Multimedia Tools and Applications (2021) 80:31783–31802
https://doi.org/10.1007/s11042-021-11190-7

# Toward a comprehensive subjective evaluation of VoIP users' quality of experience (QoE): a case study on Persian language

A. Hesam Mohseni[1] · A. H. Jahangir[2] · S. M. Hosseini[2]

Received: 23 August 2020 / Revised: 21 May 2021 / Accepted: 24 June 2021 /
Published online: 19 July 2021


**Abstract**
Quality of Experience (QoE) measures the overall quality of a service from users' point of view by considering several system, human, and contextual factors. There exist various objective and subjective methods for QoE prediction. Although the subjective approach is more expensive and challenging than the objective approach, QoE's level can be more accurately determined by a subjective test. This paper investigates various features affecting QoE by proposing a comprehensive subjective evaluation. First, we show that many unconsidered factors can significantly affect QoE. We have generated voice samples featuring different values for novel factors related to the speaker, signal, and network. Regarding the speaker, we take into account the accent and gender of Persian-speaking people. We conduct an extensive survey by employing a large number of users. Our comprehensive analysis reveals that the users' identity has a significant influence on QoE. Our experiments show that many previously studied parameters do not affect QoE in the same way for various users with different genders and accents. Finally, we show that QoE can be accurately predicted using Artificial Neural Network (ANN) and Support Vector Regression (SVR) techniques if the new identity features are taken into account.

**Keywords** Quality of experience (QoE) · Voice over IP (VoIP) · Quality of service (QoS) · Subjective test · Artificial neural network (ANN) · Support vector regression (SVR)

## 1 Introduction

The rapid growth of IP-based communication multimedia technology has encouraged more and more users to use various Voice over IP (VoIP) services. The growing number of VoIP

✉ A. H. Jahangir
jahangir@sharif.ir

[1] Faculty of Engineering, Sharif University of Technology, Intl. Campus, Kish Island, Iran

[2] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

users and their diversity have also raised their expectations of the quality of service and experience they receive. Some researchers claim that network evaluation will change from "technology-centric" to "user-centric" [15]. Therefore, the concept of Quality of Experience (QoE), which represents user satisfaction with services, has been introduced to overcome the limitations of Quality of Service (QoS).

QoS has been considered a mandatory requirement of multimedia communications such as VoIP. However, it has been noticed that user satisfaction in an IP network communication is not guaranteed only by the use of high-quality switches or routers. Other factors, such as language, accent, and noise, may also play a significant role in the quality of (voice) communication. Hence, QoE has been proposed to consider users' actual perception and measure their satisfaction independently from network equipment specifications. Nonetheless, QoE determination has many complexities, as some users' features are not well-understood. There exist objective and subjective methods for QoE prediction. Although the subjective approach is more expensive and challenging than the objective approach, QoE's accurate level can be determined only by a subjective test.

QoE metrics have not been adequately defined or considered in the literature. The majority of research studies and practical experiments on QoE typically use QoS metrics, such as throughput, delay, jitter, and packet loss, to quantify a network service and its performance. In other words, many studies on QoE are based only on QoS parameters. For example, Charonyktakis et al. have proposed a maximum tolerable end-to-end network delay of 150 ms for VoIP applications [5]. However, they have not considered that temporal statistical characteristics such as packet loss and delay can result in a different user experience quality compared to what their sole threshold predicts.

ITU-T [20] defines QoE as: "The degree of delight or annoyance of the user of an application or service." However, the definition does not address the metrics of QoE. Many factors can be considered in this area, such as the type and characteristics of the application or service, the subject, the user's expectations of the application, the user's cultural background and psychological profiles, the user's emotional state, and even socioeconomic issues. Therefore, various features of users should be considered in the evaluation of QoE. Although QoE evaluation may be complicated, time-consuming, tedious, and expensive, accurate QoE prediction is still needed to improve the performance of services.

QoE measurement and prediction are affected by three essential components: speaker, signal, and network. Considering the effect of all ingredients together on QoE can be a very complex problem. It may exist hidden mutual dependencies and non-linear relationships between these components. However, considering various subjective and objective attributes in different situations is an essential issue in the QoE prediction. These attributes can be related to the user's device, application, and network.

In this paper, we investigate various features affecting QoE by proposing a comprehensive subjective evaluation. We show various unconsidered factors such as accent, gender, and signal-to-noise ratio (SNR) can significantly affect QoE. Moreover, we show that many previously studied parameters do not affect QoE in a similar way for multiple users with different genders and accents. To prove it, we have generated voice samples featuring different values for some novel factors of the speaker, signal, and network. The voice samples are based on a well-known dataset of the Persian language. We have conducted an extensive survey by employing a large number of users. We have also categorized users according to their gender and accent, and then, we evaluated the impact of various features on QoE for each category of users and speakers separately. This kind of evaluation has not been addressed in previous

works. Our analysis reveals that the users' identity has a significant influence on QoE. Finally, we show that if we consider the new identity features in the QoE prediction methods that are based on Artificial Neural Network (ANN) and Support Vector Regression (SVR) techniques, we can achieve an accurate QoE prediction solution.

This paper is organized as follows: In Section 2, we touch on QoE measurement methods. Section 3 briefly discusses related works. The proposed approach is described in Section 4. Section 5 presents the results of our evaluations and discusses them. Finally, we conclude the paper in Section 6.

## 2 A review of VOIP users' quality of experience

Subjective and objective approaches to QoE measurement are briefly introduced in the following sections.

### 2.1 Subjective methods

In the subjective approach, evaluations data are directly collected from users' opinions. Standardization bodies have presented methodologies and recommendations for conducting subjective tests. For example, ITU-T P.800 is a well-known standard for VoIP subjective evaluation [18]. The recommendation defines a method for measuring users' QoE based on a value called Mean Opinion Score (MOS). MOS is widely used for subjective voice/video quality assessment, where human test subjects grade their overall experience on the Absolute Category Rating Scale (ACR). The scale typically comprises five choices which are: "5″ for "excellent", "4″ for "good", "3″ for "fair", "2″ for "poor", and "1″ for "bad".

In a VoIP subjective test, subjects listen to a recorded voice passed through various network conditions, and they grade its quality using an opinion scale (between 1 and 5). After collecting the opinions, MOS, which is the average of all the participants' scores, is calculated.

### 2.2 Objective methods

Objective methods are used for QoE evaluation without involving human subjects. Objective methods are categorized into signal-based and parameter-based techniques.

A signal-based method sends a voice file through a system under test and compares the output voice with the source file. The *Perceptual Evaluation of Speech Quality* (PESQ) [21] is a well-known signal-based tool that has been widely deployed in the industry and academic research studies [14, 16]. PESQ is the state-of-the-art technique for objective prediction of perceived quality and claims that it has the highest correlation with subjective measurements [14]. However, some published case studies and reports reveal that PESQ has significant disadvantages yet. It is an intrusive method and cannot predict the perceived quality of live stream data. It also cannot perform a comprehensive evaluation of transmission quality. PESQ can only measure the effects of one-way speech noise distortions on speech quality.

A parameter-based method predicts QoE by using a mathematical model based on QoS. ITU-T E-model [19, 39] is one of the best-known parameter-based methods for VoIP. E-model predicts QoE using metrics such as packet loss, end-to-end delay, voice loudness, background noise, equipment impairment, and codec robustness. E-model calculates a rate called R-factor which estimates voice quality. Thus, R-factor is mapped to MOS using a non-linear equation.

