Evaluation of BLE-based Audio Broadcasting Under Probabilistic Interference

Mathias Baert[a,*], Bart Moons[a], Jowan Pittevils[b], Yanjue Song[a], Nilesh Madhu[a], Jeroen Hoebeke[a]

[a] *imec-IDLab-UGent, Technologiepark-Zwijnaarde 126, Ghent, 9052, Belgium*
[b] *Cochlear, Schaliënhoevedreef 20/I, Mechelen, 2800, Belgium*

Abstract

Wireless Bluetooth audio communication has become an inherent part of everyday life, from listening to a podcast on our headphones to streaming music on multiple speakers. Up until now, broadcasting audio to multiple receivers has always required a proprietary implementation. Therefore, recent advances in the specification allow Broadcast Audio (BA) streams to be set up on top of Bluetooth Low Energy (BLE). To cope with the unpredictability of wireless media, audio frame retransmission opportunities are provided. However, determining the applicable number of retransmissions for a broadcast stream that is exposed to volatile environmental conditions, is a complex research challenge. This paper presents a model that is capable of simulating a BLE broadcast stream schedule, exposed to various environmental conditions while using a variable number of audio frame retransmissions. The evaluation employs several existing Packet Loss Concealment (PLC) techniques to cope with audio frame losses. The results provide insights into the impact of various frame loss patterns on the audio quality and intelligibility of broadcasted speech. The more advanced PLC techniques can handle a higher frame loss rate threshold. The analysis also shows that large audio frames requiring fragmentation exhibit a higher amount of frame loss for the same BLE packet loss rate and that overlap with in-use 802.11 channels can lead to a large variability in frame loss behavior. The model provides a baseline for the next research challenge related to continuous management of BA streams operating under volatile environmental conditions.

Keywords: LE Audio, audio quality, intelligibility, simulation, dataset

---

1. Introduction

One of the most recent additions to the Bluetooth Low Energy (BLE) ecosystem is the LE Audio paradigm [1]. It introduces standardized methods to set up synchronized multi-stream wireless audio communication, including a one-to-many approach that provides wireless encoded audio transmission from a single source to an unlimited number of receivers, i.e. Broadcast Audio (BA). This concept is marketed as Auracast [2], an entirely novel Bluetooth capability. It enables a variety of use cases, such as airport notifications, multi-language distribution systems, assistive listening systems in theatres or lecture halls, etc. However, wireless media introduce uncertainty regarding performance due to potential obstructions and interference [3]. Moreover, due to the unidirectional characteristic of a BA stream, it is more difficult for an audio source to be aware of the performance at the receivers, thus some redundancy is required. BA streams provide this via additional retransmissions of the same encoded audio frame. Since the Low Complexity Sub-band Codec (SBC) for wireless audio transmission over Classic Bluetooth has limitations regarding the audio quality it can

---

[*] Corresponding author

Email addresses: `Mathias.Baert@UGent.be` (Mathias Baert), `bamoons.Moons@UGent.be` (Bart Moons), `jpittevils@cochlear.com` (Jowan Pittevils), `Yanjue.Song@UGent.be` (Yanjue Song), `Nilesh.Madhu@UGent.be` (Nilesh Madhu), `Jeroen.Hoebeke@UGent.be` (Jeroen Hoebeke)

