UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

A VoIP call classifier for carrier grade based on Support Vector Machines

Juan Ricardo Wilches-Cortina [a], Jairo Alberto Cardona-Peña [b] & Juan Pablo Tello-Portillo [b]

[a] *Comcel S.A., Bogotá, Colombia. juanwil@gmail.com*
[b] *Departamento Ingenierías Eléctrica y Electrónica, Universidad del Norte, Barranquilla, Colombia. jacardona@uninorte.edu.co, telloj@uninorte.edu.co*

Received: November 10th, de 2016. Received in revised form: June 5th, 2017. Accepted: July 7th, 2017

Abstract
Currently, VoIP company technicians conduct tests to classify call quality as good or bad. Even though, there are automatic platforms that make test VoIP calls to classify them, they do not perform audio processing to detect False Answer Supervision (FAS), which is a common and undesirable feature of VoIP calls. In this paper, a Vector Support Machine (SVM) along with several functions used in voice recognition were implemented to emulate the human decision procedure (the task of audio classification and analysis performed by technicians). The experiments were based on the comparison between the results obtained from the current classification methods and those derived from the SVM. A 10-fold cross-validation was used to evaluate the system performance. The tests results from the proposed methodology show a better percentage of successful classification compared to a selected automatic platform called CheckMyRoutes.

Keywords: Audio analysis; pattern recognition; SVM; VoIP.

Clasificador de llamadas VoIP para calificar un proveedor basado en Support Vector Machine

Resumen
Actualmente, los técnicos de compañías de VoIP realizan pruebas y clasifican las llamadas como buenas o malas. Asimismo, existen plataformas automáticas que realizan llamadas VoIP para clasificarlas, sin realizar procesamiento de audio; proceso necesario cuando se pretende detectar el False Answer Supervision (FAS), una característica común e indeseable de las llamadas VoIP. Se implementó una Máquina de Vectores de Soporte (SVM) junto con varias funciones utilizadas en el reconocimiento de voz para emular la toma de decisiones de los humanos (tarea de clasificación y análisis de audio realizada por los técnicos). Los experimentos se basaron en la comparación entre los resultados obtenidos de los métodos de clasificación actuales y los derivados de la SVM. Se utilizó una validación cruzada de diez veces para evaluar el rendimiento del sistema. Derivado de los resultados, la metodología propuesta muestra un mejor porcentaje de clasificación exitosa comparado con una plataforma automática llamada CheckMyRoutes.

Palabras clave: Análisis de audio; reconocimiento de patrones; SVM, VoIP.

1. Introduction

Today VoIP calls are part of a large market in which VoIP carriers conduct interconnection contracts with similar companies so they can complete most of their traffic. The unregulated tariff definition of voice minutes, transported using VoIP, allows their value to be more competitive and really attractive for new companies (which join the market). These new companies tend to get away the customers from traditional ones whose tariffs are regulated [1,2]. Studies have shown that VoIP traffic continues to grow and the number of carriers offering the service has increased as well [3-5]. Therefore a large number of companies have emerged offering VoIP destinations at competitive prices.

The technical support department of a VoIP carrier, among other duties, is in charge of testing the destinations

How to cite: Wilches-Cortina, J.R., Cardona-Peña, J.A. and Tello-Portillo, J.P., A VoIP call classifier for crrier grade based on support vector machines. DYNA, 84(202), pp. 75-83, September, 2017.

offered by the suppliers of the company. This task is performed to determine their suitability to handle VoIP traffic. According to the testing result, the carrier can decide whether to sell the destination to its customers. Nevertheless, taking into account that most of the time there is a large quantity of offers received from different vendors, it is necessary to look for the best supplier whose cost-quality factor (the customer wants a termination that completes VoIP calls with good quality at a good price) better adjusts to the customer's requirements.

It should be noted that the destination of a supplier can be really volatile, i.e., there may be offers which are working good during one day at very attractive prices but then they can disappear for a long time. The promptness in which such offer is found and used, is fundamental for taking advantage of this kind of opportunities so the VoIP traffic can be completed in better economic conditions.

Technicians of VoIP carriers determine manually the classification of several VoIP calls in order to decide if the supplier is suitable for traffic completion. First, they analyze the duration of Post Dial Delay and Ring Time for each call. Then, if a call is connected (has duration), they determine whether it has voice or not, listening to the audio that comes from the end user. False Answer Supervision calls contain ring back tones, noise or dead air (no audio or environment audio) which is charged to the end user. Finally, based on the analysis of several calls, the technicians state whether the supplier is apt or no apt for traffic termination.

There is not a specific number of attempts that needs to be analyzed in order to define that a supplier will have a good or bad performance. However, the more number of calls, the more the achieved accuracy. From statistics, 30 or more attempts are needed to determine a numeric value on how the supplier will perform with traffic. However this number can be costly for the carrier if the destination has a high cost (e.g. Cuba or Satellite Phones).

The VoIP market offers a web based platform named CheckMyRoutes [6], whose goal is the automatic benchmark of a VoIP carrier destination. However, it has a high cost ($99/month) and the results are not always good. CheckMyRoutes only analyzes the signaling of the call and does not take into account the media (audio packets), originating a high amount of false positives since the audio is important for the detection of False Answer Supervision (an undesirable common VoIP issue) [7].

Since VoIP carriers have a business dynamism in which time is important for using the volatile offers of the supplier, it is necessary to accelerate the diagnosis process of possible profit opportunities. This lies in the promptness of the analysis derived from the testing. Empirically it has been shown that the suitability of a VoIP carrier for traffic termination depends on several time parameters related to the used protocol and the resulting audio coming from the answer of the end user [8,9]. Therefore, the proposed methodology aims to exploit the information which lies within the VoIP protocols and packets (signaling and media), so the grade of a particular VoIP carrier destination can be determined with little human intervention. It also has the final goal of shortening the time spent in the diagnosis of new business opportunities.

