Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 898729, 8 pages
http://dx.doi.org/10.1155/2014/898729

Research Article

An HMM-Like Dynamic Time Warping Scheme for Automatic Speech Recognition

Ing-Jr Ding and Yen-Ming Hsu

Department of Electrical Engineering, National Formosa University, No. 64, Wunhua Road, Huwei Township, Yunlin County 632, Taiwan

Correspondence should be addressed to Ing-Jr Ding; ingjr@nfu.edu.tw

Received 30 March 2014; Accepted 18 May 2014; Published 8 July 2014

Academic Editor: Teen-Hang Meen

In the past, the kernel of automatic speech recognition (ASR) is dynamic time warping (DTW), which is feature-based template matching and belongs to the category technique of dynamic programming (DP). Although DTW is an early developed ASR technique, DTW has been popular in lots of applications. DTW is playing an important role for the known Kinect-based gesture recognition application now. This paper proposed an intelligent speech recognition system using an improved DTW approach for multimedia and home automation services. The improved DTW presented in this work, called HMM-like DTW, is essentially a hidden Markov model- (HMM-) like method where the concept of the typical HMM statistical model is brought into the design of DTW. The developed HMM-like DTW method, transforming feature-based DTW recognition into model-based DTW recognition, will be able to behave as the HMM recognition technique and therefore proposed HMM-like DTW with the HMM-like recognition model will have the capability to further perform model adaptation (also known as speaker adaptation). A series of experimental results in home automation-based multimedia access service environments demonstrated the superiority and effectiveness of the developed smart speech recognition system by HMM-like DTW.

1. Introduction

Multimedia and home automation services have been popular and necessary techniques in humans' home life. Among multimedia access and home automation applications, automatic speech recognition (ASR) is an important mainstream technique and plays a kernel role for improving the interaction between home members and home devices [1]. The development of speech recognition methods with satisfactory recognition performances in multimedia and home automation applications has been a challengeable issue. This paper will propose an improved dynamic time warping (DTW) speech recognition method, called HMM-like DTW, which brings the statistical model idea of the typical hidden Markov model (HMM) into the design of conventional DTW. The presented HMM-like DTW method demonstrated its superiority in recognition accuracy in the home media access and automation application.

From the viewpoint of application scenarios, ASR techniques can be categorized into two classes, speech understanding and voice command operations. This paper focuses on the aspect of the voice command operation of ASR. Human-machine interactions and media device operations by voice commands are extremely proper in a home environment [2]. For example, voice-command-based recognition operation can increase the convenience of humans' home life in home device control and home media access. Speech recognition using voice commands not only will save a lot of time and manpower but also is helpful for automatic recognition operations without any human operators. However, speech recognition is encountering a lot of challenges due to too many unexpected variables and adverse factors, such as the variety of accents and speech habits on testing speakers [3]. The testing speaker utters the same words for the operated voice command, but these uttered commands will not have exactly the same result so that speech recognition with the correct recognition outcome in each recognition test will be hard to achieve. To overcome this problem, related works on speech recognition enhancements have been quite common in the recent years, and most of those studies aimed at increasing the reliability of the recognition result by improving the recognition system [4] or reducing the mismatch phenomenon between a new speaker and the speech recognition system by performing machine learning schemes [5] or adaptive designs [6] on original speech recognition system.

The current mainstream speech recognition methodologies are hidden Markov models (HMM) [7], artificial neural network (ANN) [8], and DTW [9, 10]. HMM and ANN are categorized into the class of model-based recognition methods, and DTW belongs to the feature-based recognition category technique. Compared with model-based speech recognition, feature-based speech recognition does not involve adopting a statistical model. Training a classification model in advance is not required for feature-based speech recognition and therefore this method is generally considered a conceptually simple and direct recognition technique. DTW, belonging to the dynamic programming (DP) methodology [9], is a type of feature-based speech recognition. Although lots of ASR-related studies focus on HMM and ANN techniques, DTW still has its technical position due to the low complexity recognition calculations and high recognition accuracy, which will be the necessary factor in multimedia and home automation applications [10]. Nowadays, the popular DTW speech recognition has been seen to be largely utilized in the sensing-based applications [11], such as the Microsoft Kinect sensing device.

For model-based HMM or feature-based DTW speech recognition, the most important technical issue is how to effectively increase the recognition rate. In fact, improving the recognition performance of a speech recognition system has been a challenging problem. In HMM speech recognition, speaker adaptation (SA), sometimes also known as HMM model adjustments, has been widely used for overcoming the problem [6]. Speaker adaptation in HMM speech recognition will continually tune the statistical model parameters of HMM such as mean and covariance parameters using the information of the speaker's uttered data, and therefore the recognition system will not be strange to the speaker again after a series of model parameter adjustments [6]. For the feature-based DTW speech recognition technique, however, such speaker adaptation methodology cannot be employed due to the lack of a statistical model. Although related investigations on improving DTW speech recognition have been conducted in recent years [12, 13], most of these DTW-related studies have either developed improved template-matching algorithms [12] or provided modified schemes for a DTW operation optimization framework [13] for increasing the robustness of the recognition system. Speaker adaptation studies on DTW speech recognition are extremely rare.

