International Journal of Economics, Commerce and Management
United Kingdom Vol. III, Issue 5, May 2015
http://ijecm.co.uk/ ISSN 2348 0386

CUSTOMER IDENTIFICATION THROUGH VOICE BIOMETRIC INDEX AT CALL CENTERS USING LEARNING ALGORITHMS

Monireh Hosseini
K. N. Toosi University of Technology, Tehran, Iran

Elnaz Nasirzadeh
M.Sc. Student (Information Technology Engineering)
Khaje Nasir Toosi University of Technology, Tehran, Iran
e.nasirzadeh15@gmail.com

Abstract

In recent years, communication approach in managing relationship with customer has been increasingly considered due to increasing competition, market maturity and rapid development of business technology. Therefore, companies have established call centers in order to manage customer support and to provide information asked by the customers. One of the main requirements of promoting service quality in call centers is recognizing customers during calls. The main purpose of the study is introducing a method for recognizing customers in call centers through voice features of the customers. For reaching the purpose, different systems of customer identification in call centers and features of audio signal have been studied. Then, various learning algorithms used in customer identification systems have been introduced. Finally, the proposed method of the manuscript for Customer identification through voice biometric is introduced. The advantage of recommended method is that this method can recognize them even when there is little information of customer voice. Experimental results showed that the recommended method in speaker identity confirmation has higher rate of recognition than other similar methods using other algorithms.

Keywords: Learning algorithms, Speaker recognition, Call center, Recognition

INTRODUCTION

In recent years, the success of organizations in competitive business environments strongly depends on the ability of their skillful and correct management in internal and external relationships (Berry, 1995). Increasing trend of competition and decrease of customer loyalty led companies to be converted from product-oriented to customer-centric. Given the communication development and also growing need of organizations to integrated and close relationship with their customers, call centers have been established that not only provide the possibility of permanent relation of customers with organizations but also provide complete tools for organizational management. Call centers cause easy access to information and services of organizations, regardless of time and place limitations and preventing time waste. These centers combine a set of the audio and video communications with computerized facilities and provide all types of automatic and semi-automatic communication services along with accurate and given communication with customers of organization.

Call centers play the main role in providing social and public services, especially in order to introduce a culture of solving public problems and are of essentials of today societies. Indeed, these centers consist of large and widespread network of computerized and telecommunication along with related software. Various hardware and software interfaces in system of call centers provide this possibility to integrate various communication channels such as telephone, fax, SMS, voice mail and email and through software, have established one management along with creating necessary relationships(So, 2007).Operators are able to do their activities automatic and semi-automatic through existing software in their own work stations, see necessary information on their screen less than a fraction of a second and to equip themselves for accountability to customers.

In present world, great volume of public and private call centers have been established by banks, insurance companies, equipment seller companies and service companies that play the main role to meet more satisfaction of customers. In past years, small and large organizations that realized the importance of call centers, tried to transfer hidden value-added of these systems to their organizations by investing in establishing such centers (McNally,2007). But not all organizations were successful in this trend because how to choose, implement and application methods of these centers are of main importance to benefit them. Thus, these call centers cannot be known as one contact point of customers with organization and consider them as a separate part of organization body that there is no enough interaction between them and other parts of organization. But call centers are a puzzle-like part of an organization and influence all parts of that organization (Abdullateef, et al, 2014).

In previous communication methods with customers in call centers, there was no complete trust in being genuine when talking about customers and thus most of items such as changing password and money transfer or electronic deposit did not meet in these centers. But in situations under which customers were considered genuine, more services can be offered to them. Therefore, in present study, usage method of voice biometric index has been recommended in order to solve the problem in recognizing real customer in call centers.

Present paper has been prepared in this way: in section two, a review of works done to study customer identification in call centers has been provided and in section three, all types of learning algorithms have been investigated. In section four, a methodology to recognize customer has been explained and also all types of recognition systems of speaker and features of audio signal have been studied. Finally and in section five, paper has been concluded.

SPEECH PROCESSING OF CUSTOMER IN CALL CENTERS

Here, various fields of speech processing have been explained and Speaker recognition area has been chosen to be investigated in order to have a review of previous researches in this field. It should be noted that there is no obvious difference between two words of “voice” and “speech” in literature of voice processing.

Voice Biometric Index

In order to recognize customers, their biometric index can be used. Biometric indexes are through which we can recognize people. These indexes are classified in two physical and behavioral characteristics categories.

