3 = A_2(cf), \quad w^2 = w^4 = 0 \quad \text{and}\\
&\tau = \frac{w^1 \cdot f_1(cf, N) + w^3 \cdot f_3(cf, N)}{w^1 + w^3}\\
&\ \ = \frac{cf_2 - cf}{cf_2 - cf_1} \cdot (a_1 \cdot cf + b_1 \cdot N + c_1) + \frac{cf - cf_1}{cf_2 - cf_1} \cdot (a_3 \cdot cf + b_3 \cdot N + c_3).
\end{aligned}
$$

Since the values of $\{a_1,\ b_1,\ c_1\}$ together with the values of $\{a_3,\ b_3,\ c_3\}$ have been already re-calculated at Step 6 and Step 8 respectively, a new value for cf_2 can now be obtained by tuning for a higher P^i value than in Step 8.

Step 10: Update the value of cf_1 such that $cf_1 : cf_2 = 1 : 3$.

Step 11: Re-estimate the parameter N_2 under the conditions, $cf_1 \le cf \le cf_2$ and $N_1 \le N \le N_2$, where

$$
\begin{aligned}
&A_1(cf) = \frac{cf_2 - cf}{cf_2 - cf_1}, \quad A_2(cf) = \frac{cf - cf_1}{cf_2 - cf_1}, \quad B_1(N) = \frac{N_2 - N}{N_2 - N_1}, \quad B_2(N) = \frac{N - N_1}{N_2 - N_1},\\
&w^1 = \frac{cf_2 - cf}{cf_2 - cf_1} \cdot \frac{N_2 - N}{N_2 - N_1}, \quad w^2 = \frac{cf_2 - cf}{cf_2 - cf_1} \cdot \frac{N - N_1}{N_2 - N_1},\\
&w^3 = \frac{cf - cf_1}{cf_2 - cf_1} \cdot \frac{N_2 - N}{N_2 - N_1}, \quad w^4 = \frac{cf - cf_1}{cf_2 - cf_1} \cdot \frac{N - N_1}{N_2 - N_1} \quad \text{and}\\
&\tau = \frac{cf_2 - cf}{cf_2 - cf_1} \cdot \frac{N_2 - N}{N_2 - N_1} \cdot (a_1 \cdot cf + b_1 \cdot N + c_1)\\
&\quad + \frac{cf_2 - cf}{cf_2 - cf_1} \cdot \frac{N - N_1}{N_2 - N_1} \cdot (a_2 \cdot cf + b_2 \cdot N + c_2)\\
&\quad + \frac{cf - cf_1}{cf_2 - cf_1} \cdot \frac{N_2 - N}{N_2 - N_1} \cdot (a_3 \cdot cf + b_3 \cdot N + c_3)\\
&\quad + \frac{cf - cf_1}{cf_2 - cf_1} \cdot \frac{N - N_1}{N_2 - N_1} \cdot (a_4 \cdot cf + b_4 \cdot N + c_4).
\end{aligned}
$$

With both $\{a_1,\ b_1,\ c_1\}$ and $\{a_2,\ b_2,\ c_2\}$ already re-determined at Step 6, $\{a_3,\ b_3,\ c_3\}$ and $\{a_4,\ b_4,\ c_4\}$ already re-calculated at Steps 8 and 7 respectively, cf_1 and cf_2 already re-estimated at Steps 10 and 9, a new value for N_2 can now be obtained by tuning for a higher P^i value than in Step 10.

Step 12: Update the value of N_1 such that $N_1 : N_2 = 1 : 3$,

$$
\begin{aligned}
&\delta = \frac{|P^i - P^*|}{P^*}, \quad /\ {*P^*} :\ \text{desired recognition rate } {*}/\\
&P^0 = P^i.
\end{aligned}
$$

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

4. **Experiments and Results.** The experimental settings and results of the proposed SVM-embedded FLCMAP speaker adaptation algorithm are respectively reported in the following subsections.

4.1. **Database and experiment design.** The speech signal was sampled at 8 kHz. The analysis frames were 30 ms wide with a 20 ms overlap. For each frame, a 24-dimensional feature vector was extracted. The feature vector for each frame was composed of a 12-dimensional mel-cepstral vector and a 12-dimensional delta-mel-cepstral vector.

