IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 24, 2018, accepted September 8, 2018, date of publication October 29, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871072

Speech Quality Assessment in Wireless VoIP Communication Using Deep Belief Network

EMMANUEL T. AFFONSO[1], RODRIGO D. NUNES[1], RENATA L. ROSA[1], GABRIEL F. PIVARO[2], AND DEMÓSTENES Z. RODRÍGUEZ[1], (Senior Member, IEEE)

[1]Federal University of Lavras, Lavras 37200-000, Brazil
[2]National Institute of Telecommunications, Brazil

Corresponding author: Renata L. Rosa (renata.rosa@dcc.ufla.br)

This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo under Grants 2015/25512-0 and 2015/24496-0, in part by the Federal University of Lavras, and in part by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais.

ABSTRACT Nowadays, the voice over Internet protocol (VoIP) communication service is widely adopted, and it counts with many users across the world. However, the users' quality of experience is not guaranteed because the voice signal quality can be affected by several degradations that happen in the network infrastructure. Thus, it is relevant to have a global speech quality assessment method that considers both wired and wireless networks to provide reliable results. In this paper, several network scenarios that consider different packet loss rates (PLRs) and wireless channel models are implemented in which the impaired signals are evaluated using the algorithm described in ITU-T Recommendation P.862. Preliminary results showed a relationship between both fading and PLR parameters and the global speech quality index. However, the P.862 algorithm is not viable in real VoIP scenarios. The ITU-T Recommendation P.563 describes a non-intrusive speech quality assessment method; nevertheless, its results are not confident. In this context, the main objective of this paper is to propose a non-intrusive speech quality classification model based on a deep belief network (DBN) that considers the wired and wireless impairments on the speech signal. Experimental results demonstrated a high correlation between the proposed model based on the DBN and P.862 algorithm, reaching a F-measure of 97.01%. For validation, the non-intrusive P.563 algorithm is used; the proposed model and P.563 reached an average accuracy of 96.14% and 72.12%, respectively. Furthermore, subjective tests were carried out, and the proposed DBN model reached an accuracy of 94%.

INDEX TERMS Speech quality, degradation, wireless network, wired network, ITU-T P.862, MOS, deep neural networks.

I. INTRODUCTION

The traffic amount of the Voice over Internet Protocol (VoIP) service has increased over the past years due to several factors, such as the network capacity, the enhanced digital processing techniques for voice signals [1], the high number of mobile devices, and the lower cost for VoIP calls. These factors turn VoIP a very advantageous and attractive service, specially for commercial uses. According to [2], the VoIP service will be responsible for 21% of the total mobile voice traffic by the end of 2020.

Currently, the communication systems to support an end-to-end VoIP call use wired and wireless network infrastructures. Therefore, the speech signal can be affected by various degradation factors, which impact on the speech quality at the reception point. The quality degradation in a wired network can occur due eventual packet losses, which may be associated with overloaded routers and other network infrastructure problems [3]. In wireless systems, other degradation factors may occur, such as, obstacles between the transmitter and receiver, reflecting agents, signal amplitude variations, among others [4].

Two popular wireless channel models are the Rayleigh and Rician fading channels [5]–[7]. An important characteristic associated to a channel model is the Doppler shifts that is related to the channel frequency-variation in a radio communication, in which the transmitter or the receiver are moving relative to each other [8].

In order to improve their annual incoming, communication service operators need to increase the number of subscribers. In this sense, a mechanism of speech quality assessment of a communication service is necessary [9]. This fact encourages the service providers to search fast and inexpensive methods

to accomplish this goal; and computational methods have these characteristics.

There are different speech quality assessment methods, most of them described in International Telecommunication Union (ITU) recommendations. In general, these methods can be classified into subjective and objective. Subjective tests are carried out in a laboratory environment. Their result, known as Mean Opinion Score (MOS), are the most confident, but the test conduction is expensive and time-consuming. On the other hand, the objective methods use an algorithm to predict a MOS index value.

The objective models, based on speech signals, are subdivided in intrusive and non-intrusive methods [10]. Algorithms that only use a signal correspond to non-intrusive method; and algorithms that use both reference and impaired signal are considered as intrusive method. The Perceptual Assessment of Speech Quality (PESQ) described in ITU-T Recommendation P.862 [11] and the Perceptual Objective Listening Quality Assessment (POLQA) [12] algorithm are intrusive methods. The ITU-T Recommendation P.563 is the most representative non-intrusive objective metric, but it does not have a satisfactory performance [13], [14] in lossy networks.

In recent years, solutions based on Deep Neural Networks (DNN) have been applied in different solutions that involve speech signals, such as, medical applications [15], speech recognition [16], speech classification [17], packet loss concealment method [18], among others. In those solutions, different speech signal parameters are extracted from speech samples, and this information is used on the DNN. However, these studies, based on DNN, do not perform speech quality assessment considering the degradations caused by network impairments; specifically, degradations originated by fading in a wireless network. In this arena, the main contribution of this work is to propose a non-intrusive speech quality classifier model with a high accuracy, based on a DNN architecture that uses five speech quality classes. For this, a large speech database is built to be used as test material, in which different packet loss rate (PLR) values and fading channel parameters are applied over original speech files. Thus, a DNN architecture compounded by a Deep Belief Network (DBN) with a Softmax function is proposed, which extracts automatically the features of the impaired signal to perform a speech quality assessment. It is worth noting that the proposed model estimates a speech quality class and it does not determine a MOS index value; the quality classes are based on the 5-point MOS scale introduced in the Absolute Category Rating (ACR) methodology [19]. For validation purposes, the non-intrusive P.563 algorithm is also used. The results obtained by the proposed speech quality assessment model outperformed the P.563 results; in this case, the P.862 results are used as ground-truth. Furthermore, to validate the performance of the proposed model in more realistic manner, subjective tests are also performed and the results are compared with the classification results generated by the DBN.

TABLE 1. 5-point MOS scale – absolute category rating.

Speech Quality	MOS Index Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

The remainder of this paper is structured as follows. Section II introduces some concepts and related works about speech quality assessment, deep neural networks and degradations, such as packet loss rate and fading channel models. Section III presents the test methodology used in this work that includes the construction of a speech data base, the proposed model and subjective test of speech quality assessment. The results are presented in Section V and then, some discussions are described in Section VI. Finally, the conclusions are presented in Section VII.

II. RELATED WORK

In this section, studies and concepts regarding the speech quality assessment methods, DNN and degradation factors in the speech signal quality are treated.

A. SPEECH QUALITY ASSESSMENT METHODS

As stated before, speech quality assessment methods can be classified in two groups, subjective and objective methods.