Biometrics is defined as a process that validates each characteristic of people in order to recognize them. Physical biometric indexes refer to a set of characteristics of body shape such as fingerprint, iris, face, DNA. These characteristics mainly begin when a child is born and sometimes they exist even before birth and remain unchangeable till the end of life time. But second category that is related to behavioral characteristics, in fact are resulted from human behavior including walking, sound and rhythm of typing that can show characteristics.

Parameters of biometric system are of important issues should be considered to use biometric indexes. There are parameters in all biometric systems that introduce features of systems. These parameters are provided in table 1.

Table1: Parameters to describe biometric features of systems

Parameter	Description
False acceptance rate(FAR) or false match rate (FMR)	This parameter determines the acceptance possibility of fake user than real user. Indeed, it shows the percentage of invalid inputs that have been accepted falsely. This parameter should be as small as possible.
False rejection rate (FRR) or false non-match rate (FNMR)	This parameter shows how well a real person is not accepted falsely and is known as a fake user. In fact, it shows the percentage of valid inputs that are rejected falsely. This parameter should be also as small as possible.
Equal error rate (EER) or crossover error rate (CER)	The decrease of FAR results in the un intentional increase of RRR. Where these two are equal, error rate is also equal. The small amount of this parameter shows that system has better sensitivity and equality and the system with the least EER is known as the most accurate system.
Failure to enroll rate (FER or FTE)	The error possibility that may occur during sampling to record in database of correct recognition.

Speaker Recognition as One of Speech Processing Fields

Speech processing is growing fast as one of subcategories of signal processing. In this section, various methods of speech processing have been investigated. Speech processing has seven main categories including: Speech recognition, speaker recognition, speech coding, speech synthesis, speech enhancement, speech compression. Explanation of each category has been provided in table 2.

Table 2: Various fields of speech processing

Field of Speech Processing	Description	Main Usage	Resources
Speech recognition	In this field, operations done related to speech content analysis of voice signal and its conversion in a way that it can be readable by computer.	Voice user interface	In an article by Anusuya & Katti (2009), investigations have been conducted and main issues and developments in a field of speech recognition have been collected during last 60 years.
Speaker recognition	Here, the mission is to identify speaker	Using in call centers of Barkley's banks in order to identify customers	An article by Campbell (1997) consists of issues about choosing features and random modeling. Another glancing overview has been done in a work by Bimbot

			et al (2004) in which issues about normalization methods and speaker recognition program can be found. New set of chapters of various books about different aspects of categorizing speaker are provided in articles by Muller (2007).
Speech coding	It is a program including data and digital audio signal compression that uses voice processing techniques to model audio signal and make a string of bits with the help of data storage algorithms.	-Mobile phone -VoIP	In an article by Vuppala et al (2010) there are issues about the effect of speech coding on test-independent recognition of speaker.
Speech synthesis	In this field, production of artificial voice of human has been considered that is produced voice similar to human voice by computer with the help of medical information such as sounds sending out of vocal cords through software and hardware	All sound producing machines such as: MacInTalk by Apple	During recent years, a wide range of speech coding techniques have been developed that are studied a comprehensive vision of these approaches and feature of coded speech signals in an article by Mattheyses & Verhelst (2015).
Speech enhancement	In this field, quality enhancement and comprehensibility of voice signals are considered through methods like voice noise reduction.	-Mobile phone -VoIP -remote control systems -hearing aids	It is mentioned in a book by Loizou (2013),"details of speech enhancement and noise estimation algorithms", and strategies to validate efficiency of speech enhancement algorithms have been explained to increase quality and comprehensibility of speech.
Speech compression	In this field, issues about telecommunication are important in a way to compress data in order to transfer, receive store or hear more volume of information.	-Mobile phones -Video conferences	In a book by Sayood (2012), strategies to compress speech have been provided. In an article by Magboub et al (2011), also various techniques of speech compression have been investigated.

After investigations, we concluded that there are numerous studies in a field of speech processing but a few of them considered speaker recognition accurately and most of researchers took into account a field of speech recognition and issues related to speech such as coding and compression.

Thus, the work field of this article has focused on speaker recognition and we are going to identify customer through biometric index of his voice in call centers using techniques of this field. Speaker recognition systems are placed in two categories of speaker identification systems and speaker verification systems in terms of usage method. Of all researches and studies have been conducted on speaker recognition, they mostly focus on speaker identification systems. Therefore, present article once again focuses on speaker identification systems.