2. Related work

There has been a lot of work on how to measure Quality of Service (QoS) and Quality of Experience (QoE) in VoIP networks. The first aspect focuses mainly on the measurement of network parameters such as delay, jitter and packet loss, especially inspired by the mechanisms used in conventional telephone networks [41], in which the perceived Speech Quality is the most important QoS metric. In [40] an interesting survey of the VoIP technology is presented. Measuring QoE is more difficult because it depends heavily on certain subjective aspects of difficult measurement, as indicated in [32-34].

The ITU defines QoE as: "*A measure of the overall acceptability of an application or service, the perceived subjectively by end-user*". Unlike telephone networks, in network VoIP network conditions can constantly vary in regard to what the perception in one moment may be respect to another. In [32] a QoE analysis is done in a period of 120 days, and it was found that in addition to the technical metrics (delay, jitter), aspects such as: device type, operating system, user interaction time, may affect the Measurement of QoE. Specifically in that study, it was shown to tend to decrease slightly with time. A similar result is confirmed in [33] where it is suggested that the measurement can be carried out in-situ, instead of performing it in controlled environments, due to the great diversity and quantity of final devices that can affect the measurement of QoE. Additionally, there may be even social aspects that affect such measurement, as indicated in [34]. In this article we develop a non-intrusive algorithm based on Machine Learning to perform the prediction of QoE based on some mathematical models such as E-Model [42], WFL (Weber Fechner Law) [43] and PESQ (Perceptual Evaluation of Speech Quality) [44]. The results were very satisfactory. In [35], a PESQ modification is performed using the NPESQ (new Perceptual Evaluation of Speech Quality) algorithm to measure the impact of some metrics on the degradation of VoIP networks.

More detailed studies on QoE can be found in [36,37] and [39], where a tour is made of the main measurement techniques. In [38] we review the relationship of QoE vs the price paid by the user, and focuses mainly on the content of the point of view of the provider.

Research papers like [10,11] focus their objectives on the quality of service offered by VoIP compared to that offered by PSTN. Works [12,13] use the information contained within the interconnection protocols and aim to evaluate the performance of the VoIP network but not a particular destination of a VoIP carrier. Related to voice recognition and voice activity detection (VAD), works like [14,15] compare several algorithms used for VAD. The research [15] includes for time domain the Short-Time Average Magnitude Function (AM), Short-Time Average Zero-Crossing Rate (ZCR), Short-Time Auto Correlation Function, Short-Time Average Magnitude Difference Function (AMDF) and for frequency domain Fourier analysis. Results show that each algorithm has its own advantages and disadvantages. Nevertheless, the combination of them can help diminish their disadvantages [15]. The research paper [30] uses support vector machines (SVM) as the soft-computing technique along with robust features of audio signals in order to classify their frames as ACTIVE or INACTIVE. In [16] features based on energy and Zero Crossing Rate are used in order to classify audio frames, by means of rules, into one of the following three classes: Voiced, Unvoiced or Silence.

In this work is proposed a methodology for VoIP call classification which helps carriers grade the destinations of their suppliers. Some audio features used in voice recognition algorithms along with a SVM were implemented in order to discern between voiced and unvoiced audio signals.

3. Proposed approach

This work took into consideration the communication structure presented in [6,8,17], along with the methodologies proposed by [18, 19] for voice quality in VoIP networks. Fig. 1 shows a block diagram of the proposed methodology which adds the Parameter Extraction and Analysis block to the approach of CheckMyRoutes. Basically the approach is based on a database which stores the test numbers used for calling the destination, then a traffic generator originates the VoIP calls (e.g. asterisk) which are sent to the destination and finally the grade is determined. These tasks will be further explained in the next subsections.

3.1. Database

In order to grade a VoIP carrier offer, it is necessary to call the particular destination (using the gateway of the local carrier) and emit an opinion according to the connectivity result. If this is performed through various carriers and destinations, a database which contains an organized list of suppliers, destinations and test numbers is needed. This database contains a list of test numbers which need to be dialed in order to determine whether the particular carrier is suitable for handling the VoIP traffic, e.g., Supplier John Doe offers Colombia Mobile Tigo to the Local Carrier, a number of calls is then dialed to 57300xxxxx numbers which are previously stored in the database.

3.2. Traffic Generator

An important part in the automation of call grading is the generation of traffic that will be sent to the supplier of the local carrier. It becomes necessary the communication between a *Softswitch* and the database described above so the calls can be routed accordingly. A *Softswitch* performs call control. This is related to functions which process a phone call and offer characteristics of telephony (ring back tone, busy tone, off-hook or dial tone, caller ID, etc.). A *Softswitch* also coordinates the routing of signaling messages which are necessary between networks that wish to communicate.

As it is shown by Fig. 1, the traffic generator sends calls to the local carrier, which are then routed to their corresponding suppliers by the gateway of the local carrier (there is no direct interconnection between the traffic generator and the suppliers). The softswitch used in this work was Genband Nextone.