The subjective tests are conducted in a laboratory environment and following a specific procedure [19], in which an assessor indicates the perceived speech quality. At the end of the tests, the average result, known as MOS, is determined. The ITU-T Recommendation P.800 [19] introduces different test methodologies to perform subjective evaluation of speech quality in telephony services, they are: conversation-opinion test, listening-opinion tests, and interview and survey tests. These methods are applicable for any type of degradation, such as, error transmission, circuit and environmental noise, distortion arising from packet switching, codecs distortions, among others.

In this work, we use the ACR methodology that is used for listening-opinion tests; this method is well-established and uses the 5-point MOS listening-quality scale, presented in Table 1. The speech signals, used as test material, are simple and short sentences that were impaired by different degradation types. Also, it is recommendable that assessors have not participated in similar tests for at least the previous six months, and do not know the sentences used in the tests.

Objective models try to predict a MOS index value using and algorithm. According to the kind of the algorithm input, objective methods are classified mainly in three models, based signal, parametric and hybrid models [20].

Models based on speech signal are classified in intrusive and non-intrusive methods. The P.862 recommendation, is the most popular intrusive method for narrow band signals [21]–[23]. PESQ compares an original signal $X(t)$

with a degraded signal $Y(t)$ that is the result of passing $X(t)$ through a communication system. The output of PESQ is a prediction of the perceived quality, a Listening Quality MOS-like score, that would be given to $Y(t)$ by subjects in a subjective listening test. In the PESQ, the quality scale from 0.5 to 4.5 is used. It is important to note that the ITU recommendation P.862.1 [24] describes a mathematical relation to approximate the PESQ output to the quality scale described in Table 1. The index value obtained by P.862.1 is known as MOS-Listening Quality Objective (MOS-LQO). The PESQ has a good performance in several scenarios, which contain degradation factors, such as errors distortions, channel codification, packet loss, delay variation effect, among others [25]. In 2014, the ITU-T Rec. P.863 was launched as a evolution of PESQ. The P.863 algorithm incorporates current industry requirements, and in particular it allows the quality assessment of narrow-band to super-wideband speech signals [8]. It is worth noting that a license is necessary to use the P.863 algorithm implementation.

Currently, the P.563 recommendation is the most accepted standardized objective non-intrusive metric. The quality score predicted by the P.563 algorithm is related to the perceived quality of a speech signal at the reception point. Basically, the P.563 algorithm works identifying the main distortion class of the degraded signal, and after applying a speech quality model, it returns the MOS index value expressed in the same quality scale presented in Table 1. Nowadays, a new ITU-T standard for non-intrusive speech quality assessment for wideband and super wideband is in development [26].

The most representative standardized parametric model is the ITU-T Rec. G.107 [27], which introduced the E-model algorithm that can be useful to telecommunication network planners. The quality score of the E-model is the rating factor R that uses a 100-point scale for narrowband signals that has a correlation with the scale presented in Table 1. The ITU-T Rec. G.107.1 [28] describes the wideband E-model algorithm that is used in different studies [29]. However, it is important to note that E-model algorithm only considers wired network parameters.

Finally, hybrid models use as algorithm input, the speech signal and network parameters to determine a MOS index [30].

B. DEEP NEURAL NETWORK

The human brain is a biological model with a high processing power. Understand its functionality and create algorithms to represent a biological neural network, by a conventional computer, has been a hard task [31]. The brain is a non-linear and parallel mechanism, with the ability to organize its structural constituents, the neurons, in order to perform faster processing than the existing fastest digital computer [32].

The Artificial Neural Networks (ANN) are mathematical models that resemble biological neural structures and they have the computational capacity acquired through learning and generalization [33]. The ANN, seen as adaptive machine, is defined as a distributed parallel processing [32] composed of simple processing units, that have natural propensity to store experimental knowledge.

The ANN architecture is defined by the number of layers, the number of neurons by layer and connections between the neurons build the network topologies [32], [34]. There are many models for the implementation of a structure of ANN, such as the Self-organizing Map (SOM), Radius Basis Function (RBF), Least Mean Square (LMS), Multi-Layer Perceptron (MPL), among others [32], [33].

The DNN is an ANN and it has been applied in many areas in the recent years, such as in images [35], [36] and speech signal studies. A DNN algorithm is composed of multiple processing layers to learn data representations through multiple abstraction levels. It is important to note that the DNN algorithms have obtained excellent results to classify speech patterns [37], [38].

A DNN is a feed-forward ANN, which presents more than one layer of hidden units between its inputs and outputs. The DNN can be trained by back-propagating derivatives using a cost function, which measures the discrepancy between target and actual outputs, that are produced in each training case [39]. A DNN presents hidden layers and many units per layer which makes the DNN capable of modeling complex and non-linear relationships between inputs and outputs [39]; such ability is very important to determine high-quality acoustic models [40].

In a DNN, the relation between the input feature ($\boldsymbol{t}$) and output of the first hidden layer h_1 is given by

$$\boldsymbol{h}_1 = H\left(\boldsymbol{W}_1\boldsymbol{t} + \boldsymbol{i}_1\right), \tag{1}$$

where $\boldsymbol{W}_1$ and $\boldsymbol{i}_1$ correspond to weight matrix and bias vector, respectively, in the first layer. $H(\cdot)$ represents an activation function, which determines the neurons' output.

The relation between the current and next hidden layer is expressed by

$$\boldsymbol{h}_n = H\left(\boldsymbol{W}_n\boldsymbol{h}_{n-1} + \boldsymbol{i}_n\right), \quad n = 2, \ldots, N \tag{2}$$

In the DNN, N represents the total number of layers.

In order to execute classifications or regressions, $H(\cdot)$ is applied on the output layer.

$$\hat{\boldsymbol{y}} = H\left(\boldsymbol{h}_N\right) \tag{3}$$

where $\hat{\boldsymbol{y}}$ represents the DNN output. The Softmax function [41], a multinomial logistic regression, can be used in classification tasks, forming a powerful classification method [42] to recognize patterns in a speech signal.

In the speech signal processing, in order to estimate with more accuracy the DNN parameters, a sufficient training data is necessary. Otherwise, a pre-training process can be performed by a DBN to work the training data limitation [43], [44]. A set of Restricted Boltzmann Machine (RBM) models forms a DBN model. The DBN has been widely used as generative models in studies about speech signals [45], [46].

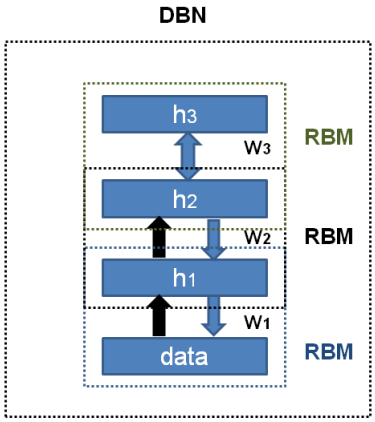


FIGURE 1. DBN Architecture formed by RBMs, indicating the weight W_i of the hidden unit h_i.