Although E-model can be utilized in network planning, its accuracy and validity are questionable [38].

The difficulty of developing, modeling, and deploying objective methods is still a challenging problem due to the large space of parameters. Moreover, any addition or deletion of parameters imposes the need for new tests to fine-tune the model or derive new statistical models for QoE prediction [4].

## 3 Related work

In this section, several related works on QoE - not limited to only VoIP applications - are studied. Alreshoodi et al. [1] proposed a QoE prediction method for video streams using QoS parameters. The correlation between QoS parameters and QoE has been determined by adopting a subjective and an objective approach. The relationship between QoS and QoE has been mentioned as an important issue in the literature. Fiedler et al. [12] proposed a mapping function between QoS parameters and QoE which led to an IQX-based relation between them. In their proposed method, a MOS grade report determines QoE using QoS parameters such as latency and jitter. The MOS grade has been tuned using an approach like the PESQ's determination. A QoE determination model based on only QoS parameters has also been proposed by Kim et al. [26]. They have shown a significant correlation between QoS and QoE. In a recent study [17], a QoE prediction method has been proposed based on a mapping between QoE and QoS parameters and machine learning techniques.

Although QoS parameters have a significant impact on QoE, they are not the only effective parameters. In addition to network parameters, other parameters can also have a considerable impact on QoE. For example, SNR can have a more significant effect on QoE in comparison to QoS parameters such as bit rate [11]. In this paper, parameters such as accent, signal-to-noise ratio, and voice amplitude are taken into account. QoE is a subjective concept that is strongly related to users' perception and experience; hence, a QoE determination method should be evaluated based on subjective tests instead of objective tests. Many other researchers have not considered this important issue. For example, in paper [36], the QoE evaluation of BATMAN routing protocol for VoIP services has been only evaluated by simulation. Nihei et al. [31] have proposed a QoE measurement index that utilizes latency spikes. Although QoE of VoIP in mobile networks was maximized by choosing a proper CODEC, the evaluation was based on E-model, which is an objective approach, and users' opinions were not considered [31].

QoE concept is based on the user's opinion and is affected by its characteristics. ITU-T has recommended that when quality is evaluated, users' nationality, culture, and language should be considered into account [22]. Some research studies are concerned with the impact of users' language and culture on QoE. For instance, E-model has been improved according to Thai culture by Daengsi et al. [8]. They have claimed that since E-model is developed in the West, it is not well-tuned for Thai people and the standard packet loss rate and delay value recommended by E-model are not appropriate for them. In another interesting study [40], the opinions of Thai users on the quality of G.711, G.722, and G.729 were considered by a subjective test. Despite the broadest bandwidth feature of the G.722 codec, the paper has shown that there was not any considerable quality difference among the three codecs according to the opinions of Thai people. Thus, it is possible to choose a lower bandwidth codec for some languages and decrease resource consumption without any considerable quality degradation. This result shows, once more, the importance of QoE evaluation for a specific language.

QoE evaluation based on users' opinions has been considered in many other studies. Chen et al. [6] investigated the impact of various bit rates of a specific codec on Skype. They proposed a method that determines the relationship between QoE and the rate of frequency change of the bit rate. QoE prediction model has been developed by using a subjective test involving 127 users. In the paper [6], this model could present a more accurate QoE prediction closer to the users' opinion than the PESQ. The Skype developers can deploy the model, predict QoE, and tune the bit rate based on users' views, but the paper has not declared the network parameters and some other QoE affecting factors.

MLQoE is another user-centric VoIP QoE prediction method which is based on machine learning [5]. The training phase of the technique is done using users' opinions. Four well-known machine learning algorithms are used in this method: ANN, Decision Trees (DTs), SVR, and Gaussian Naive Bayes (GNB). The paper [5] shows that MLQoE achieves a more accurate QoE prediction compared to PESQ and E-model. The QoE of VoIP calls using machine learning techniques has also been considered in one other recent research [7]. Although in paper [7] a system-based enhanced E-model QoE objective test is compared with a subjective test, other parameters such as users' characteristics and signal properties have not been taken into account.

CaQoEm [29] is a context-aware QoE modeling and measurement method for mobile systems. It is based on Bayesian Networks (BN) and the context space model. BN can effectively determine the relation between context parameters and QoE [29]. In addition to the typical network properties, various context attributes and users' characteristics such as their "mood" and "location" are involved in the context space mode. This model includes context state and context attributes, such as user satisfaction and technology acceptance, and determines QoE using all other underlying features. Although the paper has considered various context properties using BNs to add new parameters, it has not profoundly studied and analyzed them in its test procedure. As an example, in all of its 29 tests, participants had almost the same age. Moreover, there is no discussion on properties such as "mood."

Ochet et al. [32] have proposed a QoE prediction method for real-time multimedia applications in vehicular networks. Interestingly, in the proposed QoE prediction method, users' gender, position (city or highway), and screen resolution have been considered. However, their method is based on the objective approach and is evaluated using only a simulation. In a recent work [33], the factors affecting VoIP reliability are examined. The role of packet loss, delay and jitter on QoS and consequently QoE, is mentioned as the most influential.

## 4 Proposed approach

In this paper, a subjective QoE evaluation is proposed by taking users' opinions and applying them to the Persian language. Although the subjective test approach to QoE evaluation has been considered in some articles [7, 29], however, they have not considered all the parameters affecting QoE. In this paper, the effects of human factors such as accent, gender, and education within voice samples and subjects, which were not mentioned in previous related works, are considered for QoE evaluation. In other words, our QoE assessment approach does not rely only on QoS parameters. A rich test sample set illustrating a combination of network, signal, and human properties have been generated to consider all essential factors on QoE evaluation and prediction processes.

### 4.1 Framework

A software-based VoIP system has been set up to generate test samples. Network parameters can be changed to create audio files that comply with different network conditions. To set up the system, a VoIP Elastix server [9], whose core is based on Asterisk [2], and an open-source SIP server are used.

Our TestBed network is shown in Fig. 1. The wired network is selected here to eliminate nondeterministic and non-controllable factors of wireless network fluctuations such as delay and packet loss in the evaluation process. Thus, the network parameters are thoroughly under control and are easily adjusted using a network emulator such as NetEm [30]. NetEm inserts delay and packet loss into voice packets and measures the respective effects.

In our TestBed network setup, there are three PC-based systems: *Host1*, *Host2*, and *SIPServer*. *Host1* initiates a VoIP session to *Host2* and sends a voice sample (audio file) via a SIP connection. *Host2* accepts the session, receives and records the audio file. The recorded audio files represent the voice samples that are affected by various network conditions.

As mentioned before, NetEm is used to set network parameters such as delay, jitter, and packet loss rate. NetEm software is utilized in the three systems, as shown in Fig. 1. NetEm can apply network settings for each end system port. It can also adjust the delay of output packets with an average error rate lower than 1% of the target delay when the emulating delays are more than 50 ms [25]. A delay lower than 150 ms is an acceptable delay in VoIP communications, which does not degrade quality [5].

### 4.2 Making the dataset

First, a set of voice files should be collected to produce VoIP voice samples. Voice samples have been compiled with different characteristics in terms of properties such as the speaker's accent, SNR, and amplitude. Hence, in addition to the underlying network, i.e. features created by the simulator, the effect of voice signal features on QoE can also be studied.