preserve after compression, especially at lower encoding bit rates (EBRs), the Bluetooth SIG developed the Low Complexity Codec (LC3) [4]. LC3 demonstrates a much higher preserved audio quality using the same EBRs that are typically used in the older SBC and has a sufficient audio quality on much lower EBRs as well [5]. In order to provide multi-stream audio and include multiple retransmissions per stream, lower EBRs can be helpful to reduce the airtime a single transmission opportunity requires. The main contribution of the paper is the design and implementation of a simulation model that can be used to study the performance of a BA stream in diverse environments, while employing various configurations and different audio coding and concealment strategies on the stream. The paper presents the conclusions of a first evaluation of the model. The associated dataset and executable of the model are provided in [6]. The paper also includes a sample page [7] in order to substantiate the categorization of speech audio quality that is established during the evaluation. The model mimics how an encoded audio file is packetized for transmission by a broadcaster and depacketized into a decoded audio file at a receiver. It can apply diverse forms of packet loss on the scheduled stream of BLE packets. Packet Loss Concealment (PLC) techniques can then be applied during the depacketizing process to mitigate the impact of frame loss on the received encoded audio. The paper employs the model to examine various BLE packet loss scenarios, influenced by a Packet Error Rate (PER) probability per frequency channel. The evaluation is performed for two audio codecs: LC3 and its enhanced version LC3 Plus [8]. The latter extends LC3 with more advanced PLC algorithms. The remainder of the paper is organized as follows. Section 2 provides an overview of the different aspects of the LE audio stack that enable BA. Section 3 provides a short overview on previous work regarding Bluetooth-based wireless audio streams and the impact of external interference on broadcast streams and BLE traffic in general. Section 4 describes the design of the model. Section 5 presents the input parameters and instrumental audio scoring metrics employed during the evaluation of the simulation model. Section 6 contains an evaluation of the model applied to several packet loss scenarios that could be observed in a real deployment. Section 7 applies the model to measured PER probabilities for BLE channels overlapping with Wi-Fi traffic. Lastly, Section 8 concludes the paper and provides an outlook on future work.

2. Primer on BLE-based Audio Broadcasting

2.1. Broadcast Isochronous Group

Broadcasters can set up a Broadcast Isochronous Group (BIG) that consists of one or more Broadcast Isochronous Streams (BISes). Data is sent in periodically occurring BIG events. The time between two subsequent BIG events is called the ISO_Interval (ISI) [9]. One BIG event contains a BIS event for each BIS in the group and each BIS event provides one or more data transmission opportunities. The implementation of the model is currently limited to BIGs with a single stream so the remainder of this section solely describes BIG functionality in the context of a single BIS. Data transmission opportunities occur at a predefined schedule so that each observer knows when the BIG event has ended. Consequently, every BIG observer is able to use received data at the same time, without being synchronized to other observers separately. Data communication in a BIS event consists of subevents (SEs) occurring at fixed time slots. Each BIS event contains a fixed number of SEs (NSE) and every SE provides an opportunity to transmit a single BIS Protocol Data Unit (PDU). The BIS event is divided into equally sized groups. The first group contains original data transmissions. The Burst Number (BN) indicates the number of new PDUs that can be transmitted per BIS event and thus defines the number of SEs in a single group. The remaining groups contain either retransmissions of the data sent in the first group

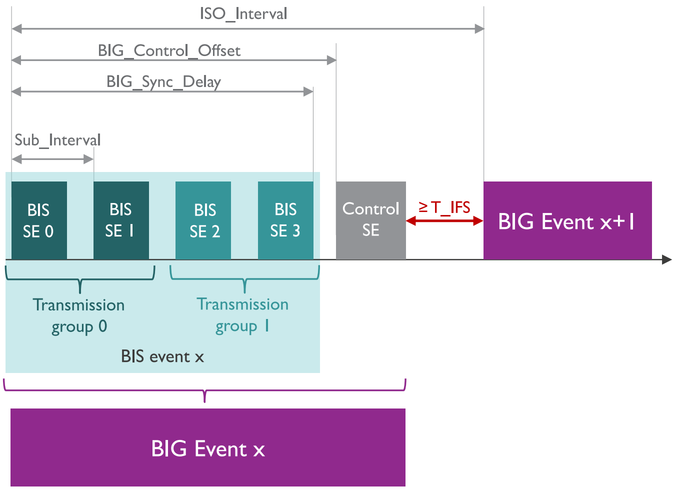


Figure 1: One BIG event (NSE = 4 and BN = 2).