3.3. Sniffer

This task belongs to the capture derived from the traffic generation and sharing of packets. Its goal is to store relevant

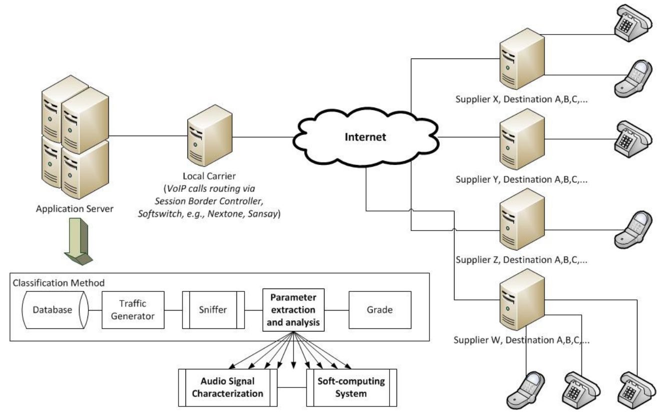

Figure 1. Communication diagram and proposed methodology.
Source: The authors.

information in order to grade call connectivity. *Sniffers* are used to capture network data without redirecting or altering them. They can be found as software applications or hardware devices completely separated from a *Softswitch*, although they can be part of it as well. The sniffer used for collecting data was Wireshark.

3.4. Parameter extraction and analysis

After the packet capture, it is necessary to extract and analyze some parameters which serve to grade the call connectivity. This task corresponds to the emulation of the reasoning made by the members of the technical support department of the local carrier (network operations center, NOC).

Based on the experience of VoIP carrier technicians, the grade of the destination of a carrier depends on several factors. A survey was performed in the company Estel Communications (a VoIP carrier) in order to determine which parameters are used to grade the call connectivity. Since the test results were stored in the e-mail account of the network operations center (NOC), each message had in its body the main parameters used for deciding whether the tests were successful or not. In 308 tests made by technicians, requested in the first days of March 2012 (Property of Estel Communications® used only for educational purposes), the parameters used for grading the performance of the carrier regarding call connectivity are shown in Fig. 2. These parameters are described below for the Session Initiation Protocol (SIP):

- PDD (post dial delay): PDD corresponds to the time difference between the INVITE message sent by the client and the answer of type 18X (session description) sent by the server. It is also the time difference between the INVITE message and any message of type 4XX or 5XX which indicate an issue with the session establishment. It is usually given in seconds by most platforms [8]. It is used to measure the time the server takes to respond a request of a session establishment. As the PDD becomes higher, the connection is said to have a poor performance [9,19].
- Ring Time: The ring time corresponds to the time difference between the 18X message received by the

client from the server and the message 200 OK sent from the server to the client, letting this last know that the call was accepted by the termination. This 200 OK message is followed by an acknowledging message (ACK) sent by the client to the server. During this time (after the first 18X is sent) RTP packets are shared and an alert starts to be heard on the client side (ring back tone or music tone) [8]. This parameter measure the time which the end user takes to answer the call. It can indicate whether there is an issue with the communication. For instance, if there are several calls sent to a particular destination whose ring time is long enough for different numbers, it will mean that the communication is not good [9,19]. VoIP also can generate a fake ring back tone to the client which can lead to undesirable channel occupation.

- Call Duration: This parameter is defined by the time difference between the ACK message sent by the client, which follows the 200 OK message, and the BYE message sent by any of the peers letting the other one know that the call has ended [8]. This parameter can point out if there is a communication issue, e.g., when there are several calls sent to the same destination (different numbers) and their duration is very low (less than 5 or 3 seconds).
- Audio Quality: This parameter belongs to the human perception of the audio which comes as a result of call conversation. For the presented survey it was considered only when RTP packets were transmitted, i.e., when the call rang, did not ring, went to voicemail, did not present noise, was answered by a person, by an IVR or was being billed.
- Release Code: This parameter is taken into consideration to discard signaling problems in VoIP calls. According to [8] SIP protocol emits several messages to alert the client about the process of each call (ringing, no channel available, on hold, busy, etc.).

Fig. 2 represents a survey made at Estel Communications on which parameters are the most relevant for technicians (people with expertise in the area of VoIP call grade); the parameters PDD (post dial delay), Ring Time and Audio Quality, are the most relevant for them. Call Duration and Release Code were not that relevant but were also taken into consideration. In order to grade Audio Quality, NOC members need to establish whether the audio from the end user has voice or not. Since soft-computing can emulate the reasoning of a human, it is possible to emulate the decision made by NOC members. First a vector of features can be extracted for the VoIP calls. Then they can be classified by means of a pattern recognition algorithm. The latter is one of the several applications of Support Vector Machines (SVM) [20-22].

The classes used to classify VoIP calls are shown in Table 1.

3.5. Grade

Subsequent to parameter analysis, a grade for the destination of the carrier needs to be concluded. This is done by taking into consideration the classification of several calls sent to the gateway of the supplier, and

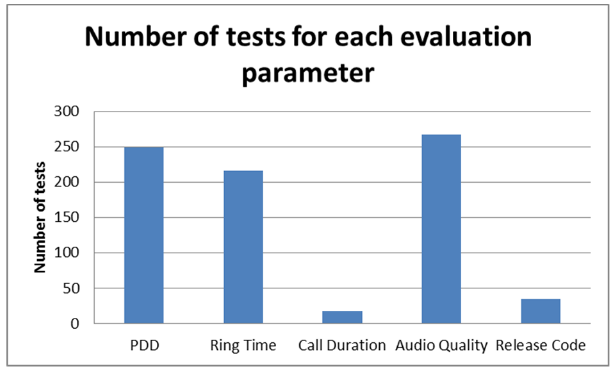

Figure 2. Survey of Estel Communications.
Source: The authors.

Table 1.
Output Classification Classes.