According to ITU's recommendation for subjective tests [18], the sentences of a voice sample must be simple, short, intelligible, and irrelevant so that the listener focuses only on the quality of the voice. We have used the voice samples of the FarsDat dataset [10], which not only is a credible dataset for the Persian language but also complies with ITU's standards [3, 35]. FarsDat has a structure similar to TIMIT, one of the most credible voice datasets of the

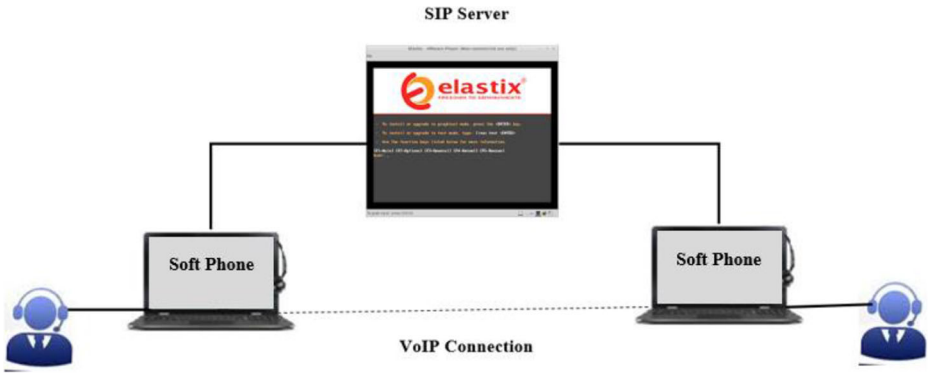


**Fig. 1** VoIP connection

English language. TIMIT has been achieved by the joint efforts of research groups from several institutions such as the Massachusetts Institute of Technology (MIT) and Texas Instruments (TI) [13]. FarsDat includes acoustically-balanced sentences that have been spoken by various speakers. The speakers differ in their gender, age, education, and accent.

Based on ITU recommendations, each played voice file should include 2 to 5 sentences on different subjects, although users usually recall the last 15 s of a voice sample [18]. At first, each voice file of the FarsDat dataset contains 10 sentences, which lasts 10 to 22 s. However, after getting a feedback from the participants, we noticed that the voice files' duration is relatively too long to make a reliable assessment of voice quality. Participants perceived that a sentence in a voice file was sometimes distorted and sometimes good. Therefore, the voice files were shortened so that each of them contained only 2 sentences and lasted 3 to 8 s. Every participant listens to 50 voice files and declares his/her opinion on their quality. The total test procedure lasted 15 min at most for each participant, which complies with the recommended range of ITU. This duration, which is lower than the 20-min voice samples of English or some international languages, has been chosen for our audience's convenience because the length of Persian language sentences is typically shorter than other common languages.

We examine some features of voice signals that have not been sufficiently addressed in previous works on QoE. For this purpose, the FarsDat data set has been selected to involve characteristics such as accent, age, gender, education, signal-to-noise ratio, and sound amplitude in each voice file so that QoE can be evaluated with more details.

The selected voice files have such properties: Their signal-to-noise ratio is uniformly distributed among six different intervals ranging from 20 dB to 40 dB. There is an equal number of male and female voices. The voice files have various audio amplitudes. The accent in each voice file is either "Tehrani" or "Azeri," which are the most prevalent accents among Persian-speaking people. Iran's capital city is Tehran, and the accent "Tehrani" refers to people that speak the Persian language as their mother language and do not speak other languages and dialects. In some Iran regions (such as Ardabil and Azerbaijan provinces), people speak the Azerbaijani language. This language is the second popular language in Iran. These people speak the Persian language with a particular accent called "Azeri." Hence, the term "Azeri accent" is used in this paper to differentiate it from the Azerbaijani or Turkish languages talked in the Republics of Azerbaijan and Turkey, respectively.

The voice file samples should pass through the VoIP network under various signal conditions. The TestBed network in Fig. 1 is used for this purpose. All hosts in this network are equipped with the NetEm network emulator. *Host1* uses a Softphone and a virtual microphone to send voice files to *Host2*. The voice quality is changed based on different network qualities, which are emulated by NetEm. Finally, *Host2* records the received voices and playbacks for evaluation by the subjective test participants.

To investigate the effects of QoS parameters on QoE, various network conditions are emulated using combinations of different packet loss rates, packet corruption rates, delays, and delay variances (jitter). Packet loss and corruption rates vary from 0% to 20% in steps of 5%. For example, when we set the packet loss rate equal to 10% and transfer a voice from *Host1* to *Host2* via the network, 10% of voice packets are deliberately and randomly dropped (by NetEm), and the received voice file is labeled with the specific feature and stored in *Host2*.

Although some references consider packet corruption as one of the main causes of packet loss, it is important to distinguish between it and packet loss for a more accurate analysis. In this paper, "packet loss" refers to packets that have been lost on their way to the destination due to link congestion or other disturbances. "Packet corruption" refers to packets that have

reached the destination imperfectly or incorrectly, for instance, with a CRC error. Regarding the timing features, we choose various constant delays from 100 ms to 300 ms in steps of 50 ms. To generate samples with jitter, we change the propagation delay with a normal distribution in our emulation. The standard deviation of the distribution varies from 0 ms to 30 ms in steps of 5 ms. NetEm has also emulated these network conditions.

All the voice files are sent from *Host1* and recorded at *Host2* after emulating each of the mentioned conditions. In total, 800 test samples have been generated. Each test sample is in accordance with a featured vector representing the speaker's accent and gender, the signal quality, i.e. SNR and amplitude, and the network parameters, i.e. packet loss rate, packet corruption rate, delay, and jitter. Finally, the users' opinions are analyzed using this feature vector.

### 4.3 Subjective test environment

The ITU reference [18] has also prescribed the environmental conditions where participants should listen to the test samples, such as the maximum acceptable noise and the room's adequate dimensions. We used a well-equipped particular audio studio in our university, in accordance with the recommendations.

To prepare participants for the experiment, they have been asked to listen to two samples before starting the test. One of them was a very high-quality voice sample, and the other one was a very low-quality sample. Each participant was also asked to enter his/her personal information, such as gender, age, education, and ethnicity, in a questionnaire form. Hence, the users could be classified, and QoE could be evaluated according to a particular user. This approach is a new idea of this study. A total of 257 participants contributed to the evaluation. Each participant designated his/her opinion regarding 50 different voice files. Therefore, 12,850 results were collected.

## 5 Evaluation

In this section, first, the impact of the voice sample features is examined. Then, methods for predicting QoE according to these features are introduced, and their accuracy is evaluated.

### 5.1 The impact of features

The overall impact of each feature on QoE can be evaluated by calculating MOS. Figure 2 shows the estimated MOS value versus each feature for both our subjective and PESQ methods. The subjective results are based on the opinions of all the 257 subjects on the 800 test samples. As shown in the figure, when packet loss increases, MOS decreases linearly with a significant slope. A similar result can be observed regarding the corruption feature. It means MOS is highly dependent on packet loss and corruption features.

While the delay is expected to impact QoE significantly, it did not affect MOS in our experiments because a listener cannot make out easily delays in an oral test. While PESQ results suggest that jitter does not significantly affect MOS, our subjective evaluation shows that MOS considerably decreases when the jitter increases.

SNR is also supposed to have a significant impact on QoE, but since the SNR of the samples of FarsDat is higher than 23 dB (the signal is clean), we have not observed its impact

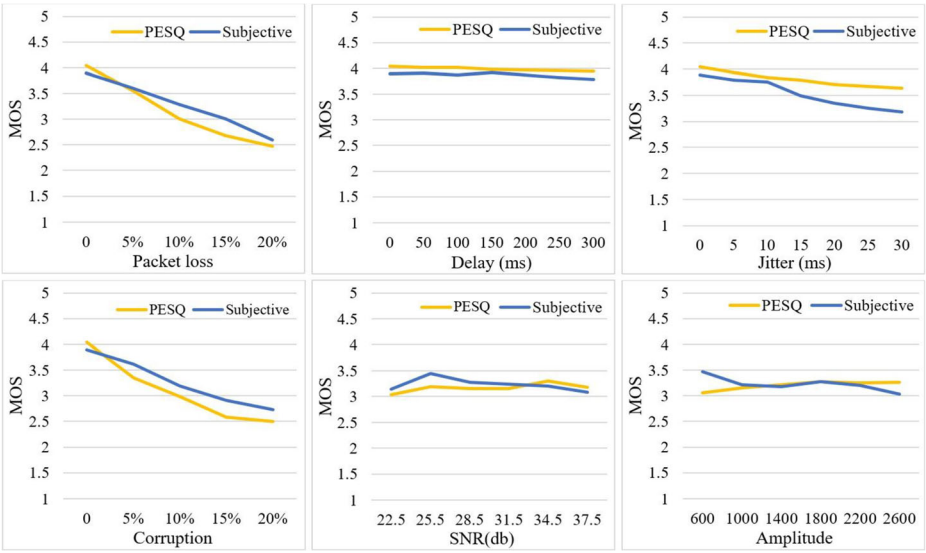


**Fig. 2** Estimated MOS value versus each feature

on MOS. The small fluctuations of the SNR plot are due to the samples' difference in other features such as packet loss. We observed a similar behavior regarding the amplitude feature. Although the speech signal of speakers differs in amplitude, it does not affect subjects' opinions. The amplitude values are depicted in the plot using a relative unit based on the sample with the highest amplitude.

While MOS shows each feature's overall impact on QoE, we conducted a Chi-square test to investigate the features' effect more deeply. In this test, the Pearson value ($p$ value) of each feature is calculated, and the opinions, which are the index of QoE, are evaluated. The p value of each feature reveals how much QoE is dependent on it.

The procedure is as follows: suppose $x$ is a feature and $y$ is the opinion. Two hypotheses can be imagined:

$H_0$: $x$ and $y$ are independent
$H_A$: $x$ and $y$ are dependent

To find out which of the hypotheses is true, $\chi 2$ is calculated according to the following relation:

$$\chi 2 = \sum \frac{(observed - expected)^2}{expected} \tag{1}$$

$\chi 2$ is compared with the Chi-square distribution to determine the relationship between these two variables. The $p$ value is also calculated, which probably indicates that hypothesis $H_0$ is correct, meaning that the two variables are independent.

If $p$ value <0.01, $H_0$ is strongly rejected, and consequently, $x$ and $y$ are highly dependent. If $0.01 < p$ value <0.05, $x$ and $y$ are moderately dependent. A p value between 0.05 and 0.1 indicates weak evidence for the dependence between the two variables. And finally, a p value higher than 0.1 reveals no evidence of dependence between the two variables.

We used SPSS for calculating the $p$ value of each feature. The results are shown in Table 1. As shown in Table 1, the $p$ value for the Packet Loss and Corruption features is lower than 0.01, meaning that QoE is strongly dependent on these two features. The $p$ value of the jitter is between 0.01 and 0.05, indicating a moderate impact on QoE. Since other features have a high p value, they do not have any considerable effect on QoE. Although SNR is supposed to have an important impact on QoE, no significant impact by SNR on QoE has been observed in our experiments since the SNR of the FarsDat samples is higher than 20 dB (which is a proper SNR).

As mentioned before, one of our aims is to analyze the relation between user features and QoE. Hence, we categorized the subjects (participants) and speakers according to their gender and accent to figure out the impact of features on QoE for each category of them. We also divided the collected opinions into four groups according to the gender of the speaker (male/female) and the participant (male/female) to determine the impact of gender on QoE. Moreover, we can figure out the opinion of each gender on the voice of the same or opposite gender. $P$-values of each feature are shown in Table 2 as a function of the gender of the speaker and participant. We notice from Table 2 that packet loss and corruption features have a high correlation with QoE regardless of gender. However, the $p$ value for the jitter differs interestingly between male and female participants. The p value for female participants (with both male and female speakers) is higher than 0.1, which means the women's opinions (QoE) is independent of jitter. However, the $p$ value for male participants is between 0.01 and 0.05, which indicates that QoE for men is moderately dependent on the jitter. Therefore, we can deduce that women's opinion is less sensitive to jitter than men's opinion.

Another interesting result is the impact of SNR on female speakers for the opinion of female participants. Unlike the three other cases in which QoE is independent of SNR, the p value of the mentioned case indicates a weak correlation between QoE and SNR when both speaker and participant are female. The amplitude feature has the same properties.

To investigate the impact of accent on QoE, the participants are divided into three groups according to their accent: Tehrani, Azeri, and others. It should be noted that the pronunciation of each voice sample is either Tehrani or Azeri. However, the articulation of a participant can also be any other accent of Persian-speaking people. The opinions of the three listener groups are separately analyzed for each of the two speaker voice accents. The $p$ value for the three participant groups' opinions on Tehrani and Azeri accents voice features is shown in Tables 3 and 4, respectively. In other words, the tables show the $p$ value of each feature as a function of the accent of the speaker and the participant.

Interestingly, as shown in Table 3, the dependency of QoE and packet loss differs among the participant groups. The packet loss feature is highly correlated with QoE for

**Table 1** p value of all participants for all samples

| Feature | Description | p value |
|---|---|---|
| Accent | Accent of Speaker | 0.337 |
| Gender | Gender of Speaker | 0.087 |
| SNR | Signal to Noise Ratio | 0.661 |
| Amplitude | Voice Amplitude | 0.699 |
| Packet Loss | Network Packet Loss | 0.001 |
| Delay | Network Delay | 0.083 |
| Jitter | Network Jitter | 0.020 |
| Corruption | Packet Corruption | 0.001 |

**Table 2** p value of results according to the gender of speakers and participants

| Feature | S: F<br>P: F<br>RF: 0.27 | S: F<br>P: M<br>RF: 0.24 | S: M<br>P: F<br>RF: 0.28 | S: M<br>P: M<br>RF: 0.21 |
|---|---|---|---|---|
| Accent | 0.110 | 0.406 | 0.150 | 0.215 |
| SNR | 0.061 | 0.498 | 0.255 | 0.639 |
| Amplitude | 0.082 | 0.498 | 0.255 | 0.639 |
| Packet Loss | 0.001 | 0.001 | 0.001 | 0.001 |
| Delay | 0.365 | 0.569 | 0.336 | 0.028 |
| Jitter | 0.138 | 0.016 | 0.143 | 0.020 |
| Corruption | 0.001 | 0.001 | 0.001 | 0.001 |

*S* Speaker, *P* Participant, *F* Female, *M* Male, *RF* Relative frequency

Tehrani-accent participants, but QoE is weakly dependent on packet loss for Azeri-accent participants. QoE and packet loss are moderately dependent for the third group.

Different dependencies between the corruption feature and each of the three accent groups can also be seen in Table 3. Although Table 1 indicates that QoE depends highly on the corruption feature, the dependencies vary from one accent group to another. QoE is moderately dependent on the corruption feature for Tehrani or Azeri accent, but the dependency is strong for the other group.

An impressive result is also observed for the jitter feature in Table 3. The $p$ value is higher than 0.1 for Tehrani and Azeri accents, which means that QoE is independent of jitter for them. However, when it comes to other accents, the p value is lower than 0.01 which indicates a strong dependency between QoE and the jitter feature for the other group of participants, who are the majority of our participants.

Significant results can also be achieved based on speakers' Azeri accent, as shown in Table 4. The p value for the packet loss parameter in Table 4 is similar to that of Table 3. The participants with the Azeri accent are less sensitive to packet loss than the other two groups. The results in Table 4 for the corruption feature are also similar to the results of Table 3. QoE is strongly dependent on the corruption feature for Tehrani and Azeri accents, while there is a moderate dependency on other participants' accents.

QoE is independent of the jitter feature based on the opinion of the Azeri accent. Nonetheless, it has a moderate dependency on the jitter feature for the Tehrani accent and remarkably a high dependency for the third group. The following results can be concluded

**Table 3** $p$ value of the results according to the accent of participants for Tehrani speakers

| Feature | S: Tehrani<br>P: Tehrani<br>RF: 0.28 | S: Tehrani<br>P: Azeri<br>RF: 0.10 | S: Tehrani<br>P: Other<br>RF: 0.62 |
|---|---|---|---|
| Gender | 0.914 | 0.322 | 0.436 |
| SNR | 0.297 | 0.375 | 0.662 |
| Amplitude | 0.261 | 0.375 | 0.620 |
| Packet Loss | 0.001 | 0.087 | 0.014 |
| Delay | 0.341 | 0.592 | 0.743 |
| Jitter | 0.350 | 0.567 | 0.005 |
| Corruption | 0.035 | 0.044 | 0.006 |

*S* Speaker, *P* Participant, *RF* Relative frequency

**Table 4** p value of results according to the accent of participants for Azeri speakers

| Feature | S: Azeri<br>P: Tehrani<br>RF: 0.28 | S: Azeri<br>P: Azeri<br>RF: 0.09 | S: Azeri<br>P: Other<br>RF: 0.63 |
|---|---|---|---|
| Gender | 0.138 | 0.130 | 0.405 |
| SNR | 0.104 | 0.072 | 0.605 |
| Amplitude | 0.236 | 0.088 | 0.605 |
| Packet Loss | 0.001 | 0.086 | 0.004 |
| Delay | 0.548 | 0.878 | 0.257 |
| Jitter | 0.083 | 0.270 | 0.008 |
| Corruption | 0.005 | 0.009 | 0.032 |

*S* Speaker, *P* Participant, *RF* Relative frequency

from Tables 3 and 4: (1) Jitter has no impact on QoE for Azeri-accent participants regardless of the speaker's accent. (2) QoE is independent of the jitter feature for Tehrani-accent participants if the speaker has a Tehrani accent. Meanwhile, there is a moderate dependency between QoE and jitter feature when the speaker's accent is Azeri. (3) QoE has a high sensitivity to the jitter feature for other accents participants. (4) SNR has a moderate impact on QoE only for Azeri-accent participants. In contrast, no considerable dependency has been observed between SNR and QoE for other accents. However, it should be noted that for FarsDat samples, the SNR feature values are higher than 23 dB (i.e. clean signals); hence, further studies should be carried out to check the dependency for lower SNRs. Finally, the key result driven from our test experiments is that the various network parameters have different effects on the evaluation of each group of participants.

## 5.2 QoE prediction

We aim to evaluate QoE using network features. Although it is impossible to determine an exact value for QoE because of its dependency on users' opinions, QoE is predicted and evaluated based on its dependence on users' opinions and the influence of various network parameters. Machine learning techniques are the best candidates for predicting QoE according to users' opinions since QoE prediction is so complex that it cannot be formulated easily based on network features and parameters. In the following, some machine learning techniques are used for QoE evaluation.

MOS can be used to compare the results of a subjective test with objective test results. Although QoE is typically represented using MOS, users' opinion, i.e. user score from 1 to 5, can also be considered. Hence, both MOS and user opinions have been considered for QoE prediction.

### 5.2.1 MOS prediction

We use ANN and SVR to predict MOS based on the mentioned network features. Although some previous works have proposed similar techniques, our method can achieve higher accuracy by taking into account more new features. As described in the following paragraphs, the new features such as SNR and accent can significantly improve the accuracy of an ANN-based or SVR-based QoE prediction method. The impact of each network feature on the accuracy of evaluation is also considered.

The ANN [27] is implemented by three hidden composed layers. These fully connected layers consist of 13, 50, and 20 nodes, respectively. "ReLU" is used for activating all these three layers. A dropout of 0.25 is applied to the first and second layers. The only node of the output layer yields the final score. The ANN uses Adam's optimization algorithm as an optimizer and Categorical cross entropy as a loss function. After completing the learning phase, QoE can be predicted by the ANN.

Our second QoE prediction method is based on SVR. It should be noted that SVR is a version of Support Vector Machine (SVM), which is used for regression analysis. Regarding the SVR method, choosing its kernel function is the most critical step in transforming input data into our intended form. We have used the Radial Basis Function (RBF) which is typically deployed in learning algorithms such as SVR [37]. The function is defined for two samples $x$ and $x'$ as:

$$K\left(x, x'\right) = exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \tag{2}$$

In relation 2, the term $\|x-x'\|^2$ is the squared Euclidean distance between two features, and σ is a free parameter determining the spread of kernel.

Figure 3 shows the absolute error rate of our prediction methods (both ANN and SVR) in comparison to PESQ. In the PESQ method, each sample file is compared with the original voice file to obtain a quality indicator between 1 and 5 for comparison. ANN-based and SVR-based methods in which new features are taken into account achieve significantly higher accuracy than PESQ, as shown in Fig. 3.

Moreover, Table 5 compares our methods' absolute error with the absolute error rate reported by MLQoE [5]. Similar to our approach, MLQoE has conducted a subjective test and used machine-learning techniques. It should be noted that since we have used a different dataset (for the Persian language), the results cannot be directly compared in a plot. However,

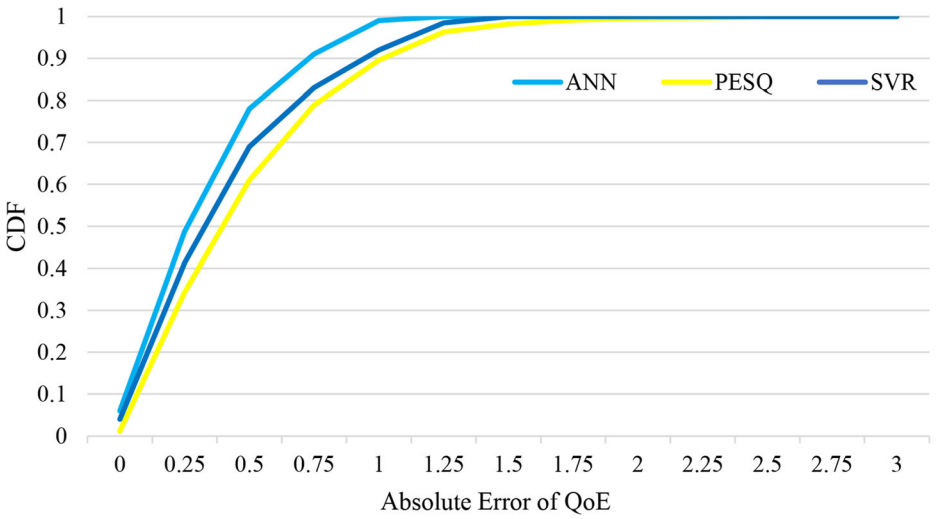


**Fig. 3** The absolute error derived from ANN, SVR, and PESQ: Confusion matrix for the five-choice opinions

the mean, median, and standard deviation of the errors can be compared to make their difference clear. As shown in Table 5, both our ANN and SVR-based techniques have yielded less error than MLQoE.

### 5.2.2 Opinion prediction

As mentioned before, while most papers consider MOS, predicting user opinions is also appealing [7]. To predict a user opinion, classification techniques are necessary. There are various well-known machine learning methods for this purpose [24, 28]. We have used three well-known machine learning methods of data mining, namely Decision Tree (DT), Naive Bayes (NB), and Rule Induction (RI). We have used the majority of users' opinions (rather than their average opinions) to determine the QoE index. We have also used the 10-fold Cross-Validation (CV) for all cases. The samples are divided into ten folds, and for each fold, 70% of samples are used for the training phase, and the other 30% are used for the testing phase.

The accuracy of opinion prediction can evaluate the correctness of each method. An opinion (with a score from 1 to 5) may be correctly predicted, or it may be guessed incorrectly as one of the other four possible scores. Accuracy is defined as the ratio between the numbers of correctly predicted samples and the total number of samples. Table 6 shows a $5 \times 5$ confusion matrix. Each element of the matrix is indexed by $n_{ij,}$ representing the prediction of opinion $j$ when the true opinion is $i$. It is evident that $n_{ii}$ indicates the number of samples that have been correctly predicted, i.e. the true opinion is $i$, and the value $i$ has also been predicted.

Although the ITU standard recommends the range of 1 to 5 for voting scores [18], simpler voting scoring conditions from 1 to 3 and even from 1 to 2 are usual and adequate [7]. It is often difficult for people to choose or decide between 5 options. Voting for one of the three scores 1 (bad), 2 (fair), and 3 (good) is likely more comfortable for them. For a two-choice option, a user evaluates the voice sample as acceptable or not. Thus, according to the tolerable complexity, the 5-score users' opinion is converted to a 3- or 2-score matrix. The 5-score evaluation table is converted to a 3-score table by mapping opinion scores 1 and 2 to (a new) score 1, 3 to score 2, and 4 and 5 to score 3, respectively. In the same way, a 5-score opinion table is converted to a 2-score table by mapping opinion scores 1 and 2 to score 1, and 3, 4, and 5 to score 2, respectively. Tables 7 and 8 represent the confusion matrices for the three- and two-choice opinions, respectively.

The accuracy of evaluation is calculated based on three relations from 3 to 5. $T_i$ represents the cases that have been correctly classified, and $F_i$ represents the incorrectly classified cases. $S$ represents the number of conditions, which can be 5, 3, or 2.

$$\mathrm{Ti} = \mathrm{nii} \tag{3}$$

**Table 5** Mean, Median and Standard deviation of absolute error for different methods

| Method | Mean | Median | Standard deviation |
|---|---|---|---|
| ANN | 0.38 | 0.31 | 0.28 |
| SVR | 0.47 | 0.39 | 0.37 |
| MLQoE [5] | 0.73 | 0.68 | 0.66 |

**Table 6** Confusion matrix for the five-choice opinions

| | | Predicted score | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Observed score | 1 | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{14}$ | $n_{15}$ |
| | 2 | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{24}$ | $n_{25}$ |
| | 3 | $n_{31}$ | $n_{32}$ | $n_{33}$ | $n_{34}$ | $n_{35}$ |
| | 4 | $n_{41}$ | $n_{42}$ | $n_{43}$ | $n_{44}$ | $n_{45}$ |
| | 5 | $n_{51}$ | $n_{52}$ | $n_{53}$ | $n_{54}$ | $n_{55}$ |

$$\mathrm{Fi} = \sum_{\substack{j=1 \\ j \neq i}}^{S} \mathrm{nij} \tag{4}$$

$$\text{Accuracy}\ (\%) = \frac{\sum_{i=1}^{S} \mathrm{Ti}}{\sum_{i=1}^{S} (\mathrm{Ti} + \mathrm{Fi})} \times 100 \tag{5}$$

The calculated values for accuracy are shown in Table 9. As expected, the accuracy value has increased when the number of states (or groups) gets lower. This is because the number of incorrect predictions decreases with the number of conditions (choices). Lesser conditions mean fewer and larger groups yielding better matching. It should be noted that the 5-choice opinions have been mapped to 3- and 2-choice opinions in this calculation. Obviously, better accuracy could be achieved if the poll of users' opinions could be rerun based on one of the three options: bad, fair, and good.

As Table 9 shows, the Decision Tree method achieves better accuracy than two other machine learning methods. The Rule Induction method is also more accurate than the Naive Bayes method. We have used Receiver Operating Characteristic (ROC) curve to compare these three machine learning methods [34]. This curve is drawn based on the True Positive Rate (TPR) ratio to False Positive Rate (FPR) in various states. The closer a curve is to the upper left corner of the diagram, the higher the accuracy. Figure 4 shows the ROC curves for these three methods. As shown in Fig. 4, the Decision Tree method has higher accuracy than the other two methods.

Paper [7] has used machine learning techniques to predict the opinion of a user. It has also analyzed both five-choice and two-choice opinions. Table 10 compares the accuracy of our method with the results of this paper. Our method has achieved higher accuracy, especially for five-choice opinions.

**Table 7** Confusion matrix for the three-choice opinions

| | | Predicted score | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Observed score | 1 | $n_{11}$ | $n_{12}$ | $n_{13}$ |
| | 2 | $n_{21}$ | $n_{22}$ | $n_{23}$ |
| | 3 | $n_{31}$ | $n_{32}$ | $n_{33}$ |

**Table 8** Confusion matrix for the two-choice opinions

| | | Predicted score | |
|---|---|---|---|
| | | 1 | 2 |
| Observed score | 1 | $n_{11}$ | $n_{12}$ |
| | 2 | $n_{21}$ | $n_{22}$ |

In a recent work [23], ML methods have been used to evaluate classification. Its results show that the Decision Tree method is more accurate than other machine learning methods. In [23], the audio samples are classified according to the types of noise pattern and the amount of SNR into one of 8 classes according to the accuracy of the prediction.

### 5.3 Discussion

Our evaluation, which is based on the effectiveness of features, shows that the packet loss rate and corruption parameters play a significant role in QoE evaluation. Although previous related papers have proposed that the packet loss rate is the main factor in determining QoE, the corruption feature has not been considered. Despite the recent improvements in data transmission networks that have led to significant reductions in corruption rates, the corruption of packets is still an important parameter of VoIP applications due to the use of UDP protocol and diverse environments and networks.