LEARNING ALGORITHMS

Basically, learning ability is the main feature of intelligent system. A system that can learn is more flexible and can be programmed more easily. So then it can be accountable better about new issues and equations. If computers can recognize patterns trustfully and infer about real world, they can help people more using higher efficiency. In this regard, numerous activities have been done in a field of learning algorithms and it has been more than half a century that studies have been focused on this issue. Main methods of learning are divided in two categories of supervised and unsupervised (Han et al, 2011).

Supervised methods include methods of classification and regression. There are some predetermined categories in classification methods and after receiving new input, it is determined that this input relates to which one of previous categories. Regression method includes determination of continuous reply variable (dependent variable) in terms of amount of independent variables. Regression is similar to classification but with one difference, dependent variable is discrete in classification. Unsupervised methods also determine the description of general features of data. The purpose of description is to find data patterns that are definable for human. One of common unsupervised method is clustering which seeks to recognize limited number of clusters to describe data.

In present study, the focus is on supervised learning and of all subcategories, classification has been used for investigation. To this end, learning algorithms to classification have been studied in a field of speaker recognition and summary of results have been provided in table 3.

Table 3: A review of conducted researches in a field of speaker recognition

Method	Conducted Researches
Artificial neural network	This method is used for various classification issues such as speaker recognition (Farell et al, 1994).Prediction is in a form of lawless result and data should be numeral. One of potential advantages of neural networks is that we can combine features extraction and speaker modeling in a network and this way, feature extraction model (regardless speaker) and mode of speaker will be combined optimally (Heck & Weintraub, 1997). They also can be used to combine various sub-systems (Tong et al, 2006).
Support vector machine (SVM)	This method is of strong tools of classification that has been used recently in a field of speaker identification. In some articles, this method has been used along with Gaussian combined model to enhance accuracy (Campbell et al, 2006). SVM method has been used for high level features (Campbell et al, 2004). One of reasons for popularity of this method is its high efficiency in generalization to total in order to have an ability to classify unseen data.
Vector quantization (VQ)	This model which is known as gravity center model is one of the easiest speaker modeling methods (Hautamaki et al, 2008). This method has been introduced in 1980 to identify speaker (Soong et al, 1985). Root vector quantization in data compression has been explained in an article by Gersho & Grey (1991). Even though VQ has been used for techniques to increase calculation speed (Kinnunen et al, 2006) and reduction of necessary implementation operations (Saastamoinen et al, 2005), it can enhance accuracy in combination with other models (Hautamaki et al, 2008).

As can be seen in table 3, there are various methods for classification and we used the most famous one, artificial neural network. These networks are a pattern to process information and are built with imitation of biological neural networks like human brain. Key element of this pattern is a new structure of information processing system and includes many elements that work together co-ordinately with strong internal communication to solve problems. Artificial neural networks transfer knowledge or hidden law beyond data to network structure by processing experimental data and this is called learning.

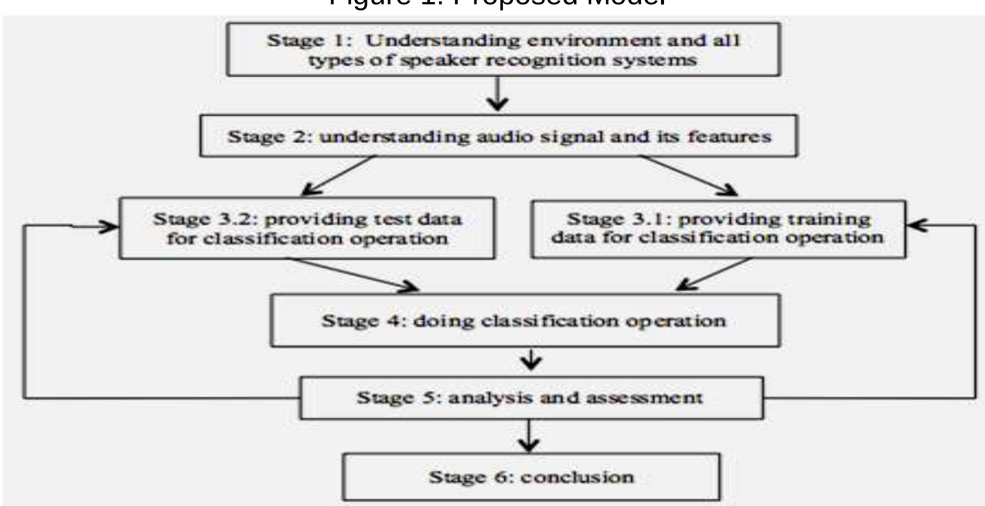
In other word, artificial neural network is a data processing system which is inspired by human brain and has made many small processors to process data that behave as a consistent and parallel network to solve a problem. In such networks, data structure is designed with the help of programming knowledge that can act as a neuron and is called node. Then network is thought by creating a network between these nodes and conducting a learning algorithm. The easiest type of neural network consists of two incoming and out coming layers that serves as input-output system and value of incoming neurons are used to calculate the value of outcoming neurons with a activation function. Nodes have two status: active (on or one) and passive (off or zero) in this memory or neural network and each edge (synapsis or relation between nodes) have same weight. Edges with positive weight leads to stimulate or activate next passive node and edges with negative weight inactivate next integrated node (if it is active). In multi-layer neural networks, hidden layers also used in addition to incoming and out coming layers, because these hidden layers lead to improve the performance of these networks. In this study, two types of neural networks (feed-forward & probabilistic) have been investigated. Feed-forward neural network consists of several layers, where data are transferred unidirectionaly.