In the author's previous work [5], speaker learning for DTW speech recognition has been explored where the learning strategy is interpolated into traditional DTW. Under the scheme, the DTW system is additionally equipped with the developed machine learning approaches for modifying the database containing referenced templates of speech patterns [5]. However, the fundamental structure of DTW in [5] is almost still the same as that of conventional DTW except the additionally given machine learning scheme for the database of DTW referenced templates, both of which still belong to feature-based recognition techniques. The DTW system learning performance by the developed work in [5] will still be largely restricted due to the essence of invariable feature-based template matching and the lack of a statistical recognition model when performing recognition. In order to solve the problem, this paper presents an HMM-like DTW approach, which is to thoroughly change the fundamental structure of DTW operations by establishing an HMM-like recognition model. By transforming feature-based into model-based recognition methodologies, the developed HMM-like DTW in this work will behave as the modeling technique of HMM speech recognition and therefore will have all benefits of HMM model-based speech recognition category techniques including the above-mentioned speaker adaptation techniques used in model-based speech recognition. Different to the improved DTW approach in [5], the proposed HMM-like DTW in this work is essentially a modeling recognition technique, and the developed HMM-like recognition model for DTW will provide a crucial framework for the development of possible speaker adaptation techniques on DTW speech recognition. The popular HMM speaker adaptation techniques [14, 15] with proper modifications will be able to be extended to the proposed HMM-like DTW herein, which can effectively solve the problem of learning restriction of developed DTW machine learning in [5]. In summary, the proposed HMM-like DTW approach in this study has several advantages compared with those without

(i) better performances in recognition accuracy and more flexibility in recognition system alignments,

(ii) a statistical HMM-like classification model with the ability of model adjustments for recognition performance improvements as compared with those enhanced DTW methods that only aim at dynamical programming design of template matching of acoustic features (e.g., [12, 13]),

(iii) more convenience and greater efficiency for further extensions of speaker adaptation, compared with those feature-based DTW system learning methods (e.g., the machine learning method for just the adaptive design of the DTW referenced template database [5]).

The remainder of this paper is organized as follows. Section 2 details the theoretical formulation of DTW speech recognition. Section 3 introduces the concept of hidden Markov model that is employed in the developed HMM-like DTW, followed by the formulation of HMM-like DTW speech recognition, containing model initialization of DTW referenced templates, recursive model training of DTW referenced templates, and recognition estimates of the established HMM-like DTW model in the testing phase. Section 4 presents the experiment results where the effectiveness and performance of presented HMM-like DTW are demonstrated, compared with conventional DTW. Finally, Section 5 provides concluding remarks.

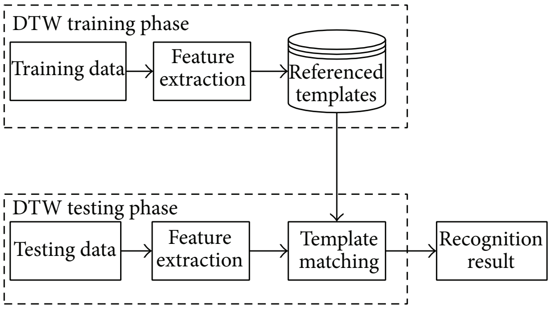


Figure 1: Frameworks of DTW-based speech recognition.

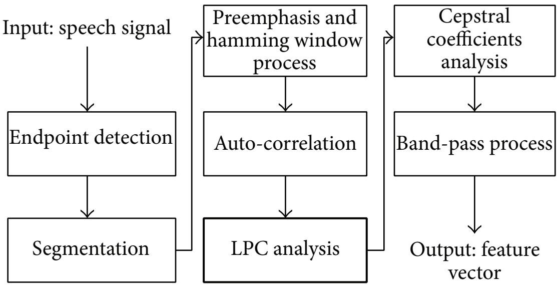


Figure 2: Feature extraction of speech signals.

2. Speech Recognition by DTW

The conventional DTW speech recognition procedure will be illustrated in this section. As mentioned before, DTW is belonging to dynamic programming category techniques. DTW speech recognition combines both time-warping and template-matching calculations for achieving the purpose of speech pattern recognition [9].

The framework of DTW speech recognition is depicted in Figure 1. As shown in Figure 1, DTW speech recognition contains two phases, the training phase and the testing phase. In the training phase of DTW, the main work is to establish the database of reference templates, which could be employed to complete the template matching work in the DTW testing phase. The primary mission of the DTW testing phase is to perform template matching between the testing template and the reference template. When computing the similarity degree between the testing template and the reference template, the low distortion between the two of them suggests a high similarity degree. As could be seen in Figure 1, feature extraction is an important and crucial procedure for such DTW feature-based recognition method. DTW template matching attempts to find an optimal comparison path between the testing template feature vector and the referenced template feature vector. Figure 2 shows the feature extraction procedure indicated in Figure 1. At the end of feature extraction, the input speech signal will be transformed into the parameter of speech features, LPC parameters of the time domain, or linear predicted cepstral coefficient (LPCC) parameters of the frequency domain. This paper adopts the LPCC parameter to be the feature of speech signals in the DTW template matching work.