The delay feature has not been observed to affect QoE because listeners cannot logically experience it in an oral test. However, the jitter feature has a significant impact on QoE due to the loss of voice quality.

SNR is one of the new features that we have considered. SNR can have a significant effect on the quality of voice signal, and users can also perceive it. However, we did not observe its impact on QoE in users' opinions due to the relatively high SNR level of our voice samples (taken from FarsDat where SNR is between 23 and 43 dB). Nonetheless, based on our preliminary experiments, we observed when SNR is less than 20 dB, then the effect of noise on voice quality and, consequently, QoE is very prominent. In such a situation, the voice signal could be lost in the background noise, and its amplitude (energy) is not enough to be correctly perceived. This issue should be considered more profoundly in detail as future work.

The other important point is the effect of other features on the quality. Features such as gender and different accents of the Persian language are the other new concepts considered in this paper. Previous works have not addressed such features. When each of the features is considered separately, the results of Table 1 show that in this case, there is no significant relationship between them and QoE. A diverse set of features has been chosen to better predict QoE based on these features' interaction. In other words, while each of the features is not effective on its own, their combination into one feature set can lead to a more accurate

**Table 9** Accuracy results

| Method | 5 states | 3 states | 2 states |
|---|---|---|---|
| Decision Tree | 62.38% | 71.22% | 84.21% |
| Naive Bayes | 50.63% | 67.47% | 78.13% |
| Rule Induction | 57.83% | 68.44% | 80.32% |

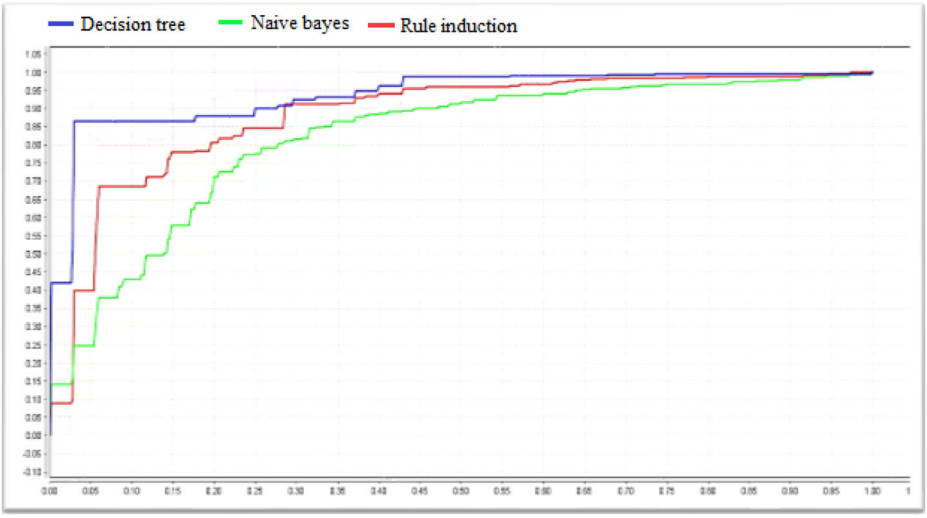


**Fig. 4** ROC of decision tree, naive Bayes, and rule induction

evaluation or prediction of QoE. For this purpose, in this paper, users are divided into several categories based on their gender and accent to realize their distinct impact on QoE. This is very important because most of the previous works took into account the average opinion of all users, which meant that some important features on QoE were not considered. However, this paper shows that by analyzing and considering more different users' and network characteristics, a more accurate evaluation of QoE can be estimated and obtained.

## 6 Conclusion

In this paper, the importance of subjective tests for QoE evaluation has been studied. We used a subjective method to predict QoE more accurately, whereas in previous related work on QoE, subjective methods have been used only to validate objective methods. We considered three different characteristic factors: human (gender and accent), voice signal (SNR and amplitude), and network (packet loss rate, delay, jitter, and corruption rate). The proposed scheme was evaluated using an extensive poll of users. Based on the data evaluation, several significant correlations have been observed between QoE and features, such as packet loss rate, packet corruption rate, and jitter. The other features had less correlation with QoE.

Moreover, we divided users into several groups according to their gender and accent features, and then, we repeated the experiments. Under the circumstances, we observed that QoE is sometimes affected by the features based on network, signal and speaker parameters. In other words, it was observed that if gender and accent features are used along with the other

**Table 10** Comparison of the accuracy of opinion prediction

| Method | Five-choice opinions | Two-choice opinions |
|---|---|---|
| Our method | 62.38% | 84.21% |
| Paper [7] | 51.21% | 83.06% |

main features, QoE could be more accurately evaluated. ANN-based and SVR-based implementations show that QoE prediction can be accurately evaluated by using all the above features.

One of our ongoing works is to investigate the impact of environmental noise on QoE. This factor directly affects SNR, which can, in turn, affect QoE. To this end, samples with various SNR values lower than 20 dB should be generated in different network conditions, and then, the subjective test should be applied again to clarify the effect of SNR factor on QoE more accurately.

As another future work, QoE can be predicted according to the relation between different groups or classes of listeners and speakers. According to each user group, various QoE indicators can be evaluated instead of using a single QoE indicator for all users. In other words, based on the user's group, the values of configurable features will be customized by the system to fulfill users' quality of experience.

## References


1. Alreshoodi M, Danish E, Woods J, Fernando A, De Alwis C (2015) Prediction of perceptual quality for Mobile video using fuzzy inference systems. IEEE Trans Consum Electron 61(4):546–554. https://doi.org/10.1109/TCE.2015.7389811
2. Asterisk (2018) Free and open source framework for building communications and applications. https://www.asterisk.org. Accessed 3 Dec 2018
3. Bijankhan M, Sheikhzadegan J, Roohani M, Samareh Y, Lucas C, Tebyani M (1994) Farsdat-the speech database of Farsi spoken language. Proceedings of 5th Australian international conference on speech science and technology, Dec. 1994, Perth, Australia, pp. 826–831. https://www.researchgate.net/publication/292798168_The_speech_database_of_Farsi_spoken_language
4. Brooks P, Hestnes B (2010) User measures of quality of experience: why being objective and quantitative is important. IEEE Netw 24(2):8–13. https://doi.org/10.1109/MNET.2010.5430138
5. Charonyktakis P, Plakia M, Tsamar dinos I, Papadopouli M (2016) On user centric modular QoE prediction for VoIP based on machine learning algorithms. IEEE Trans Mob Comput 15(6):443–1456. https://doi.org/10.1109/TMC.2015.2461216
6. Chen S, Chu C, Yeh S, Chu H, Huang P (2014) Modeling the QoE of rate changes in Skype - SILK VoIP calls. IEEE/ACM Trans Networking 22(6):1781–1793. https://doi.org/10.1109/TNET.2013.2286624
7. Cipressi E, Merani M (2020) An effective machine learning (ML) approach to quality assessment of voice over IP (VoIP) calls. IEEE Netw Lett 2(2):90–94. https://doi.org/10.1109/LNET.2020.2984721
8. Daengsi T, Wuttidittachotti P (2013) VoIP Quality Measurement_ Enhanced E-model Using Bias Factor. IEEE Global Communicatons Conference (GLOBECOM), Dec. 2013, Atlanta, Georgia, USA https://doi.org/10.1109/GLOCOM.2013.6831258
9. Elastix (2019) Unified Communications Server. https://sourceforge.net/projects/elastix. Accessed 20 April 2019
10. FarsDat (2019) Speech database of Farsi spoken language. https://www.rcdat.com/node/54. Accessed 14 March 2019
11. Fez I, Belda R, Guerri J (2020) New objective QoE models for evaluating ABR algorithms in DASH. Comput Commun 158:126–140. https://doi.org/10.1016/j.comcom.2020.05.011
12. Fiedler M, Hossfeld T, Tran-Gia P (2010) A generic quantitative relationship between quality of experience and quality of service. IEEE Netw 24(2):36–41. https://doi.org/10.1109/MNET.2010.5430142
13. Gatofrod J, Lamel L, Fisher W, Fiscus J, Pallett D, Dahlgren N (1993) DARPA TIMIT acoustic phonetic continuous speech Corpus. National Institute of Standards and Technology (NIST). https://www.researchgate.net/publication/243787812_TIMIT_Acoustic-phonetic_Continuous_Speech_Corpus
14. Goudarzi M, Sun L, Lfeacheor E (2009) PESQ and 3SQM measurement of voice quality over live 3G networks. Proceedings of the Measurement of Speech, Audio and Video Quality in networks (MESAQIN), June 2009, Prague, Czech Republic. http://wireless.feld.cvut.cz/mesaqin2009/papers/5_PESQ_3SQM_over_3G_final.pdf