Probabilistic neural networks are in fact statistical implementations that are used generally as classification method in learning. These networks are unsupervised feed-forward networks with four layers: incoming, pattern, total and out coming. A probabilistic function such as Gaussian is used for each pattern node. Weights of network are updated according to incoming patterns. Then patterns are classified using the closest and adjacent function according to Gaussian classification. Median and variance of each node are updated during training to minimize the distance between patterns and their closest classification. Further theoretical details of such networks can be found in an article by Ganchev et al (2007).

PROPOSED METHOD TO IDENTIFY CUSTOMER

Nowadays, voice and data entry to call centers are increasingly done through new set of technology. This valuable data and voice set of customers can be used to improve services. As mentioned, one of the ways to use received data from customers is using their voice. Thus after customer calls, their voices have been recorded and it is used to identify them in their next calls. In fact after receiving new call, system compares received voice through methods will be discussed later with other voices available in database and in case of compatibility, it provides customer information to an operator who is giving service to that customer. Proposed model for this research has been shown in figure 1.

Figure 1: Proposed Model



In the following, each stage of proposed model will be studied. Broadly speaking, there are two systems of speaker recognition that are studied in the section below. On one hand, to use customer voices, first we have to choose characteristics of audio signals and studied them as inputs of learning algorithms in the third section and after implementing mentioned algorithms, we determine customer identity due to output.

Types of Speaker Recognition Systems

Speaker recognition systems are placed in two categories of speaker identification systems and speaker verification systems based on usage method. In speaker identification system, individual declares that he is registered user of system by choosing or entering the name of one of special users. In this regard, This is system to compare vocal features of this person with those saved for registered user and by using this result can accept or reject this claim.in speaker verification system, individual does not declare to have an identity of a special user and this is system to recognize him among registered users of system or checks his vocal features are not consistent with none of registered users.

Although there is no significant difference between these two systems, but today due to many customer-centric organizations and companies, usages of speaker verification systems are more considerable than speaker identification systems. Therefore in this study, speaker verification systems have been used. Customer identity is recognized by using database of customers' voices after receiving new calls and matching new input to existing data in database.

Choosing Features of Audio Signal

Efficiency of learning algorithms strongly depends on features of selected input. Speech signal also has various features, not all of them are needed to be used to recognize speaker. Therefore, various features of audio signal should be studied in order to consider the most appropriate ones as inputs of learning algorithms. An ideal feature of audio signal should have following qualifications (Rose, 2002):

1. Should be resistant to noise and distortion.
2. Naturally and frequently occurs in speech.
3. It can be easily measured from the speech signal.
4. It is difficult to counterfeit and to imitate it.
5. Should not be influenced by human health and long term changes of voice.

In addition to what mentioned above, the numbers of these features are also important and not to be exceeded from given numbers, because traditional statistical models such as Gaussian mixed model cannot manage many data features (Reynolds, et al, 2000). The number of training samples required increases exponentially with the number of features. However, working with less features, reduces computation cost. From Physical characteristics point of view, features are divided into 4 categories and are explained in table 4.

Table 4: Feature classification in terms of physical characteristics

Title	Description
Short-term spectral characteristics	These characteristics, as their names imply, are computed from short frames of 20 to 30 ms
Features of voice source	These characteristics describe voice source. (Larynx and trachea opening)
Spectral-temporal characteristics	These characteristics are computed according to frames with more than tens or hundreds of milliseconds, for example, include pronunciation and tune.
High level characteristics	These characteristics try to achieve profile of speaker's conversation such as the use of words (for instance, ok, oh yes!)