The DTW template matching operation between the testing LPC feature and the referenced LPC feature is described herein. The testing utterance is composed of T frames and an arbitrary frame (a feature vector), denoted by t. The reference template consists of R frames and the arbitrary frame, indicated as r. The distortion between the T and R frames can be represented as $d[T(t), R(r)]$. The starting point and the end point of the comparison path are $(T(1), R(1)) = (1, 1)$ and $(T(M), R(M)) = (T, R)$, respectively. Based on these DTW operational settings, the DTW distance, d, from the optimal comparison path can be derived using (1). The arbitrary frame t in the testing data is generally not equal to the arbitrary frame r in the indices reference template [9]. Consider

$$D = \min \sum_{m=1}^{M} d\left(T\left(m\right), R\left(m\right)\right). \tag{1}$$

Assuming that the point $(T(0), R(0)) = (0, 0)$ and $d(0, 0) = 0$, the accumulated distance that selects the optimal source path can be represented as

$$\min D\left(t, r\right) = \min_{(t-1, r-1)} \left\{\min D\left(t-1, r-1\right) + d\left(t, r\right)\right\}, \tag{2}$$

where $\min D(t, r)$ is the shortest distance from the starting position to position (t, r). In the testing recognition of DTW, the recognition outcome is the label of the referenced template with the smallest value of $\min D(t, r)$.

Note that, in the previous work on DTW enhancements [5], machine learning schemes to drive the DTW recognition system to be adaptive with a new speaker are to provide proper management on the database of referenced templates (see Figure 1). However, such scheme in [5] will still encounter inefficiency and ineffectiveness on system adaptation due to the lack of a statistical model. A modeling technique for DTW, HMM-like DTW, will be presented in the following section.

3. The Proposed HMM-Like DTW Approach for Speech Recognition

This section describes the proposed improved DTW, HMM-like DTW, for speech recognition. At the beginning of this section, the basic methodology of HMM will be primarily introduced.

3.1. Hidden Markov Model (HMM). HMM is a statistical probability model, which is composed of a series of state transitions. HMM is essentially a hidden Markov chain that could be used to simulate and then model acoustic signals. All frames in the state will have the same characteristics in a Markov chain. In the methodology of HMM, the probability model is employed to describe the pronunciation characteristics of a segment of uttered speech signals. In this uttering process of a speaker, the segment of acoustic signal will be viewed as a continuous state transition in a Markov model. HMM state transition will be the primary work in an

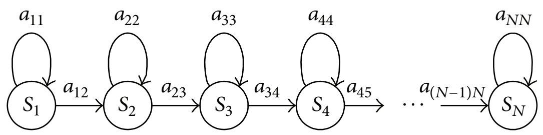


Figure 3: Left-to-right HMM state transition schemes in speech recognition applications.

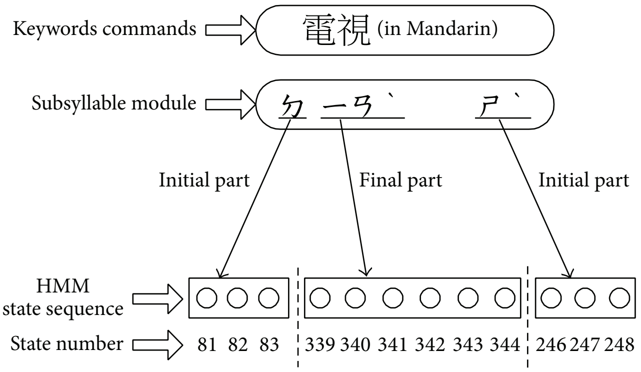


Figure 4: HMM state sequence of the keyword voice command "電視" in Mandarin.

HMM-based speech recognition system. Figure 3 illustrates the frequently used left-to-right state transition in HMM speech recognition. As shown in Figure 3, there are N states in total in the HMM model; the term a_{ij} denotes the state transition between the state i and the state j. Only two ways of state transitions could be done in the HMM model of Figure 3, staying at the same state or going to the next state.

HMM-based speech recognition is usually used in the keywords-spotting voice command operation applications. As shown in Figure 4, the keywords voice command "電視" (pronounced in Mandarin) is modeled as the HMM state sequence composed of 12 states, two 3-state initial parts and one 6-state final part. In Mandarin speech recognition using HMM, the subsyllable method is used to establish the HMM model of each keyword voice command. In general, there are 3 states in the initial part and 6 states in the final part. In this work, the proposed HMM-like DTW approach will establish the acoustic model for each of the DTW referenced template database using HMM-like left-to-right state sequences of the keyword voice command, which will be described in detail in the following section.