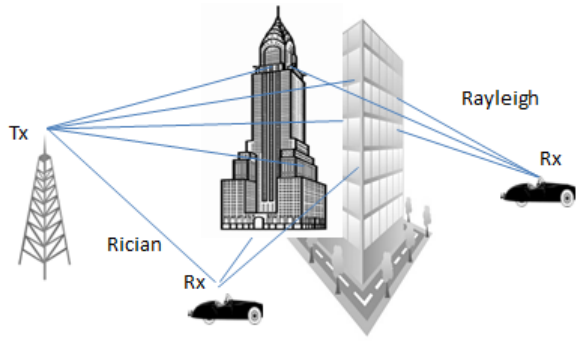


FIGURE 2. Wireless communication system with LOS and NLOS between the transmitter and receiver, representing the Rician and Rayleigh fading channel models, respectively.

Fig. 1 presents the DBN formed by RBMs, with the weight W_i of the hidden unit h_i.

Although the DNN algorithms has been used in several speech applications, there is a scarce number of studies regarding the analysis of speech quality in an environment with the presence of wireless network degradation, such as fading, which commonly occur in a communication system. In [18], a packet loss concealment method using DNN is proposed, but the speech quality is not treated. Affonso *et al.* [47] introduce a speech quality classification method that only considers wired network impairments, such as PLR. Monika and Rama [48] use Neural Network with Hidden Markov Model in speech transmission, but they do not consider the network parameters effects in the communication. In [49] a speech enhancement technique to improve noise corrupted speech via DNN is proposed, but degradations that correspond to communication systems are not considered. Martin-Donas *et al.* [50] propose a speech enhancement DNN-based solution for smartphones. Xie *et al.* [15] evaluate pathological speech quality using acoustical parameters, extracted from speech samples, through a DNN. Bhamre and Kulkarni [51] apply an DNN to map the relationship between a noise and a reference speech signal, considering different acoustical environments.

C. DEGRADATION FACTORS: PACKET LOSS RATE AND FADING CHANNEL MODELS

As stated before, the degradation of the speech signal quality can occur in any stage of a communication system.

In the wired network, the most common impairment factor is the PLR [52], which can occur due to overloaded transmission channels. A PLR distribution can be determining using the Gilbert-Elliot model [53] that is represented by:

$$a = P(q_t = B|q_{t-1} = G) \quad b = P(q_t = G|q_{t-1} = B) \tag{4}$$

where, a is the probability to pass from a bad state (B) that represents a packet loss to a Good state (G) that indicates a success in the packet delivery; b is the probability to pass from G to B state; and q_t and q_{t-1} represent the states at the instants t and $t-1$, respectively. It is important to note that a same PLR value can be implemented using different packet losses distributions. This manner, different burst intensities are configured [47]. A PLR distribution model is determined with the variation of a and b, as follow:

$$PLR = \frac{a}{a+b} \tag{5}$$

On the other hand, in wireless networks other speech quality impairment factors are found. In a real communication scenario, usually the end-points of a communication are in different places and with many obstacles between them, as in the cities, where the user-device can be inside of a establishment or between buildings. Thus, it is probably that there is no line of sight (LOS) between the sender and the receiver; the communication is established by the mirroring of the waves or by diffraction around the objects [8]. To simulate a real wireless channel there are different channel models. In this work, two channel models, which are very accepted in the literature, are used. The Rayleigh fading channel model, in which there is no LOS, and the signal spreads among the various obstacles till reach the receiver. In Rician model, there is a LOS among sender and receiver, but also different phases and amplitudes of the signal arrive at the receiver. Fig. 2 shows a typical urban scenario where there is a LOS and NLOS situations that are described by Rician and Rayleigh fading models, respectively.

The fading can occur due to several conditions [4], [54]–[56], such as the presence of obstacles in the signal path, various paths formed by reflectors, channel frequency-variation, among others. Some physical phenomena that occur in wireless communication are multipath, reflection, diffraction, mirroring and absorption. These phenomena lead to fading in the signal traveling through the air and can be divided into two types: the large and small scale [8], [57], [58].

Large-scale fading occurs in longer distances from the transmitter. The power of the received signal decreases as the distance increases due to the path loss or obstacles in the path, such as buildings, vegetation and mountains. Atmospheric

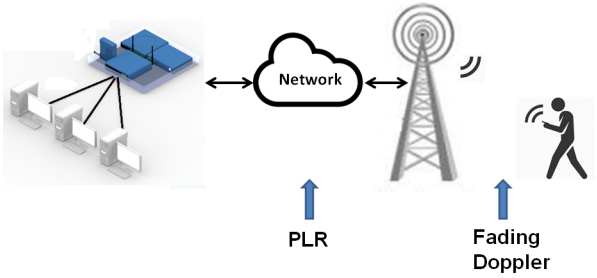


FIGURE 3. Experimental test scenario for voice quality evaluation, representing degradation in the wired and wireless network.

phenomena such as rain, snow, hail also contribute to the fading; the shadowing factor is also another known problem.

In small-scale fading scenario, the signal suffers variation in its amplitude and frequency [59]. Two phenomena that are related to the fading are the Doppler and the multipath. In multipath effect, the signal travels by several paths between the transmitter and the receiver due to reflections and refractions when it encounters obstacles. This phenomenon generates the reception of several versions of the signal, presenting different amplitudes and phases.

On the other side, the Doppler effect occurs due to the relative movement between the transmitter and receiver, suffering a frequency variation of the signal received [8], [60]. The Doppler shift can be calculated according to:

$$fq' = fq\left(\frac{V \pm V_o}{V \pm V_s}\right) \tag{6}$$

where, fq' represents the expected frequency, fq is the emitted frequency, v is the constant of propagation speed, V_o and V_s represent the receiver and transmitter speed, respectively.

III. TEST METHODOLOGY

In this work, the degraded speech signals consider impairments caused by wired and wireless network. Fig. 3 illustrates the proposed scenario for the realization of the experiments.

In the wired network, different PLR values were applied on the original speech files; for each file and each PLR, the tests were repeated 10 times. Then, each impairment file was evaluated by PESQ and a MOS index was calculated. Considering the probabilistic nature of the PLR formulation, in which a lesser PLR value can obtain a higher MOS score value and vice-versa; the impairment file with the MOS index closer to the average, for each PLR value, was separated and used in the subsequent phase.

The Wav2rtp Open Source software [61] was used for application of different PLR values and distribution models according to equations (4) and (5). This software is able to convert a ".wav" file into a RTP data stream, generating a new degraded file.

In the wireless network, the fading effect was performed considering the Rayleigh and Rician channel models, implemented in MATLAB software, in which different Doppler shifts are considered. The channel model with a specific configuration was applied in each resulting files with PLR, and the experiments were also repeated 10 times. Later, the algorithm ITU-T P.862 was applied in each speech file and its MOS index was calculated. Thus, different degradation scenarios with PLR and fading channel models were created.