Feature selection depends on field work, computing resources, the number of speech data available in the database. For researchers who are willing to start their research in the field of speaker recognition, it is recommended to use the short- term spectral characteristics, as the calculation of these features is simple and highly efficient. High level characteristics have higher power but it is difficult to separate them and is easier to fabricate. For example, professional counterfeiters tend to manipulate intonation in order to imitate voice of speaker. High level characteristics require complex systems such as automatic speech recognition systems. Thus, in general, there is not the best set of characteristics that can be used for various researches in a field of speaker recognition. Therefore, in order to select features, we should consider the ability of speaker recognition on one hand and power and practicality of characteristics on the other hand. In this article, we used features of voice source. Air in the mouth and nasal cavity fluctuates in a range of frequencies. Frequency spectrum can be produced by Fourier transform of a signal. The obtained results are usually presented as amplitude and phase. These frequencies are influenced by size and shape of vocal cords, tongue and lip positions. Resonance of vocal cords is investigated in terms of formant frequencies of vowels. Since vocal cords in men are 15% longer than vocal cords in women, audio signal of men has lower formant frequencies than women. Resonated frequencies of vocal cords are known as formants with validated features to recognize speaker automatically and speech synthesis. The location of these resonances in the frequency spectrum depends on the form and shape of vocal cords. Since the physical structure of vocal organs is of each speaker characteristics and differences between speakers can be found in various locations of formant frequencies. These resonances that influence the whole shape of frequency spectrum are called formant. A few of these formant frequencies can be sampled with a suitable rate and can be used to recognize speaker. These features are usually used in combination with other characteristics. In this paper, formants along with WP entropy are used to recognize speaker and in our proposed method, formants and WP parameters are considered as inputs of neural network.

Formants are defined as peak of vocal frequency spectrum or hearing resonance of human vocal cords. In a spectroscope with wide band, formants can be shown as black bars. Voice of each speaker can be modeled as time-variable linear system. Using this linear system of n order and calculating its reverse, model of linear system spectrum can be created by audio signal. The use of small order may lead to noisy resonances while using big order can create artificial harmonics. For small values of n, resonance is noisy whereas for big n, some of peaks in fact instead of formants show harmonics. If we consider n about 1ms, the we have 5 formants and it seems it is a suitable choice for n order. Five initial resonant vocal frequencies (e.g. F1, F2, F3, F4, and F5) are distinguishable separately for each speaker and therefore can be used as a characteristic of those speakers. One of general method of wavelet decomposition is WP. The reason to use WP entropy is that we can extract extra features of frequencies through Shannon entropy.

Provision of Training Data and Test for Classification

In present study, 80 customers have participated recording and for each, this process has been repeated at least 15 times. The age range was 20 to 40 years and among them, 46 were male and 34 are female. The process of recording has been done in a room with closed doors and without any external noise. Speech signals with sound card with 3 kHz spectral frequency and sampling frequency of 8 kHz have been recorded. These audio signals contain noises that will change obtained results of various learning methods. For this purpose, first there is required the operation on them to eliminate noises. To this end, we obtained fast Fourier transform (FTT) from audio signals, then because of 3 kHz extend of recorded voice and 8 kHz noise frequency, this noise has been eliminated, applying a low-pass filter and saved signal without any noise. Then among these cleaned-up signals, 70% for training data and 30% remaining signals have been used to test data.

Performing Classification

As discussed in section Choosing Features of Audio Signal, choosing appropriate features of input data or engineering features is a way to increase the efficiency of learning algorithms. But in order to maximize the efficiency and accuracy, it is better to have learning algorithms less dependent on features of input data. For this purpose, we used two methods of feed-forward neural network and probabilistic neural network to classify available data in database after preparation of initial data which include voice of customers.

We examined different values for the parameters of the neural network to study necessary conditions to achieve the best efficiency. The best obtained results of this network are provided via parameters that are shown in table 5.