3.2. HMM-Like DTW. The basic idea of statistical HMM models introduced in the previous section will be incorporated into the design of the HMM-like speech recognition system. Figure 5 depicts the framework of the proposed HMM-like DTW speech recognition, which is different to conventional feature-based DTW and is belonging to a model-based technique. As could be seen in conventional DTW of Figure 1 and in the developed HMM-like DTW of Figure 5, the primary work of HMM-like DTW is to model the DTW system by establishing the HMM-like acoustic model for each of DTW referenced templates of keywords voice commands. HMM-like DTW contains mainly two design phases, the training phase to model DTW referenced templates and the testing phase to use the established acoustic models of TW referenced templates for performing the recognition of the test utterances. The training phase design of HMM-like DTW will be provided in Sections 3.2.1 and 3.2.2, which primarily describe model initialization and recursive model training of DTW keywords referenced templates, respectively. Section 3.2.3 describes how HMM-like DTW with established acoustic models of DTW keywords templates in Sections 3.2.1 and 3.2.2 is used for recognition calculations in the testing phase. As could be seen in Figure 1 and Figure 5, proposed HMM-like DTW changes template matching of conventional DTW as model recognition estimating.

3.2.1. Model Initialization of DTW Referenced Templates. The proposed HMM-like DTW will perform the model initialization first in the beginning of the model training phase. Model initialization of DTW referenced templates is to establish the initial model for certain keyword voice command template. The initial model will be represented as the HMM-like state sequence. Figure 6 shows the averaged segmentation procedure for model initialization of certain DTW keywords command template. In the model initialization of the DTW referenced template, averaged segmentation is an important task for establishing the initial state sequence. Averaged segmentation divides each of the training data into a series of acoustic segments with the same numbers of acoustic frames. As shown in Figure 6, N states are set for certain keywords voice command "打開電視," pronounced in Mandarin, where the DTW referenced template "打開電視" is modeled as the state sequence with N states. Each of N states denotes the characteristics of a series of acoustic frames within certain segment of continuous time and therefore is represented as the corresponding averaged frame segmentation information of the training data. For example, the state S_1 in Figure 6 reveals the statistical information of frames of N training data, $Training\text{-}data_1, Training\text{-}data_2, \ldots,$ and $Training\text{-}data_n$, at the first time interval. The state S_1 is derived using (3) as follows:

$$S_1 = \frac{(f_{11} + f_{12}) + (f_{21} + f_{22} + f_{23}) + \cdots + (f_{n1} + f_{n2})}{2 + 3 + \cdots + 2}. \tag{3}$$

The initial model, the state sequence with N states, of certain DTW keywords command referenced template will be further reestimated for achieving the optimal recognition performance using a recursive model training procedure, which will be presented in the following section.

3.2.2. Recursive Model Training of DTW Referenced Templates. Model initialization of DTW referenced templates is to establish the initial model for each of DTW keyword voice command template. These initial models are further tuned for achieving the optimal performance on recognition accuracy. The developed recursive model training procedure in HMM-like DTW is depicted in Figure 7. As could be seen in

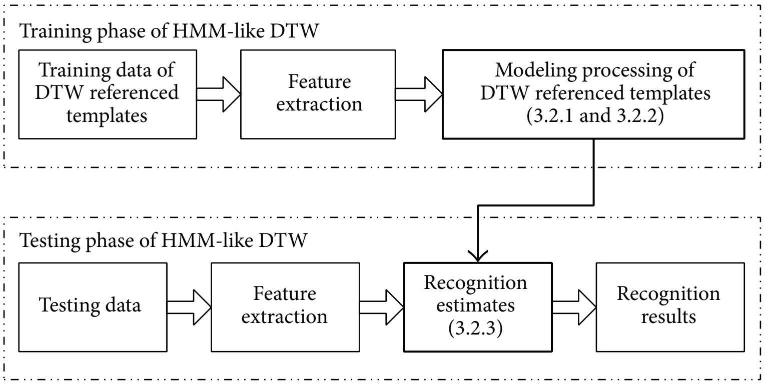

Training phase of HMM-like DTW
Training data of DTW referenced templates
Feature extraction
Modeling processing of DTW referenced templates (3.2.1 and 3.2.2)
Testing phase of HMM-like DTW
Testing data
Feature extraction
Recognition estimates (3.2.3)
Recognition results

Figure 5: Frameworks of model-based HMM-like DTW speech recognition systems.