Consequently, with the knowledge of resulting MOS for each degradation scenario, an DNN can be trained to discover the relation between the fading and packet loss parameters and the voice quality index.

TABLE 2. Characteristics of original speech files.

Speech Database	Min. / Max. Length	Average silence
ITU-T Rec. P. Sup. 23 [63]	7 - 10 s.	45%
ITU-T Rec. P.862 [62]	8 - 10 s.	41%

A. CONSTRUCTION OF A SPEECH DATABASE CONSIDERING PLR AND FADING

In order to determine a speech quality classification model, an impaired speech database (DB) was built. It is important to note that a DB that considers degradations caused by fading effects is not available in the current literature.

40 unimpaired speech files were considered to create the speech samples to be used as test material; 20 files from [62] and the remaining from [63]. All the speech samples have a maximum and a minimum percentage of silence of 80% and 20%, respectively, as recommended in [11]; and the sampling rate considered for all the speech samples is 8 kHz, because our study is regarding narrow-band signals. The main characteristics of the original speech files are presented in Table 2.

The objective is to create speech samples containing wired and wireless network degradations. Thus, the process to obtain these samples can be divided in three phases:

- In the wired network simulation, the PLR values of 0.5%, 1%, 3%, 5%, 7%, 10%, 15% and 20% were applied in the 40 speech original files. As stated before, the application of PLR was repeated 10 times for the different rates in each file, resulting in 3,200 impairment files, in which 80 different versions of the same file were obtained. The MOS index of each impaired file was obtained using the ITU-T P.862 algorithm. Then, an average value for each 10 repetitions was determined, and the speech file with the MOS index closest to each average value were separated. Thus, 320 files were selected for the next step.
- Later, in the wireless network simulation, the Rayleigh and Rician channel models were used, and configured using 10 different Doppler shifts (Hz): 0, 5, 10, 15, 20, 50, 75, 100, 150 and 200. The 320 files, obtained from the first phase, passed trough to each channel model and Doppler shift, and each simulation was repeated ten times, resulting in total 64, 000 impaired files. Thus, each speech file contain degradation caused by PLR and fading.

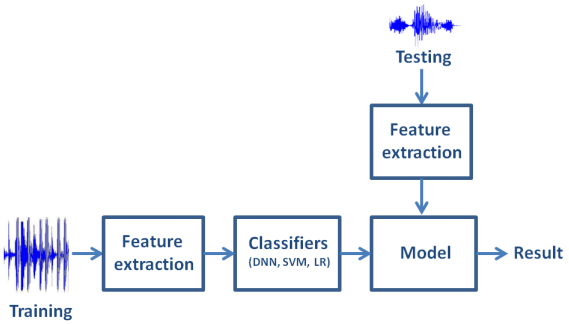


FIGURE 4. Flowchart for the Speech Quality Assessment Classification by Deep Belief Networks.

- Finally, a MOS index of each impaired file was determined using the PESQ algorithm

Thus, the impairment is related to the PLR and fading channel model with several Doppler values. All these speech files were used in the training and testing phases of the proposed model based on DBN.

B. PROPOSED MODEL USING A DEEP BELIEF NETWORK

Fig. 6 presents the flowchart for the speech quality assessment classification.

The speech signal features were automatically extracted by the proposed DBN classifier. These features correspond to the Zero-crossing rate (ZCR) parameter in the temporal domain [64], 13 Mel-Frequency Cepstrum Coefficients (MFCC) static features (12 MFCCs and log energy) and the first and second derivatives of the static features [65], twenty FFT Power Spectrum, the spectral centroid, the spectral roll-off and the spectral flux. In total, 63 features from the speech signal over 25 ms of frames with 10 ms overlap are extracted.

Once the features were extracted, the DBN is trained and different classifiers are tested. Linear Regression (LR) and Support Vector Machine (SVM) classifiers were used with the aim of performance comparison with the DBN containing the Softmax function. The LR model was used because it is a classification model very used to maximize the conditional log-likelihood [66]. The SVM classifier was used because it represents, in a good manner, how to separate different classes in the training step [67]. In the training phase the DBN and the LR and SVM classifiers are considered only for performance evaluation.

In order to obtain the final DBN model, the correlation between the classification performance and different DNN architecture configuration were tested in the training phase. The variables were the number of hidden layers and neurons in each hidden layer. From one to six hidden layers, with each hidden layer containing 50, 100, 150, 200, 250 and 300 neurons were tested.

The contrastive divergence (CD) algorithm was used to estimate the parameters in the RBM, which is an efficient approximate training procedure, making it suitable as building blocks for learning DBNs [68].

TABLE 3. Speech quality classes according to MOS index values.

Speech Quality Class	Perceived quality (ACR Scale)	MOS index values
Class-A	Excellent	5.00-4.00
Class-B	Good	3.99-3.00
Class-C	Fair	2.99-2.00
Class-D	Poor	1.99-1.00
Class-E	Very Poor	0.99 and lower

The parameters of the first RBM were used for estimating a DBN model, which were estimated by using a training data. A learning algorithm trained other RBM by using as the input data the activation of the previous layer. This mechanism continued until reached the last layer of RBM, forming the DBN. The Softmax function was used in the top of the DBN model, and the back-propagation was applied for estimating the parameters of the model.

The result obtained by the proposed DBN model corresponds to a speech quality class. Thus, five quality classes are defined according to the 5-point MOS scale, which are introduced in Table 3.

In the training phase, the precision, recall, and F-measure were used as performance assessment metrics to compare the results obtained by our proposed model with other classification methods. These performance metrics are very used in classification systems of different applications [69]–[71].

The accuracy is also a performance assessment metric and it was used in our experiments to compare the results obtained by different network configurations and different classifier algorithms. The accuracy is given by (7):

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}, \tag{7}$$

where, T_p and T_n represent the number of true positive and true negative classification in the data sample, respectively. Conversely, F_p and F_n represent the number of false positive and false negative classification, respectively.

Additionally, the confusion matrix was used to compare the results obtained by the proposed solution and the P.563 algorithms. The MOS scores obtained by P.563 were grouped into a specific quality class described in Table 3.

C. SUBJECTIVE TESTS IN A LABORATORY ENVIRONMENT

Once the proposed model was defined, its performance was evaluated by subjective tests carried out in a laboratory environment that accomplished the acoustic requirements described in [19]. 63 assessors participated of the subjective listening tests, consisting of 29 women and 34 men; they did not have experience in speech quality assessment tests, and they also reported no hearing problems.