Class	Description
1	Call Rejected
2	Possible FAS
3	Call Not Connected with High RBT or High PDD
4	Call Not Connected with Low RBT and Low PDD
5	Call Connected with High RBT or High PDD
6	Call Connected with Low RBT and Low PDD

Source: The authors.

giving an opinion related to the pertinence of the supplier for traffic termination. There is not a precise number of calls that should be analyzed. This is decided by the person who performs the tests. When the amount of destinations or suppliers that need to be tested is low, five or six calls are sent by technicians to the gateway of the supplier. However, when the number is large, fewer calls are sent in order to accelerate the testing process and emission of results. This can be misleading since a low amount of calls might not represent the successful operation or failure of the destinations of the suppliers. Also as stated before, for statistics, the more the test samples, the more reliable the outcome.

4. Results and evaluation

The proposed methodology aims to compare itself with expert knowledge (knowledge acquired by technicians of VoIP carriers). A total of 2750 calls (with their corresponding audio when possible) were extracted from CheckMyRoutes platform. These calls were classified into each of the six classes of Table 1 by means of expert knowledge (people with expertise in the area of VoIP call grade). First, classes which did not involve audio were classified, based on time parameters (Class 1, Class 3 and Class 4). Then, classes that involved audio (Class 2, Class 4 and Class 5) were classified based on the analysis of the audio. Here the expert determined whether the audio from the end user had voice or not. Thus, comparing the classification given by the expert with that given by the CheckMyRoutes, it was possible to determine that the accuracy of CheckMyRoutes on identifying FAS was 68% and the accuracy on determining calls properly connected was 77%.

In order to emulate the human decision making for FAS detection, audio features were analyzed. Derived from this task, some parameters that served as inputs to a soft-computing classifier were established. Then, cross-validation was computed as a way to objectively conclude the accuracy of the proposed classifier. These tasks are described in the next subsections.

4.1. Audio analysis

VoIP carrier technicians are the ones who analyze the audio of the calls and determine whether it has voice or not. The standard low level features for audio analysis described in [23,24] can help to computationally solve this problem with little human intervention.

From the 2750 total calls 1468 have audio files. They are *.wav* files with 8 kHz sampling rate. They do not always have the same duration. In order to determine whether the audio file had voice or not, standard low level features were applied. Each file was segmented into 25ms audio frames (stationary audio) [24] and a 50% overlap was used [25]. Also low amplitude samples were replaced by zero since they seem to be related to background noise [16]. This was done by taking the mean of the average absolute amplitude of the first 25ms of each audio file, since in this time frame no voice should be present in a call (there is a delay in starting the call conversation and greeting). This resulted in replacing by zero the amplitudes according to the equation $|s(n)| < 0.004$, where $s(n)$ is the original digital audio signal.

The main concern in this part of the research was to differentiate ring back tones from voice. The variability of the standard low level features values resulted to be higher when voice was present in the audio file, since ring back tone signals are sinusoids. Therefore, second order statistics of the features could help in order to differentiate voiced from unvoiced signals [23].

A SVM for pattern recognition was used for classification, based on the 2750 samples and the second order statistics of the standard low level features. This SVM could also help to establish which parameters were the best suited for the classification task. Therefore, trial and error tests were implemented increasing the number of parameters until the highest possible accuracy was found. A computer running Linux (Fedora 17, 64 bits, Beefy Miracle distribution) with Intel Core i5-2410M CPU @ 2.30 GHz (4 cores) and 6 GB of RAM was used for simulation. The library LIBSVM [26] was used along with MATLAB for carrying out the experiments.

The input parameters for the system after correlation coefficients analysis (to discard any high linear dependency among parameters) are listed below:

1. PDD.
2. Ring Time.
3. Duration.
4. Standard deviation of the Short-time Average Energy.
5. Standard deviation of the Short-time Zero Crossing Rate.
6. Standard deviation of the Spectral Roll Off.
7. Standard deviation of the Short-time Average Energy (samples higher than zero).
8. Standard deviation of the Short-time Zero Crossing Rate (samples higher than zero).
9. Skewness of the Short-time Average Energy (samples higher than zero).
10. Skewness of the Short-time Zero Crossing Rate (samples higher than zero).
11. Skewness of the Spectral Roll Off (samples higher than zero).
12. Skewness of the Spectral Centroid (samples higher than zero).
13. Standard deviation of the derivative of Short-time Average Energy (samples higher than zero).
14. Standard deviation of the derivative of Short-time Zero Crossing Rate (samples higher than zero).
15. Standard deviation of the derivative of Spectral Roll Off (samples higher than zero).
16. Standard deviation of the Spectral Flux.

4.2. Support vector machine

It is known that in order to use soft-computing for classification, it is necessary to have similar amount of samples per class so the system can be trained well [27,28]. The amount of samples per each class are shown in Table 2. Prior to selecting SVM to perform the classification task, Artificial Neural Networks were considered. However, in order to avoid the risk of overtraining and time performance due to hidden layers and number of neurons [31], SVM were selected. A SVM was implemented to classify VoIP calls, based on the classes shown in Table 1 with sixteen parameters as inputs.

Two SVM approaches were validated using 10-fold Cross-validation [28]. First, a six class approach was validated for the classes presented in Table 1. Second, a two class approach was validated that confronted Class 2 (Possible FAS) against Class 5 (Call Connected with High PDD or High RBT) and Class 2 (Possible FAS) against Class 6 (Call Connected with Low PDD and Low RBT). The radial basis function (RBF) kernel was used for the SVM. This kernel has two parameters (C,γ) that are not known beforehand [29]. Therefore, a grid search was conducted in order to find the parameter values that outcome the best accuracy for the problem stated. Trying exponentially growing sequences of C and γ is a practical method to identify good parameters (for example, $C = 2^{-5}, 2^{-4}, 2^{-3}, \ldots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-14}, 2^{-13}, \ldots, 2^{3}$) [26].