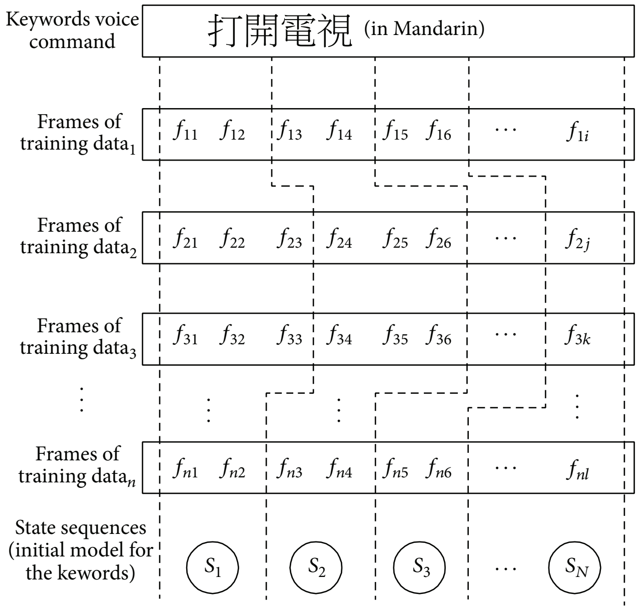

Keywords voice command
打開電視 (in Mandarin)
Frames of training data$_1$
f_{11} f_{12} f_{13} f_{14} f_{15} f_{16} ⋯ f_{1i}
Frames of training data$_2$
f_{21} f_{22} f_{23} f_{24} f_{25} f_{26} ⋯ f_{2j}
Frames of training data$_3$
f_{31} f_{32} f_{33} f_{34} f_{35} f_{36} ⋯ f_{3k}
Frames of training data$_n$
f_{n1} f_{n2} f_{n3} f_{n4} f_{n5} f_{n6} ⋯ f_{nl}
State sequences (initial model for the kewords)
S_1 S_2 S_3 ⋯ S_N

Figure 6: Averaged segmentation for model initialization of certain DTW keywords command template.

Figure 7, the Viterbi algorithm is employed to carry out resegmentation of acoustic frames of training data. After doing the Viterbi algorithm, the new model is estimated in the iteration. It is noted that in this training procedure a checking process of the index ξ is performed to verify the performance of the trained state sequence model. The index ξ is the Euclidean distance and is determined using (4) as follows:

$$\xi = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{k} \left(X_{ij} - \widetilde{X}_{ij}\right)^2}, \tag{4}$$

where ξ is the error value between the current and the last state sequence models; X_{ij} denotes the Gaussian mean value of the jth dimension of the ith state of the current new state sequence model trained in this iteration; $\widetilde{X}_{ij}$ is the Gaussian mean value of the jth dimension of the ith state of the past old state sequence model trained in the last iteration. Note that the ideal value of ξ is expected to approach zero in this recursive model training of DTW referenced templates. However, such ideal trained model is hard to be established in the real training situation. The threshold T is set to decide if the value of the calculated ξ is acceptable for model parameter convergence in the recursive model training. When the value of ξ is limited to be smaller than the value of the preset threshold T, the overall recursive training procedure is finished and then the estimated state sequence model of DTW reference templates will have the highest performances in recognition accuracy in the test phase.

3.2.3. Recognition Estimates of HMM-Like DTW in the Testing Phase. As mentioned in the previous section, when finishing recursive model training of DTW referenced templates, trained state sequence models for the corresponding DTW keywords templates could be used for online speech recognition in the testing phase. An HMM-like DTW speech recognition system with M keywords command templates in the conventional DTW referenced template dataset will have M trained state sequence models for each of the DTW referenced templates. When performing the recognition estimate of HMM-like DTW in the testing phase, the likelihood degree between each of those M trained state sequence models and the input test utterance of a new test speaker will be calculated. The label of the trained state sequence model with the highest value of the likelihood degree will be the recognition outcome. In this work, a Viterbi-like approach is developed for performing the likelihood degree estimates.

Figure 8 depicts the operation of the presented Viterbi-like approach in the HMM-like recognition test phase. Viterbi-like approach belongs to the category of dynamical programming in essence, and therefore a global optimization result will be calculated when completing the overall path (P) programming. In this work, the score function $\delta_t(i)$ is defined as in (5), given the observed set of T speech frames, $O = \{o_1, o_2, \ldots, o_T\}$,

$$\delta_t(i) = \max_{s_1, s_2, \ldots, s_N} P\left(s_1, s_2, \ldots, s_N = S_i, o_1, o_2, \ldots, o_T \mid \lambda\right), \tag{5}$$