The tests were conducted for 11 weeks, and during this period, the test room was with the same acoustic characteristics, without noise or disturbance sound. At the beginning of the subjective tests, a supervisor explained the instructions to the assessors, who listened some degraded speech sequences for a better understanding of the test procedure;

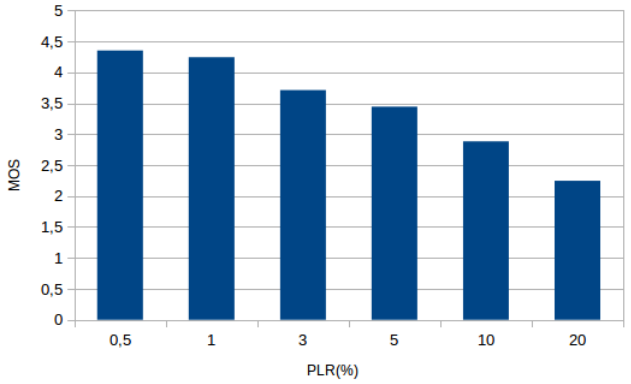


FIGURE 5. Impact of different PLR values on speech quality index (MOS).

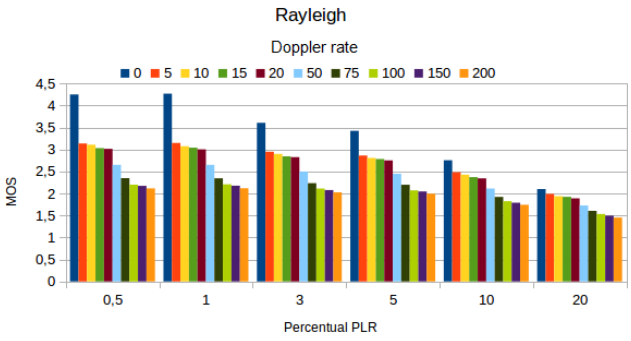


FIGURE 6. Impact of Rayleigh fading channel with different Doppler Shift values on speech quality index (MOS).

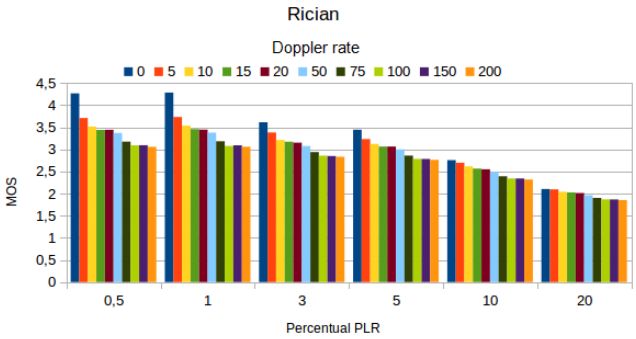


FIGURE 7. Impact of Rician fading channel with different Doppler Shift values on speech quality index (MOS).

those speech sequences were not used in the next test phases. Later, the assessors scored each speech file using the 5-point MOS scale described in Table 1. The tests were performed in a computer. Each speech file received at least 15 scores from different assessors. The tests were performed individually and without a time limit.

Comparative studies were performed between the MOS index results obtained in the subjective tests and the classification performed by the proposed method. For this comparison the quality classes presented in Table 3 were used.

Additionally, the non-intrusive ITU-T Recommendation P.563 was also used to determine a MOS index of each speech files used in the proposed method. Thus, a performance comparison of these non-intrusive models is performed.

IV. RESULTS

The preliminary test results show a relationship between MOS index values and the degradation factors considered in this work, specifically, PLR, fading channel model and Doppler shift. Note that the values of these degradation factors are known and the MOS scores correspond to the PESQ algorithm results.

Fig. 5 shows how the speech quality, represented by a MOS index value, decreases when PLR values increase. In this test scenario, only the degradation in wired networks is taken into account.

Fig. 6 shows the impact of Doppler shift values on the MOS index values considering the Rayleigh fading channel. Note that the impact of the same PLR values considered in Fig. 5 are also added to the effects of the Rayleigh channel model.

Similarly, the impact of Doppler shift values on the MOS index values, considering the Rician fading channel, is presented in Fig. 7.

As can be observed in Figures 6 and 7, the Rayleigh channel model has a higher negative impact on the speech quality than the Rician channel model.

Based on the results presented in Figures 5, 6 and 7, the study of the speech signal parameters of impairment files is performed via DBN. It is worth noting that the proposed solution in this work is an non-intrusive speech quality

TABLE 4. Accuracy obtained by DBN with respective number of layer and neurons.

		Number of neurons					
		50	100	150	200	250	300
Hidden Layers	1	79.1%	79.2%	79.1%	79.1%	79.0%	79.1%
	2	90.3%	91.1%	90.1%	91.3%	90.2%	90.5%
	3	92.5%	93.1%	93.0%	92.1%	92.0%	91.8%
	4	92.8%	94.5%	94.0%	93.6%	93.7%	93.2%
	5	93.9%	**96.1**%	94.3%	93.8%	92.8%	92.2%
	6	93.1%	94.6%	94.1%	93.8%	93.7%	93.5%

TABLE 5. Accuracy, recall and F-measure of DBN, LR and SVM classifiers.

	LR (%)	SVM (%)	DBN (%)
Accuracy	95.33%	92.50%	96.14%
Recall	81.06%	88.76%	97.89%
F-measure	87.62%	90.59%	97.01%

classification model; then, it is useful in scenarios in which the PLR and fading problems occur in the communication system, but their values are unknown.

Table 4 presents the accuracy obtained by DBN using different architecture configurations. The numbers of hidden layers varied from 1 to 6, and the numbers of neurons were: 50, 100, 150, 200, 250 and 300. In order to determine the accuracy, the results obtained by the PESQ algorithm were considered as ground-truth. It is important to note that 80% of the total number of impaired speech files (51, 200) were used

TABLE 6. Confusion matrix for speech quality classification (in percentage) using the proposed model and P.563 algorithm.

Speech Qual. Class	P. Model / P.563 Class-A	P. Model / P.563 Class-B	P. Model/ P.563 Class-C	P. Model / P.563 Class-D	P. Model / P.563 Class-E
Class-A	97.23 / 51.13	2.30 / 9.53	0.47 / 18.16	0.0 / 16.45	0.0 / 4.73
Class-B	1.25 / 1.21	95.86 / 58.13	2.93 / 16.25	0.0 / 17.34	0.0 / 7.07
Class-C	0.0 /0.31	1.17 / 4.26	95.70 / 78.24	2.85 / 12.15	0.27 / 5.04
Class-D	0.0 / 0.08	0.55 / 0.23	2.23 / 4.73	94.96 / 83.13	2.27 / 11.84
Class-E	0.0 / 0.0	0.0 / 0.0	0.78 / 3.75	2.30 / 6.29	96.95 / 89.96

in the training phase, in order to guarantee data independence in the next phase.

