Analysis of the Dependency of Call Duration on the Quality of VoIP Calls

Jan Holub[iD], Michael Wallbaum, Noah Smith, and Hakob Avetisyan

Abstract—This letter analyses call detail records of 16 million live calls over Internet-protocol-based telecommunications networks. The objective is to examine the dependency between average call duration and call quality as perceived by the user. Surprisingly, the analysis suggests that the connection between quality and duration is non-monotonic. This contradicts the common assumption, that higher call quality leads to longer calls. In light of this new finding, the use of average call duration as an indicator for (aggregated) user experience must be reconsidered. The results also impact modeling of user behavior. Based on the finding, such models must account for quality since user behavior is not fully inherent, but also depends on external factors like codec choice and network performance.

Index Terms—ACD, call detail record, call duration, Internet protocol, VoIP, voice quality, speech codecs, telephony.

I. Introduction

IT IS widely assumed that longer call durations indicate better call quality. Indeed, this dependency between call quality and average call duration (ACD) was reported for a mobile network in 2004 [1]. However, the study was conducted in times when most of the mobile calls were charged based on their duration so that users were motivated to keep calls as short as possible. Other conditions, such as the transition to IP-based packet switching and the introduction of new codecs, have also changed, which motivates a second look at the relation between call quality and duration.

Telephone calls carried over IP networks are affected by technical impairments, influencing the users' subjective perception of the call. Common technical impairments include coding distortion, packet loss, packet delay and its variations (jitter). The relation between the amount of each impairment and the final quality as perceived by a service user is not simple, as impairments can mask each other; or two impairments, each unnoticeable by itself, can multiply their effect and become subjectively annoying.

Monitoring systems analyzing live calls in telecommunications networks apply algorithmic models that attempt to estimate the subjective quality based on objective measurements of selected technical impairments. Commercial monitoring products for voice over IP (VoIP) services often use derivatives of the E-model defined in ITU-T G.107 [2] to estimate

call quality. In some parts of the industry, e.g., in international wholesale business, the ACD is used as a cost-effective indicator of subjective call quality. The underlying assumption is that higher call duration means better user experience.

Call duration, meaning the time difference between call establishment and call termination, serves as input parameter in many network- and service-models [3]. It is influenced by a number of factors [4], particularly by the calling and called party situation, amount of information to be exchanged, social circumstances [5], gender of the call parties [6] or their nationalities [7] and is generally of great interest when examining large amounts of network data [8], [9].

II. Background

This letter is based on call detail records (CDR) produced by a commercial non-intrusive VoIP monitoring system measuring the quality of real calls in the network of a communication service provider. The system method analyses the Session Initiation Protocol (SIP) signaling messages as well as the flow of Real-time Transport Protocol (RTP) packets, their interarrival times and the information contained in the protocol headers. For every five-second segment of each RTP flow, the system generates a quality summary with several hundred metrics. Each summary contains basic information, such as the source/destination IP addresses, the used codec, as well as details about packet losses, interarrival times and the estimated quality. Estimates for the subjective quality are calculated using the E-Model with information about the packet loss, jitter, and the used codec as input. The E-Model yields an R-factor value for every time slice, which is mapped to an estimated MOS (Mean Opinion Score) value. MOS is the commonly used metric for subjective call quality. Finally, the system marks 'critical' five-second segments which suffer from burst loss or excessive jitter. Specifically, a five-second segment is marked as 'critical' if more than three packets are lost in sequence or if the packet interarrival time exceeds the packet rate by 40 ms or more.

The monitoring system's CDRs describe the characteristics of each call from the signaling and media quality perspective. They summarize the five-second data, e.g., by storing the minimum, average and maximum R-factor and MOS for each call direction. Other quality metrics provided by the CDRs are described in the next section.

III. Data Set Characteristics

The data was provided as a database of CDRs, with each database record corresponding to one call in the network. The following parameters were used for the analysis: call duration, used audio codec, critical minute ratio (CMR) per media direction and average R-factor/MOS per media direction. CMR is

TABLE I
DISTRIBUTION OF CODECS IN THE ANALYZED DATA SET

Codec	Number of CDRs	Percentage
G.711	10,998,417	68.1%
G.729	4,108,152	25.4%
AMR-NB 12.2k	1,044,612	6.5%

TABLE II
BEST QUALITY ACHIEVED AND ACD PER CODEC

Codec	Max. R-Factor	Max. MOS	ACD/[s]
G.711	93	4.41	262
AMR-NB 12.2k	86	4.23	210
G.729	82	4.10	181
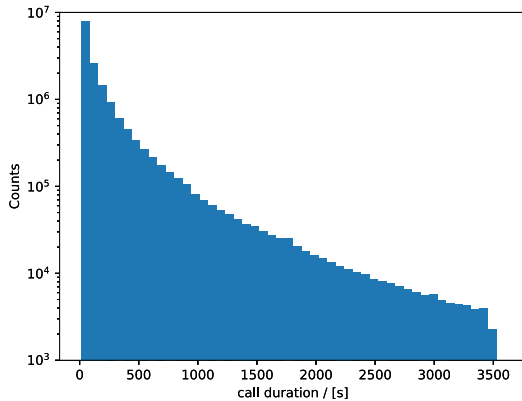
Fig. 1. Distribution of the call duration in the analyzed data set.

Fig. 2. Distribution of call durations depending on Critical Minute Ratio.

a proprietary metric provided by the monitoring system. It is calculated as the ratio of 'critical' five-second segments over all segments. For example, a CMR of 10% states that one out of ten time slices were affected by critical loss or jitter.

All quality data is available separately for media streams sent by the calling party (A-party) and the called party (B-party). However, the analysis only considers the worst direction per call, i.e., the direction with smallest R-factor and greatest CMR.

The raw data set contains nearly 30 million mobile, international and domestic calls. The calls used 22 different types of codecs including G.711, G.729, G.722, G.723.1 and various modes of AMR-NB and AMR-WB. The majority of these codecs were however used so rarely that the respective CDRs were excluded from further analysis. The following CDRs were not considered for the analysis:
- calls which did not use the top three codecs G.711, G.729 or AMR-NB 12.2k,
- calls which lasted less than 10 s or more than one hour,
- calls where at least one direction was impacted by duplicate packets.

At the end of filtering process, the data contained more than 16 million CDRs. The codec distribution of the data set is shown in Table I.

Figure 1 shows the distribution of call durations. It closely matches a log-normal distribution, that is often used to model call duration distributions. The ACD over the entire (filtered) data set is 220 s.

IV. DATA ANALYSIS

A. Dependency on Codec Quality

The three audio codecs in the analyzed data set are all narrowband codecs. G.711 [10] is a widely used codec based on pulse code modulation. The audio signal is sampled at 8 kHz using 8 bits per sample, which leads to a net bitrate of 64 kbit/s.

The samples are companded by one of two logarithmic functions, called A-law and μ-law. G.711 μ-law is mainly used in North America and Japan; A-law is preferred in the rest of the world. No distinction is made in the following, as all characteristics relevant for this letter are identical.

The second codec G.729 [11] uses conjugate structure algebraic-code-excited linear prediction (CS-ACELP) to compress speech frames of 10 ms. Annexes to the basic G.729 define several variants and extensions, e.g., for discontinued transmission or alternative bit rates. The analyzed data set contains only the variant compliant to G.729 Annex A and B – other variants are currently not used by commercial VoIP services. The sampling frequency is 8 kHz with 16 bits per sample. An encoded frame consumes 10 bytes yielding a net bitrate of 8 kbit/s.

The AMR codec described in [12] is a narrowband audio compression scheme using an ACELP coding scheme. It offers multiple bit rates ranging from 4.75 to 12.2 kbit/s and is widely used in GSM and UMTS. The only rate that is included in the data set is AMR-NB 12.2k.

Table II shows how the ACD depends on the calls' main codec. To mask out the impact of network conditions on the ACD, the table only considers calls with CMR=0%, i.e., without any critical packet loss or jitter.

The data shows that ACD correlates with the best possible user experience that can be achieved by a codec. For example, calls using G.711 are more than 40% longer on average than calls using G.729. The higher a codec's maximum R-factor/MOS, the higher the ACD.

B. Dependency on Transport Quality

The codec employed by a call is one technical aspect that can be controlled by a communication service provider. The other technical parameter is the amount of packet loss and

TABLE III
DEFINITION OF CATEGORIES OF SPEECH TRANSMISSION QUALITY

R-factor	MOS	Speech quality	User satisfaction
≥ 90	≥ 4.34	Best	Very satisfied
≥ 80	≥ 4.03	High	Satisfied
≥ 70	≥ 3.60	Medium	Some users dissatisfied
≥ 60	≥ 3.10	Low	Many users dissatisfied
≥ 50	≥ 2.58	Poor	Nearly all users dissatisfied

jitter, i.e., the network's transport performance. Figure 2 shows the impact of the CMR on the ACD. The data points represent the window centers and the window radius is five; error bars correspond to one standard deviation.

The ACD drops sharply from 230 s to 114 s for CMR=35%. This drop could be expected as more time slices are impacted by packet loss and jitter, which has a negative impact on the user experience. The severity of the drop is however surprising.

It should be noted that the CMR measures the distribution of severe impairments over the duration of a call. This is not equivalent to measuring the overall amount or intensity of packet loss and jitter. For example, RTP streams which lose three packets in sequence every five seconds, have a CMR of 100%. At a packet rate of 20 ms this corresponds to a packet loss ratio of about 1%, which is not much by conventional wisdom. In contrast, a call where half of one stream's packets are lost in sequence would yield a CMR of only 50%.

The sharp drop implies that even low impairment levels have a significant impact on the ACD. Beyond CMR=35% the ACD shows unexpected behavior as it rises again to 148 s for CMR=55%, drops down to 115 s for CMR=75% and finally rises to 153 s. Since voice quality depends on the level of packet loss and jitter, this behavior apparently contradicts the findings in [1] and common industry assumptions, namely that call duration is a monotonic function of the user experience. Yet, the CMR is a metric describing the technical quality of RTP streams, not the actual user experience. The next section looks into the connection between user experience and ACD.

C. Dependency on Estimated User Experience

As discussed before, the user experience is estimated using the E-model [2], which yields an R-factor value. This value is in the range of 0 to 100, where 0 represents extremely bad quality and 100 very high quality. Table III, based on G.107 [2], relates the E-model ratings R and their corresponding MOS to categories of speech quality and user satisfaction. Note that this mapping is only valid on the narrowband scale, i.e., when only narrowband codecs such as G.711, G.729 and AMR-NB are considered.

Figure 3 shows the ACD as a function of the average MOS of a call's worst stream. The average is calculated from the individual MOS values of the worst stream's five-second time slices. The data points represent the window centers with a window radius of 0.5.

The ACD drops from 222 s for MOS=4.25 to an absolute minimum of 133 s only to rise again to 190 s for MOS=1.25. It must be underlined that the absolute low of the ACD at MOS=2.25 is located just below the area of the MOS scale
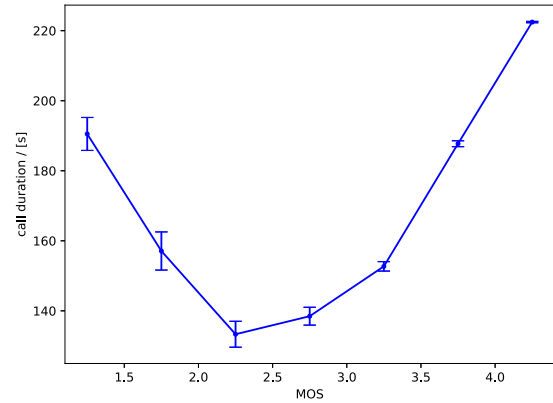

Fig. 3. Distribution of call durations depending on the average MOS.

where - according to Table III - 'nearly all users [are] dissatisfied'. For MOS < 2.58 one can assume that practically all call parties are dissatisfied. A rising ACD despite increasingly dissatisfied users is an unexpected result.

V. DISCUSSION

The previous section showed that the average duration of millions of VoIP calls generally depends on call quality. This is true for different definitions of quality, i.e., when defining quality via the main codec, the networks' transport performance (CMR) and the user experience (MOS).

The common assumption, substantiated in [1], is that average call duration is a monotonic function of quality, i.e., better user experience leads to longer durations. For moderate to high call quality this assumption is confirmed by the underlying data. Yet, the results for low call quality are surprising, since call durations also increase as quality gets worse. This suggests that the simplistic presumption about the link between quality and ACD may no longer be true. At least for contemporary VoIP-based services, the ACD appears to be a non-monotonic function of call quality. For example, given Figure 3 one cannot decide whether an ACD of 190 s indicates horrible or good user experience.

This unexpected finding has practical impact on the telecommunications industry. Specifically, mobile and international wholesale service providers may currently be working under wrong assumptions. Both types of service providers frequently have to deal with low-quality calls, e.g., because of poor air interface conditions or because of problematic routes to developing countries. The non-monotonic dependency of call duration and quality renders the ACD useless as a measure of user experience; service operators that solely rely on ACD as quality metric are likely to make wrong decisions. For example, actual mobile network issues may go undetected, when the ACD is relatively high or – even worse – international traffic is switched to a route with seemingly better quality. Interestingly, in wholesale business knowledge about the non-monotonic behavior could even be exploited for financial benefit. As the ACD is considered the main quality metric for routes (with impact on price), it may pay off to deliberately degrade moderate-quality routes to increase the ACD.

Consequently, other metrics, such as CMR or the average MOS, are needed to complement ACD.

Another area that is impacted by the findings of this letter is modeling of user behavior in terms of the call duration distribution, e.g., as described in [13]. Obviously, such models must also account for quality since the behavior of a user or user group is not entirely inherent to the user (group), but also depends on external factors like codec choice and network transport quality. New user models must consider the unexpected behavior of the ACD.

There are multiple potential reasons for the observed ACD increase for heavily compromised quality. One of them is the obvious need of word and sentence repetition when communicating over bad channels. Another one might be the Lombard effect [14] – an involuntary tendency to speak louder and slower under uncomfortable listening situations. Also, high packet loss and jitter may lead to stronger channel coding schemes and deeper de-jitter buffer adoption which adds delay to the communication path. If the call parties avoid double-talk situations, such delay is directly added to the overall call duration, multiplied by twice the number of role swaps (talker/listener) during the call [15]. Conversations in English language exhibit 103 swaps on average during a three minute call [16]. Assuming the additional swap delay increases by 60 ms, then a call, which would last three minutes under perfect conditions, is prolonged by 12.4 s through this effect only. In practice, the need for repetitions under adverse conditions will increase the number of swaps, which further adds to the call duration.

VI. CONCLUSION

This letter explored the dependency between the average duration and quality of VoIP calls. More than 16 million live calls were analyzed by a commercial non-intrusive monitoring system. The main contribution of this letter is that it reveals a surprising non-monotonic dependency between call quality and average call duration. The common assumption, that better user experience leads to increasing call duration, is only valid for moderate to high call quality. Under this condition the findings of an earlier study [1] could be confirmed, although charging models, traffic types, quality measurement methods and geographical regions differ. However, under conditions of low quality the dependency changes, i.e., lower quality yields longer call durations, which contradicts common expectations. The inflection point roughly corresponds to quality that dissatisfies virtually all users.

The following conclusions can be drawn:
- The ACD is not an indicator for user experience.
- Codec choice and network transport performance influence the call duration.

Further studies with traffic using a wider variety of codecs need to be performed, to confirm that the codec quality influences the average call duration even for unconventional codecs. Specifically, traffic using wideband codecs needs to be examined, to determine if there is a quality saturation beyond which call duration is not impacted. Furthermore, the dominating factors of IP transport quality on call duration need to be analyzed, i.e., which impairment patterns lead to changes in the ACD. Finally, empirical studies need to investigate the reasons for increasing call duration under adverse conditions.

REFERENCES

[1] J. Holub, J. Beerends, and R. Smid, "A dependence between average call duration and voice transmission quality: Measurement and applications," in *Proc. Wireless Telecommun. Symp.*, 2004, pp. 75–81.

[2] *The E-Model: A Computational Model for Use in Transmission Planning*, Int. Telecommun. Union, Geneva, Switzerland, ITU Recommendation G.107 (06/15), Jun. 2015.

[3] Z. Yang and Z. Niu, "Load balancing by dynamic base station relay station associations in cellular networks," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 155–158, Apr. 2013.

[4] V. D. Blondel *et al.*, "A survey of results on mobile phone datasets analysis," *EPJ Data Sci.*, vol. 4, no. 1, p. 10, Dec. 2015.

[5] Y. Dong, J. Tang, T. Lou, B. Wu, and N. V. Chawla, *How Long Will She Call Me? Distribution, Social Theory and Duration Prediction*. Berlin, Germany: Springer-Verlag, 2013, pp. 16–31.

[6] G. Friebel and P. Seabright, "Do women have longer conversations? Telephone evidence of gendered communication strategies," *J. Econ. Psychol.*, vol. 32, no. 3, pp. 348–356, Jun. 2011.

[7] D. Goodman and R. Nash, "Subjective quality of the same speech transmission conditions in seven different countries," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 7. Paris, France, 1982, pp. 984–987.

[8] J. Kim *et al.*, "Modeling cellular network traffic with mobile call graph constraints," in *Proc. IEEE Win. Simulat. Conf. (WSC)*, Phoenix, AZ, USA, Dec. 2011, pp. 3165–3177.

[9] M. Seshadri *et al.*, "Mobile call graphs: Beyond power-law and lognormal distributions," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD)*, Las Vegas, NV, USA, 2008, pp. 596–604.

[10] *Pulse Code Modulation (PCM) of Voice Frequencies*, Int. Telecommun. Union, Geneva, Switzerland, ITU Recommendation G.711 (11/88), Nov. 1988.

[11] *Coding of Speech at 8 kbit/s Using Conjugate Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)*, Int. Telecommun. Union, Geneva, Switzerland, ITU Recommendation G.729 (06/12), Jun. 2012.

[12] "Mandatory speech CODEC speech processing functions; AMR speech codec; general description," 3GPP, Sophia Antipolis, France, Rep. 26.071, 1999.

[13] P. O. S. Vaz de Melo, L. Akoglu, C. Faloutsos, and A. A. F. Loureiro, *Surprising Patterns for the Call Duration Distribution of Mobile Phone Users*. Berlin, Germany: Springer-Verlag, 2010, pp. 354–369.

[14] S. A. Zollinger and H. Brumm, "The Lombard effect," *Current Biol.*, vol. 21, no. 16, pp. R614–R615, 2011.

[15] J. Holub and O. Tomiska, "Delay effect on conversational quality in telecommunication networks: Do we mind?" in *Wireless Technology: Applications, Management, and Security*, S. Powell and J. P. Shim, Eds. Boston, MA, USA: Springer, 2009, pp. 91–98. [Online]. Available: https://doi.org/10.1007/978-0-387-71787-6_6, doi: 10.1007/978-0-387-71787-6_6.

[16] "Speech and multimedia transmission quality (STQ); adaptation of the ETSI QoS model to better consider results from field testing LQO and delay," ETSI, Sophia Antipolis, France, Rep. ETSI TR 103 121, Rev. 1.1.1, Mar. 2013.