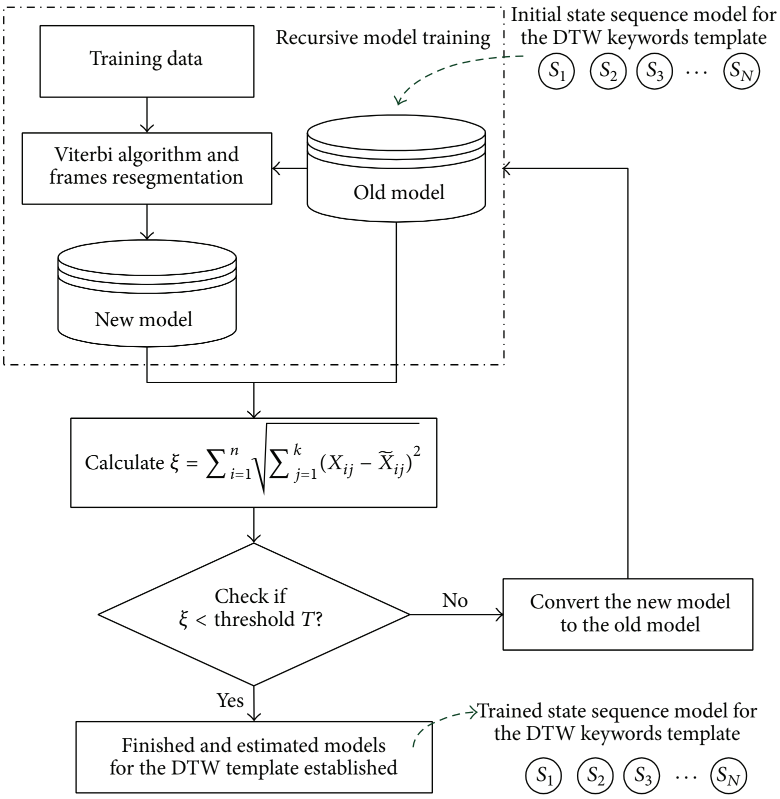

Initial state sequence model for the DTW keywords template
S_1 S_2 S_3 ⋯ S_N
Training data
Recursive model training
Viterbi algorithm and frames resegmentation
Old model
New model
Calculate $\xi = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{k}(X_{ij}-\widetilde{X}_{ij})^2}$
Check if $\xi <$ threshold T?
No
Convert the new model to the old model
Yes
Finished and estimated models for the DTW template established
Trained state sequence model for the DTW keywords template
S_1 S_2 S_3 ⋯ S_N

Figure 7: The developed recursive model training procedure in proposed HMM-like DTW.

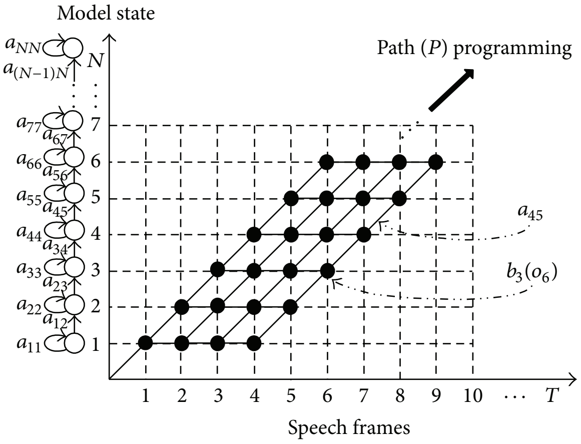


Figure 8: Recognition calculations of HMM-like DTW by the Viterbi-like method in the test phase.

where $\delta_t(i)$ has the largest probability at time t and at state S_i; λ is the trained model for each of DTW keywords referenced templates as mentioned in the previous section. $\delta_{t+1}(i)$ is computed as follows using $\delta_t(i)$ by induction:

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i)\, a_{ij}\right] b_j(o_{t+1}), \qquad (6)$$

where a_{ij} is the state transition probability of going from state S_i to state S_j; $b_j(o_{t+1})$ denotes the Gaussian distribution probability of the observed frame o_{t+1} given the state S_j. The Viterbi-like approach in this study is to solve the iterative procedure of (5) and (6) and the state sequence that has the maximum likelihood will be searched if one keeps tracking of all the states which maximize (5).

4. Experiments and Results

The proposed HMM-like DTW speech recognition is performed in the application of multimedia and home automation services. The HMM-like DTW speech recognition system adopts the voice command operation mechanism where a set of DTW keywords referenced template models is established in advance. Table 1 shows the voice command set containing 8 keywords that denote command operations of noticing the strong wing (the index a), opening the light (the index b), showing the temperature (the index c), turning off the air conditioner (the index d), adjusting the temperature (the index e), turning on the TV set (the index f), turning up the volume (the index g), and selecting the TV channel (the index h).

In the HMM-like DTW speech recognition experiments, the sampling rate of speech signals is 44.1 KHz; the resolution of the speech sample is set as 16 bits; the number of channels is one (i.e., mono settings); for each acoustic frame, the frame size is set as 20 ms with a 10 ms frame overlap; the LPCC feature is adopted on feature extraction, and each feature parameter of the acoustic frame is composed of the 10-dimension linear prediction cepstrum parameters. The

Table 1: The voice command set of keywords in the HMM-like DTW speech recognition system.

Index of keywords	Keywords (in Mandarin)
a	强風
b	請開燈
c	二十度
d	關掉空調
e	調整温度
f	打開電視
g	放大音量
h	選擇電視頻道

Table 2: Numbers of states of HMM-like DTW for each DTW keywords template model and the corresponding recognition performance in the training phase.