Table 2.
Amount of Samples per Class

Class	Number of Samples
1	362
2	538
3	433
4	446
5	480
6	491

Source: The authors.

4.2.1. Six Class approach

A SVM was used for call classification of the six classes already presented in Table 1. Each call had 16 parameters. Those parameters were used for training and testing the SVM. The SVM output was the class to which each input call belonged. From the grid search the best cross-validation overall accuracy was 93.96% and the best C and γ were 2^{15} and 2^{-13} respectively.

The result of the 10-fold Cross-validation confusion matrix (combining the results of 100 tests matrices) is shown in Fig. 3. It can be seen that for most classes the accuracy was higher than 90%. Although FAS detection was 82.7% (target class 2), it increases the accuracy of CheckMyRoutes (68% presented in section 4). Also it increases the accuracy for classes of calls that connected properly. From this result, it can be said that the SVM is more trustworthy for the classification of properly connected calls. However FAS detection accuracy is better compared with CheckMyRoutes.

4.2.2. Two class approach

For this approach, the SVM was only used for classification between Class 2 (Possible FAS) and Class 5 (Call Connected with High PDD or High RBT) and between Class 2 (Possible FAS) and Class 6 (Call Connected with Low PDD and Low RBT). The decision for the implementation of this approach was based on the fact that a similar amount of calls was available for Class 2, Class 5 and Class 6. Therefore, if the classes 5 and 6 were combined to form one single class, there would be a mismatch in the number of samples since the new class would have about two times the amount of samples for Possible FAS. The user can establish a threshold for PDD and Ring Back Tone values. Each carrier can define what for them is a low or high PDD and a low or high Ring Back Tone. For each call the same 16 parameters were considered. Those parameters were used for training and testing the SVM. The SVM output classified the calls in Class 2, Class 5 or Class 6 depending on the user threshold.

Two grid searches were performed. The first one involved the SVM for classification of Class 2 and Class 5. The values of C and γ that gave the best overall 10-fold Cross-validation were 2^{4} and 2^{-8} respectively. The second grid search involved the SVM for classification of Class 2 and Class 6. The values of C and γ that gave the best overall 10-fold Cross-validation were 2^{4} and 2^{-8} respectively. Particularly for this case, the values of C and γ for these two different SVMs resulted to be the same.

The results of 10-fold Cross-validation are shown in Fig. 4 and Fig. 5. It can be seen that FAS detection for this approach increased the one given by the six class approach. Also it can be seen that the differentiation between Class 2 (Fig. 4 Class 1) and Class 5 (Fig.4 Class 2) is more difficult to discern by the SVM than the differentiation between Class 2 (Fig. 5 Class 1) and Class 6 (Fig. 5 Class 2) given that the accuracy is better when classifying Class 2 vs Class 6. This result may be related to an inner relationship that PDD, RBT or Duration of the calls may have with FAS, but it is not apparent.

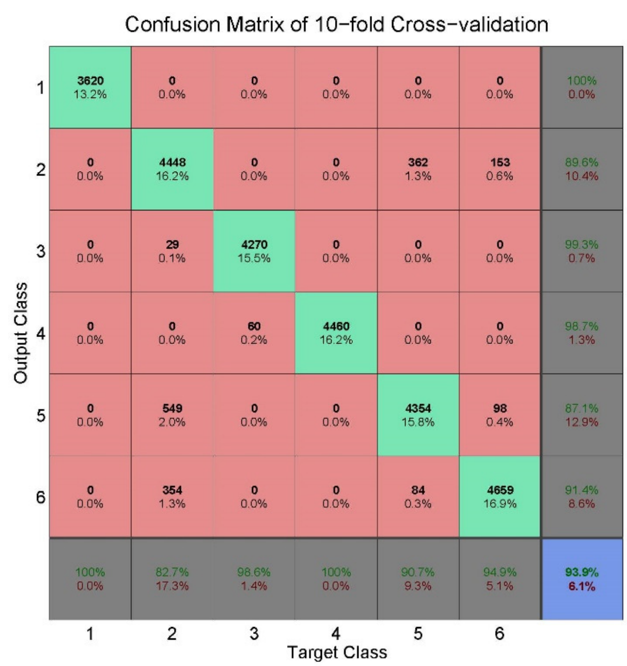

Figure 3. 10-fold Cross Validation Confusion Matrix.
Source: The authors.

5. Discussion

Table 3 shows the accuracy of the classifier compared with CheckMyRoutes for each of the SVM approaches already described. We can see that CheckMyRoutes classifies Classes 3 and 4 and Classes 5 and 6 as one.

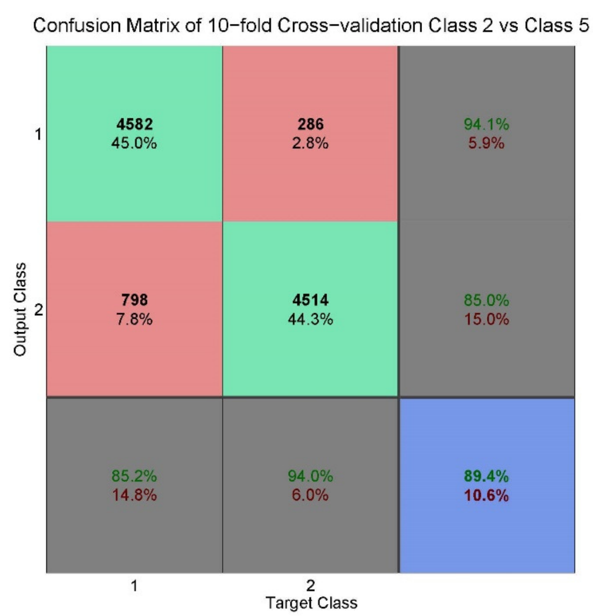

Figure 4. 10-fold Cross-validation Confusion Matrix (Class 2 vs Class 5).
Source: The authors.

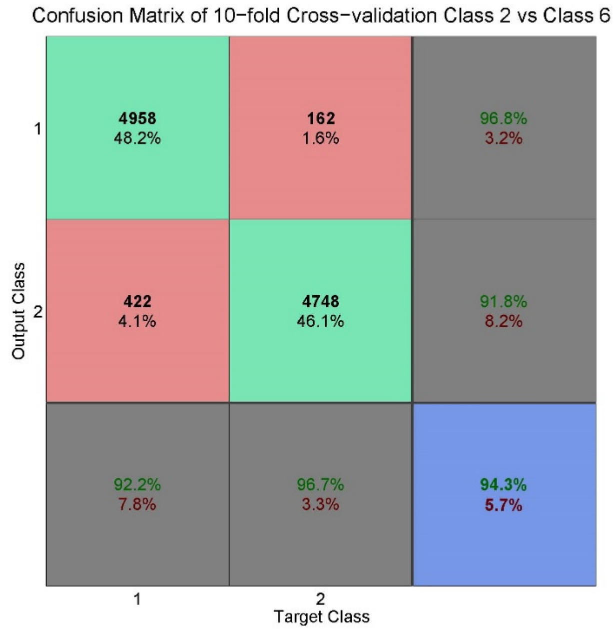

Figure 5. 10-fold Cross-validation Confusion Matrix (Class 2 vs Class 6).
Source: The authors.

Table 3.
Comparison of Accuracy of CheckMyRoutes and Proposed Approach.

Class	Description	CheckMy Routes	Six Class SVM	Two Class SVM
1	Call Rejected	100%	100%	100%
2	Possible FAS	68%	82.7%	88.6%
3, 4	Call Not Connected	100%	99.4%	100%
5, 6	Call Connected	77%	92.8%	95.4%

Source: The authors.

Therefore, in order to compare the accuracy of the proposed classifier, the combination of accuracies for classes 3 and 4 and classes 5 and 6 were computed. From the six class approach, it can be seen that the proposed SVM classifier is superior to that given by CheckMyRoutes. For the two class approach, the overall accuracy of the classifier was computed by the combination of the correct classified calls for Class 2 and the combination of the correct classified calls for Classes 5 and 6. It can be seen that the two class approach increases the overall accuracy of CheckMyRoutes and the six class approach as well.

Call grade using the proposed methodology also improves the efficiency of VoIP call testing, in terms of time, by using computational technologies to bypass human analysis. Supposing that 10 calls are sent to the destination carrier, a human takes around 10 minutes to perform this task. However, when using a computational technology that can dial the 10 calls simultaneously and analyze by itself the result of the calls, around two minutes are spent. Table 4 summarizes this comparison. The Parameter Extraction value is the mean of the time spent analyzing each one the 2750 total calls. The SVM analysis value was computed as the average time that an SVM takes to analyze each of the 2750 calls (16 parameters) times 10. Finally, the call grade was determined as the time MATLAB computes a weighted

Table 4.
Time Spent Testing and Analyzing (10 calls).

Human Testing	CheckMy Routes	Proposed	approach
≈ 10 min	≈ 1 min	Call dialing and storing	≈ 1 min
		Parameter extraction	≈ 0.6425 s
		SVM analysis	≈ 2.26e-04 s
		Grade	≈ 5.31e-03 s

Source: The authors.

mean. Now supposing that the commercial department of the VoIP carrier needs to test 30 destinations with the constraint of testing 10 calls per destination, using the same math, a technician would need 300 minutes to test for the best scenario (test numbers available, fast dialing, acute hearing, no pause between tests, etc.). However, CheckMyRoutes would take, for the same number of tests, at most four minutes supposing that it only supports testing 100 calls simultaneously. The proposed system would take at most five minutes. This result shows the gain in time when automating the testing.

Table 4 shows that the time spent by the proposed approach to provide a result is higher than the time spent by CheckMyRoutes. However, despite that the proposed approach takes more time in providing a grade for the carrier, it is not considerable (a few more seconds at most). Also the time that the system takes for analysis, can be really low compared to the time a technician would need to verify when CheckMyRoutes gives a wrong analysis (e.g. good classification for a destination that in reality presented FAS).

5. Conclusions and future works

In this work a SVM which performed the processing of the parameters for call connectivity grade was implemented. This SVM also provided a better accuracy value for classification of unknown samples compared with common used methodologies previously described. This accuracy could be objectively determined by means of cross-validation.

The proposed methodology increases the accuracy for FAS detection and properly connected calls detection compared to the automatic platform CheckMyRoutes. The usage of audio processing features along with Support Vector Machines helped with this improvement. From six class classification, the proposed methodology gave an 82.7% of accuracy for properly FAS detection. From two class classification the proposed classifier gave an 88.6% of accuracy for FAS. CheckMyRoutes gave 68% of accuracy for this type of calls. On the other hand, the six class approach gave a 93.8% when classifying properly connected calls. Two class approach gave 95.4% of accuracy for this case. CheckMyRoutes gave a 77% for properly connected calls. It can be concluded also, that despite the proposed approach takes more time than CheckMyRoutes to provide a result, the improvement in accuracy is worth the delay. Since CheckMyRoutes is not so trustworthy with its results, technicians need to step up in order to have certainty that a particular destination has FAS or is working properly. This human involvement is highly avoided with the proposed system since its results are more reliable. As shown in Table 4, the implementation of the proposed system reduces considerably the time taken for testing compared with human testing. This can become a long term economic savings for the carrier since the human resources can reduce their workload for such a tiresome task.

Future works should focus on the implementation of a different soft-computing technique such as decision trees or artificial neural networks. Also, there may be some other audio signal parameters that could help improve the accuracy of the current system for FAS detection. It is also recommendable to use more audio samples that may help the training of the soft-computing classifier. This system can also be implemented as a software platform and be tested with real VoIP traffic.

Acknowledgements

We gratefully would like to thank the Universidad del Norte and the Colombian Government's *Departamento Administrativo de Ciencia, Tecnología e Innovación* (COLCIENCIAS) for their support on this project. This work has been supported by COLCIENCIAS via the *Jóvenes Investigadores e Innovadores 2013* program.

References

[1] Cricelli, L., Di Pillo, F., Gastaldi, M. and Levialdi, N., Wholesale competition in the international telecommunications system. Networks and Spatial Economics, 5(3), pp. 261-277, 2005. DOI: 10.1007/s11067-005-3035-1

[2] Nucciarelli, A., Gastaldi, M. and Levialdi, N.G., Asymmetric competition and collection rates differentials: determinants of prices in international telephone service markets. NETNOMICS: Economic Research and Electronic Networking, 11(3), pp. 315-329, 2010. DOI: 10.1007/s11066-009-9042-5

[3] Beckert, S. and Mauldin, A., Telegeography international telecom trends seminar. Telegeography, PTC2009. 2009.

[4] Beckert, S. and Schoonover, E., Global traffic, bandwidth and pricing trends and wholesale market outlook. Telegeography, PTC07, 2007.

[5] Beckert, S., Stronge, T. and Schult, R., Telegeography international telecom trends seminar. Telegeography, PTC2010, 2010.

[6] CheckMyRoutes, CheckmyroutesTM, [online]. 2012. [accessed on June 2012]. Available at: http://www.checkmyroutes.com.

[7] Tata. False answer supervision. [online]. [accessed on July 2014]. Available at: http://www.tatacommunications.lk/providers/voice/false-answer-supervision. 2014.

[8] Collins, D., Carrier grade voice over IP. McGraw-Hill, 2001.

[9] Tektronix., Common VoIP service quality thresholds: Tektronix reference chart. [online]. 2007. [accessed on June 2012].

[10] De, B.S., Joshi, P.P., Sahdev, V. and Callahan, D., End-to-end voice over IP testing and the effect of QoS on signaling. On System Theory, 2003. Proceedings of the 35th Southeastern Symposium on. IEEE, pp. 142-147, 2003. DOI: 10.1109/SSST.2003.1194546

[11] Broß, J.F. and Meinel, C., Can VoIP live up to the QoS standards of traditional wireline telephony?. Telecommunications, 2008. AICT'08. Fourth Advanced International Conference on. IEEE, pp. 126-132, 2008.

[12] Voznak, M. and Rozhon, J., SIP back to back user benchmarking. wireless and mobile communications (ICWMC), 2010 6th International Conference on. IEEE, pp. 92-96, 2010.

[13] Subramonian, V., Cheung, E. and Karam, G., Automated testing of a converged conferencing application. Automation of Software Test, 2009. AST'09. ICSE Workshop on. IEEE, pp. 101-105, 2009. DOI: 10.1109/IWAST.2009.5069047

[14] Prasad, R.V., Sangwan, A., Jamadagni, H.S., Chiranth, M.C., Sah, R. and Gaurav, V., Comparison of voice activity detection algorithms for VoIP. Computers and Communications, 2002. Proceedings. ISCC Seventh International Symposium on. IEEE, 2002. pp. 530-535, 2002.

[15] Yang, X., Tan, B., Ding, J., Zhang, J. and Gong, J., Comparative study on voice activity detection algorithm. In: Electrical and Control Engineering (ICECE), 2010 International Conference on. IEEE, pp. 599-602, 2010. DOI: 10.1109/iCECE.2010.153

[16] Nica, A.C., Automatic Silence/Unvoiced/Voiced classification of speech using a modified teager energy feature. Energy (RMSE), [online]. 4(1), pp. 2. Available at: http://dl.acm.org/citation.cfm?id=1981459.1981470.

[17] Ohrtman, F., Softswitch: Architecture for VoIP. McGraw-Hill Professional, 2007.

[18] Conway, A.E., A passive method for monitoring voice-over-IP call quality with ITU-T objective speech quality measurement methods. In: Communications, 2002. ICC 2002. IEEE International Conference on. IEEE, pp. 2583-2586, 2002. DOI: 10.1109/ICC.2002.997309

[19] Minacom., Common voip service quality thresholds: Minacom reference chart. [online]. 2006. [accessed on June 2012]. Available at: http://www.teracomtraining.com/ref/voip-service-quality-metrics-and-thresholds.pdf

[20] Shukla, A., Tiwari, R. and Kala, R., Real life applications of soft computing. CRC Press, 2010. DOI: 10.1201/EBK1439822876

[21] Boswell, D., Introduction to support vector machines: University of Carlifornia. San Diego, [online]. [accessed on June 2012]. Available at: http://dustwell.com/PastWork/IntroToSVM.pdf. 2002.

[22] Ivanciuc, O., Applications of support vector machines in chemistry. Reviews in Computational Chemistry, 23, 291 P., 2007.

[23] McKinney, M.F. and Breebaart, J., Features for audio and music classification. In: ISMIR. pp. 151-158, 2003.

[24] Lu, L., Zhang, H.J. and Li, S.Z., Content-based audio classification and segmentation by using support vector machines. Multimedia systems, 8(6), pp. 482-492, 2003. DOI: 10.1007/s00530-002-0065-0

[25] De Poli, G. and Mion, L., Algorithms for sound and music computing. Creative Commons, 2006.

[26] Hsu, C.W., Chang, C.C. and Lin, C.J., A practical guide to support vector classification. [online]. 2003. Available at: http://www.csie.ntu.edu.tw/~cjlin/papers.htm.

[27] Konar, A., Artificial intelligence and soft computing: Behavioral and cognitive modeling of the human brain. CRC Press, 1999. DOI: 10.1201/9781420049138

[28] Witten, I.H. and Frank, E., Data mining: Practical machine learning tools and techniques. 3rd ed. Morgan Kaufmann, 2011.

[29] Chang, C. and Lin, C., LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27 P., 2011.

[30] Ramırez, J., Yélamos, P., Górriz, J.M., Segura, J.C. and Garcıa, L., Speech/non-speech discrimination combining advanced feature extraction and SVM learning. Proceedings of Interspeech 2006, 2006.

[31] Gazzah, S. and Ben, N., Neural networks and support vector machines classifiers for writer identification using arabic script. Int. Arab J. Inf. Technol., [online]. 5(1), pp. 92-101, 2008. Available at: http://dblp.uni-trier.de/db/journals/iajit/iajit5.html#GazzahA08.

[32] De Pessemier, T., Stevens, I., De Marez, L., Martens, L. and Joseph, W., Analysis of the quality of experience of a commercial voice-over-IP service. Multimedia Tools and Applications, 74(15), pp. 5873-5895, 2015. DOI: 10.1007/s11042-014-1895-4

[33] Ganjam, A., Sekar, V. and Zhang, H., In-Situ quality of experience monitoring: The case for prioritizing coverage over fidelity. White-paper submission, CMU/Conviva, 2., 2016.

[34] Papadopouli, M., On user-centric QoE prediction for VoIP and video based on machine-learning. White-paper submission, U. of Crete/ICS-FORTH, 2, 2015.

[35] Zhang, W., Chang, Y., Liu, Y. and Tian, Y., Perceived QoS assessment for Voip networks. In: Communication Technology (ICCT), 15th IEEE International Conference, pp. 707-711, 2013.

[36] De Pessemier, T., Stevens, I., De Marez, L., Martens, L. and Joseph, W., Quality assessment and usage behavior of a mobile voice-over-IP service. Telecommunication Systems, 61(3), pp. 417-432, 2016. DOI: 10.1007/s11235-014-9961-9

[37] Jelassi, S., Rubino, G., Melvin, H., Youssef, H. and Pujolle, G., Quality of experience of VoIP service: A survey of assessment approaches and open issues. IEEE Communications Surveys & Tutorials,14(2), pp. 491-513, 2012. DOI: 10.1109/SURV.2011.120811.00063
[38] Andrews, M., Quality of experience in wireless pricing: Some research directions. White-paper submission, Alcatel-Lucent Bell Labs, 4., 2015.
[39] Carvalho, L.S.G.D. and Mota, E.D.S., Survey on application-layer mechanisms for speech quality adaptation in VoIP. ACM Computing Surveys (CSUR), 45(3), 36 P., 2013.
[40] Singh, H.P., Singh, S., Singh, J. and Khan, S.A., VoIP: State of art for global connectivity—A critical review. Journal of Network and Computer Applications, 37, pp. 365-379, 2014. DOI: 10.1016/j.jnca.2013.02.026
[41] International Telecommunication Union. Definitions of terms related to quality of service (QoS). ITU-T Rec. E.800. September, 2008.
[42] ITU, G.107: The e-model: A computational model for use in transmission planning.
[43] Reichl, P., Egger, S., Schatz, R. and D'Alconzo, A., The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment, in: Communications (ICC), 2010 IEEE International Conference on. IEEE, pp. 1-5, 2010. DOI: 10.1109/ICC.2010.5501894
[44] ITU, P.862: Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.

J.R. Wilches-Cortina, received the BSc. degree in Electronics Engineering from Universidad del Norte Barranquilla, Colombia in 2009, and the MSc. degree in Electronics Engineering in 2015 from the same university. He is currently customer experience management engineer at Claro Colombia. His current research interests are business intelligence and modeling of customer experience in telecommunications carriers.
Orcid: 0000-0003-4483-3947.

J.A. Cardona-Peña, received the BSc. degree in Electronic Engineering from Universidad del Cauca Colombia, Popayán, Colombia in 1993, and the MSc. degree in Engineering - Telematics in 2005 from the same university. He is assistant professor in Electronic and Electric Engineering department at Universidad del Norte. His current research interests Networking, SDN and telecommunications.
ORCID: 0000-0001-8406-3014.

J.P. Tello-Portillo, received the BSc. degree in Electronic Engineering from Universidad Nacional de Colombia, Manizales, Colombia in 2000, and the MSc. degree in Engineering - Industry Automation in 2006 from the same university. He is assistant professor in Electronic and Electric Engineering department at Universidad del Norte. His current research interests are image compressing, signal acquisition, processing applied to the biomedical area and telecommunications.
Orcid: 0000-0002-0503-1215.