As can be observed in Table 4, the highest accuracy was obtained in the network configuration of 5 hidden layers and 100 neurons. This configuration was used in the subsequent tests.

The performance results, in terms of accuracy, recall and F-measure metrics, of the DBN with Softmax, LR and SVM classifiers are presented in Table 5. These results represent the average value obtained in each one of the five quality classes.

Results presented in Table 5 demonstrated that the proposed DBN with Softmax classifier outperforms the other classifiers considering all the performance metrics used. Thus, the proposed speech quality classifier considers the Softmax.

In the testing phase, 20% of all impairment files were used. In this phase, the performance of our proposed model is also compared considering the results obtained by the current non-intrusive quality algorithm described in the ITU-T Recommendation P.563. In order to compare the results, each MOS score determined by P.563 is attributed according to each quality class presented in Table 3. The results are presented in Table 6 using the confusion matrix format. As stated before the PESQ results are used as ground-truth.

Finally, in the subjective test, 50 extra impairment files were used as test material, which are different to those used in the training and testing phases of the proposed model. It is worth noting that 10 speech files correspond to each quality class defined in Table 3, and there are speech samples with MOS scores closer to the border of the classes. Experimental results demonstrated that 47 speech files were correctly classified, that represents a global accuracy of 94%.

V. DISCUSSIONS

In case of a wired network, we focus on the PLR, because it is the parameter that represents the most impairment factor in a wired network [47], [52], [72]. The impact of wireless channel models and the Doppler shift on speech communication quality is not widely discussed in current literature. The results obtained in this work shows the relations between degradation parameters and quality speech assessment. By means of this relation, it is possible to emphasize the importance of considering the fading and Doppler shift in the wireless communication research area. In this context, a DB of speech samples impaired by wireless network problems is relevant to analyze the performance of speech quality assessment algorithms, because the samples represent more realistic scenarios.

The experimental results obtained demonstrated that the proposed model for speech quality assessment by the DBN can be applied in any realistic communication scenario and achieved a high accuracy. In Table IV, it is possible to observe that DBN structure with five hidden layers, each layer with 100 neurons, reached the best results, demonstrating that a larger number of neurons does not increase the accuracies of DBN. Though the DBN model is possible to apply it in a realistic scenario without computational complexity.

As can be observed in Table V, the DBN with the Softmax function, that is a multi-classifier, reached the best accuracy (96.14%). Thus, the results obtained are very closer to the intrusive PESQ algorithm. Also, it can be observed that the other classifiers reached an accuracy greater than 92% indicating that speech signal characteristics belong to any class are well defined.

The performance comparison of the proposed DBN model and the P.563 algorithm, which results are presented in Table VI, demonstrated that the proposed model largely outperforms the P.563 results. Furthermore, in subjective tests the proposed model reached an accuracy of 97% that is a confident result.

VI. CONCLUSION

In this work, a study of impairments on voice communication services caused by PLR and fading effects is presented. For this, 64, 000 impaired speech samples were created, each of them considering different degradation levels. In order to classify the different speech qualities, in a non-intrusive manner, a new classification model based on DBN was proposed. Different number of layers and neurons by layer are tested to determine the best performance of the DBN. The results obtained by DBN model with Softmax classifier outperformed the LR and SVM algorithms, reaching an accuracy of 96.14% considering the P.862 results as ground-truth.

The subjective tests, carried out in a laboratory environment, validate the performance of the proposed method based on DBN, reaching an accuracy of 94%. Also, the proposed model outperformed the most representative non-intrusive quality metric, ITU-T P.563, that reaches 72.12%.

Furthermore, the proposed speech quality classification model, for being non-intrusive, can be used in real applications of speech communications. The quality class

determined by the proposed DBN model can be useful for telephone network operators, a range of MOS values can be enough to satisfy the providers necessities.

ACKNOWLEDGMENT

The authors wish to thank both the Engineering and the Computer Science Department at the Federal University of Lavras for the motivation at the research in the telecommunication area.

REFERENCES

[1] J. Slavata and J. Holub, "Evaluation of objective speech transmission quality measurements in packet-based networks," *Comput. Standards Interfaces*, vol. 36, pp. 626–630, Mar. 2014.

[2] Cisco Inc. (Feb. 2016). *Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015—2020*. [Online]. Available: https://www.cisco.com/c/dam/m/en_in/innovation/enterprise/assets/mobile-white-paper-c11-520862.pdf

[3] J. Saldana, J. Fernández-Navajas, J. Ruiz-Mas, E. V. Navarro, and L. Casadesus, "The utility of characterizing packet loss as a function of packet size in commercial routers," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2012, pp. 346–347.

[4] S. Yuhe and X. Jie, "New solutions of VoIP on multi-hop wireless network," in *Proc. IITA Int. Conf. Control, Automat. Syst. Eng. (CASE)*, Zhangjiajie, China, Jul. 2009, pp. 199–202.

[5] S. A. Obeidat and V. R. Syrotiuk, "An opportunistic cross-layer architecture for voice in multi-hop wireless LANs," *Int. J. Commun. Syst.*, vol. 22, no. 4, pp. 419–439, Apr. 2009.

[6] J. C. F. Li and S. Dey, "Outage minimisation in wireless relay networks with delay constraints and causal channel feedback," *Eur. Trans. Telecomm. Banner*, vol. 21, no. 3, pp. 251–265, Apr. 2010.

[7] D. S. Michalopoulos and T. A. Tsiftsis, "Performance analysis of wireless multihop diversity systems," *Int. J. Commun. Syst.*, vol. 21, no. 9, pp. 955–969, Sep. 2008.

[8] S. Haykin and M. Moher, *Modern Wireless Communications*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[9] D. Z. Rodriguez, J. M. Sousa, and G. Pivaro, "Apparatus and method for evaluating voice quality in a mobile network," U.S. Patent 9 078 143 B2, Jul. 7, 2015, pp. 1–15.

[10] J. Benesty, M. M. Sondhi, and Y. A. Huang, *Springer Handbook of Speech Processing*. Secaucus, NJ, USA: Springer-Verlag, 2007.

[11] Document Rec. P.862, ITU-T, 2007. Acessed: Feb. 1, 2014. [Online]. Available: http://www.itu.int/rec/T-REC-P.862/en

[12] *Perceptual Objective Listening Quality Assessment (POLQA)*. document ITU-T Rec. P.863, Sep. 2014. [Online]. Available: http://www.itu.int/rec/T-REC-P.863-201409-I

[13] J. Polacký and P. Počta, "An analysis of the impact of packet loss, codecs and type of voice on internal parameters of P.563 model," in *Proc. IEEE 10th Int. Conf. Digit. Technol. (DT)*, Zilina, Slovakia, Jul. 2014, pp. 281–284.