Before conducting comparative experiments to show the effectiveness of the SVM-embedded FLCMAP adaptation method proposed in Section 3, it is necessary to establish the initial SI models. The database MAT400 sub-database DB3 [31], which comprises 4800 utterances from native Mandarin speakers, was used to develop the initial SI models in the form of a set of HMM parameters. In Mandarin, an utterance may contain one to several syllables, and each syllable consists (as HMM states) of a three-state initial part and a six-state final part. Thus, the HMM of a Mandarin utterance includes the HMMs of the constituent syllables, which in turn include an HMM of three states for the initial part (if it is present) and an HMM of six states for the final part [32], totaling 440 states in the SI models.

In the training phase of the SVM of hybrid SVM-FLC MAP and the proposed SVM-embedded FLCMAP, the data used for training the SVM hyperplane parameters were collected from 10 speakers. From each of the 10 speakers, 300 utterances of city names (10 utterances for each of 30 cities) were requested as training data. Thirty SVM models were trained in total, each of which corresponds to each of the 30 city names. For the SVM model of a city name, the training data are composed of two parts: 1) the content of the city name consisting of 100 utterances (10 utterances for each of the 10 speakers), and 2) the content of the other 29 city names, used as the training data of anti-models and consisting of 100 utterances (selected from the remaining 2900 utterances).

In the training phase of the fuzzy controller, a new group of 10 speakers was selected for utterance recording. Thirty utterances of city names (one utterance for each of 30 cities) as adaptation data for establishing SA models and 60 utterances of city names (two utterances for each of the 30 cities) as tuning data for the fuzzy controller of the FCMAP, hybrid SVM-FLC MAP, and the proposed SVM-embedded FLCMAP were collected from each of the 10 speakers.

In the recognition experiments, adaptation and testing data were gathered from a new group of 10 speakers entirely different from the previous groups. The adaptation data consisted of 30 utterances from each speaker (one utterance for each of 30 cities). The testing data consisted of 60 utterances from the speakers, each uttering 30 city names twice. For each speaker, 2, 6, 10, 14, 18, 22, 26 and 30 utterances were selected from their 30-utterance adaptation data for SI model adaptation, and eight sets of SA models were established per speaker. Eighty SA models were thus prepared and used for performance comparison among MAP, FCMAP, hybrid SVM-FLC MAP, and the proposed SVM-embedded FLCMAP adaptations.

4.2. **Experiment results.** The averaged recognition performance of 10 speakers of the conventional MAP with various settings of τ is shown in Table 1. According to Table 1, the conventional MAP has better results when τ is fixed at 30. Thus, the value of τ is chosen in the conventional MAP for comparison. For the recognition experiment with SVM-embedded FLCMAP adaptation, eight adapted models were constructed, using 2, 6, 10, 14, 18, 22, 26 and 30 adaptation utterances from each of the 10 speakers, and τ for each of the eight adaptation sets were calculated using (8) with the derived cf parameter (from (5), with $P = 2$, 6, 10, 14, 18, 22, 26 or 30) and $N_{utterances} = 2$, 6, 10, 14, 18, 22, 26 or 30. FLC hyperparameters had already been determined during the training phase. Figure 7 shows the average recognition rate of the proposed SVM-embedded FLCMAP against the hybrid SVM-FLC MAP, FCMAP, and the conventional MAP with a fixed τ. The proposed

Table 1. Average recognition rates (%) of the conventional MAP with various τ

Value of τ	Average recognition rate (%)								
	Numbers of utterances for adaptation								
	0	2	6	10	14	18	22	26	30
5	93.00	87.0	87.1	88.0	89.1	89.8	91.3	93.2	94.5
10	93.00	87.0	87.2	88.1	89.2	89.8	91.5	93.2	94.5
15	93.00	87.1	87.5	88.3	89.3	90.0	91.5	93.3	94.5
20	93.00	87.1	87.3	88.0	89.3	90.0	91.5	93.3	94.8
25	93.00	87.5	87.5	88.3	89.5	90.1	91.7	93.3	94.8
30	**93.00**	**88.0**	**88.0**	**89.0**	**89.5**	**91.5**	**92.0**	**93.5**	**95.0**
35	93.00	87.3	87.5	88.0	89.1	90.1	91.5	93.3	94.5
40	93.00	87.3	87.3	87.5	89.0	90.5	91.5	93.3	94.5
45	93.00	87.0	87.1	87.5	89.1	90.2	91.3	93.3	94.3
50	93.00	87.0	87.1	87.3	89.0	90.1	91.3	93.2	94.3

