DRX-based energy-efficient supervised machine learning algorithm for mobile communication networks

David E. Ruiz-Guirola | Carlos A. Rodríguez-López | Samuel Montejo-Sánchez | Richard Demo Souza | Muhammad Ali Imran

1 Department of Electronics and Telecommunications, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
2 Programa Institucional de Fomento a la I+D+i, Universidad Tecnológica Metropolitana, Santiago, Chile
3 Electrical and Electronics Engineering Department, Federal University of Santa Catarina, Florianópolis, Brazil
4 School of Engineering, University of Glasgow, Glasgow, UK

Correspondence
Samuel Montejo-Sánchez, Programa Institucional de Fomento a la I+D+i, Universidad Tecnológica Metropolitana, Santiago, Chile.
Email: smontejo@utem.cl

Funding information
CNPq; FONDECYT, Grant/Award Number: 11200659; FONDEQUIP, Grant/Award Number: EQM180180

Abstract
The continuous traffic increase of mobile communication systems has the collateral effect of higher energy consumption, affecting battery lifetime in the user equipment (UE). An effective solution for energy saving is to implement a discontinuous reception (DRX) mode. However, guaranteeing a desired quality of experience (QoE) while simultaneously saving energy is a challenge; but undoubtedly both energy efficiency and the QoE have been essential aspects for the provision of real-time services, such as voice over Internet protocol (VoIP), voice over LTE, and mobile broadband in 4G networks and beyond. This paper focuses on human voice communications and proposes a Gaussian process regression algorithm that is capable of recognizing patterns of silence and predicts its duration in human conversations, with a prediction error as low as 1.87%. The proposed machine learning mechanism saves energy by switching OFF/ON the radio frequency interface, in order to extend the UE autonomy without harming QoE. Simulation results validate the effectiveness of the proposed mechanism compared with the related literature, showing improvements in energy savings of more than 30% while ensuring a desired QoE level with low computational cost.

1 INTRODUCTION

With the ever increasing traffic growth of wireless communication systems, energy efficiency becomes a very important and challenging issue [1]. 5G and 6G wireless networks promise a notable number of innovative services and technologies [2, 3], whose incorporation may increase even more the power dissipation in the network. Energy demand from battery supplied devices, needed to tolerate new software applications, exceed substantially the real technology capacity from batteries. As we move toward new generation networks (NGN), user experience also depends on his ability to save energy. Energy efficiency at the base station (BS) has attracted a lot of attention from the industry and the research community [4], as the BS contributes with a much larger energy consumption than the user equipment (UE). However, energy efficiency at the UE is also extremely important to prolong the battery lifetime. In this sense, discontinuous reception (DRX) [5] is one of the energy-saving solutions that has proven effective on several generations of wireless networks. DRX mechanisms prolong the battery lifetime of the devices by turning OFF/ON the radio frequency interface (RFI), while considering the nature and quality of the service executed [6].

Voice over Internet protocol (VoIP) applications and related protocols, such as session initiation protocol (SIP) [7] and real-time transport protocol (RTP) [8], revolutionised traditional voice calling technology. They brought up to the level of computer networks a service that traditionally ran in another network [9]. Resting on a long-prototyped technology, the popularity of VoIP dates back to approximately the mid-1990s [10], hence VoIP devices can be considered among the earliest Internet of things (IoT) members [11]. Moreover, their integration with the current IoT is gaining significant momentum.
lately, also due to the quest for controlling devices and services remotely by voice [10].

In today’s competitive world, VoIP quality of experience (QoE) can pave the way to success for any business [12]. VoIP is increasingly common among providers and is used in diverse industries. As stated by the business consulting company Frost & Sullivan, mobile VoIP alone (like WhatsApp and Viber, excluding fixed VoIP) constituted a 30 billion dollars business worldwide by 2015 (up from just 600 million dollars in 2008), and this trend continues. Because of this, Frost and Sullivan is incentivizing all mobile providers to start applying VoIP [12]. In November 2017, the number of concurrent VoIP users on Skype exceeded 300 million. Research has shown that there are approximately 200 million VoIP hard phone subscribers in the world and the number of mobile VoIP users exceed 150 million [13]. Undoubtedly, the deployment of high-speed networks has contributed to the increasing growth and popularity of VoIP [12].

5G and 6G networks have expanded their focus, including not only human-centred communications but also many use cases of machine-type communications (MTC) [14, 15]. Since MTC devices are usually battery-operated without frequent human intervention, energy efficiency becomes extremely important. Based on the MTC traffic parameters, in [16] a radio resource control protocol, that makes use of DRX, was designed by taking into account the unique characteristics of MTC devices. In [17] and [18], DRX mechanisms associated with MTC services are discussed. In [17], the average power consumption of the wake-up enabled MTC device is modelled by using a semi-Markov process and then optimized through a delay-constrained problem. In [18] the authors propose a new online learning based DRX mechanism, which aims at improving energy efficiency for MTC services by adapting to different traffic patterns.

IoT applications emerge as new and innovative ways to collaborate, communicate and interact, both from a human and machine perspective [19]. In this sense, as the natural mode of communication, the integration of speech/voice and telephony into IoT applications can offer a versatile method to provide human interaction, communication, and control [19]. Successful integration of IP telephony and IoT technologies is mutually beneficial and even essential for both fields [20]. Such an advancement would enable numerous new IoT applications and products with voice-awareness, which can provide a flexible user experience in a more economical way than traditional methods. Ultimately, all these new applications and features will result in a broader and smarter ecosystem [20].

However, human communication based on voice over IP (VoIP) continues to be one of the main demanded services from wireless networks [25], be it for one-to-one conversations or for online meetings. It is a feature of voice communications that the signal is composed of active periods separated by pauses or silence periods [26]. Interrupting the transmission during pauses or silence intervals has very little impact on the quality-of-experience (QoE), which could be exploited by a DRX mechanism. By conditioning the switching to the OFF state with the beginning of a silence period, DRX is able to save energy without deteriorating the QoE, provided that the turning to the ON state is synchronized with the start of the activity period. However, we must take special care not to prolong the OFF state beyond the beginning of active speech. Therefore, there is an operational trade-off between energy savings and QoE.

Several DRX approaches for VoIP communications can be found in the literature, as for instance [22–24, 27–29]. In [27], the authors propose a modified power-saving mechanism (PSM) that reversely applies the state transition of legacy LTE PSM by considering the attributes of network propagation delay, whereas in [28] it is considered an adaptive DRX method that utilizes service and terminal acknowledgement provided by the deep packet inspection mechanism and modifies the DRX settings on the fly. In [29], the authors model the DRX mechanism using an n-state Markov chain, and evaluate the energy consumption and the delay based on the transition probabilities between states, but unfortunately the proposed method does not guarantee a high level of QoE. Furthermore, [27–29] do not exploit the active/silence periods characteristic of VoIP sessions to reduce energy consumption. In [22], the authors exploit the silence periods of a VoIP session to achieve greater energy savings in WiFi networks, but without taking into account the effect of burst losses nor ensuring a required QoE value.