Numbers of states (N)	Recognition rates (%)
50	**70.6%**
40	67.5%
30	49.4%
20	23.1%
10	25.0%

HMM-like DTW speech recognition experiment is divided into two phases, the training phase that establishes the state sequence model for each of DTW keywords referenced templates and the testing phase to evaluate the recognition performance of proposed HMM-like DTW.

In the training phase, a training dataset for establishing HMM-like DTW models is made. Ten males and 10 females are requested for uttering. Each of the 10 males and 10 females is asked to make 5 utterances for each of the 8 keywords, and therefore there are 800 utterances in total for training these 8 models of keywords, 100 utterances for each of the 8 keywords models. Table 2 shows numbers of states (N) of HMM-like DTW set for each DTW keywords template model and the corresponding recognition performance. Observed from Table 2, when the number of states is set improperly, the recognition rate of HMM-like DTW will be very dissatisfactory. Among all state settings, HMM-like DTW with the state setting $N = 50$ performs best on the recognition accuracy, which achieves 70.6%. HMM-like DTW with $N = 50$ will be chosen to be compared with conventional DTW in the testing phase.

In the testing phase, the collected 10 males and 10 females are requested again to make the additional utterances for the testing experiments. There are 160 utterances in total for the test experiment, 20 utterances for each of the 8 keywords models. Note that these 160 utterances are completely different from those 800 utterances in the training phase. Table 3 shows the recognition performance comparisons of proposed HMM-like DTW with $N = 50$ and conventional DTW. As could be seen in Table 3, the proposed HMM-like DTW with the developed HMM-like modeling scheme is apparently more competitive than conventional DTW with only simple template matching. HMM-like DTW has a better recognition performance than conventional DTW, which is about 6.8%.

Table 3: Performance comparisons of proposed HMM-like DTW with $N = 50$ and conventional DTW on the recognition accuracy.

Keywords commands	Recognition rates	
	HMM-like DTW ($N = 50$)	Conventional DTW
Index *a*	70%	70%
Index *b*	55%	85%
Index *c*	65%	70%
Index *d*	70%	80%
Index *e*	75%	50%
Index *f*	70%	60%
Index *g*	60%	40%
Index *h*	100%	55%
Average	70.6%	63.8%

5. Conclusions

In this paper, the HMM-like DTW method is proposed for speech recognition. Proposed HMM-like DTW provides a statistical model recognition strategy for traditional feature-based DTW template matching using the kernel concept of hidden Markov model. The proposed HMM-like DTW will be able to further carry out model adaptation as HMM. Speech recognition experiments in the application of home automation-based multimedia access services showed that the presented HMM-like DTW with the appropriately designed acoustic model is obviously more competitive than conventional DTW without any statistical models on the recognition accuracy.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] L. Ceccaroni and X. Verdaguer, "Agent-oriented, multimedia, interactive services in home automation," in *Proceedings of the 2nd European Workshop on Multi-Agent Systems*, 2004.

[2] J. Zhu, X. Gao, Y. Yang, H. Li, Z. Ai, and X. Cui, "Developing a voice control system for ZigBee-based home automation networks," in *Proceedings of the 2nd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC '10)*, pp. 737–741, September 2010.

[3] V. Young and A. Mihailidis, "Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: a literature review," *Assistive Technology*, vol. 22, no. 2, pp. 99–112, 2010.

[4] I.-J. Ding, "Speech recognition using variable-length frame overlaps by intelligent fuzzy control," *Journal of Intelligent and Fuzzy Systems*, vol. 25, no. 1, pp. 49–56, 2013.

[5] I. J. Ding, C. T. Yen, and Y. M. Hsu, “Developments of machine learning schemes for dynamic time-wrapping-based speech recognition,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 542680, 10 pages, 2013.

[6] K. Shinoda, “Acoustic model adaptation for speech recognition,” *IEICE Transactions on Information and Systems*, vol. 93, no. 9, pp. 2348–2362, 2010.

[7] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[8] S.-H. Chen and Y.-R. Wang, “Tone recognition of continuous Mandarin speech based on neural networks,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 2, pp. 146–150, 1995.

[9] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[10] P. G. N. Priyadarshani, N. G. J. Dias, and A. Punchihewa, “Dynamic time warping based speech recognition for isolated Sinhala words,” in *Proceedings of the 55th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS ’12)*, pp. 892–895, August 2012.

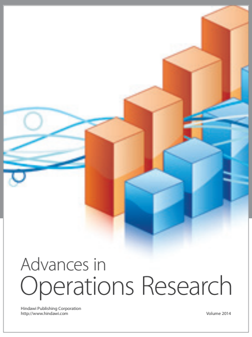
[11] J. Wu, J. Konrad, and P. Ishwar, “Dynamic time warping for gesture-based user identification and authentication with Kinect,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2371–2375, 2013.