[14] M. Abareghi, M. M. Homayounpour, M. Dehghan, and A. Davoodi, "Improved ITU-P.563 non-intrusive speech quality assessment method for covering VOIP conditions," in *Proc. IEEE 10th Int. Conf. Adv. Commun. Technol. (ICACT)*, Gangwon-Do, South Korea, Feb. 2008, pp. 354–357.

[15] S. Xie, N. Yan, P. Yu, M. L. Ng, L. Wang, and Z. Ji, "Deep neural networks for voice quality assessment based on the GRBAS scale," in *Proc. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, San Francisco, CA, USA, Sep. 2016, pp. 2656–2660.

[16] S. Kundu, G. Mantena, Y. Qian, T. Tan, M. Delcroix, and K. C. Sim, "Joint acoustic factor learning for robust deep neural network based automatic speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 5025–5029.

[17] J. Tao, S. Ghaffarzadegan, L. Chen, and K. Zechner, "Exploring deep learning architectures for automatically grading non-native spontaneous speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 6140–6144.

[18] B.-K. Lee and J.-H. Chang, "Packet loss concealment based on deep neural networks for digital speech transmission," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 2, pp. 378–387, Feb. 2016.

[19] *Methods for Subjective Determination of Transmission Quality*, document ITU-T Rec. P.800, Jun. 1996. [Online]. Available: http://www.itu.int/rec/T-REC-P.800/en

[20] S. Möller, W.-Y. Chan, N. Côté, T. H. Falk, A. Raake, and M. Wältermann, "Speech quality estimation: Models and trends," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 18–28, Nov. 2011.

[21] H. Gamper, L. Corbin, D. Johnston, and I. J. Tashev, "Synthesis of device-independent noise corpora for speech quality assessment," in *Proc. IEEE Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Xian, China, Sep. 2016, pp. 1–5.

[22] W. Zhou, Q. He, Y. Wang, and Y. Li, "Sparse representation-based quasi-clean speech construction for speech quality assessment under complex environments," *IET Signal Process.*, vol. 11, no. 4, pp. 486–493, Jun. 2017.

[23] Y. S. E. Ali, V. Parsa, P. Doyle, and S. Berkane, "Disordered speech quality estimation using the matching pursuit algorithm," in *Proc. 30th Can. Conf. Elect. Comput. Eng. (CCECE)*, Windsor, ON, Canada, Apr./May 2017, pp. 1–5.

[24] *Mapping Function for Transforming P.862 Raw Result Scores to MOS-LQO*, document ITU-T Rec. P.862.1, Nov. 2003. [Online]. Available: www.itu.int/rec/T-REC-P.862.1/en

[25] A. W. Rix, J. G. Beerends, D.-S. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality—Technology and applications," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 1890–1901, Nov. 2006.

[26] *Technical Requirement Specification Proposals for Scope of Single-Ended Perceptual Evaluation of Listening Quality (P.SPELQ)*, document, May 2015. [Online]. Available: http://www.itu.int/md/T13-SG12-150505-TD-GEN-0724/en

[27] *The E-Model: A Computational Model for Use in Transmission Planning*, document ITU-T Rec. G.107, Jun. 2015. [Online]. Available: https://www.itu.int/rec/T-REC-G.107

[28] *Wideband E-Model*, document ITU-T Rec. G.107.1, Jun. 2015. [Online]. Available: https://www.itu.int/rec/T-REC-G.107.1/en

[29] M. A. Raja, A. Jagodzinska, and V. Barriac, "On losses, pauses, jumps, and the wideband E-model," *IEEE Access*, vol. 5, pp. 16130–16148, 2017.

[30] Y. Han and G.-M. Muntean, "Hybrid real-time quality assessment model for voice over IP," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2015, pp. 1–6.

[31] H. Larijani and K. Radhakrishnan, "Voice quality in VoIP networks based on random neural networks," in *Proc. 9th Int. Conf. Netw. (ICN)*, French Alps, France, Apr. 2010, pp. 89–92.

[32] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[33] D. W. Patterson, *Artificial Neural Networks: Theory and Applications* (Prentice-Hall Series in Advanced Communications). Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[34] S. Desai, E. V. Raghavendra, B. Yegnanarayana, A. W. Black, and K. Prahallad, "Voice conversion using artificial neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3893–3896.

[35] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolutional neural networks," in *Proc. Comput. Vis. Pattern Recognit.*, vol. 6, no. 2, May 2015, pp. 1–13.

[36] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930–1943, Aug. 2013.

[37] T. N. Sainath *et al.*, "Multichannel signal processing with deep neural networks for automatic speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 5, pp. 965–979, May 2017.

[38] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 1, pp. 7–19, Jan. 2015.

[39] G. E. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Oct. 2012.

[40] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural Comput.*, vol. 22, no. 12, pp. 3207–3220, 2010.

[41] F. Zang and J.-S. Zhang, "Softmax discriminant classifier," in *Proc. 3rd Int. Conf. Multimedia Inf. Netw. Secur.*, Nov. 2011, pp. 16–19.

[42] I. McLoughlin, H. Zhang, Z. Xie, Y. Song, and W. Xiao, "Robust sound event classification using deep neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 3, pp. 540–552, Mar. 2015.

[43] S. M. Siniscalchi, T. Svendsen, and C.-H. Lee, "An artificial neural network approach to automatic speech processing," *Neurocomputing*, vol. 140, pp. 326–338, Apr. 2014.

[44] R. G. Guimarãesm, R. L. Rosa, D. De Gaetano, D. Z. Rodríguez, and G. Bressan, "Age groups classification in social network using deep learning," *IEEE Access*, vol. 5, pp. 10805–10816, 2017.

[45] G. E. Dahl, M. Ranzato, A.-R. Mohamed, and G. Hinton, "Phone recognition with the mean-covariance restricted Boltzmann machine," in *Proc. 23rd Int. Conf. Neural Inf. Process. Syst.*, BC, Canada, Dec. 2010, pp. 469–477.

[46] Y.-J. Hu, Z.-H. Ling, and L.-R. Dai, "Deep belief network-based post-filtering for statistical parametric speech synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 5510–5514.

[47] E. T. Affonso, R. L. Rosa, and D. Z. Rodríguez, "Speech quality assessment over lossy transmission channels using deep belief networks," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 70–74, Jan. 2018.

[48] S. Monika and A. Rama, "An efficient digital speech transmission using neural network with HMM (hidden Markov model)," in *Proc. Int. Conf. Emerg. Eng. Trends Sci.*, India, Mar. 2016, pp. 34–43.

[49] Y. Li and S. Kang, "Deep neural network-based linear predictive parameter estimations for speech enhancement," *IET Signal Process.*, vol. 11, no. 4, pp. 469–476, Jun. 2017.

[50] J. M. Martín-Doñas, A. M. Gomez, I. López-Espejo, and A. M. Peinado, "Dual-channel DNN-based speech enhancement for smartphones," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.