Each service has its particular traffic characteristics, and that is not different with VoIP. The alternation of silence and active periods is typical of voice communications, and it provides an essential feature to control the DRX mechanism in order to save energy without affecting the QoE. However, for that sake, it is of paramount importance to statistically characterize the behaviour of silence periods. Dynamic DRX mechanisms that predict packets arrival using a simple neural network, in response to traffic conditions, are proposed in [23, 24]. An algorithm for multiple beam communications is proposed in [23], which enables dynamic short and long sleep cycles. In [24], a neural network is trained for prediction of next packet arrival time based on real wireless traffic. However, [23, 24] do not focus on the distinctive features of VoIP that, according to [30], are described more accurately by means of a Gaussian process (GP). It is known that GPs can conveniently be used to specify a very flexible non-linear regression with feasible computational load, working well on small datasets and having the ability to provide uncertainty measurements on the predictions. Gaussian process regression (GPR) is non-parametric, so rather than calculating the probability distribution of parameters of a specific function, it calculates the probability distribution over all admissible functions that fit the data [31].

The method proposed in this paper performs a trade-off analysis between the average receive energy consumption and the VoIP service QoE. We propose a new DRX mechanism by a service aware supervised learning algorithm, which predicts the duration of the silence periods by means of a GP model that effectively characterizes the VoIP traffic pattern. To the best of our knowledge, this approach has not been previously proposed. The results show that our novel mechanism significantly reduces energy consumption with respect to previous DRX schemes without appreciably deteriorating QoE, while...
## TABLE 1 | Features of the addressed DRX-based mechanisms

| Features          | MTC [16–18] | 3GPP [21] | SiFi [22] | AI-DRX [23, 24] | Our Proposal |
|-------------------|--------------|-----------|-----------|-----------------|--------------|
| DRX               | Yes          | Yes       | Yes       | Yes             | Yes          |
| Traffic analysis  | Yes          | Yes       | Yes       | Yes             | Yes          |
| QoS               | No           | No        | No        | No              | Yes          |
| QoE               | No           | No        | No        | No              | Yes          |
| ML                | No           | No        | Yes       | No              | Yes          |
| Real-time services| Yes          | Yes       | Yes       | No              | Yes          |
| Silence prediction| No           | No        | Yes       | No              | Yes          |
| Application       | IoT          | VolTE     | VolIP<sup>a</sup> | Data | VolIP<sup>b</sup> & MBB |

<sup>a</sup>In SiFi, it is specifically VoIP over WiFi.

<sup>b</sup>We specify VoIP, which of course includes its application in VolTE, but does not limit it to LTE.

### TABLE 2 | Table of Acronyms

| Symbol | Description |
|--------|-------------|
| ARIMA  | Auto Regressive-Integrated-Moving Average |
| ARMA   | Auto Regressive-Moving Average |
| BS     | Base Station |
| CDF    | Cumulative Distribution Function |
| CN     | Comfort Noise |
| DRX    | Discontinuous Reception |
| ECDF   | Empirical Cumulative Distribution Function |
| GP     | Gaussian Process |
| GPR    | Gaussian Process Regression |
| IoT    | Internet of Things |
| LTE    | Long-Term Evolution |
| MBB    | Mobile Broadband |
| ML     | Machine Learning |
| MOS    | Mean Opinion Score |
| MTC    | Machine Type Communication |
| NGN    | New Generation Networks |
| NN     | Neural Networks |
| PDF    | Probability Density Function |
| PLR    | Packet Loss Ratio |
| PSM    | Power Saving Mechanism |
| QoE    | Quality of Experience |
| QoS    | Quality of Service |
| RF     | Radio Frequency |
| RFI    | Radio Frequency Interface |
| RMSE   | Root Mean Square Error |
| RTP    | Real-time Transport Protocol |
| SIP    | Session Initiation Protocol |
| SNR    | Signal to Noise Ratio |
| SVM    | Support Vector Machine |
| TTI    | Transmission Time Interval |
| UE     | User Equipment |
| VoIP   | Voice over Internet Protocol |
| VolTE  | VoIP over LTE Networks |
| 4G     | Fourth Generation Networks |
| 5G     | Fifth Generation Networks |
| 6G     | Sixth Generation Networks |

Now, the contributions of this paper are summarized as follows: (i) We propose a new DRX-based energy-efficient supervised machine learning algorithm for mobile communication networks, which saves energy by switching OFF/ON the radio frequency interface, in order to extend the UE autonomy without harming QoE; (ii) For the first time, the problem of silence intervals prediction on human voice communications is addressed using a Gaussian process model, with a prediction error as low as 1.87%; (iii) Our proposal allows an improvement in energy savings as big as 30% compared with the other DRX-based mechanisms and its energy benefits are even greater when high levels of QoE are required.

The rest of this paper is organized as follows. Section 2 introduces the system model, while Section 3 describes Gaussian process techniques to characterize the behaviour of the silence periods through regression. Section 4 designs a model that is capable of predicting the length of the silence periods. Section 5 discusses some performance metrics, while Section 6 evaluates the proposed method in terms of energy saving and QoE. Finally, Section 7 concludes the paper. Table 2 lists the acronyms and Table 3 the symbols used throughout this paper.

### 2 SYSTEM MODEL

The period in which the audio does not contain spoken words is called “a period of silence”, which can correspond to an interval between two consecutive words in a sentence or to a pause in the conversation. Intuitively, during a period of silence it is not necessary to transmit or receive any packets, resulting in energy savings, since the payload in real-time transport protocol (RTP) packets in such periods does not have significant data. In VoIP communications it is possible to distinguish RTP packets of silence from RTP packets of voice. Since we assume that the transmitter sends a comfort noise (CN) packet at the beginning of a silence period, allowing the receiver to identify the start of the silence period. The main purpose is to notify the receiver predicting silence periods with an error of less than 2%. Table 1 summarises the features of each analysed mechanism.

So that comfort noise is generated while the silence lasts. In this way, the transmitter avoids sending unnecessary packets and the listener does not think that the communication has been cut off, as happens in absolute silence. However, in our case we take advantage of that information to turn OFF the radio frequency interface, for the sake of saving energy also at the receiver side. Once the silence period is over, the first incoming packet has a time stamp that indicates an interval larger than expected for two consecutive active packets, but the sequence number has
only increased by one, thus indicating a new activity period after the silence [32].

Moreover, the use of voice activity detection techniques allows to distinguish speech from non-speech segments in an audio stream and delimits the length of the period of silence using comfort noise and normal RTP packets. Opportunistically and dynamically turning OFF/ON the RFI at the transmitter side is a trivial operation. However, the receiver cannot take advantage of the activity restart notifications while its RFI is turned OFF, then predicting with high precision the length of periods of silence at the receiver side is a considerably complex but much necessary task. Note that, increasing the OFF time causes an increase in the energy savings in the receiver. However, the risk of packet loss increases if an active period starts during the OFF time. Since, although the new packets are stored in a buffer waiting for re-establishing the connection, excessive delay causes information loss, ultimately affecting the service QoE.

Figure 1 illustrates our System Model. Without limiting the relevance of our proposal in 5G networks and beyond, we use as example a typical Voice over LTE (VoLTE) architecture by which the UE accesses an IP flow of conversational multimedia. The modules involved in the operation of the proposed scheme are represented within the UE and described in Sections 2, 3 and 4. It is necessary to clarify that we assume that the UEs involved in the communication keep a single active conversational session, even though we consider that in conference sessions it could perform in a similar way.

Finding a proper solution between the opposed aspects of energy savings and QoE is challenging, as predicting the duration of each silence period is non-trivial. Without losing the generality, in this work we assume that voice activity detection is used because most of the current codecs support it, so whenever the RFI is active we can determine when a silence period
System model: DRX mechanism aimed at saving energy considering the QoE by a service aware supervised learning algorithm, which predicts the duration of the silence periods by means of a GP model that characterizes the VoIP traffic pattern.

When trying to predict the silence duration there are three possible scenarios that are illustrated in Figure 2. First, the radio interface (RFI) is turned ON before the end of the period of silence, what is inefficient in energy saving terms. Second, the RFI is turned ON beyond the current silence period, therefore leading to packet losses, decreasing the QoE. Third, the ideal scenario, in which the system is capable of predicting the exact duration of silence, where we can save the maximum amount of energy without deteriorating QoE.

3 | GAUSSIAN PROCESS

Supervised learning [34] in the form of regression is an important element of statistics and machine learning, either for data set analysis or as part of a more complex problem. Traditionally, parametric models have been used for this purpose, which can be easily interpreted and used to advance a general theory from the model. In that sense, models like the Auto Regressive-Moving Average (ARMA) model and the Auto Regressive-Integrated-Moving Average (ARIMA) model stand out for stochastic processes. However, according to [35], when the process is only approximately governed by stochastic differential equations the parametric modelling can be weak. Besides, parametric models such as the Normal, Weibull and Poisson distributions require an assumption about the algebraic form of the relationship between two or more variables, as well as a lot of clean, complete, and uncorrelated data to be properly validated [36]. The shortcomings of these parametric techniques are caused by their limited assumptions over the underlying function $f$; such linear assumptions are inadequate to model a long-term forecasting problem and thus a more powerful model is needed. Non-parametric techniques like GPR provide a solution to the problem, as mentioned earlier. These techniques make much less restrictive assumptions over the function $f$.

Non-parametric methods do not require a prior model. They count with innumerable possibilities to configure weights and to adapt to different non-linear datasets [37]. The neural
network (NN) approach is the most popular and widely used non-parametric method. A NN can perform tasks that a linear algorithm cannot and it is usually reliable for highly dynamic and non-linear processes. However, we have preferred not to use NNs on this occasion, since among its documented limitations are [38]:

- The resulting models are black boxes for which the association between features and classes may not be easily described;
- The training process is typically lengthy, especially using the batch strategy;
- Tuning the configuration parameters, such as learning rate or selection of the activation method, can be time consuming;
- Estimating the minimum size of the training set to get accurate results is not obvious.

Alternatively, other non-parametric models, such as Support Vector Machines (SVM) and GP, can overcome these deficiencies and provide efficient and powerful classification algorithms that are capable of dealing with high-dimensional input features and with theoretical bounds on the generalization error and sparseness of the solution provided by statistical learning theory. SVM and GP models are non-parametric models based on kernel functions and Gram matrices. The advantages of GP with respect to SVM is that their predictions are truly probabilistic and they provide a measure of the output uncertainty. In addition, there exist algorithms for GP hyperparameter learning, but not for the SVM framework [39]. GPs are Bayesian non-parametric models that are becoming more popular for their capability to capture highly non-linear data relationships in tasks such as dimensionality reduction, time series analysis, and novelty detection [40]. More specifically, GP has the following strong aspects [41]:

- The prior specification of covariance function enables to accommodate a wide class of non-linear regression functions, while prior knowledge about the regression function can be incorporated;
- The model can be easily applied to address regression problems with multidimensional functional covariates;
- The model provides a natural framework for modelling a non-linear regression function and a covariance structure, simultaneously.

A GP, denoted by \( f \rightarrow GP(m, k) \), is fully specified by its mean function \( m(x) \) and covariance function \( k(x, x') \). It is a natural generalization of the Gaussian distribution, whose mean and covariance are a vector and a matrix, respectively. The main objective of using a GP is to derive certain rules on how to predict future data based on the training. Therefore, we should determine a function (after training the model) that makes accurate predictions on future data. Typically, there is vague prior information, so we use a prior hierarchical where the mean and covariance functions are parameterized in terms of hyperparameters.

First, we need a prediction function whose parameters must be optimized in order to minimize the prediction error. For this sake we use the logarithmic marginal probability,

\[
L = \log p(y|x, \Theta) = -\frac{(y-\mu)\gamma}{2\gamma} - \log (|\gamma|)^{\frac{1}{2}} (2\pi)^{\frac{n}{2}}, \tag{1}
\]

where \( \log p(y|x, \Theta) \) is the logarithmic marginal probability of \( y \) with respect to \( x \) given the hyperparameters \( \Theta \). In this case, \( y \rightarrow GP(m, k) \) and \( \mu = m(x), i = 1, 2, \ldots, n \) is the length of the training set, \( \gamma \) is the covariance matrix, \( (\cdot)^T \) is the matrix transpose. Note that (1) contains two terms, the first, \( (y-\mu)(y-\mu)^T/2\gamma \), depends on the training set as \( \mu \) and \( \gamma \) are calculated based on \( n \) training samples. The second, \( \log (|\gamma|)^{\frac{1}{2}} (2\pi)^{\frac{n}{2}} \), is a logarithmic term which penalizes the complexity and normalizes the model, as it depends on \( n \). The trade-off between penalty and adjustment of the data in the GP model is implicit. There is no weight parameter that needs to be tuned by an external method, such as cross-validation. This is a very convenient feature, as it simplifies training [31].

Now, we can find values of hyperparameters that optimize \( L \) based on their partial derivatives, \( \delta L/\delta \Theta_m \) and \( \delta L/\delta \Theta_k \), where \( \Theta_m \) and \( \Theta_k \) are the hyperparameters of the mean and covariance functions, respectively, and \( (\Theta_m, \Theta_k) \in \Theta \). Figure 3 contains a flowchart showing the hyperparameters associated with the proposed scheme and where GP is used in it. The logarithmic marginal probability (L) is used to take a measure about the difference between the predicted (expected) and empirical (observed) values of silence duration. Therefore, L is a reference used to optimise the hyperparameters values in the Gaussian process regression [31]. The values that minimise the error (\( \|L\| \)) are those used for the posterior stage of prediction.

![Figure 3: Gaussian process](image)

**Figure 3**: Gaussian process
PROPOSED SCHEME

The proposed method aims at predicting future silence periods, in order to control the DRX mechanism, saving energy without compromising the QoE. In order to achieve this goal, first we describe the mean and covariance functions that are used to characterize the statistics of the voice activity as follows in Figure 3, and thus define an appropriate GP model. Then, such model is used in an algorithm that, given a predefined reliability parameter, is able to determine how long a silence period should last. This algorithm can therefore control a DRX mechanism that is able to operate the trade-off between power consumption and QoE.

4.1 Mean and covariance functions

As mean function \( m \) to describe the behaviour of the silence periods, we use the one suggested by ITU Recommendation P.59 [42], whose parameters were estimated for silence periods of up to 5s. This Recommendation has been recently used in [43] to model the interactivity of real-world conversations with distinct levels of features and different transmission delays, proving its accuracy from an empirical study. By this mean we model the probability density function (PDF) of pause duration by two weighted geometric PDFs:

\[
m(x) = D_1 (1 - W_1)^{\rho_{x} - 1} + D_2 (1 - W_2)^{\rho_{x} - 1},
\]

where \( x = 1, 2, 3, \ldots, 5000 \) and

\[
D_1 = 0.76693 \quad W_1 = 0.89700 \\
D_2 = 0.23307 \quad W_2 = 0.99791,
\]

so that the cumulative probability distribution function \( \rho \) according to [42] can be represented as

\[
\rho(t) = 1 - \exp\left( - \frac{(T - \delta)}{(T - \delta)} \right),
\]

where \( T \) is the average duration of pauses and \( \delta \) represents the minimum pause time as to consider it a silence period. Therefore, \( T \) and \( \delta \) are the hyperparameters of the mean function that are used to optimize \( L \), in order to later find the function that predicts silence periods. For that sake, we use (1) to calculate the marginal probability and determine the generalization error; then we proceed to minimize the objective function \( L \) using optimization techniques based on the gradient [31].

As covariance function, we consider a Matern 5/2 kernel [31], widely used in Machine Learning to define statistical covariance between values at two points,

\[
k(y; \mu) = \sigma_0^2 \left( 1 + \frac{\sqrt{5}r}{\lambda} + \frac{\sqrt{5}r^2}{3\lambda^2} \right) \exp\left( -\frac{\sqrt{5}r}{\lambda} \right),
\]

where \( r = \sqrt{(y - \mu)^T(y - \mu)} \) is the Euclidean distance between \( y \) and \( \mu \), \( \sigma_0 \) is an amplitude and \( \lambda \) a scale hyperparameter that determines the relevance of the input parameter. A kernel (or covariance function) describes the covariance of the Gaussian process random variables. Together with the mean function the kernel completely defines a Gaussian process. As \( m(x) \) we can utilize any real function, however for \( k(y; \mu) \) the function should be positive definite, which implies that the matrix should be symmetric, hence invertible. The same apply for kernels, therefore any kernel function can be used as covariance function [31]. Then, we can state that \( k(x, y) = f(x) \cdot f(y) \), where \( k \) is the kernel function, \( x, y \) are \( n \) dimensional inputs, \( \cdot \) denotes the dot product, and \( f(.) \) is a map from \( n \)-dimensional to \( m \)-dimensional space such that \( m \gg n \).

Matern 5/2 kernel\(^1\) is a twice differentiable covariance. It is commonly used to define the statistical covariance between measurements made at two points that are \( d \) units distant from each other. Since the covariance only depends on distances between points, it is stationary. A lot of time this kernel works better than the standard Gaussian kernel as it is “less smooth”. It is considered the simplest and natural covariance, in spatial statistics, that works well in \( \mathbb{R}^2 \).

The partial derivatives of (4) and (5) are calculated with respect to the hyperparameters in order to obtain the values that optimize the objective function and start the training. The optimization problem is formulated as follows (See Figure 3):

\[
\hat{T}, \hat{\delta}, \hat{\sigma}_0, \hat{\lambda} = \arg \min \mathbb{V}(\hat{T}, \hat{\delta}, \hat{\sigma}_0, \hat{\lambda}) ||L||,
\]

where \((\hat{T}, \hat{\delta}, \hat{\sigma}_0, \hat{\lambda}) \in \Theta_w \) and \((\hat{\sigma}_0, \hat{\lambda}) \in \Theta_k\). The values of \( \hat{T}, \hat{\delta}, \hat{\lambda} \) and \( \hat{\sigma}_0 \) that meet this condition are calculated using the partial derivatives of \( L \) with respect to each of them

\[
\frac{\partial L}{\partial \hat{T}} = \frac{(\hat{T} - t)}{(\hat{T} - \delta)^2} \exp\left( -\frac{\delta - t}{\hat{T} - \delta} \right),
\]

\[
\frac{\partial L}{\partial \hat{\delta}} = -\left( \frac{T - t}{(\hat{T} - \delta)^2} \right) \exp\left( -\frac{\delta - t}{\hat{T} - \delta} \right),
\]

\[
\frac{\partial L}{\partial \hat{\sigma}_0} = 2\sigma_0 \left( \frac{\sqrt{5}r}{\lambda} + \frac{\sqrt{5}r^2}{3\lambda^2} + 1 \right) \exp\left( -\frac{\sqrt{5}r}{\lambda} \right),
\]

\[
\frac{\partial L}{\partial \hat{\lambda}} = \frac{\sigma_0^2 \sqrt{5}r}{\lambda^2} \left( \frac{\sqrt{5}r}{\lambda} + \frac{\sqrt{5}r^2}{3\lambda^2} - \frac{2r}{\lambda} \right) \exp\left( -\frac{\sqrt{5}r}{\lambda} \right),
\]

where \( y = k(y; \mu), i = 1, 2, \ldots, n \) is the covariance considering the training set. Then, solving the following systems of two

\(^1\) A second order function represented as a second degree polynomial with an exponential, so the degree of the function is \( 2 + 1/2 = 5/2 \).
equations with two variables each,

\[
\begin{align*}
\frac{\partial L}{\partial \gamma} &= -y^{-1}(y - \mu) \gamma \frac{\partial \rho}{\partial \gamma} , \\
\frac{\partial L}{\partial \delta} &= -y^{-1}(y - \mu) \gamma \frac{\partial \rho}{\partial \delta},
\end{align*}
\tag{11}
\]

and

\[
\begin{align*}
\frac{\partial L}{\partial \gamma} &= \frac{y^{-1}}{2} \frac{\partial \gamma}{\partial \gamma} \left( 1 + (y - \mu)(y - \mu)^T \frac{\partial \gamma}{\partial \gamma} \right), \\
\frac{\partial L}{\partial \delta} &= \frac{y^{-1}}{2} \frac{\partial \delta}{\partial \delta} \left( 1 + (y - \mu)(y - \mu)^T \frac{\partial \delta}{\partial \delta} \right),
\end{align*}
\tag{12}
\]

we obtain the trained model \( f \rightarrow \mathcal{G}(m, k) \). The procedure we used to solve these systems of equations was as follows. In either (11) or (12), we fix the variable \( \delta, \Delta \) and calculate the gradient by varying \( \gamma, \Delta \) (where \( \gamma \rightarrow f \)), then with the value of \( \gamma \), \( \Delta \) that minimizes \( L \) we vary \( \delta, \Delta \) to find the global minimum. Finally, with these values the mean and covariance functions used for the prediction are obtained.

### 4.2 Silence duration prediction

Next, based on the GP model in Subsection 4.1, we describe an algorithm that, when a silence period is detected, predicts how long such period should last under a certain statistical confidence.

Let us assume that \( X \) denotes the length of the silence period, and \( P(X \leq \Gamma) = \rho(\Gamma) \) is the probability that the silence period is less than \( \Gamma \), while the probability that a silence period is longer than \( \Gamma \) is \( 1 - \rho(\Gamma) \). Assuming that the silence period has already lasted for \( \Gamma \) seconds, then the conditional probability that this silence period lasts for longer than \( \Gamma + \Delta \) is

\[
P(X > \Gamma + \Delta | X > \Gamma) = \frac{1 - \rho(\Gamma + \Delta)}{1 - \rho(\Gamma)}. \tag{13}
\]

Using \( \rho \), the proposed Algorithm 1 is able to predict appropriate values of \( \Delta \) (new increase in time) that give a conditional probability greater than or equal to a given confidence interval \( \beta \). The inputs of the proposed algorithm are the confidence \( \beta \), the current silence duration \( \Gamma \), and the trained GP model \( f \rightarrow \mathcal{G}(m, k) \). Once a silence period is detected, the algorithm calculates how long it is expected to last \( \Delta \) and the OFF time is extended to \( \Gamma = \Gamma + \Delta \), updating the value in the cumulative distribution function. The Algorithm 1 continues calculating values of \( \Delta \) until \( \Delta > 0 \) cannot be found, either because the algorithm is not able to find a value of \( \Delta \) with a probability of occurrence \( \beta \) or that the maximum of \( f \) is reached, then the algorithm sets the RFI ON.

The model keeps updating the mean and covariance functions by maintaining the training process, but using the new incoming data as to achieve greater precision and adapt the system to changes in the traffic distribution. Note that if the training values of the successive calls are used in the continuous training process, significant energy values can be saved by maintaining a high fidelity, therefore in the proposed scheme we use the prior knowledge of silence periods to predict the future silence based on this self-similar behaviour.

The proposed Algorithm 1 can be used to control a DRX mechanism as to achieve a given trade-off between average power consumption and QoE. For that sake we determine the values of \( \hat{\beta} \) and \( \hat{\Delta} \) that minimize the energy consumption\(^2\) while keeping a fixed QoE value \( \hat{Q}_f \). Such problem can be formulated as

\[
\hat{\Delta}, \hat{\beta} = \arg \min_P \frac{\Delta \geq 0}{\min \{ \text{PLR} \}} \quad \rho(\Gamma + \Delta) \leq \hat{Q}_f
\]

where \( P \) is the average power consumption. Then, from the fixed QoE value we calculate the R factor, obtaining packet loss rate (PLR) values for reaching that QoE through Eqn. (19).

Next, we obtain \( \hat{\beta} \) and \( \Delta \) from the PLR and the CDF since, in last instance, they are inversely proportional, \( P_{\text{error}} \) (subject to PLR\(^3\)) is calculated as follows:

\[
P(\Gamma < (\Gamma + \Delta)| \Gamma) = \rho(\Gamma + \Delta) - \rho(\Gamma).
\]

where \( \hat{\beta} \) is obtained from \( \Delta \) as a simple calculation from [13]. Processing delays do not affect the performance of our scheme, since due to the low computational cost required these small delays are absorbed by the playback buffer.

### 5 PERFORMANCE METRICS

In this section we discuss two performance metrics utilized to assess the proposed method and to compare it with competing

\(^2\)Note that \( \Delta \) depends on \( \beta \), and \( \beta \) varies as we move through the CDF. Furthermore, \( \beta \) depends on the silence history.

\(^3\)Note that 15 is used to predict the duration of silence and does not consider the effect of the playback buffer, therefore in terms of QoE our proposal has an even higher performance than expected.”
schemes from the literature: energy consumption and QoE. Additionally, in order to confirm the accuracy of the predictor, we calculate the root mean square error (RMSE) as a metric of how close are the expected values from the observed empirical ones. The error is obtained as follow:

\[ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (f_i - f_o)^2}, \]  

where \( f_i \) are the predictions (expected values), \( f_o \) are the observed values (known results) and \( N \) is the sample size.

### 5.1 Energy consumption

The energy consumption of mobile devices depends on different components such as active processes, display, and transceiver. For the sake of simplicity, we adopt a simplified model that refers only to the energy consumption of the radio frequency (RF) interface, based on [44]. The RFI consumes the most energy in the active state, during which it is transmitting or receiving packets. In the OFF state, on the other hand, most RF circuits are turned off to reduce energy consumption, leading to a much smaller consumption.

Moreover, our goal is to analyze the energy savings due to the proposed scheme, not the specific energy consumption values. Hence, we use a relative energy consumption model, which takes into account the percentage of the total time that RFI is ON. Note that this consideration allows a fair and effective comparison between different DRX mechanisms, in which energy savings are achieved precisely by turning OFF the RFI. Then, first we compute \( T_a \), the time in which the RFI is ON, as

\[ T_a = \sum_{i=1}^{N_i} t_{\omega}, \]  

where \( t_{\omega} \) is the ON time per DRX cycle (time interval between the beginning of two instants \( t_{\omega} \)) and \( N_i \) is the number of cycles that are executed during the call. If \( T_r \) is the total call time, then

\[ E = \frac{T_a}{T_r}, \]  

is a normalized measure of energy when using a DRX mechanism. The smaller \( E \), the larger the energy savings.

### 5.2 Quality of experience

In recent years, QoE has become fundamental in telecommunications, taking into account the evaluation of human experience when interacting with technology in a particular context [45]. Although QoE depends on quality of service (QoS) metrics like packet loss and latency, a good (bad) QoS does not necessarily mean a good (bad) QoE [46]. QoS metrics are specified by the service and not by the user [47], while the QoE approach focuses on the user [48]. The QoE leads to a better understanding of user satisfaction with a service in a given context. In [49], the authors discuss why QoE assessment occupies a key role in various multimedia networks and applications and why it is a challenging problem.

We use the ITU-T Model E [50] to estimate VoIP session QoE by using a psycho-acoustic factor \( R \in \{0, 100\} \), with the modifications proposed in [51] to take into account the effect of burst losses. Then, the rating factor \( R \) is a scalar value defined as a linear combination of the individual impairments and is given by (19):

\[ R = R_0 - I_e - I_d - I_{eff} + A, \]  

where \( R_0 \) is the signal to noise ratio (SNR) factor, \( I_e \) represents the impairments that occur simultaneously with the voice signal, \( I_d \) is the delay factor, \( I_e \) is the equipment impairment factor, caused by the codec and packet loss, and \( A \) is the advantage factor. Moreover, \( P_{pl} \) is the packet loss ratio (PLR) considering the burst effect, \( B_{pl} \) is the robustness factor relative to packet loss. To calculate \( I_d \) we follow [52]:

\[ I_d = 2.4 \times 10^{-5} D_t + 0.11 d_0 H(d_0); \]  

where \( D_t \) is the average packet delay, \( d_0 = (D_t - 177.3) \) [ms], and

\[ H(d_0) = \begin{cases} 0 & \text{if } d_0 < 0 \\ 1 & \text{if } d_0 \geq 0. \end{cases} \]  

In order to calculate the \( R \) factor we take typical impairments and advantage values \( (I_e, A, B_{pl}, I_e, R_0) \) from [52]. See Table 4.