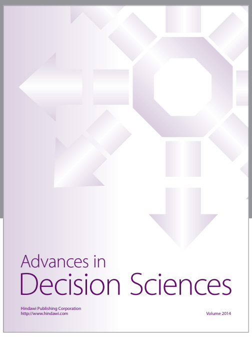
[12] X. Anguera, R. Macrae, and N. Oliver, “Partial sequence matching using an unbounded dynamic timewarping algorithm,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’10)*, pp. 3582–3585, March 2010.

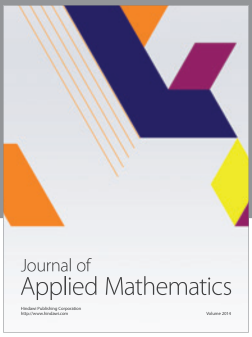
[13] X. Chen, J. Huang, Y. Wang, and C. Tao, “Incremental feedback learning methods for voice recognition based on DTW,” in *Proceedings of the International Conference on Modelling, Identification and Control (ICMIC ’12)*, pp. 1011–1016, June 2012.

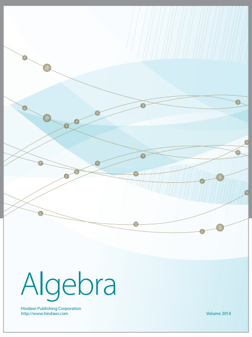
[14] I.-J. Ding, “Reinforcement of MLLR speaker adaptation using optimal linear interpolation,” *Electronics Letters*, vol. 48, no. 5, pp. 290–292, 2012.

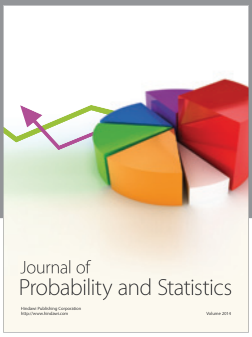
[15] B. Das, S. Mandal, P. Mitra, and A. Basu, “Aging speech recognition with speaker adaptation techniques: study on medium vocabulary continuous Bengali speech,” *Pattern Recognition Letters*, vol. 34, no. 3, pp. 335–343, 2013.


Advances in
Operations Research

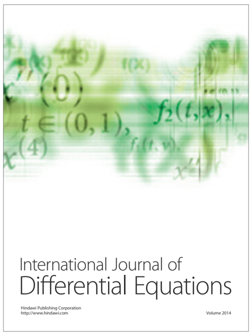

Advances in
Decision Sciences


Journal of
Applied Mathematics

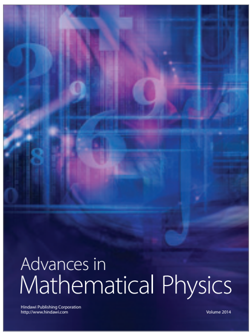

Algebra

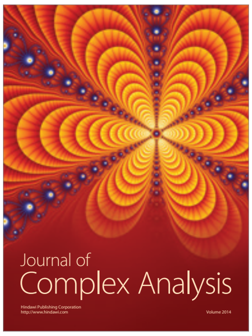

Journal of
Probability and Statistics

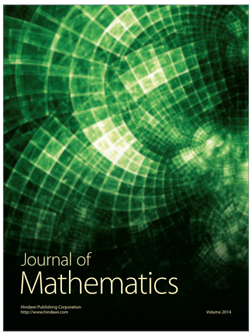
The Scientific
World Journal

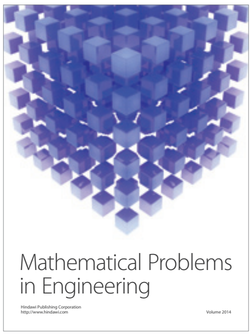

International Journal of
Differential Equations

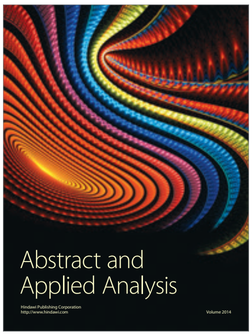
International Journal of
Combinatorics

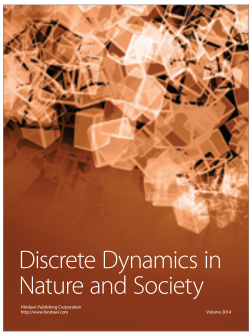

Advances in
Mathematical Physics

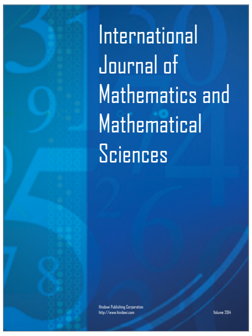

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

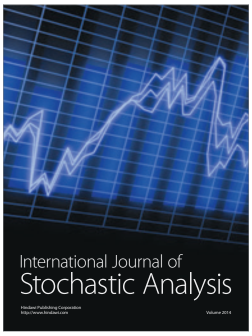

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization