Evaluation of Human Voice Biometrics and Frog Bioacoustics Identification Systems Based on Feature Extraction Method and Classifiers

Asst.Lecturer.Aws Saad Shawkat

The Great Emam University College

Abstract

Biometrics is defined as the science of recognizing human by using their personal biological characteristics, for example voice, fingerprint and signature. Biometrics approach has then been implemented for recognizing animal for the purpose of biological and ecological research and development. Due to the research on animal based recognition is still in infancy, so in this study, the evaluation on the effectiveness of the audio based biometric system approach to the bioacoustics identification system is experimented. Bioacoustics based on frog call in order to identify the frog species is employed in this study. Consequently, the well-known features used in audio based biometric system i.e. Mel-frequency Cepstral Coefficients (MFCC) is experimented as features for the frog bioacoustics based identification system. For the classification process, performances of Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Local Mean k Nearest Neighbor (LMkNN) and Fuzzy k-NN (FkNN) classifiers have been compared in this study. The performances of the biometric system and the frog bioacoustics system based on the proposed classifiers are evaluated. The best performance has been observed using FkNN classifier with the accuracy of 97% for the frog bioacoustics identification system and 93.38% for the biometric speaker identification system with 20 training data.

Keywords – Biometrics; Bioacoustics; MFCC; SVM; kNN; LMkNN; FkNN; speaker identification

التقييم على انظمة تحديد الصوتيات البشرية والصوتيات الحيوية للضفدع اعتماداً على طريقة استخراج وتصنيف الخصائص

م.م. أوس سعد شوكت حسن
كلية الإمام الأعظم (رحمه الله) الجامعة.

المستخلص
يتم تعريف القياسات الحيوية كعلم تمييز الإنسان باستخدام خصائصه البيولوجية الشخصية على سبيل المثال الصوت وبصمات الأصابع والتوقيع. ثم تم تطبيق نهج القياسات الحيوية لتمييز الحيوان لغرض البحوث البيولوجية والبيئية والتنمية. ويرجع ذلك إلى كون بحوث التمييز على أساس الحيوان لا يزال في مرحلة الطفولة، لذلك في هذه الدراسة، يتم عمل تقييم فعالية النهج القائم على نظام القياسات الحيوية الصوتية لنظام تحديد الصوتيات الحيوية. يستخدم علم الصوتيات الحيوية على أساس دعوة الضفدع من أجل التعرف على الضفادع الاخرى في هذه الدراسة. ونتيجة لذلك، يتم اختبار الميزات المعروفة المستخدمة في نظام القياسات الحيوية الصوتية مثل استخدام Mel-frequency Cepstral Coefficients (MFCC) كميزات لنظام التعرف على الصوتيات الحيوية للضفدع. أما بالنسبة لعملية التصنيف، فقد تمت مقارنة أداء Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Local Mean k Nearest Neighbor (LMkNN) and Fuzzy k-NN (FkNN)وقد تم في هذه الدراسة تقييم أداء نظام القياسات الحيوية للضفدع على أساس المصنفات المقترحة. وقد لوحظ أفضل أداء باستخدام FkNN classifier مع دقة ٩٧٪ لنظام التعرف على الصوتيات البيولوجية للضفدع و ٩٣,٣٨٪ لنظام تمييز القياسات الحيوية على المتحدث مع ٢٠ بيانات التدريب.

الكلمات الأفتتاحية: القياسات الحيوية؛ الصوتيات الحيوية؛ SVM؛MFCC ؛ kNN؛ LMkNN؛ FkNN ؛ تعريف المتحدث.

Introduction

Security and protection are the most important sciences [1]; they become very significant for the human daily life. The most common way which has been used in protecting data is by using a password or PIN codes (Personal Identification Number), [2]. This approach is the simplest level of protection for the personal information and data which can now be considered as outdated. During the progress in biometric science [3], the way of protecting information by using biometric becomes more secure. Here, parts of it, such as fingerprint, palm print, iris and voice have been used to get a better method [4-6]. This work focuses on speaker recognition which uses human voice to recognize an individual. It is one of the most secure technologies and has been an interesting research field in recent years [7]. This technology has been applied widely to the real world daily life such as telephone, communication and voice mail. Biometric word comes originally from the Greek language, which is divided into two parts, Bio and Metric or "Life measurement" [8]. It is the science of determining the identity of a person based on physical or behavioral characteristics [9], [10]. Nowadays, this science starts to become famous, reliable and trusted for people identification compared to the use of passwords, PIN codes or ID cards. This is due to the fact that PIN codes and ID cards are easy to be duplicated, forgotten and robbed. Furthermore, ID and PIN codes give poor evidence for person identification especially in the crime scenes. Another type of science working on animal's voice recognition and their calls , is called bioacoustics [11]. Many animals generate sounds either for communication or as a by-product of their living activities such as eating, moving, or flying. Automatic recognition of bioacoustics sounds is valuable for biological research and environmental monitoring applications,

particularly for detecting and locating animals [12]. Animal sound productions can be normally divided into two categories. The first in cludes as incidental sounds that result as the by-product of their activities and the latter refeos to the non-incidental sounds, which are used for communication purpose. Quite naturally, the animal species could be identified according to their sound productions. Nevertheless, manual classification of bioacoustics signals can be very ambiguous and most often rely heavily on the surveyor's expert knowledge of the group under investigation [13]. Automatic animal voice identification is done by many people who had worked on it by installing acoustical sensors coupled to a computer-assisted processing system. The recorded sound signals of the major grain insect species were digitized and stored into a reference database. A classification algorithm was developed for the automatic recognition of recorded insect noise signals by their comparison to the specific spectra of the reference database [14]. Speaker recognition system can mainly be categorized as speaker verification system and speaker identification system. Generally, it works in several steps that can be connected together [15], and figure (1) shows that in steps:

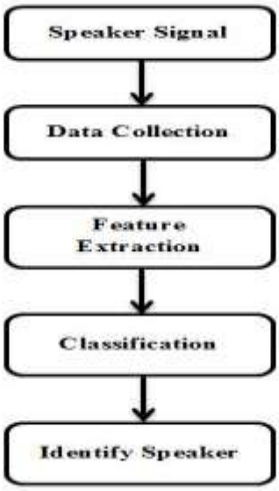


Figure 1. Speaker recognition system.

Architecture of Frog Identification System

The identification system basically consists of three parts namely, data collection, feature extraction and classification. Figure 2 shows the flowchart of the system.

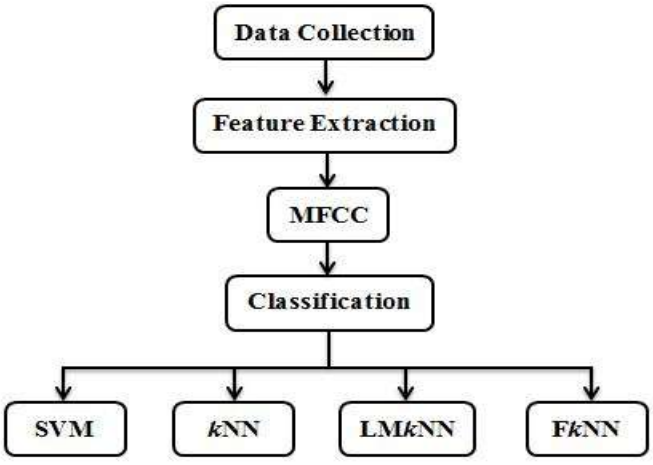


Figure 2. Identification system.

Data Acquisition

The database for the biometric human speaker identification system and bioacoustics frog call identification system are collected from two sources i.e. internet database [1, 2]. both databases are in wave form. The human database is recorded in frequency 32 kHz and 16 bits resolution, while the high resolution is not required in speech recognition. The frog database is recorded in 48 kHz and 32 bits. The frog database uses 9 types of species as shown in the table below, while the human speaker identification system used 37 speakers.

TABLE I. FROG SPECIES DATABASE

Family	**Scientific name**	**Common name**
Microhylidae	*Microhyla butleri*	Painted chorus frog
Ranidae	*Babina adenopleura*	Olive frog
	Hylarana taipehensis	Taipei brown frog
	Lithobates catesbeianus	American bullfrog
	Rana sauteri	Sauteri's brown frog
Rhacophoridae	*Polypedates braueri*	White throated tree frog
	Kurixalus idiootocus	Surface-day tree frog
Bufonidae	*Bufo bankorensis*	Taiwan common frog
Hylidae	Litoria caerulea	Green tree frog
	Litoria splendida	Magnificent tree frog

Pre-processing

Each of the syllables has undergone a series of speech processing step that is pre-emphasis, framing and windowing [3-6]. The purpose of pre-emphasis is to compress the high frequencies in a dynamic range by flattening the spectral speed; this is done to increase the signal to noise ratio (SNR). While using the first order of finite impulse response (FIR) filter for filtering the speech signal, the pre-emphasis time domain is shown in Eq. (1);

$$x'(n) = x(n) - ax(n-1) \quad (1)$$

where a is the pre-emphasis parameter and is considered as 0.9357, $x(n)$ is the input frog call and $x'(n)$ is the output of the filter. $x'(n)$ becomes a string of windowed sequence. $x_t(n)$, $t = 1, 2\ldots T$, and all of these are called frames, these frames are processed individually $x_t(n)$ and is rewritten in equation;

$$x_t(n) \equiv w(n).x_t'(n) \quad (2)$$

The purpose of using a windowing function is to minimize the discontinuities in the signal especially at the beginning and the end of each frame by adding zero outside the signal. In this study Hamming window was chosen to be used in this part, $w_H(n)$ which can be defined by the showing equation:

$$w_H(n) = 0.54 - 0.46\left(\frac{2n\pi}{N-1}\right),$$

$$\mathrm{n} = 0, \ldots, N-1 \quad (3)$$

Hamming window was chosen because the side lobes of this window are much lower compared to the other types of windows. Hamming window reduces the resolution, and it can be considered as a good choice compared with a high windo

Feature Extraction

In this paper, MFCC is selected becanse it is robust reliable to noise which makes easy to be implemented in an outdoor environment that contains interference of background noise

such as the sound of the wind, running water and other animal calls. There are 15 mel cepstrum coefficients, one log energy coefficient and three delta coefficients per frame set in the experiments.

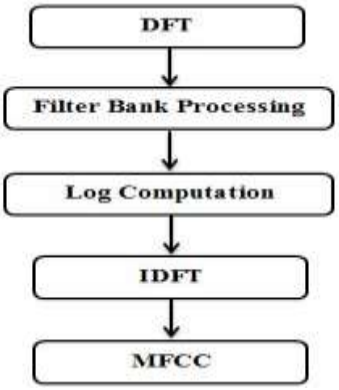


Figure 3. MFCC process.

Classification

SVM is a classifier that is based on the principle of structural risk minimization. The SVM is formulated in such a way that it is only capable of discriminating between two classes whereas most classification tasks typically involved more than two classes [7-9].Equation (4) shows how the SVM solves the problem of the linearly separable case.

$$D = \left\{\left(x^1, y^1\right),...,\left(x^L, y^L\right)\right\} \; x \in \Re^n \; y \in \{-1,1\} \qquad (4)$$

While y^1= 1 or -1 where they are indicating the class to point x_i and x^1 is a *p*-dimensional real vector. To divide the points into two groups, these are $y^1 = 1$ and $y^1 = -1$ by finding the maximum-margin hyperplane. Hyperplane can be written as a set of points x shown in this equation:

$$w.x - b = 0 \qquad (5)$$

Where *w* is the normal vector of the hyperplane, *b* is the original w for the selected hyperplane. The offset of the hyperplane can be presented by the parameter b/|| w|| from the

origin along the norm vector w. Two hyperplanes can be selected if the training data are linearly separable, and there are no points between them. After that, try to maximize the distance between them. The margin is the region bounded between the points. The hyperplanes can be explained by the next equations:

$w.x - b = 1$ and $w.x - b = -1$ (6)

To find the distance between two hyperplanes $2/\|w\|$; the value of $\|w\|$ has to be minimized. Data points must be prevented from the falling in the margin. Because of that there is a need to add *i* to each x as shown in the coming equation;

$y_i(w.x_i - b) \geq 1, For\ \ all\ 1 \leq i \leq n$ (7)

To get the optimization problem, put these entire ingredients together by minimizing (w, b), title the $(i = 1,\ldots, n)$. The optimization problem presented in the preceding section is difficult to solve because it depends on $\|w\|$, the norm of *w*, which involves a square root. Fortunately it is possible to alter the equation by substituting $\frac{1}{2}\|w\|^2$, (the factor of 1/2 being used for mathematical convenience) without changing the solution (the minimum of the original and the modified equation have the same *w* and *b*). To solve the previous constrained problem; Lagrange multipliers α is expressed in this equation;

$\min_{w,b} \max_{\alpha \geq 0} = \{\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[y_i(w.x_i - b) - 1]\}$ (8)

In this equation, look for a saddle point, and all the points can be separated as $y_i(w.x_i - b) - 1 > 0$and α_i and must be set to zero. To solve this problem we have to use quadratic programming techniques, and the solution can be shown by linear combination of the training vectors in this equation:

$w = \sum_{i=1}^{n} \alpha_i y_i x_i$ (9)

x_i is the support vectors, relied on the margin and expressed by this equation:

$$y_i(w.x_i - b) = 1 \qquad (10)$$

From equation (10) can be presented the support vector equation as;

$$w.x_i - b = {}^{1}/_{y_i} = y_i \iff$$

$$b = w.x_i - y_i \qquad (11)$$

This will allow defining the offset *b* which is more powerful to average over all the support vectors N_{SV} in this equation:

$$b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (w.x_i - y_i) \qquad (12)$$

While the fuzzy k-Nearest Neighbor process is working by setting a class membership to a vector rather than assigning the vector to a particular class. The basis of the algorithm work is by assigning the membership as a function of the vector's distance from its *k*-Nearest Nighbor and those neighbors' memberships in the possible classes.The vector membership values should provide a level of assurance to accompany the resultant classification. An example for this, if a vector is assigned to the value 0.55 membership in class number one, and the membership to class number two is 0.44, while the membership in class number three is 0.01, in this hesitant case to assign vector based on the previous numbers, but in a very confident it can show that does not belong to the third class. But later it might examine the vector to determine its classification because the vector exhibits a high degree of membership in both classes one and two. The fuzzy algorithm is similar to the crisp version in the sense that it must also search the labeled sample set for the k-NN.

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} \left(1/\|x - x_j\|\right)^{2/(m-1)}}{\sum_{j=1}^{k} \left(1/\|x - x_j\|\right)^{2/(m-1)}} \qquad (13)$$

The assigned memberships of x are influenced by the inverse of the distances from the nearest neighbors and their class memberships. The inverse distance serves to, weight a vector's membership more if it's closer and less if its farther from the vector under consideration. The labeled samples can be given complete memberships in several ways. The first way is by giving a complete membership in the known class and nonmembership in the rest of them. The second way is by assigning the samples membership based on distance from their class mean or based on the distance from labeled samples of their own class and those of the other classes then use the resulting memberships in the classifier. The two techniques have been used in this study and the results are presented. The variable m determines how heavily the distance is weighted when calculating each neighbor's contribution to the membership value by setting it to two.

k-NN considers the k nearest situations (il, i2,.. , ik) from an instance (x) and decides the most frequent class in the set that had been found (c1, c2, ... ck). Then the most frequent class is assumed to be the class of that instance (x) that had been found. In order to determine the nearest instance, k-NN technique adopts a distance metric that measures the proximity of instance (x) to k of stored instances. The distance matrix that can be used is the Euclidean Distance method. The Euclidean distance method had been used because of the similarity between the preferences of the data used in this study, which have the same influence on the distance measure between instances.

Let's define the Euclidean Distance method between two points p and q, and regarding the Cartesian coordinates, we assume that the first point p = (p1,p2,…,pn) and we assume the second point q = (q1,q2,…,qn) and these two points will be used in Euclidean N-space, and the distance from the first

point p to the second point q is shown by the coming first step of it:

$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$ (14)

n: The dimensionality of the vector input.

The prediction class of *k*-Nearest Neighbor is shown in equation (15).

$y(d_i) = arg\ \max_k \sum_{q_i \in NN} y(q_i, c_k)$ (15)

d_i: Test example

q_i: is one of the *k* nearest neighbors in the training set.

$y(q_i, c_k)$: Indicates whether q_i belongs to class c_k.

Regarding the *k* Nearest Neighbor and by changing the parameter k and make it equal to 1,3,5,7 and 9 to get a better result, and the reason of choosing k only with odd number to avoid ties votes. In this work the best k selected and applied to the four classifiers is *k* = 3

For the Local Mean *k* Nearest Neighbor (LM*k*NN) classifer, it is defined as shown in this hypothesis. $X^i = \{x_j^i | j = 1, \dots, N_i\}$ is a training sample set from class $w_i$where N_i the number of the training samples form is the class w_i and the pattern *x* is classified into class w_c as shown below:

Select the Nearest Neighbor training samples from x with a Euclidean Distance measure for each class xi. Here, a value of r must be selected by ranging from 1 to Ni.

Compute the local mean vector, y_i, using *r* Nearest Neighbor training samples, $\{x_{k_1}^i, x_{k_2}^i, \dots, x_{k_r}^i\}$:

$y^i = \frac{1}{r}\sum_{j=1}^{r} x_{k_j}^i$ (16)

Classify *x* into class w_c:

$(x - y^c)^T(x - y^c) =$

$\min_i (x - y^i)^T (x - y^i)$ (17)

Where *r* = 1 in equation (17) and the Euclidean Distance classifier when *r* = N_i. With this classifier (LM*k*NN) the

parameter r must be optimized and changed regarding each given data [10].

Calculate the distance d_j^i between the local mean vectors from equation (17) $y^i$and the test pattern *x* shown in the equation**:**

$$d_j^i = \left(x - y^i\right)^T (x - y^i) \qquad (18)$$

Computing the distance d_j^c between the class mean vector $\mu_j$and the test pattern *x* shown in equation (19).

$$d_j^c = \left(x - \mu_j\right)^T (x - \mu_j) \qquad (19)$$

Equation (20) will combine the two equations (19) and (18) as shown below:

$$d_j = d_j^l + w \times d_j^c, \ 0 \le w \le 1 \qquad (20)$$

Finally, classify the test pattern *x* into class w_c as shown in equation (21);

$$d_c = \min\{d_j\}, \ j = 1,2, \dots, M \qquad (21)$$

Experimental Results

The experiments are implemented by using Matlab R2011(b) and have been teste in Intel, 1.5 GHz 2 CPUs, 6Gb RAM and Window 7 operating system. In this experiment, the data of 49 syllables have been extracted. 20 syllables are used for training and 29 for testing. The classification accuracy is defined as;

$$SA = \frac{NA}{NT} \times 100\% \qquad (14)$$

While the SA is the system accuracy that it needs to be measured, NA is the number of the authentic data and NT is the total number of the testing data.

Performance Results

Table II shows the performance of MFCC with four classifiers in clean data with 20 testing data and 29 training for the both systems. The performances for the human speaker identification system compared with the frog call identification

system using the four classifiers and MFCC for extracting the features for both shows the differences in the results based on the number of training data that were used every time.

TABLE II. HUMAN SYSTEM USING THE FOUR CLASSIFIERS

Number of training data	SVM classifier	*k*NN classifier	LM*k*NN Classifier	F*k*NN Classifier
5	80.08%	85.55%	82.57%	87.51%
10	87.42%	90%	86.67%	91.15%
15	88.72%	92.45%	90.40%	93.10%
20	89.19%	92.45%	92%	93.38%

The performance of the four classifiers using the frog call identification system is presented by the next table;

TABLE III. FROG SYSTEM USING FOUR CLASSIFIERS

Number of training data	SVM classifier	*k*NN classifier	LM*k*NN Classifier	F*k*NN Classifier
5	82.67%	90.42%	89%	90.42%
10	83.91%	91%	91.00%	93.00%
15	87.36%	92.00%	93.00%	94.00%
20	90.42%	95.00%	95%	97.00%

Discussion

The performance of both systems is summarized in Figure 4, comparing between the two systems by using the four classifiers and one feature extraction method (MFCC) in clean data form internet database. The best performance is shown in

the frog call system, this is because the number of species that were used in the frog call identification are 9 species while the human speaker identification system used 37 speakers by using 20 testing data and 29 training for both systems. The main contribution of this study is performing two different types of data-base (i.e. human biometric data-base and frog bioacoustics data-base) by applying the same type of features and parameters that suite both systems. That is leading to unnecessary changes in the system setting and parameters in-order to work on a different type of data-base that is different in terms of frequencies and bits.

Conclusion

In this paper Mel-frequency Cepstral Coefficients (MFCC) is experimented as features for both systems, frog bioacoustics based identification system and human biometric based identification system. The classification process, for classifier, had been used, the performances of Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Local Mean k Nearest Neighbor (LMkNN) and Fuzzy k-NN (FkNN) classifiers have been compared in this study for both systems. The performances of the human biometric system and frog bioacoustics system based on the proposed classifiers are evaluated. The best performance has been observed by using FkNN classifier with the accuracy of 97% for the frog bioacoustics identification system and 93.38% for the biometric speaker identification system with 20 training data.

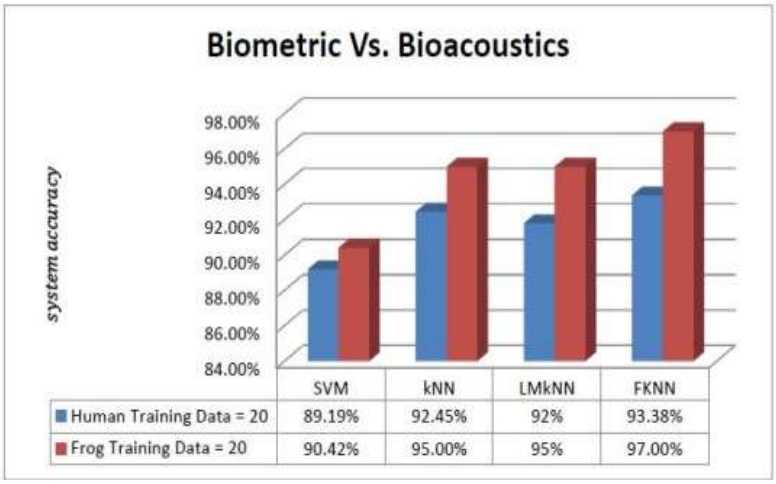


Figure 4. The performance of MFCC with the four classifiers on the both system.

Figure 6 comparesthe two systems and the two classifiers by using four divisions of training data.

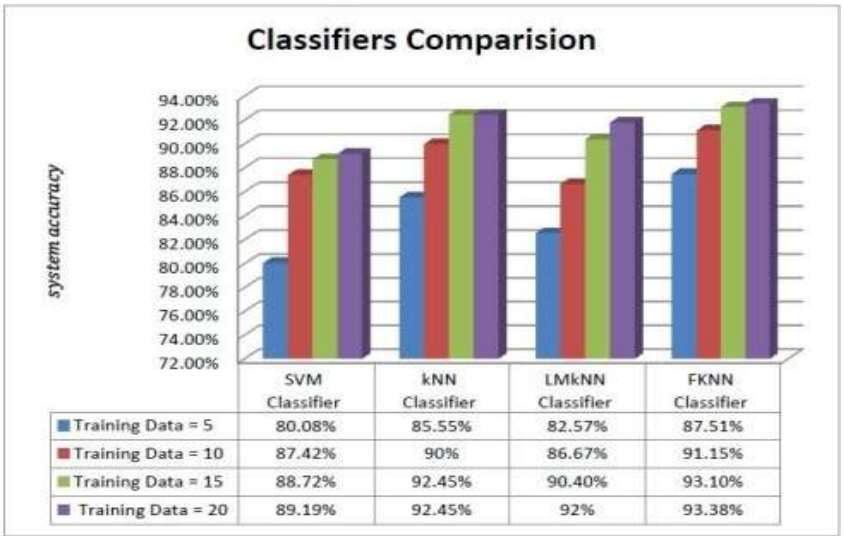
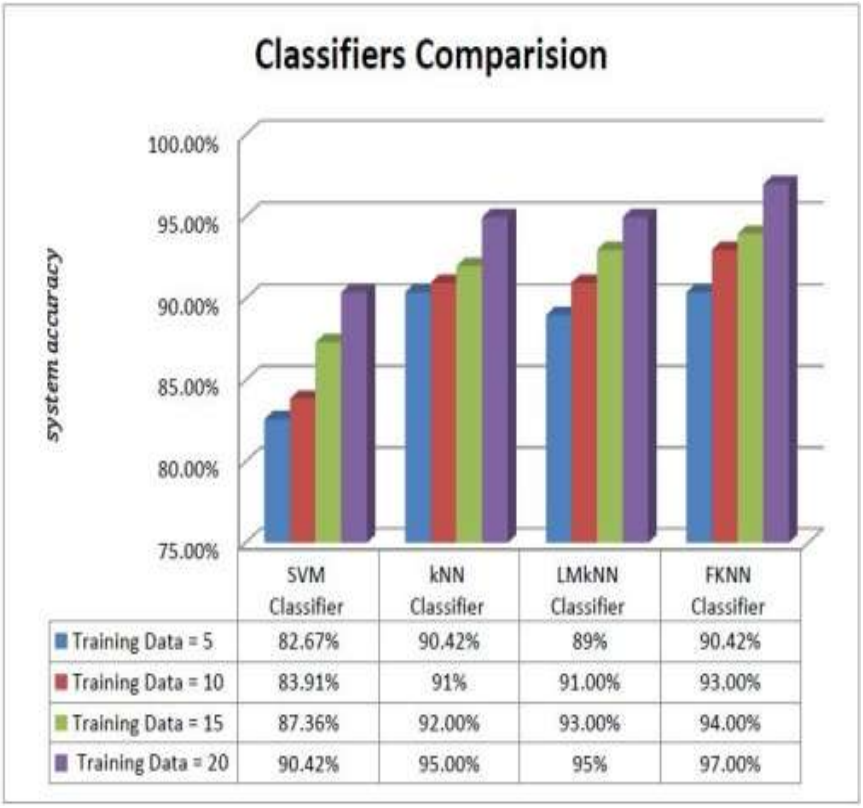


Figure 5 : Human systems using four classifiers and four sets of training data



	SVM Classifier	kNN Classifier	LMkNN Classifier	FKNN Classifier
Training Data = 5	82.67%	90.42%	89%	90.42%
Training Data = 10	83.91%	91%	91.00%	93.00%
Training Data = 15	87.36%	92.00%	93.00%	94.00%
Training Data = 20	90.42%	95.00%	95%	97.00%

Figure 6 : Frog systems using four classifiers and four sets of training data

References

[1] Inthavisas, K. and D. Lopresti, Secure speech biometric templates for user authentication. Biometrics, IET, 2012. 1(1): p. 46-54.