| Parameter | Value | Ref. |
|-----------|-------|------|
| \( t_{\omega} \) (On time) | 2, 3, 4, 5, 6 (TTI) | [29] |
| \( t_{\omega} \) (Short cycle time) | 16, 20, 32, 40, 64 (TTI) | [29] |
| \( t_{\omega} \) (Long cycle time) | 40, 64, 80, 128, 160 (TTI) | [29] |
| \( I \) (Inactivity Timer) | 50 (TTI) | [29] |
| \( \alpha \) (Step) | 2 to 16 | |
| \( \beta \) (Observed silence time) | 50 ms | [22] |
| \( \lambda \) (Prediction probability) | 0.25 to 0.60 | [22] |
| \( I_e \) (Voice signal impairments) | 0 | [52] |
| \( I_e \) (Equipment impairments factor) | 10 | [52] |
| \( A \) (Advantage factor) | 0 | [52] |
| \( B_{pl} \) (Robustness factor) | 14.1 | [52] |
| \( R_0 \) (SNR factor) | 129 | [52] |
Following (19), (20) and (21) we can see that the $R$ factor depends on the end-to-end delay and the PLR. Then, let us analyze traces in Figure 4. The traces are representative of the traffic behaviour in LTE mobile networks under different channel conditions. These traces are a mimic for essential characteristics of real communications in various networks from United States and Korea at different times and traffic patterns, which characterize the diversity of situations in LTE networks when handling voice traffic. These same four traces have been selected by 3GPP for use by that body for testing under channel impairments [53]. The study of these traces using histograms shows that the 76% of the active periods have duration below 200 ms, while for silence periods the duration of the 82% of the total silence bursts are lower than 500 ms. After analyzing these traces, we obtain that the mean delay is between 110 and 168 ms. The jitter also varies, although it can be seen that in all traces self-similar behaviour occurs. We assume in the subsequent analysis that packet loss is only caused by the excessive delay effect due to the DRX mechanism. Finally, since the traces constitute a continuous flow of packets, we insert silence periods following ITU recommendation P.59 [42], obtaining the resulting traces in Figure 4(b).

The behaviour of the $R$ factor as a function of the delay can be seen in Figure 5. End-to-end delay of less than 180 ms have little influence on the $R$ factor, however values that exceed 180 ms have a large impact. According to [54], for delays beyond 280 ms it is not possible to guarantee user satisfaction.

It is well known that audio packets must be played back at the receiver at regular intervals [55] and that, to achieve this, delay variation must be attenuated in the playback buffer. As a consequence, this step also adds delay in the process. Audio packets find an end to end ($t_e$) delay, which is composed by several factors, the variable factors are the network delay ($d_N$) and the buffer delay ($d_{bf}$). If the path is the same (without changes), $t_e$ is asymptotically constant:

$$t_e = d_{fc} + d_N + d_{bf} + d_{dec}$$

where $d_{fc}$ is the transmission delay (including encoding and look ahead delay), $d_N$ is the network delay, $d_{dec}$ is the decoding delay and $d_{bf}$ is the jitter removal buffer delay, which is variable, both $d_{bf}$ and $d_{dec}$ are part of the reception delay. The traces in [53] include $d_{fc}$, which is a constant delay, and $d_N$, which is variable. Then, in the receiver side, $d_{dec}$ is constant and has a very small value comparing with $d_{bf}$ and $d_N$, so we can dismiss it. We consider a fixed buffer size, large enough to attenuate the delay variations. Therefore, $t_e$ is asymptotically constant because $d_N + d_{bf}$ is asymptotically constant.

Based on the above considerations, the mouth to ear delay, $d_{MTE}$, is modelled as the typical $t_e$ delay plus the additional time while the RFI at the mobile is OFF. Note that while the RFI is turn OFF, the packets that are about to be sent cannot be delivered and remain in the system. Therefore, what determines their true late arrival is the playback time conditioned by the size of the playback buffer. In this work, the maximum tolerable delay ($d_{MAX}$) is considered to be of 240 ms, which according to [54] is an adequate threshold for user satisfaction, then $d_{MTE}$ for
packets to be reproduced at the UE is enclosed by $d_{MTE}$, and packets with $d_{MTE} > d_{MLV}$ are considered lost. Note that this threshold must be greater than the maximum delay in the traces and as close as possible to this value. If the delay is taken to be too small, we obtain losses that are not inherent to the DRX mechanism itself and cannot be avoided in any way. Conversely, taking too large a threshold value can directly affect QoE. Packets with a delay greater than this threshold are discarded to avoid QoE degradation. Then, we calculate the PLR as follows:

$$\text{PLR} = \frac{1}{N} \sum_{i=1}^{N} Q(d_{p_i})$$

where

$$Q(d_{p_i}) = \begin{cases} 0 & \text{if } d_{p_i} < d_{\text{max}} \\ 1 & \text{if } d_{p_i} \geq d_{\text{max}} \end{cases}$$

where $d_{p_i}$ is the total delay for each $p_i$ and $N$ is the sample size.

Finally, we calculate the corresponding mean opinion score (MOS) value, which provides a subjective measure of the impact that service failures have on the users. The MOS takes values between 1 and 5, with 1 being the lowest QoE and 5 the highest. The MOS value can be calculated from the R factor as follows:

$$\text{MOS} = \begin{cases} 1 & \text{if } R_x \leq 0 \\ F(R_x) & \text{if } 0 < R_x < 10^2 \\ 4.5 & \text{if } R_x \geq 10^2 \end{cases}$$

where

$$R_x = \frac{R}{1.29}$$

and

$$F(R_x) = 1 - 7 \times 10^{-6}(R_x^2 - 160R_x^2 + 105R_x).$$

The QoE is calculated at regular intervals in order to maintain link quality at all times. If at any time the quality falls, it is a consequence that the model describing the silence made wrong predictions or that anomalous traffic was present. To avoid miss-detection, the system stops the event prediction and returns to the first stage to derive the prediction function again. In order to validate the proposed scheme, the results obtained by simulating this algorithm are discussed in the next section. The evaluation also includes previous published algorithms for comparison purposes.

5.3 Practical implementation considerations

Switching between OFF/ON states is a regular procedure, however, the transition states result in delays that lead to an increase in power consumption. Nevertheless, without losing the generality in this paper we assume for all the algorithms that the transitions between OFF and ON states are instantaneous, what implicate an error in the computation of power consumption, but guarantee fair comparisons between them.

P59 [42] allows times for silence below 40 ms, but in this paper we only consider as silence periods those with a duration of 40 ms or higher. Therefore, in all the algorithms we dismiss silence periods longer than 40 ms. Certainly, there is an error in the response times, however, following [56] the switching times for the DRX mechanism are in ranges of just 1 ms, then at worst, the resulting error in assuming instantaneous transitions for minimum silence periods is about 1.25%, which is small compared to the other calculations.

On other hand, although we perform the algorithms in a synchronized way, where the receiver clock is locked to that of the transmitter; we would like to emphasize that the effect of offset due the lack of synchronism on the subjective quality is minimal because the time scales involved are fairly long [57], representing a negligible error, around 0.01%.

6 PERFORMANCE EVALUATION

In this section, we analyze the performance of the method proposed in this work. Direct numerical comparisons are made against [29] and [22], since these works are widely recognized in this area of study. The work in [29] aims at saving as much energy as possible, while in [22] the focus is on predicting silence periods. Thus, it is pertinent to compare the accuracy of our predictor with the one in [22] while at the same time comparing the energy saving capability of our system with that in [29].

In the DRX mechanism presented in [29] the system is ON for a time $t_{off}$ with a combination of short and long sleeping cycles of duration $t_s$ and $t_l$. The scheme starts in the ON state and switches OFF according to the traffic characteristics. If packets are not received during an interval $T_T$, then it goes to a short cycle OFF state. If packets are received, then it returns to the ON state, otherwise the short cycle counter is incremented. If $n_s$ short cycle states occur in a row, then the system enters into a long cycle state, in which it takes longer to wake up and check if packets are being received. If packets are received, the system returns to the ON state. In the following, the timers are set according to [29] and are listed in Table 4. Moreover, in [22] the authors consider the empirical cumulative distribution function (ECDF) obtained from the analysis of previous traces to predict how much time the system can be OFF. At the beginning of a silence period, it turns OFF the RFI for a predefined time $\alpha$, and then predicts the length of the silence period using the ECDF. Herein, the observed silence time $\alpha$ is set to 50 ms (as recommended in [22]) while $\beta$ varies depending on the curve of the QoE/saving trade-off. All the parameters used in the simulation are defined in Table 4, unless otherwise specified.

To perform the evaluation, we use traces from [53], available in [21] and which were obtained from real-world call logs of RTP packet arrival times collected in different wireless networks in South Korea and the United States, in combination with
ITU Recommendation P.59 [42] to characterize silence periods. We use the GP model in Subsection 4.1 to obtain a curve that describes the behaviour of the silence periods along the traces, using 75% of the sample data to perform training while the remaining 25% are used for testing. All algorithm programming and data analysis was carried using MatLab [58].

6.1 Numerical results

In Figure 6 we illustrate the performance of the proposed method in predicting the statistics of silence periods, using the VoIP traces available in [21]. Figure 6(a) models the cumulative distribution function (CDF) of silence periods from predicted and empirical results. We calculate the CDF as follows:

\[ \rho(x) = P(X \leq x), \]  

(29)

where the right-hand side represents the probability that a random variable \( X \) takes on a value less than or equal to \( x \). Note that empirical CDF is calculated from (29) as well but with the empirical values obtained from the reading of the traffic pattern and not from the prediction of the traffic pattern. Then, in order to confirm the accuracy of the predictor, we calculate the RMSE from (16).

From the visual inspection of Figure 6(a) it is clear that the proposed method can predict well the statistical behaviour of silence periods. Moreover, Figure 6(b) illustrates the error between predicted and actual silence periods per audio packet, confirming the good accuracy of the proposed method. Moreover, the root mean square error (RMSE) of the predictions is as low as 1.87%.

Next, we utilize traces from [53], which have different characteristics in terms of delay, jitter and PLR, in order to evaluate the performance of the proposed algorithm and compare it to [22, 29]. We adjust the parameters of each mechanism estimating the MOS. The range of permissible MOS values varies from 3.6 to 4.5, (where users evaluate the communication as satisfactory), to determine the trade-off curve between QoE and energy savings. Figure 7 shows the QoE/consumption curves of each of the algorithms\(^4\). The theoretical limit in Figure 7 assumes that the system is able to know exactly when the silence period begins and how long it will last, being capable of handling the OFF/ON states to locate the lower bound of the energy consumption without exceeding the established maximum threshold of packet losses. Then, while keeping a fixed QoE value:

\[ D_e = d_{max}, \]

\[ P_{pl} = \text{PLR}_{Q_e}, \]

subject to

\[ \text{QoE} = Q_e, \]

(30)

where \( D_e \) and \( P_{pl} \) were introduced previously, \( d_{max} \) and \( \text{PLR}_{Q_e} \) are the maximum possible values subject to the condition. Note that the theoretical limit is unreachable for our scheme since

\(^4\)Note that the different values of MOS (QoE) and power consumption are obtained by varying the corresponding parameters (e.g. Table 4 for [29] and [22], and adjusting \( \beta \) for Gaussian process).
the prediction of silence is made from the probability of silence learned from the traffic pattern, then our algorithm is rigorous with respect to maintaining QoE and sacrifices the savings in the presence of high probability of error.

Figure 7 considers a normalized measure of energy consumption, with respect to the case when the RFI is turned ON all the time, versus the target MOS. From the results it is clear that the proposed scheme significantly reduces the consumed energy to achieve a certain MOS when compared to [22, 29]. This is because the proposed mechanism is able to save a greater amount of energy during silence periods, by adequately estimating their occurrence and duration from the characteristics of the flowing traffic. For example, to reach a MOS equal to 3.6 the proposed method consumes about 2.25% of energy, while [22, 29] exceed 3%, leading to a relative saving of approximately 34.5%. As we increase the target MOS value, the advantage of the proposed scheme becomes even more evident. For MOS values greater than 4.1 the energy consumption in [29] increases exponentially, while when [22] operates at a MOS larger than 4.25 the consumption increases rapidly.

To facilitate visual comparison in terms of energy consumption, Figure 8 shows the energy consumption of each scheme as a function of the MOS limit values used in the simulations. We also use an intermediate value and the average consumption as a function of the MOS limit values used in the simulations.

FIGURE 8 Energy consumption vs. quality

7 CONCLUSION

In this work we introduced a method to predict silence periods through ML, reducing energy consumption with respect to previous schemes without deteriorating QoE, while predicting silence periods with an RMSE of less than 2%. We were able to solve a very complex optimization problem, the trade-off between energy savings and QoE, by introducing ML mechanisms, allowing to characterize quite accurately the behaviour of silence periods in a VoIP session to reduce the energy consumption in more than 30% without delays in processing complexity affecting the performance of our scheme due to the low computational cost required, while ensuring a desired QoE level when predicting future silence periods in which the device does not need to remain in constant ON state.

In future work we will focus on two-way conversation with a four state model to characterise the voice patterns, as well as on interaction and integration between speech/voice into IoT applications. In addition, traffic-aware DRX schemes for machine-type communications will be proposed to improve energy efficiency.

ACKNOWLEDGMENTS

This work has been partially supported in Brazil by CNPq, as well as in Chile by ANID FONDECYT Iniciación No. 11200659, SCC-PIDI-UTEM FONDEQUIP-EQM180180, Fund of Scientific and Technological Equipment code L318-04 (UTEM), and Collaborative Research Activities between PIDI/UTEM and FIE/UCLV.

