A Modern Quality of Service Evaluation Approach of VOLTE Calls Focused on Packet Delays

İsmail Yıldız
Business Intelligence Solutions – Digital Analytics
Turkcell Technology R&D Inc.
Istanbul, Turkey
e-mail: ismail.yildiz@turkcell.com.tr

Abstract — **Long-Term Evolution (LTE) coverage in mobile networks is becoming more widespread day by day. As a consequence of this situation, Voice over LTE (VOLTE) calls in LTE networks have increasingly become more important compared to traditional Circuit Switch (CS) voice calls. Also, smartphone usage grew rapidly in the last few years and VOLTE support is a vital feature of smartphones. Additionally, Mobile Network Operators (MNO) tend to promote VOLTE service since VOLTE technology has lower operational costs than traditional voice service in the long run and promises improved sound quality. Although there are several straightforward Key Performance Indicators (KPIs) such as call drop rate, jitter, packet delay, packet loss rate, etc. in order to measure VOLTE quality of service, evaluation of the measured values and revealing real perceived quality of service by subscribers is complex. Experienced VOLTE quality of service value depends on many factors. Determining factors and their significance is central to measuring and improving perceived VOLTE quality of service. This paper deals with the evaluation problem and proposes a new approach to evaluate perceived VOLTE quality of service experienced by subscribers.**

Keywords— Expected quality of VOLTE call; evaluation approach; packet delay; user experience; perceived quality.

I. INTRODUCTION

It is generally believed that communication between people far away started with smoke in the very early ages. Centuries later, the telegraph appeared as a crucial invention in peoples' lives. In the following decades, other new types of communication devices and technologies were invented such as the telephone, the radio, the computer, the Internet and the cell phone. Finally, smartphones were produced and have become widespread, which was the beginning of a new smart era. In this era, people increasingly tend to use smartphones as small pocket computers and access the Internet with their mobile devices mostly using the Global System for Mobile Communications (GSM) network.

In the smartphone era, almost every person owns a smartphone in order to increase their life quality. Some people use it for its original purpose, which is connectivity, but now smartphones have become more than that: they can be used for social media, navigation, gaming, business, and even health. People can also customize their smartphones according to their needs.

Intensive smartphone and mobile application usage fall short in fulfilling people's download and upload needs. As a result, 4G or LTE technologies have been proposed for data access. Since LTE is an advanced technology and allows higher data consumption and speed, traditional voice calls became less important. Instead, VOLTE technology was invented which can enable better sound quality. VOLTE is a Packet Switch based technology, which differs from the traditional Circuit Switch voice infrastructure.

VOLTE is becoming more popular day by day. MNO companies also encourage customers to use VOLTE for better user experience and lower infrastructural costs in the long term. VOLTE holds the promise of providing users with clear, high-definition call quality combined with much greater spectral efficiency and capacity compared with conventional circuit-switched calls over legacy 2G and 3G networks [4]. This means that it is easier for a caller not only to hear clearly what the person at the other end of the line is saying, but also to detect their tone of voice, giving a much richer overall experience. Global System for Mobile Communications Association (GSMA) announced that, by February 2018, 127 Mobile Network Operators in 63 different countries launched the VOLTE service. Also, there are more than 1218 mobile phone models that support and are ready to use VOLTE services [5].

There are several straightforward KPIs including call drop rate, jitter, packet delay, packet loss rate measuring VOLTE quality of service. But evaluation of measured values and hitting perceived quality of service by subscribers is difficult and these KPIs are not sufficient by themselves. Additionally, there is no globally accepted standard process for this purpose.

In this paper, the problem of VOLTE quality of service evaluation is discussed in detail. Previous market insights and measured values are combined in order to propose a modern approach to solve this problem. Finally, an expected call delay approach is devised and the approach is tested, which leads to successful results.

The rest of the paper is structured as follows. In Section II, we present the problem definition, including supportive figures. In Section III, exploratory data analysis and data preparation are performed. In Section IV, factors affecting experienced call delay by GSM subscribers are investigated in detail. Finally, our work is concluded in

Section V and future directions are discussed in Section VI.

II. PROBLEM DEFINITION

As VOLTE technology is becoming more widespread and gradually replaces traditional voice service, measuring VOLTE and improving quality of service has become a standard requirement for GSM companies. In fact, the traditional approach that is measuring call drop numbers and rates is still valid, but it does not provide continuous scoring of perceived quality of service. Since a call is either dropped or it succeeds, there is nothing to measure between these two extreme sides. Therefore, call drop is not considered in the scope of this study.

The Mean Opinion Score (MOS) method is another voice quality scoring method, which has been used for decades. In MOS method, subjects judge and score the quality of voice ranging from 1 (lowest quality) to 5 (highest quality). An average MOS value is derived from individual subject scorings. However, using the MOS method in the telecom industry has a limitation, namely, it is unclear what threshold values should be applied to identify problems or determine acceptability [3]. Which MOS value is good, or good enough? It is difficult to find the correct threshold value.

VOLTE technology is a packet switching based technology so, generally, KPIs used in data service are also valid for VOLTE service. The most commonly used KPIs for measuring VOLTE quality of service are jitter, packet loss ratio and packet delay during a VOLTE call [1].

Packet delay, which is also called latency in some resources, is the total time it takes a VOLTE packet to travel from a source location to a destination location. Jitter is a measure of fluctuations in packet delays during a VOLTE call. Lastly, packet loss ratio is the ratio of the number of packets that are completely lost and never reached the destination versus the total number of packets sent [8].

Since networking has improved over time and VOLTE is an evolved technology, packet losses, and jitter problems are handled successfully in modern networks. But packet delays still remain a big issue because packets must physically travel from a source location to a destination location and an acceptable travel time should be guaranteed. Among these three basic KPIs, packet delay is used in the scope of this study in order to model user experience.

In the remaining part of the article, we use the term call delay to denote the average of all packets' delay belonging to a single VOLTE call. Although it is easy to define and calculate call delay, it is not so easy to evaluate measured call delay values. It is a complicated issue to make inferences from measured call delay values. As an example, is the measured experience of 300 milliseconds call delay acceptable? What about 250 milliseconds call delay? Is it a good experience or bad experience?

The main problem is that evaluation of measured values (such as 300 milliseconds) is not straightforward and there is no globally accepted reference point. In fact, worldwide references may also not give us the real picture, because each VOLTE network of each Mobile Network Operator has its own parameters, settings and establishment. In order to overcome this evaluation problem, one common methodology is comparing individual subscriber's call delay figures with the whole network's averages.

Fig. 1 shows the average call delay values measured in Turkcell mobile network in November 2017. Because of market competition and commercial limitations, exact call delay values cannot be revealed. Instead, values are multiplied by a private multiplier and general trend is depicted. The unit of multiplied call delay values is defined as "D" in the remaining part of this paper. Regarding Fig. 1, average call delay values vary daily and it is difficult to interpret this figure in terms of experience evaluation. Comparing individual subscriber's call delay figures with the whole network's figure will be a misleading approach.

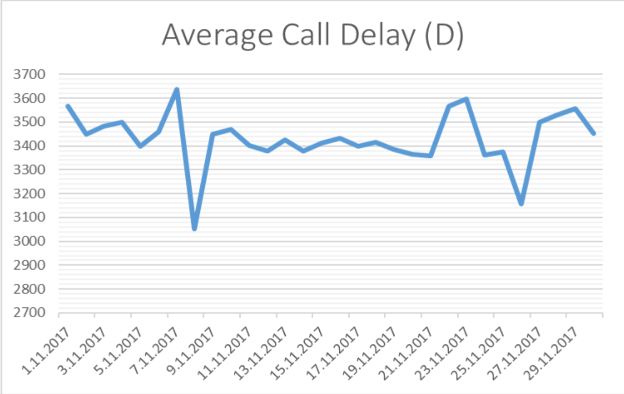

Fig. 1. Daily average call delay trend of the whole network

Another common methodology is the location-based comparison, where individual subscriber's call delay figures are compared to his/her city or district's call delay figure. As we mentioned previously, these simple methodologies may result in wrong interpretations. VOLTE experience should be evaluated call by call and generalizations will lead to incorrect results.

III. EXPLORATORY DATA ANALYSIS & DATA PROCESSING

VOLTE call analysis data set is selected as November 2017 when VOLTE usage maturated as a major service for Turkcell in Turkey. Fig. 2 shows the daily trend of VOLTE call numbers. Monthly total number of calls is 302.673.551 which is pretty huge data.

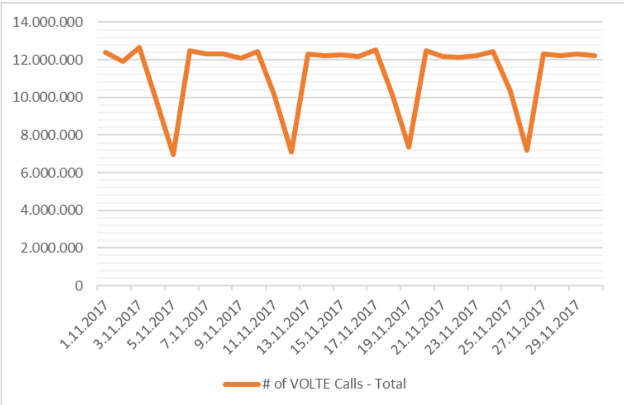

Fig. 2. Total VOLTE call count trend

Since this article is focused on continuous VOLTE call scoring and detection of affecting factors, all VOLTE calls should not be directly included in the detail analysis set. Firstly, VOLTE calls targeted to other Mobile Network Operator and originated from other Mobile Network Operator should be eliminated because other operator's network settings may be different. Secondly, dropped calls should be eliminated because they present bad extreme experience and may mislead the study. Fig. 3 shows VOLTE call numbers daily trend after this reduction and monthly total number of calls became 178.350.133.

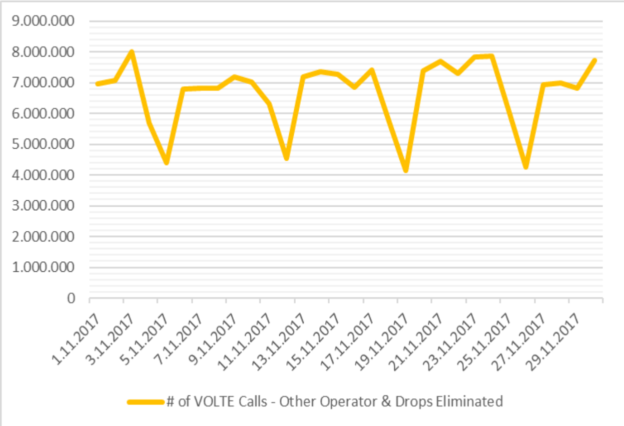

Fig. 3. Reduced VOLTE call count trend

As a rule of basic statistics, outliers are detected and eliminated from data set. In order to eliminate outliers, first standard deviation (σ) is calculated as 1.571,33 D and mean (μ) is calculated as 3.286,77 D. Values greater than $\mu+3\sigma$ and less than $\mu-3\sigma$ are accepted as outliers and removed from data set. Fig. 4 shows VOLTE call numbers daily trend after outlier elimination and monthly total number of calls became 174.783.131. It seems that the general trend behavior did not change after filtering operations.

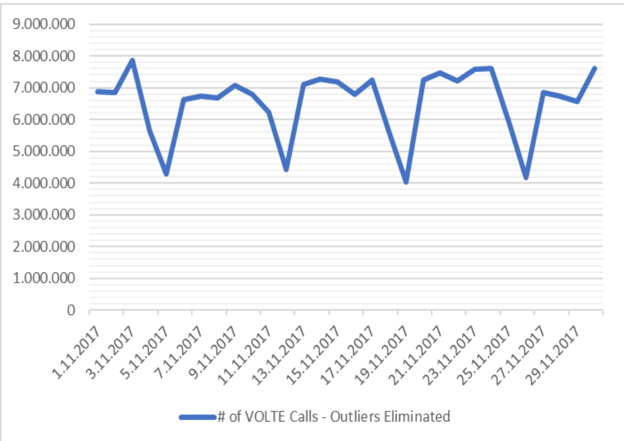

Fig. 4. Outlier eliminated VOLTE call count trend

Fig. 5 shows call delay histogram, which can be assumed as a normal distribution.

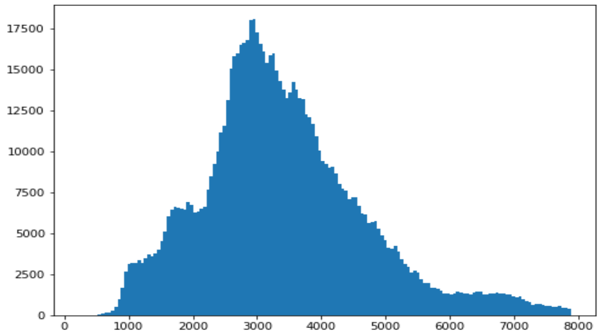

Fig. 5. Call delay histogram

IV. Investigation of Factors

In this section of the article, investigations of various factors that are assumed to affect perceived VOLTE quality of service of subscriber are performed.

First, the distance between caller subscriber and called subscriber must be the dominant factor affecting transmitted packet delay. By nature, data packets propagate physically through the network and arrive at the end point. The distance is assumed as geographical distance as the crow flies between attached base stations of caller subscriber and called subscriber. This distance is named as call distance in the remaining of this article and it is measured in kilometers.

Fig. 6 shows the average call delay correlation analysis with the distance between attached base stations of caller subscriber and called subscriber. Not surprisingly, average call delay and call distance are strongly correlated with Pearson r coefficient equal to 0.71 [6]. So, it is proved that distance between attached base stations of caller subscriber and called subscriber is a major factor affecting experienced call delay.

Correlation analysis and visualization depicted are performed using Python's Pandas, Seaborn, Matplotlib and Numpy packages.

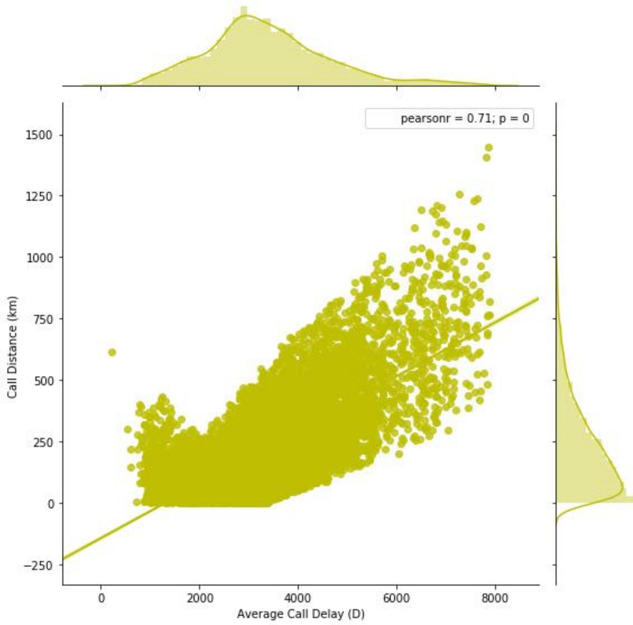

Fig. 6. Correlation and effect of call distance

Second, audio codec used may be another important factor affecting experienced call delay in November 2017. Fig. 7 shows the average call delay trend of AMR and AMR Wideband codec types. In Turkcell VOLTE infrastructure, these two types of codec are used for VOLTE service. Average call delay of AMR Wideband codec is absolutely higher than AMR codec. Therefore, codec type is another factor affecting experienced call delay.

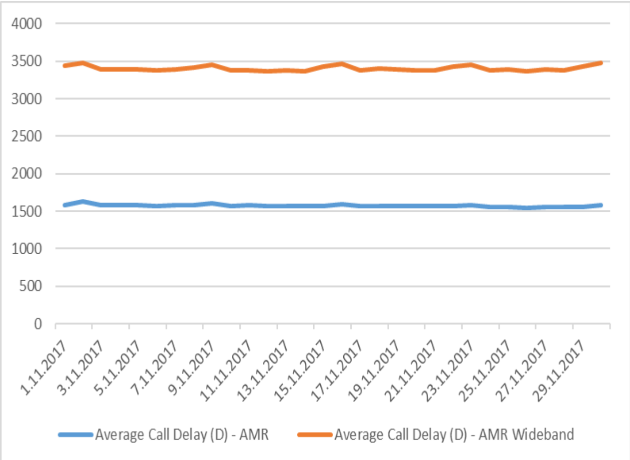

Fig. 7. Effect of audio codec used

Another factor may be the duration of VOLTE call. Since call duration is continuous data set, statistical analysis is performed and percentile results are given in Table I [2]. Call durations divided into three main groups as Short, Normal, Long using division points at 33rd percentile and 66th percentile.

TABLE I. CALL DURATION PERCENTILES

Percentile	*Call Duration (Sec)*
Minimum	0,5
1th percentile	4,1
25th percentile	21,9
33rd percentile	27,4
50th percentile	38,5
66th percentile	57,2
75th percentile	77,8
99th percentile	890,4
Maximum	7.235,1

Fig. 8 shows the call delay trend of clustered call durations in November 2017. In this figure, the first finding is that short calls significantly has lower average call delay than other two clusters. Long calls have higher average call delay than normal calls in general. However, there are some fluctuations and in a few days normal calls have a bit higher or close averages than long calls. Although there are few extreme cases, call duration is positively correlated with call delay, so call duration category is a factor affecting experienced call delay.

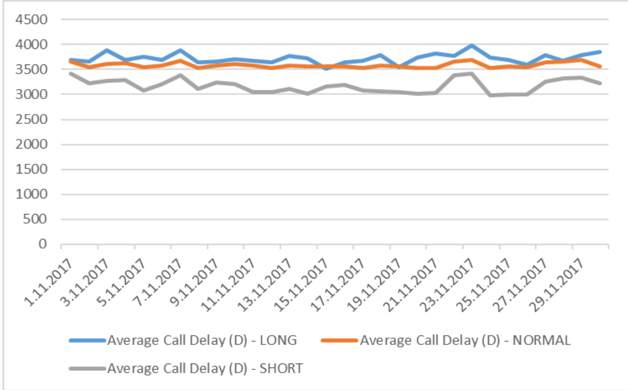

Fig. 8. Effect of call duration category

Mobile devices that are mostly smartphones vary in terms of hardware features such as CPU, storage, memory, camera. These physical features affect the responsiveness of the phone in other words user experienced speed when some action is triggered. Similar to traditional computers mobile devices have their own operating system. Most widespread ones are IOS and Android. Also, some other less used operating systems exist such as Windows 8, Symbian, etc.

Since operating system controls every process executed in a mobile device, type of operating system may be another factor affecting VOLTE quality of service. Fig. 9 shows the average call delay trend of VOLTE calls grouped by mobile device operating system of caller subscriber in November 2017. Less used operating systems are grouped as other in statistical analysis. As depicted in the figure, Android operating systems have minimum average call delay almost every day whereas other operating systems have the higher call delay. Other operating systems have higher average call delay than IOS. It should also be noted that other operating systems' average call delay trend is not steady and is very fluctuating. Therefore, mobile device operating system of caller subscriber is a good factor in experienced call delay.

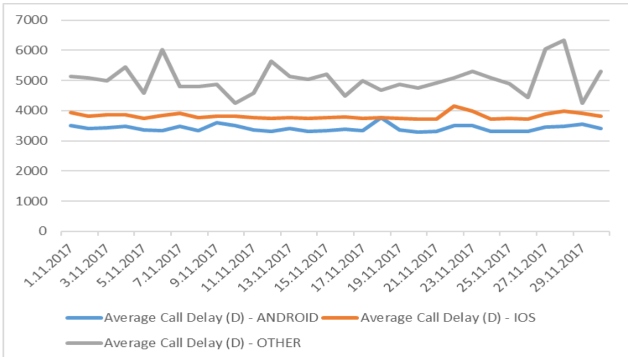

Fig. 9. Effect of mobile device operating system of caller subscriber

Similar to mobile device operating system of caller subscriber, mobile device operating system of called subscriber should be another variable to be investigated. Fig. 10 shows the average call delay trend of VOLTE calls grouped by mobile device operating system of called subscriber in November 2017. As depicted in the trend graph, mobile device operating system of called subscriber is also determining factor in experienced call delay.

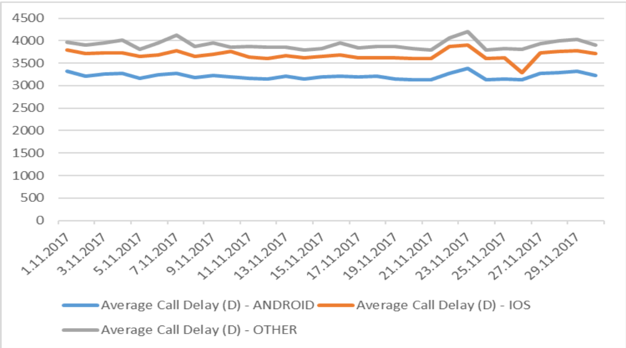

Fig. 10. Effect of mobile device operating system of called subscriber

During the investigation of factors, it is revealed that the more mobile subscribers are the worse quality of service they experience. Fig. 11 shows the call delay trend of VOLTE calls grouped by caller subscriber mobility in November 2017. It is obvious that VOLTE calls of moving subscribers have higher average call delay than VOLTE calls of not moving subscribers. Therefore, the mobility of caller subscriber is an indirect but a significant factor affecting experienced call delay.

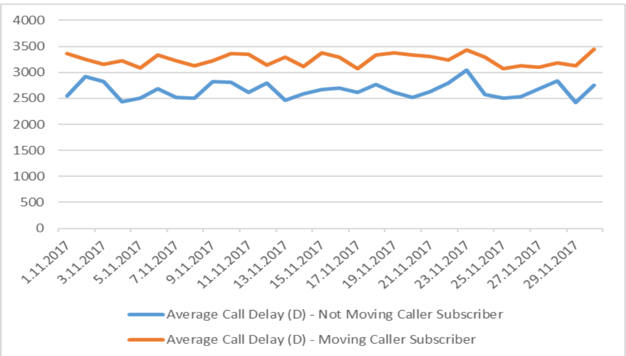

Fig. 11. Effect of mobility of caller subscriber

Since a call is two ended concept, if the mobility of caller subscriber affects VOLTE experience, it is expected that the mobility of called subscriber should also affect it.

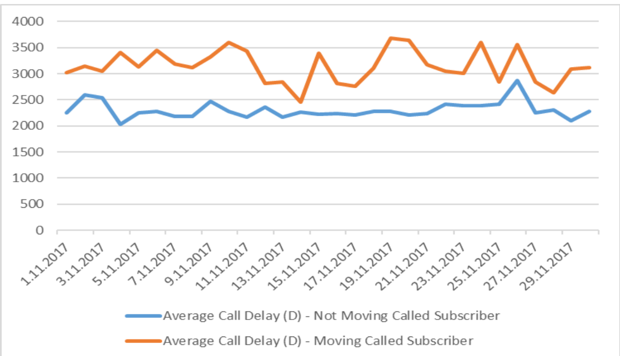

Fig. 12. Effect of mobility of called subscriber

Fig. 12 shows the call delay trend of VOLTE calls grouped by called subscriber mobility in November 2017. The average call delay of moving subscriber is apparently fluctuating. But there is no doubt that VOLTE calls of moving called subscribers have higher average call delay than VOLTE calls of not moving called subscribers. As a consequence, the mobility of called subscriber is an indirect but a determining factor affecting experienced call delay.

V. Proposal of New Approach

In the previous section, we have proved that experienced call delay depends on numerous factors which are:

- Audio codec used (AMR, AMR Wideband)
- Call duration category (Short, Normal, Long)
- Mobile device operating system of caller subscriber (Android, IOS, Other)
- Mobile device operating system of called subscriber (Android, IOS, Other)
- Mobility of caller subscriber (Moving, Not Moving)
- Mobility of called subscriber (Moving, Not Moving)

When VOLTE quality of service experience is simplified down to call level, similar to data sessions [7], each call has unique characteristics formed by 6 major factors mentioned above. In other words, either AMR or AMR Wideband is used in VOLTE call. Similarly, caller subscriber is either moving from one point to another point or is not moving (remain in a fixed location) during a VOLTE call.

TABLE II. SAMPLE CHARACTERISTICS

VOLTE calls	***Audio codec***	***Call duration category***	***Mobile device operating system of caller subscriber***	***Mobile device operating system of called subscriber***	***Mobility of caller subscriber***	***Mobility of called subscriber***
VOLTE call 1	AMR	Short	IOS	Android	Moving	Moving
VOLTE call 2	AMR	Short	Android	Other	Not moving	Moving
VOLTE call 3	AMR	Short	Other	IOS	Moving	Moving
VOLTE call 4	AMR	Medium	IOS	Android	Not moving	Moving
VOLTE call 5	AMR	Medium	Android	Other	Moving	Moving
VOLTE call 6	AMR WB	Medium	Other	IOS	Not moving	Not moving
VOLTE call 7	AMR WB	Long	IOS	Android	Moving	Not moving
VOLTE call 8	AMR WB	Long	Android	Other	Not moving	Not moving

Table II shows 9 different VOLTE call characteristics as an example. Since the characteristic of a VOLTE call is formed by 6 major factors, we reach 324 (3x3x3x3x2x2) possible combinations, therefore 324 distinct VOLTE call characteristics.

Each of the 324 distinct session characteristics has its own conditions, dependencies and stories. We have come up with the concept of Expected Call Delay that is ideal delay value of any VOLTE call based on call characteristics. The formula for Expected Call Delay is devised as follows:

$$Expected\ Call\ Delay = K_i * Log\ (CD) \quad (1)$$

where K_i is a constant value for related VOLTE call characteristic and CD is call distance. Call distance is the measured distance between two parties of VOLTE call, where actual coordinates of subscribers' attached base stations are taken into consideration. For each VOLTE call characteristic, a detailed statistical analysis is performed and constant K values are calculated separately. Basically, successful VOLTE call samples are taken into consideration. Actual call delay and call distance values can be measured from these samples. Then constant K_i values are calculated according to the proposed equation after basic statistic operations including outlier elimination. So 324 different K values are prepared for the equation.

For any VOLTE Call, Expected Call Delay is calculated by measuring the distance between calling and called subscribers, taking logarithm and then multiplying by related constant K value.

In order to test the success of proposed VOLTE quality of service evaluation approach, which is Expected Call Delay concept, real subscriber complaints are used as test set. Our test method was that whether the proposed approach could detect unhappy subscribers or not. Happy subscribers generally do not give feedback. In other words, happy subscribers do not tend to call GSM call center and thank for high level of quality of service. In the test, all subscribers who made a complaint about voice call related problems are taken into account. In order to make a correct correlation analysis, only VOLTE calls of the complaining subscribers before actual complaint time are included in the test set.

During the analysis, expected call delay values, according to the proposed approach, are calculated for each VOLTE call in the test set and compared with measured actual call delay values. Then, each call is classified as good, fair and bad experience based on the comparison of its actual call delay value and its expected call delay value. Table III shows classification details where ECD is calculated expected call delay value and ACD is measured actual call delay value of the related call.

TABLE III. CALL CLASSIFICATIONS

Rule	***Class***
ACD > ECD + (%5 * ECD)	Bad
ECD + %5 * ECD ≥ ACD ≥ ECD - %5 * ECD	Fair
ECD - %5 * ECD > ACD	Good

Results showed that our approach classified %87.77 of the calls as Bad experience, % 9.41 of the calls as Fair experience and %2.82 of the calls as Good experience. Therefore, 87.77% of individual subscribers who complained about voice calls have been classified as Bad experience within complaint period. As a result, it can be said that the proposed Expected Call Delay approach is successful in evaluating subscriber perceived quality of service and detecting Bad experience.

VI. CONCLUSION AND FUTURE WORK

This paper presented a detailed analysis on VOLTE quality of service measures and evaluation problem. Based on analysis results, a new concept of expected call delay was proposed and a formula was created.

The solution was easy to use and very practical in detecting individual subscriber's VOLTE quality of service perception. Initial success rate (which is 87.77%) of the expected call delay concept and formula to reflect the subscriber's perceived VOLTE service experience was very high.

Currently, the solution is applied to evaluate VOLTE service experience of a sample set of subscribers in Turkcell and an initial feedback on our solution is very positive.

In the future, there are several factors to be investigated whether they affect customer experience on VOLTE or not. Firstly, weather condition especially the average temperature and temperature changes during a day may be an important factor. Secondly, eliminated VOLTE calls originated from or terminated by other Mobile Network Operators should be investigated. Although other operator's network is another network and it has its own differential settings, this investigation may give some clues about customer VOLTE quality of service perception. Thirdly, the business of base station and attached cell of caller subscriber may be another factor. The simultaneous number of active VOLTE call at the same time may act as a factor. Similarly, the business of base station and attached cell of called subscriber may be investigated. Lastly, this study is conducted on one month's data. Although hundreds of millions of VOLTE calls are analyzed and a very strong proposal is formed, analyzing one or two year's data may give us more detailed results and idea about seasonality effect on VOLTE quality of service.

ACKNOWLEDGMENT

I am thankful to Turkcell as a visionary company that significantly supports R&D studies. Additionally, special thanks to my managers for their encouragement for performing this study.

REFERENCES

[1] R.Tabany and Chris G. Guy, "An End-to-End QoS performance evaluation of VoLTE in 4G," 10th International Conference on Wireless and Mobile Communications, ICWMC 2014.

[2] M. J. Dixon and F. J. Massey, Introduction to Statistical Analysis, McGraw Hill, 1957.

[3] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," Multimedia Systems, vol. 22, pp. 213-227, 2016.

[4] VoLTE faces up to the call quality challenge, 2016. Retrieved from https://continualexperience.com/volte-faces-up-to-the-call-quality-challenge/

[5] Realising revenue potential with Voice over LTE, 2018. Retrieved from https://www.gsma.com/futurenetworks/technology/volte/

[6] N. R. Draper and H. Smith, Applied Regression Analysis, Wiley & Sons, 1998.

[7] İ.Yıldız and S. Aydeniz, "Throughput evaluation approach for GSM networks," 14th International Conference on Wireless and Mobile Communications, ICWMC 2017, pp. 1-6.

[8] B. Kiziltan, M. Khan, and F. M. Velotti, "Voice over IP - WLAN, 3G and LTE issues," Wireless Network, Chalmers 2011, pp. 1-7.