Table 5: Parameters of artificial neural network

Parameter	**Description**
Type of network	Feed-forward post-propagation error
Number of layers	3: input, hidden, output
Number of neurons of each layer	12 in input layer, 30 in middle layer and 4 in output layer
Training function	Levenberg-Marquardt
MSE	10^{-5}
Activation function	Sigmoid
Number of courses	200
Maximum of validation failure	5
Minimum of gradient efficiency	10^{-10}
Initial μ	10^{-2}
Increase factor of μ	10
Reduction factor of μ	0.1
Maximum of μ	10^{10}

Analysis and Evaluation

In order to select the best method for the customer identification, the results of the classification of learning algorithms should be evaluated. Comparison indexes of various methods of classification include: accuracy of a model which depends on true predictions done by that model, speed that is related to build and use of model, robustness defines the ability of model to deal with unusual data or lost values, interpretability shows the comprehensibility of model by others and finally, compactness of model that is very important in motivating its use. In order to evaluate, accuracy and effectiveness of system has been measured on 80 speakers whose voices have been saved in database. In every performance, 50% of signals were chosen from a class from where correct speakers have been selected and remaining 50% was chosen from another class .In order to describe the rate of recognition we used to evaluate results of classification, it is needed to explain four concepts that are provided in table 6.

Table 6: Explanation of used concepts to calculate the rate of recognition

Element	Description
True positive	It shows the correct observations that have been diagnosed truly by classification method.
True negative	It shows the false observations that have been diagnosed truly by classification method.
False positive	It shows the correct observations that have been diagnosed untruly by classification method.
False negative	It shows the false observations that have been diagnosed untruly by classification method.

As has been presented in table 6, there is this possibility that two statuses of false positive and false negative lead to have two types of false occurrence that are called false positive error (FPE) and false negative error (FNE), respectively. The amount of FPE equals total numbers of signals that are known as spurious signals, divided by total number of test signals and FNE equals total number of true signals divided by total number of signals. Due to definitions of these two errors, the rate of recognition is obtained by equation (1):

Rate of recognition= 100% - (FNE+FPE)

Obtained results were compared with two methods used in similar articles. In one of them, MFCC method along with neural network has been used (Ganchev et al, 2005) that we showed it here as MFNN-NN. In second article, LPC method has been used with neural network (Bennani& Gallinari, 1995) that we showed it as LPC-NN. Comparison results are provided in table 7. It should be noted that 3% increase of recognition rate in FW-NN method in compared to MFCC-NN method and 19% increase in compared to LPC-NN method shows the improvement of this rate in proposed method of article.

Table7: Recognition rate in order to evaluate experimental results of various learning algorithms

Method	% recognition rate	% FNE	% FPE
FW-NN	80.24	12.3	6.79
P-NN	79.01	13.37	6.98
MFCC-NN	77.32	14.23	8.11
LPC-NN	61.88	15	23.12

In subsequent experiments, speaker recognition has been tested. In this experiments, 70% of signals each category for training data and remaining 30% for test data were used. The method proposed in this paper has been compared to two new methods that used wavelet transform: one of these methods is to use genetic wavelet packet with artificial neural network (Engin, 2007) that we showed it as GWP-NN. Next method is discrete wavelet transform (at level 5) using extraction method of proposed feature in section 'Choosing Features of Audio Signal' and artificial neural network that we showed it here as DWT-NN. The results have been provided in table 8.

Table 8: Results of recognition rate for related methods

Recognition method	Number of speakers	% recognition rate (dependent to vowels)	% recognition rate (independent of vowels)
FW-NN	80	88.12	81.54
P-NN	80	87.54	81.1
MFCC-NN	80	79.66	74.03
LPC-NN	80	66.73	59.45
DWT-NN	80	80.66	79.71
GWP-NN	80	85.47	80.07

As shown in Table 8, according to the results, the proposed method has the best results feed-forward neural network in two statuses dependent on vowels with recognition rate of 88.12 and also independent of vowels with recognition rate of 81.54. Therefore, using the proposed method we can accurately classify received customer voices and after receiving new voice through phone, there will be recognition match and customer will be recognized if a similar voice existed in database.

CONCLUSION

With the development of communication and also increasing need of organizations to work closely with their customers, call centers have been created that led to easy access to information and service of companies, regardless of time and place limitations and they prevent from time wasting. In previous methods of communication with the customer in call centers, there was no full trust in having true customers and consequently, so many things such as changing password and money transfer or e-transfer were not performed in these centers. But when there is a full trust in having true customers, there will be more services to offer customers. Hence, in this paper it is recommended to use voice biometric index to solve the problem of true customer recognition in call centers. In this method, first some of features of audio signals have been extracted using formants and Shannon entropy, and then these features were considered as inputs of neural network. In this method, learning algorithms of feed-forward neural network and probabilistic neural network were used for classification process. Then proposed method was compared with some similar methods that used other learning algorithms (GWP-NN, DWT-NN, LPC-NN, MFCC-NN). Results indicated that proposed method has higher recognition rate than other methods. Therefore, we can use voice biometric index in call centers applying this method and as soon as hearing customer voice in these centers, he will be identified and according to his needs, appropriate service will be offered based on his history. Moreover, since this index is unique for each customer, it can be used as inimitable and same password and sometimes unchangeable one. There will be no problems such as forgetting password or losing it. In addition, there is no need to train normal users and also there is less possibility of theft in this method.

REFERENCES

Abdullateef, A. O., Mokhtar, S. S. M. & Yusoff, R. Z., (2013). Linkages between CRM technologies, knowledge applications and first call. *International Journal of Electronic Customer Relationship Management,* 7(1): 68-86.

Abdullateef, A. O., Muktar, S. S. M., Yusoff, R. Z. & Ahmad, I. S. B., (2014). Effects of Customer Relationship Management Strategy on Call Centre's Employee Intention to Quit: Evidence from Malaysia Call Centers. *Procedia - Social and Behavioral Sciences,* 305–315.

Anusuya, M. A. &Katti, S. K., (2009).Speech Recognition by Machine, A Review. *International Journal of*

Computer Science and Information Security, 181-205.

Beleites, C. et al., 2013. Sample size planning for classification models. AnalyticaChimicaActa, 760: 25-33.

Benesty, J., Sondhi, M. M. & Huang, Y., (2008). In: Springer Handbook of Speech Processing. s.l.:Springer.

Bennani, Y. & Gallinari, P., (1995). Neural networks for discrimination and modelization of speakers. Speech Communication, 17(1-2): 159–175.

Berry, L. L., (1995). Relationship marketing of services; Growing interest, emerging Perspectives. *Journal of Academy of Marketing Science,* 236-245.

Bimbot, F. et al., (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing,* 430-451.

Burton, D., (1987). Text-dependent speaker verification using vector quantization source coding. IEEE Trans. Acoustics, Speech and Signal Processing, 35(2): 133–143.

Campbell, J., (1997). *Speaker recognition: a tutorial.* s.l., s.n., p. 1437–1462.

Campbell, W. et al., (2004). Phonetic speaker recognition with support vector machines.s.l. Advances in Neural Information Processing Systems 16.

Campbell, W. et al., (2006). Support vector machines for speaker and language recognition. Computer Speech and Language, 20(2-3): 210–229.

Campbell, W., Sturim, D. & Reynolds, D., (2006). Support vector machines using GMM supervectors for speaker verification. IEEE Signal Processing Letters, 13(5): 308–311.

Engin, A. (2007). A new optimum feature extraction and classification method for speaker recognition: GWPNN. Expert Syst. Appl, 32: 485–498.

Farrell, K., Ramachandran, R. & Mammone, R., (1994). Speaker recognition using neural networks and conventional classifiers. IEEE trans.on Speech and Audio Processing, 2(1): 194–205.

Ganchev, T., Fakotakis, N. & Kokkinakis, G., (2005).Comparative evaluation of various MFCC implementations on the speaker verification task. s.l., in Proc. of the SPECOM-2005, 191-194.

Ganchev, T., Tasoulis, D., Vrahatis, M. & Fakotakis, D., (2007). Generalized locally recurrent probabilistic neural networks with application to text-independent speaker verification. *Neurocomputing,* 70(7–9): 1424–1438.

Gersho, A. & Gray, R., (1991). Vector Quantization and Signal Compression. Boston, s.n.

Han, J., Kamber, M. & Pei, J., (2011). Data mining cocepts and techniques.3 ed. s.l.:Morgan Kaufmann.

Hautamaki, V. et al., (2008). Maximum a Posteriori estimation of the centroid model for speaker verification. *IEEE Signal Processing Letters 15:* 162–165.

Heck, L., Konig, Y., Sonmez, M. & Weintraub, M., (2000). Robustness to telephone handset distortion in speaker recognition by discriminative feature design. *Speech Communication,* 31: 181-192.

Heck, L. & Weintraub, M., (1997). Handset-dependent background models for robust text-independent speaker recognition. Munich, In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 1997) 1071-1074.

He, J., Liu, L. & Palm, G., (1999). A discriminative training algorithm for VQ-based speaker identification.IEEE Trans. on Speech and Audio Processing, 7(3): 353–356.

Hilbert, M. & López, P., (2011). The World’s Technological Capacity to Store, Communicate, and Compute Information. *Science (journal),* 332(6025): 60–65.

Jain, A., Hong, L. & Pankanti, S., (2000). Biometric Identification. Communications of the ACM, 43(2): 90-98.

Jain, A. K. & Ross, A. (2008). Introduction to Biometrics. In: *Handbook of Biometrics.* s.l.:Springer: 1-22.

Kinnunen, T., Karpov, E. & Franti, P., (2006). Real-time speaker identification and verification.IEEE Trans. Audio, Speech and Language Processing, 14(1): 277–288.

Kinnunen, T. & Li, H., (2010). An Overview of Text-Independent Speaker Recognition: from Features to Supervectors. Speech Communication*,* 52(1): 12–40.

Lapidot, I., Guterman, H. & Cohen, A., 2002.Unsupervised speaker recognition based on competition between selforganizing maps. s.l., s.n.: 877–887.

Loan, C. V., (1992). Computational Frameworks for the Fast Fourier Transform. s.l.:Society for Industrial and Applied Mathematics.

Loizou, P. C., (2013). Speech Enhancement: Theory and Practice. 2 edition ed. s.l.:CRC Press.

Magboub, H. M., Ali, N., Osman, M. A. & Alfandi, S. A., (2011). Multimedia speech compression techniques. s.l., IEEE: 9-11 .

Mattheyses, W. & Verhelst, W., (2015). Audiovisual speech synthesis: An overview of the state-of-the-art. Speech Communication: 182–217.

McNally, R. C., (2007). An exploration of call centre agents' CRM software use, customer orientation and job performance in the customer. Journal of Financial Services Marketing, 2(12): 169-184.

Mohri, M., Rostamizadeh, A. & Talwalkar, A., (2012). Foundations of Machine Learning.s.l.:The MIT Press.

Muller, C., 2007. Speaker Classification I: Fundamentals, Features and methods. Lecture Notes in Computer, 4343.

Mutch, J. & Lowe, D. G., (2008). Object class recognition and localization using sparse features with limited receptive fields. International Journal of Computer Vision, 80(1): 45–57.

Obleser, J., Leaver, A. M., Van Meter, J. & Rauschecker, J. P., (2010). Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. In Front. Psychol24.

Pinto, N., DiCarlo, J. & Cox, D., (2008). Establishing good benchmarks and baselines for face recognition. s.l.,s.n.

Ramírez, M. A. & Minami, M., (2011). Technology and standards for low-bit-rate vocoding methods. The

Handbook of Computer Networks, 2: 447–467.

Reynolds, D. A., Quatieri, T. F. & Dunn, R. B., (2000). Speaker verification using adapted gaussian mixture models. Digital Signal Processing, 10(1-3): 19–41.

Roland, P., (1991). Introductory signal processing. Singapore: World Scientific.

Rose, P., (2002). Forensic Speaker Identification.London: CRC Press.

Saastamoinen, J., Karpov, E., Hautamaki, V. & Franti, P., (2005). Accuracy of MFCC based speaker recognition in series 60. EURASIP Journal on Applied Signal Processing, 17: 2816–2827.

Sahoo, S. K., Choubisa, T. & Prasanna, S. M., (2012). Multimodal Biometric Person Authentication: A Review. IETE Technical Review, 29(1).

Sayood , K., (2012). Introduction to Data Compression.4 edition ed. s.l.:Morgan Kaufmann.

Sinha, P., (2009). Speech processing in embedded systems. Springer, p. 9.

Soong, F., Rosenberg, A., Rabiner, L. &Juang, B. H., (1985). A vector quantization approach to speaker recognition. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85, 10: 387 - 390.

So, S., (2007). An Empirical Analysis on the operational Efficiency of CRM call centers in Korea. *International Journal of Computer Science,* 7(12):171-178.

Sra, S., Sebastian, N. & Wright, S. J., (2012). Optimization for Machine Learning. s.l.:MIT Press.

Strang, G., (1994). Wavelets. American Scientist, 82(3): 250-255.

Tong, R. et al., (2006). Fusion of acoustic and tokenization features for speaker recognition. In: Q. Huo, B. Ma, E.

S. Chng & H. Li, eds. Chinese Spoken Language Processing. Singapore: Springer Berlin Heidelberg: 566-577.

Vuppala, A. K., Rao, K. S. & Chakrabarti, S., (2010). Effect of speech coding on speaker identification. Kolkata, IEEE: 1 - 4.