SVM-embedded FLCMAP and the other three adaptation methods have an adaptive learning curve. For the conventional MAP and FCMAP, when the amount of training data is insufficient (e.g., two adaptation utterances), the recognition rate is low, even lower than the baseline. Conversely, the recognition rates of the proposed SVM-embedded FLCMAP and hybrid SVM-FLC MAP are equal to or greater than the baseline when the amount of training data is insufficient. Among all four adaptation methods against adaptation data impropriety, the proposed SVM-embedded FLCMAP shows the greatest performance. Furthermore, as the amount of training data increases, the recognition performance of the conventional MAP, FCMAP, and hybrid SVM-FLC MAP improves gradually and is above baseline but still inferior, compared with the SVM-embedded FLCMAP. The SVM-embedded FLCMAP consistently outperforms the other three adaptation methods in all conditions of adaptation utterances. Therefore, the SVM-embedded FLCMAP has a superior adaptive learning curve among all four adaptation approaches. However, FCMAP and hybrid SVM-FLC MAP adaptations perform well when adaptation data are sufficient (e.g., adaptation utterances = 30), but still perform worse than the proposed SVM-embedded FLCMAP. The causes of such behavior may be attributed to the matching of HMM components in the recognition phase not being adapted appropriately when using the hybrid SVM-FLC MAP, or adapted but also "polluted" by some substandard utterances when using FCMAP.

When using scarce and inferior adaptation data (e.g., two adaptation utterances with $cf = 0.3$), the performance of conventional MAP adaptation with various τ values is shown in Figure 8. Increasing τ tends to improve the performance. Figure 9 shows the performance of conventional MAP adaptation with various τ values for cases of sufficient and high-grade adaptation data (e.g., 30 adaptation utterances with $cf = 8.2$). A tendency that increasing the value of τ causes the recognition rate to decline was also observed. These further justify the rationale supporting the design of the SVM-embedded FLCMAP adaptation method.

5. **Conclusions.** The quality of conventional MAP speaker adaptation relies significantly on the adaptation data acquired from a new speaker. This study proposed an SVM-embedded FLCMAP method for further enhancing MAP to account for the quality and the quantity of adaptation data in addition to MAP adaptation tasks. The developed SVM-embedded FLCMAP uses an SVM-embedded fuzzy logic controller to adjust

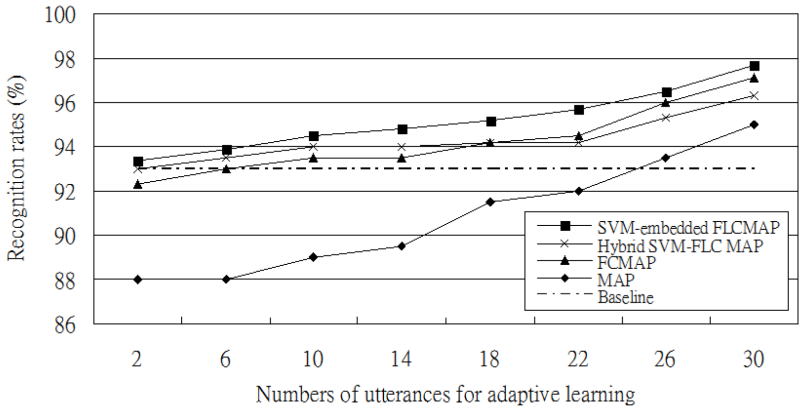

FIGURE 7. Average recognition rates by 10 speakers using SVM-embedded FLCMAP, Hybrid SVM-FLC MAP, FCMAP and conventional MAP adaptation methods

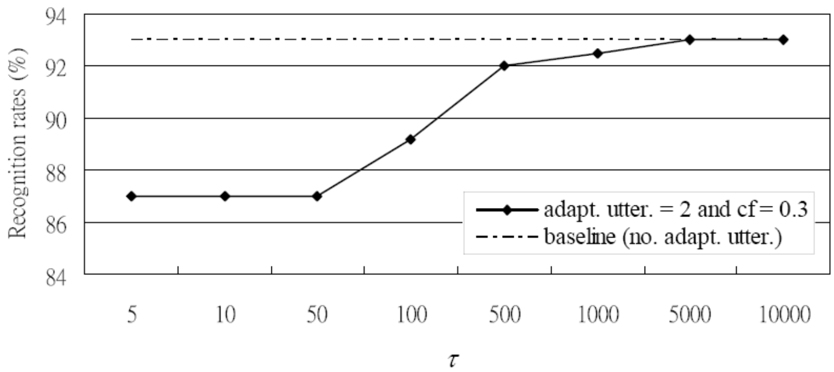

FIGURE 8. The number of adaptation utterances = 2 and cf = 0.3 (conventional MAP testing experiments)