15. Hamam A, Saddik A (2013) Quality of experience evaluation for haptic applications. IEEE Trans Instrum Meas 62(12):3315–3322. https://doi.org/10.1109/TIM.2013.2272859
16. Hoene C, Dulamsuren-Lalla E (2004) Predicting performance of PESQ in case of single frame losses. Measurement of speech and audio quality in network (MESAQIN 2004), June *2004,* Prague, Czech Republic. https://pdfs.semanticscholar.org/0276/1adbdde231150099f8bdbf6c90dc50b61734.pdf
17. Hu Z, Yan H, Yan T, Geng H, Liu G (2020) Evaluating QoE in VoIP networks with QoS mapping and machine learning algorithms. Neurocomputing 38:63–83. https://doi.org/10.1016/j.neucom.2019.12.072
18. ITU-T (1996) Methods for subjective determination of transmission quality. ITU-T Recommendation P.800. https://www.itu.int/rec/T-REC-P.800-199608-I
19. ITU-T (2015) International telephone connections and circuits –The E-model: a computational model for use in transmission planning. ITU-T Recommendation G.107 https://www.itu.int/rec/T-REC-G.107
20. ITU-T (2017) Vocabulary for performance, quality of service and quality of experience. ITU-T Recommendation P.10/G.100. https://www.itu.int/rec/T-REC-P.10-201711-I
21. ITU-T (2018) Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs-corrigendum 2. ITU-T Recommendation p862. https://www.itu.int/rec/T-REC-P.862-201803-I!Cor2/en
22. ITU-T Study Group 12 (2009) Methods, tools and test plans for the subjective assessment of speech, audio and audiovisual quality interactions. ITU-T. http://www.itu.int/ITU-T/studygroups/com12/sg12-q7.html. Accessed 12 Jan 2009
23. Jaiswal R, Hines A (2020) Towards a Non-Intrusive Context-Aware Speech Quality Model. June 2020 31st Irish signal and systems conference (ISSC), Letterkenny, Ireland, https://doi.org/10.1109/ISSC49989.2020.9180171
24. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with applications in R. Springer, New York
25. Jurgelionis A, Laulajainen J, Hirvonen M, Wang A (2011) An empirical study of Netem network emulation functionalities. Proceedings of 20th international conference on computer communications and networks (ICCCN), Maui, Hawaii, USA, July 2011. https://doi.org/10.1109/ICCCN.2011.6005933
26. Kim HJ, Choi SG (2013) QoE assessment model for multimedia streaming services using QoS parameters. Multimed Tools Appl 27(3):2163–2175 https://link.springer.com/article/10.1007/s11042-013-1507-8
27. Mitchell TM (1997) Artificial neural networks in: machine learning. McGraw-Hill, pp 81–127
28. Mitchell TM (1997) Bayesian learning in: machine learning. McGraw-Hill, pp 154–200
29. Mitra K, Zaslavsky A, Ahlund C (2015) Context-aware QoE modelling, measurement and prediction in mobile computing systems. IEEE Trans Mobile Comput 14(5):920–936. https://doi.org/10.1109/TMC.2013.155
30. Netem (2018) Linux Network Emulator. https://wiki.linuxfoundation.org/networking/netem. Accessed 30 May 2018
31. Nihei K, Satoda K, Yoshida H (2016) A QoE Indicator and a Transmission Control Method for VoIP on Mobile Networks Considering Delay Spikes. IEEE International Conference on Communications Software, Services and Multimedia Applications Symposium (ICC 2016), May 2016, Kuala Lumpur, Malaysia 10.1109/ICC.2016.7511338
32. Oche M, Noor R, Chembe C (2017) Multivariate statistical approach for estimating QoE of real-time multimedia applications in vehicular ITS network. Comput Commun 104:88–107. https://doi.org/10.1016/j.comcom.2016.12.022
33. Parkash Roy O, Kumar V (2020) A Survey on Voice over Internet Protocol (VoIP) Reliability Research. 6th International Conference on Computers Management & Mathematical Sciences (ICCM 2020), Nov. 2020, Arunachal Pradesh, India. https://iopscience.iop.org/article/10.1088/1757-899X/1020/1/012015
34. Rapid Miner (2020) Rapid Miner documents. https://www.docs.rapidminer.com. Accessed 14 April 2020
35. Sameti S, Veisi H, Bahrani M, Babaali B, Hosseinzadeh K (2011) A large vocabulary continuous speech recognition system for Persian language. EURASIP Journal on Audio, Speech and Music Processing 1–12. https://link.springer.com/article/10.1186/1687-4722-2011-426795
36. Sanchez-Iborra R, Cano M, Garcia-Haro J (2014) Performance evaluation of BATMAN routing protocol for VoIP services - a QoE perspective. IEEE Trans Wirel Commun 13(5):4947–4958. https://doi.org/10.1109/TWC.2014.2321576
37. Steinwart I, Christmann A (2008) Support vector machines. Springer, USA
38. Takahashi A, Kurashima A, Yoshino H (2006) Objective assessment methodology for estimating conversational quality in VoIP. IEEE transactions on audio. Speech Lang Process 14(6):1984–1993. https://doi.org/10.1109/TASL.2006.883261
39. Wuttidittachotti P, Daengsi T (2017) VoIP-quality of experience modeling: E-model and simplified E-model enhancement using bias factor. Multimed Tools Appl 76:6:8329–8354. https://link.springer.com/article/10.1007/s11042-016-3389-z

40. Wuttidittachotti P, Daengsi T, Preechayasomboon A (2013) VoIP Quality of Experience: A Study of Perceptual Voice Quality from G.729, G.711 and G.722 with Thai Users Referring to Delay Effects. Fifth International Conference on Ubiquitous and Future Networks (ICUFN), July *2013*, Da Nang, Vietnam. https://doi.org/10.1109/ICUFN.2013.6614850