REFERENCES

1. Zhang, S., et al.: First 20 Years of Green Radios. IEEE Transactions on Green Communications and Networking 4(1), 1–15 (2019)
2. Tariq, F., et al.: A speculative study on 6G. arXiv:1902.06700 (2019)
3. Samdanis, K., Taleb, T.: The road beyond 5G: A vision and insight of the key technologies. IEEE Netw. 34(2), 135–141 (2020)
4. Auer, G., et al.: How much energy is needed to run a wireless network?. IEEE Wirel. Commun. 18(5), 40–49 (2011)
5. Yang, S.-R., Lin, Y.-B.: Modeling UMTS discontinuous reception mechanism. IEEE Trans. Wirel. Commun. 4(1), 312–319 (2005)
6. Gödör, I.: Revitalizing DRX for enhanced mobile sleep modes. Comput. Commun. 138, 98–105 (2019)
7. Rosenberg, J., et al.: SIP: session initiation protocol (2002). RFC 3261, RFC Editor. https://tools.ietf.org/html/rfc3261 (2002) [2020-05-15].
8. Schulzrinne, H., et al.: RTP: a transport protocol for real-time application. RFC 3550, RFC Editor. https://tools.ietf.org/html/rfc3550 (2003) [2020-05-15].
9. Mehra, P., Udani, S.: Voice over IP. IEEE Potentials 20(4), 36–40 (2001)
10. Pepper, R.: “The History of VoIP and Internet Telephones”, Retrieved from https://getvoip.com/blog/2014/01/27/history-of-voip-and-internet-telephones/ [2020-05-15].
11. Biondi, P., Bognaresi, S., Bella, G.: Voip can still be exploited-badly. In: Fifth International Conference on Fog and Mobile Edge Computing (FMEC), Paris (2020)
12. Montazerolghaem, A., Yaghmaee, M.H., Leon-Garcia, A.: Green cloud multimedia networking: NFV/SDN based energy-efficient resource allocation. IEEE Transactions on Green Communications and Networking, 4(3), 873–889 (2020)
13. Ahmad, I., Ranka, S.: Handbook of Energy Aware and Green Computing, CRC Press, Boca Raton (2012)
14. Bockelmann, C., et al.: Massive machine-type communications in 5G: physical and mac-layer solutions. IEEE Commun. Mag. 54(9), 59–65 (2016)
15. Mahmood, N.H., et al.: Six Key Features of Machine Type Communication in 5G. In: 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi (2020)
16. Qi, Y., et al.: Semi-persistent RRC protocol for machine-type communication devices in LTE networks. IEEE Access 3, 864–874 (2015)
17. Rostami, S., et al.: Optimized wake-up scheme with bounded delay for energy-efficient MTC. In: IEEE Global Communications Conference (GLOBECOM), Waikoloa (2019)
18. Zhou, J., et al.: Online learning-based discontinuous reception (DRX) for machine-type communications. IEEE Internet of Things Journal 6(3), 5550–5561 (2019)
19. Brown, A.: The role of voice in IoT applications, Whitepaper, Strategy Analytics, Newton (2016)
20. Karaagac, A., et al.: Viot: Voice over Internet of Things. In: 2020 Global Internet of Things Summit (GIoTS), Dublin (2020)
21. EVS permanent documents. 2015. https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/EVS_Permanent_Documents [2020-05-15].
22. Szabó, G., et al.: Service aware adaptive DRX scheme. In: IEEE Globecom Workshops, Austin (2014)
23. Memon, M.L., et al.: Deep-DRX: A framework for deep learning-based discontinuous reception in 5G wireless networks. Transactions on Emerging Telecommunications Technologies 30(3), e3579 (2019)
24. Memon, M.L., et al.: Deep-DRX: A framework for deep learning-based discontinuous reception in 5G wireless networks. Transactions on Emerging Telecommunications Technologies 30(3), e3579 (2019)
25. Navarro-Ortiz, J., et al.: A survey on 5G usage scenarios and traffic models. IEEE Communications Surveys & Tutorials 22(2), 905–929 (2020)
26. Zhang, X., et al.: A covert channel over volte via adjusting silence periods. IEEE Access 6, 9292–9302 (2018)
27. Shin, W., Lee, J.-R., Choi, H.-H.: Energy-delay tradeoff analysis and enhancement in LTE power-saving mechanisms. In: IEEE International Conference on Consumer Electronics (ICCE), Berlin (2013)
28. Szabó, G., et al.: Service aware adaptive DRX scheme. In: IEEE Globecom Workshops, Austin (2014)
29. Tseng, C.-C., et al.: Delay and power consumption in LTE/LTE-A DRX mechanism with mixed short and long cycles. IEEE Trans. Veh. Technol. 65(5), 1721–1734 (2016)
30. Simonovici, A.-M., et al.: VoIP traffic modelling using gaussian mixture models, gaussian processes and interactive particle algorithms. In: 2015 IEEE Globecom Workshops, San Diego (2015)
31. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning, MIT press, Cambridge (2006)
32. Zoëpf, R.: Real-time transport protocol (RTP) payload for comfort noise (CN). RFC 3389, RFC Editor. https://tools.ietf.org/html/rfc3389 (2002) [2020-05-15].
33. Lin, X.-H., et al.: On exploiting the on-off characteristics of human speech to conserve energy for the downlink VoIP in WiMAX systems. In: IEEE 7th International Wireless Communications and Mobile Computing Conference, Istanbul (2011)
34. Angarita-Zapata, J.S., Masegosa, A.D., Triguero, I.: A taxonomy of traffic forecasting regression problems from a supervised learning perspective. IEEE Access 7, 68 185–68 205 (2019)
35. Lázaro, G., Georgy, O., Harro, W., Machine Learning for Financial Engineering, Vol. 8 World Scientific, Singapore (2012)
36. Gorod, A., et al.: Evolving Toolbox for Complex Project Management. CRC Press, Boca Raton (2019)
37. Swastanto, B.A.: Gaussian process regression for long-term time series forecasting. Dissertation. Delft University of Technology (2016)
38. Nicolas, P.R., Scala for Machine Learning. Packt Publishing, Birmingham (2015)
39. Markow, K., Matsui, T.: Music genre and emotion recognition using Gaussian processes. IEEE Access 2, 688–697 (2014)
40. Koriyama, T., Kohayashi, T.: Statistical parametric speech synthesis using deep Gaussian processes. IEEE/ACM Transactions on Audio, Speech, and Language Processing 27(5), 948–959 (2019)
41. Shi, J.Q., Choi, T., Gaussian Process Regression Analysis for Functional Data. CRC Press, Boca Raton (2011)
42. Speech, A.C.: ITU-T recommendation P. 59. ITU, March (1993)
43. Michael, T., Möller, S: Simulating turn-taking in conversations with delayed transmission. In: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 1st virtual meeting (2020)
44. Somasundaram, S., et al.: System level information for discontinuous reception, cell reselection and rach, US Patent 9,173,167, 27 Oct 2015
45. Laghari, K.U.R., Connelly, K.: Toward total quality of experience: A qoe model in a communication ecosystem, IEEE Commun. Mag. 50(4), 58–65 (2012)
46. Mohanan, V.M., Aldmour, I.: Network selection problems-QoE vs QoS who is the winner? Computer Engineering and Applications Journal 4(1), 1–18 (2015)
47. Agbonta, F., Liotta, A.: QoE-aware QoS management. In: Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, Linz (2008)
48. Liout, E., et al.: Quality of experience management in mobile cellular networks: key issues and design challenges. IEEE Commun. Mag. 53(7), 145–153 (2015)
49. Akhtar, Z., et al.: Why is multimedia quality of experience assessment a challenging problem? IEEE Access 7, 117 897–117 915 (2019)
50. ITU- T Recommendation G. 107, The E-model, a computational model for use in transmission planning (2003)
51. Zhang, H., et al.: Packet loss burstiness and enhancement to the E-Model. In: IEEE Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Network, Towsen (2005)
52. Möller, S., et al.: Impairment factor framework for wide-band speech codecs. IEEE Transactions on Audio, Speech, and Language Processing 14(6), 1969–1976 (2006)
53. Atti, V., et al.: Improved error resilience for volte and VoIP with 3GPP EVS channel aware coding. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane (2015)
54. Norell, L., Parsons, E., Synnergren, P.: Telephony services over LTE end-to-end. Eric Parsons and Per Synnergren, Ericsson Review (2010)
55. Kummar, S., Puram, B.M.P.: Bandwidth management. US Patent 10,587,523, 10 Mar 2020
56. Rostami, S., et al.: Wake-up radio based access in 5G under delay constraints: Modeling and optimization. IEEE Trans. Commun. 68(2), 1044–1057 (2019)
57. Ogunjumobi, T., Narasimha, M.J.: Speech over VoIP networks: advanced signal processing and system implementation. IEEE Circuits Syst. Mag. 12(2), 35–55 (2012)
58. MATLAB version 9.3.0.713579 (R2017b), The Mathworks, Inc., Natick, Massachusetts (2017)