[2] Jain, A.K. Biometric recognition: How do I know who you are? in Signal Processing and Communications Applications Conference, 2004. Proceedings of the IEEE 12th. 2004.

[3] Susan, S. and S. Sharma. A Fuzzy Nearest Neighbor Classifier for Speaker Identification. in Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on. 2012.

[4] Femila, M.D. and A.A. Irudhayaraj. Biometric system. in Electronics Computer Technology (ICECT), 2011 3rd International Conference on. 2011.

[5] Adán, M., et al., Biometric verification/identification based on hands natural layout. Image and Vision Computing, 2008. 26(4): p. 451-465.

[6] Nanni, L. and A. Lumini, An experimental comparison of an ensemble of classifiers for biometric data. Neurocomputing, 2006. 69(13–15): p. 1670-1673.

[7] Sumithra, M.G. and A.K. Devika. A study on feature extraction techniques for text independent speaker identification. in Computer Communication and Informatics (ICCCI), 2012 International Conference on. 2012.

[8] Duane Blackburn, et al., Biometrics History,. Biometrics History, . 2006, USA National Science and Technology Council (NSTC) 27.

[9] Anil, J., Arun, Ross, Karthik Nandakumar,, Introduction to Biometrics. 2011 ed, ed. Springer New York Dordrecht Heidelberg London. 2011, New York, USA: Springer

[10] Kumari, R.S.S., S.S. Nidhyananthan, and A. G, Fused Mel Feature sets based text-independent speaker identification using Gaussian Mixture Model. Procedia Engineering, 2012. 30(0): p. 319-326.
[11] Chesmore, D., Automated bioacoustic identification of species. Anais da Academia Brasileira de Ciências, 2004. 76(2): p. 436-440.
[12] Lee, C.-H., et al., Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. Pattern Recognition Letters, 2006. 27(2): p. 93-101.
[13] Han, N.C., S.V. Muniandy, and J. Dayou, Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Applied Acoustics, 2011. 72(9): p. 639-645.
[14] Chesmore and Nellenbach. Acoustic methods for the automated detection and identification of insects. in International Symposium on Sensors in Horticulture 562. 1997.
[15] Farrell, Mammone, and Assaleh, Speaker recognition using neural networks and conventional classifiers. Speech and Audio Processing, IEEE Transactions on, 1994. 2(1): p. 194-205.
[16] database, F.c. http://learning.froghome.org/.
[17] database,F.c., http://www.frogwatch.org.au/?action=animal.list.
[18] Wu, J.-D. and B.-F. Lin, Speaker identification based on the frame linear
predictive coding spectrum technique. Expert Systems with Applications, 2009. 36(4): p. 8056-8063.
[19] Wei HAN, C.-F.C., Chiu-Sing CHOY and Kong-Pang PUN, An Efficient MFCC Extraction Method in Speech Recognition. 2006.

[20] Furui, S. and T. Kobayashi. Introduction of the METI project; development of fundamental speech recognition technology. in Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on. 2007.
[21] Karpov, E., Real-Time Speaker Identification. M.Sc Thesis. University of Joensuu, 2003.
[22] TianYuan, L., et al. Application of least squares support vector machine in futures price forecasting. in Electronics Computer Technology (ICECT), 2011 3rd International Conference on. 2011.
[23] Lei, W. and Y. Yong. Training One-class Support Vector Machines in the Primal Space. in Electronic Computer Technology, 2009 International Conference on. 2009.
[24] Shuxia, L., S. Pu, and L. Xianhao. Compact Fuzzy Multiclass Support Vector Machines. in Natural Computation, 2008. ICNC '08. Fourth International Conference on. 2008.