[51] P. D. Bhamre and H. H. Kulkarni, "Speech enhancement using deep neural network," *Int. Res. J. Eng. Technol.*, vol. 03, no. 7, pp. 1348–1354, Jul. 2016.

[52] F. J. Suárez, A. García, J. C. Granda, D. F. García, and P. Nuño, "Assessing the QoE in video services over lossy networks," *J. Netw. Syst. Manage.*, vol. 24, no. 1, pp. 116–139, Jan. 2015.

[53] A. Bildea, O. Alphand, F. Rousseau, and A. Duda, "Link quality estimation with the Gilbert-elliot model for wireless sensor networks," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug./Sep. 2015, pp. 2049–2054.

[54] H. Lee, S. Byeon, B. Kim, K. B. Lee, and S. Choi, "Enhancing voice over WLAN via rate adaptation and retry scheduling," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2791–2805, Dec. 2014.

[55] K. S. Shanmugan, "Simulation-based estimate of QoS for voice traffic over WCDMA radio links," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Beijing, China, Sep. 2009, pp. 1–4.

[56] S. Choudhury and J. D. Gibson, "Payload length and rate adaptation for multimedia communications in wireless LANs," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 796–807, May 2007.

[57] V. J. Arokiamary, *Mobile Communication*. Technical Publications, 2009.

[58] J. D. Gibson, *The Mobile Communications Handbook*, 2nd ed. Boca Raton, FL, USA: CRC Press, 1999.

[59] J. D. Gibson, *The Communications Handbook*. Boca Raton, FL, USA: CRC Press, 2002.

[60] A. Khare, K. Trivedi, and S. Dixit, "Effect of doppler frequency and BER in FFT based OFDM system with Rayleigh fading channel," in *Proc. IEEE Students' Conf. Elect., Electron. Comput. Sci. (SCEECS)*, Bhopal, India, Mar. 2014, pp. 1–6.

[61] Wav2rtp. *A Tool to Generate RTP Data Packets*. [Online]. Available: http://wav2rtp.sourceforge.net.

[62] *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs*, document ITU-T Rec. P.862, Feb. 2001. [Online]. Available: www.itu.int/rec/T-REC-P.862/en.

[63] *Coded-Speech Database*, document ITU-T Rec. Sup. 23, Feb. 1998. [Online]. Available: www.itu.int/rec/T-REC-P.Sup23-199802-I

[64] H. Lin and Z. Ou, "Switching auxiliary chains for speech recognition," *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 568–571, Aug. 2007.

[65] D. Chazan, R. Hoory, G. Cohen, and M. Zibulski, "Speech reconstruction from mel frequency cepstral coefficients and pitch frequency," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, Jun. 2000, pp. 1299–1302.

[66] F. E. Harrell, Jr., *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY, USA: Springer, 2001.

[67] C. Campbell and Y. Ying, *Learning With Support Vector Machines* (Synthesis Lectures on Artificial Intelligence and Machine Learning). San Rafael, CA, USA: Morgan & Claypool, 2011.

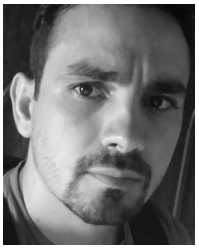
[68] A. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 14–22, Jan. 2012.

[69] S. Goyal, R. K. Chauhan, and S. Parveen, "Spam detection using KNN and decision tree mechanism in social network," in *Proc. 4th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Dec. 2016, pp. 522–526.

[70] M. P. Pushpalatha and M. R. Pooja, "A predictive model for the effective prognosis of asthma using asthma severity indicators," in *Proc. Int. Conf. Comput. Commun. Inform. (ICCCI)*, Jan. 2017, pp. 1–6.

[71] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Proc. Eur. Conf. Speech Commun. Technol.*, Rhodes, Greece, Sep. 1997, pp. 1895–1898.

[72] A. Raake, "Short- and long-term packet loss behavior: Towards speech quality prediction for arbitrary loss distributions," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 1957–1968, Nov. 2006.



EMMANUEL T. AFFONSO received the B.S. degree in computer science from UNIFOR-MG, Brazil, in 2009, the master's degree in network management from the Federal University of Lavras (UFLA), in 2012, the master's degree in web systems development from the Pontifical Catholic University of Minas Gerais in 2012, and the M.S. degree from UFLA in 2017. He is currently pursuing the Ph.D. degree with the Federal Center of Technology Education of Minas Gerais. His current research interests include voice quality assessment, quality of experience in multimedia services, wireless communications, and artificial neural network algorithms.

RODRIGO D. NUNES received the B.S. degree in computer science from UNIFOR-MG, Brazil, and the M.S. degree from the Federal University of Lavras in 2017. He is currently an IT Analyst with the Federal Center of Technology Education of Minas Gerais. He has a solid knowledge in software development based on 12 years of professional experience working with systems for solutions in desktop, Web, and mobile environment. His current research interests include voice quality assessment, quality of experience of multimedia services, and recommendation systems.

RENATA L. ROSA received the M.S. degree from the University of São Paulo in 2009 and the Ph.D. degree from the Polytechnic School, University of São Paulo, in 2015. She is currently an Adjunct Professor with the Department of Computer Science, Federal University of Lavras, Brazil. Her current research interests include computer networks, artificial intelligence algorithms, recommendation systems, telecommunication systems, wireless networks, and quality of service and quality of experience in multimedia services.

GABRIEL F. PIVARO received the M.Sc. degree in electrical engineering from the University of São Paulo, São Paulo, Brazil, in 2008, and the Ph.D. degree in electrical engineering from the State University of Campinas, São Paulo, in 2016. From 1999 to 2012, he was with Motorola, Nokia, Huawei, and Claro (América Móvil), where he focused on radio access networks. From 2015 to 2016, he was a Research Scholar with Rice University, Houston, TX, USA. He is currently with the Radiocommunications Reference Center, National Institute of Telecommunications Minas Gerais, Brazil, focusing on 5G networks, channel sounding, and cognitive radio. He received the Teacher Certification in mathematics from the São Paulo State University in 2005. His research interests are 5G networks, random matrix theory applied to multi-in multi-out communications, and wireless communications.

DEMÓSTENES Z. RODRÍGUEZ (M'12–SM'15) received the B.S. degree in electronic engineering from the Pontifical Catholic University of Peru, and the M.S. and Ph.D. degrees from the University of São Paulo in 2009 and 2013, respectively. He is currently an Adjunct Professor with the Department of Computer Science, Federal University of Lavras, Brazil. He has a solid knowledge in telecommunication systems and computer science based on 15 years of professional experience in major companies. His research interests include quality of service and quality of experience in multimedia services, artificial intelligence algorithms, and architect solutions in telecommunication systems.