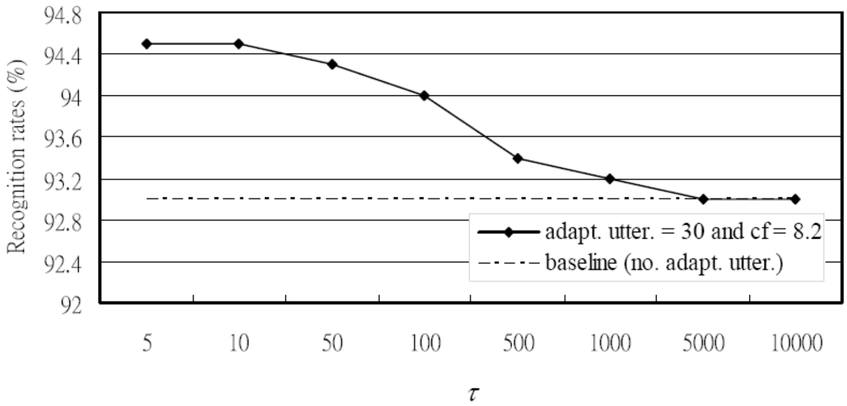

FIGURE 9. The number of adaptation utterances = 30 and cf = 8.2 (conventional MAP testing experiments)

the adaptation speed parameter τ of MAP. In the adaptation framework of the SVM-embedded FLCMAP, the quality information of adaptation data can be acquired first from the support vector machine analysis, which together with the amount of adaptation data, is then input into the fuzzy logic control mechanism to be inputs of fuzzy inference.

The performance evaluation experiments in this research demonstrated the effectiveness of the proposed SVM-embedded FLCMAP speaker adaptation method.

Acknowledgment. This research is partially supported by the National Science Council (NSC) in Taiwan under grant NSC 101-2221-E-150-084. The author also gratefully acknowledges the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] I.-J. Ding, A hybrid SVM-FLC approach to MAP speaker adaptation for speech recognition, *ICIC Express Letters*, vol.5, no.2, pp.349-354, 2011.
[2] I.-J. Ding, Enhancements of maximum likelihood eigen-decomposition using fuzzy logic control for eigenvoice-based speaker adaptation, *International Journal of Innovative Computing, Information and Control*, vol.7, no.7(B), pp.4207-4222, 2011.
[3] K. Shinoda, Acoustic model adaptation for speech recognition, *IEICE Transactions on Information and Systems*, vol.E93-D, no.9, pp.2348-2362, 2010.
[4] M. Nakayama and S. Ishimitsu, Speech support system using body-conducted speech recognition for disorders, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4255-4266, 2009.
[5] X. Wang, J. Lin, Y. Sun, H. Gan and L. Yao, Applying feature extraction of speech recognition on VoIP auditing, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1851-1856, 2009.
[6] T. Guan and Q. Gong, A study on the effects of spectral information encoding in mandarin speech recognition in white noise, *ICIC Express Letters*, vol.3, no.3(A), pp.415-420, 2009.
[7] C. H. Lee, C. H. Lin and B. H. Juang, A study on speaker adaptation of the parameters of continuous density hidden Markov models, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.39, no.4, pp.806-814, 1991.
[8] J. L. Gauvain and C. H. Lee, Maximum a *posteriori* estimation for multivariate Gaussian mixture observations of Markov chains, *IEEE Transactions on Speech and Audio Processing*, vol.2, no.2, pp.291-298, 1994.
[9] C. J. Leggetter and P. C. Woodland, Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models, *Computer Speech and Language*, vol.9, pp.171-185, 1995.
[10] M. J. F. Gales, The generation and use of regression class trees for MLLR adaptation, *Tech. Rep. CUED/F-INFENG/TR263*, Cambridge University, 1996.
[11] R. Kuhn, J.-C. Junqua, P. Nguyen and N. Niedzielski, Rapid speaker adaptation in eigenvoice space, *IEEE Transactions on Speech and Audio Processing*, vol.8, no.6, pp.695-707, 2000.
[12] B. Mak, S. Ho and J. T. Kwok, Speedup of kernel eigenvoice speaker adaptation by embedded kernel PCA, *Proc. of International Conference on Spoken Language Processing*, pp.2913-2916, 2004.
[13] B. Mak, J. T. Kwok and S. Ho, Kernel eigenvoice speaker adaptation, *IEEE Transactions on Speech and Audio Processing*, vol.13, no.5, pp.984-992, 2005.
[14] B. Zhou and J. Hansen, Rapid discriminative acoustic model based on eigenspace mapping for fast speaker adaptation, *IEEE Transactions on Speech and Audio Processing*, vol.13, no.4, pp.554-564, 2005.
[15] B. Mak, R. Hsiao, S. Ho and J. T. Kwok, Embedded kernel eigenvoice speaker adaptation and its implication to reference speaker weighting, *IEEE Transactions on Audio, Speech, and Language Processing*, vol.14, no.4, pp.1267-1280, 2006.
[16] Y. T. Juang, K. C. Huang and I. J. Ding, Speaker adaptation based on MAP estimation using fuzzy controller, *Pattern Recognition Letters*, vol.24, no.15, pp.2807-2813, 2003.
[17] 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.
[18] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning*, vol.20, no.3, pp.273-297, 1995.
[19] C.-F. Lin and S.-D. Wang, Fuzzy support vector machines, *IEEE Transactions on Neural Networks*, vol.13, no.2, pp.464-471, 2002.
[20] 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.
[21] R. Yager and D. Filev, *Essentials of Fuzzy Modeling and Control*, Wiley, 1994.

[22] C. Li, J. Yi and D. Zhao, Design of interval type-2 fuzzy logic system using sampled data and prior knowledge, *ICIC Express Letters*, vol.3, no.3(B), pp.695-700, 2009.

[23] J. Yen, R. Langari and L. A. Zadeh, *Industrial Applications of Fuzzy Logic and Intelligent Systems*, IEEE Press, New York, 1995.

[24] S. Kermiche, M. L. Saidi, H. A. Abbassi and H. Ghodbane, Takagi-Sugeno based controller for mobile robot navigation, *Journal of Applied Science*, vol.6, no.8, pp.1838-1844, 2006.

[25] M. C. M. Teixeira, G. S. Deaecto, R. Gaino, E. Assunção, A. A. Carvalho and U. C. Farias, Design of a fuzzy Takagi-Sugeno controller to vary the joint knee angle of paraplegic patients, *Proc. of International Conference on Neural Information Processing*, pp.118-126, 2006.

[26] J.-I. Takahashi and S. Sagayama, Vector-field-smoothed Bayesian learning for fast and incremental speaker/telephone-channel adaptation, *Computer Speech and Language*, vol.11, pp.127-146, 1997.

[27] Y. Li, S. S. Narayanan and C.-C. J. Kuo, Audiovisual-based adaptive speaker identification, *Proc. of the IEEE International Conference on Acoustic, Speech and Signal Processing*, vol.5, pp.812-815, 2003.

[28] P. C. Woodland, Speaker adaptation: Techniques and challenges, *Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding*, pp.85-90, 1999.

[29] J. L. Gauvain and C. H. Lee, Speaker adaptation based on MAP estimation of hmm parameters, *Proc. of the IEEE International Conference on Acoustic, Speech and Signal Processing*, vol.2, pp.558-661, 1993.

[30] J. G. Wilpon, L. R. Rabiner, C. H. Lee and E. R. Goldman, Automatic recognition of keywords in unconstrained speech using hidden Markov models, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.38, no.11, pp.1870-1878, 1990.

[31] H. C. Wang, MAT – A project to collect mandarin speech data through telephone networks in Taiwan, *Comput. Linguist. Chinese Lang. Process.*, vol.2, pp.73-89, 1997.

[32] C. H. Lin, C. H. Wu, P. Y. Ting and H. M. Wang, Frameworks for recognition of mandarin syllables with tones using sub-syllabic units, *Speech Communication*, vol.18, no.2, pp.175-190, 1996.