or pre-transmissions of data that will be part of the first group in a future BIS event. The NSE should be an integer multiple of BN, to achieve equally-sized groups. The total number of groups (Group Count (GC)) can be calculated as $\frac{\mathrm{NSE}}{\mathrm{BN}}$. The Immediate Repetition Count (IRC) parameter indicates how many groups contain retransmissions of data in the first group. It should be higher than 0 but not greater than GC. The remaining GC - IRC groups contain pre-transmissions of future BN-based data. Based on the Pre-Transmission Offset (PTO), each remaining group determines for which future BIS event it will provide pre-transmissions. The pre-transmissions functionality is out of scope so the remainder of the paper assumes IRC == GC at all times. Figure 1 illustrates the internal scheduling of a single BIG event. The Sub_Interval (SUI) parameter indicates the time between consecutive SEs in a BIS event. The minimum value for SUI depends on the on-air time required to send a BIS PDU of maximum payload size (MPT). Two parameters impact the MPT: the maximum number of bytes for a single BIS PDU (maxPDU, can be up to 251 bytes) and the physical data rate (PHY). The specification indicates a minimum time interval of 150 µs between the end of a SE and the beginning of the subsequent one, denoted as T_MSS. The minimal required time for the SUI is MPT + T_MSS. SEs can contain PDUs of less than maxPDU bytes payload as well. Figure 1 also illustrates the BIG_Sync_Delay (BSD) and BIG_Control_Offset (BCO) parameters. Synchronized observers require a common baseline to, for instance, synchronize playback. The BSD indicates the maximum possible length for the data portion of a BIG event, calculated as (NSE - 1) * SUI + MPT. The BIG event can also contain an optional control SE. The BCO indicates when an observer can expect the control SE (if present) and shall be greater than or equal to NSE * SUI. Currently, the control SE can be used to inform observers of an updated channel map for the Channel Selection Algorithm (CSA) or a pending termination of the BIG. In general, the BIG should always be able to close the current BIG event at least 150 µs (T_IFS) before the next one starts. Each BIS PDU contains a 2-byte header to indicate the payload length, identify its context (Data or Control) and indicate whether a Control SE is present in the current BIG event. The BIS PDU is sent as payload of a generic BLE Link Layer (LL) packet. The header of an LL packet includes a 4-bytes access address. It acts as an additional identification of the ongoing communication paradigm (i.e. BLE connection, BIS, etc.). The creation process of a BIG includes the generation of a random Seed Access Address (SAA), based on a given set of requirements [9]. The SAA is used to generate an access address for the BIS, based on a BIS identification number (starting from 1). The evaluation in this paper is limited to unencrypted BIS PDUs broadcasted on 1 Mbps Uncoded PHY so the total header overhead is 10 bytes. BIGs maintain a 39-bit bigEventCounter that starts at 0 for the first BIG event and is incremented by 1 for every subsequent BIG event. Each SE in a BIG employs a 2 MHz frequency channel, chosen from 37 distinct center frequencies between 2404 MHz and 2478 MHz, to transmit a BIS PDU [9]. The channel selection can also be limited by (temporarily) excluding channels from the selection set. The BIG maintains the current selection set in the channel map, consisting of minimally 2 used channels. The channel selection for all SEs in a single BIG event is based on the CSA #2. It consists of two phases (the first SE (FSE) and the subsequent SEs (SSEs)). The channel for the FSE is uniformly selected out of all available channels in the channel map [10], using a calculation that employs pseudo-random numbers based on the BIS access address and the current bigEventCounter. Next, each SSE in the same BIS event uses a similar calculation, that also ensures that the chosen channel is never the same as the previous SE.

2.2. Isochronous Adaptation Layer

BIGs send data periodically at a given ISI. However, the data generation rate at the upper layer(s) (e.g. audio frames being encoded at a certain rate and within a certain size per frame), does not necessarily match the capabilities of the LL. Therefore, the Isochronous Adaptation Layer (ISOAL) is placed between the upper layer(s) and the BIG, in order to fragment an upper layer data packet into multiple PDUs if it does not fit into a single PDU and recombine the PDUs back into a single upper layer data packet at a receiver. The upper layer data packets are referred to as Service Data Units (SDUs) and the data generation rate follows the SDU_Interval (SDI). In case the ISI of the BIG is equal to or an integer multiple of the SDI and the frames are generated at a constant rate, unframed PDUs can be employed. Otherwise, an additional time offset needs to be added to the PDUs, that bridges the unknown gap between the associated SDU's generation point at the broadcaster and its arrival at the observer. Framed PDUs can be used to achieve this. The current implementation of the model is limited to unframed PDUs so the remainder of this section only describes that aspect of the ISOAL in detail. The BIG defines an additional parameter maxSDU, which indicates the maximum upper layer SDU size it can process. Fragmentation of an SDU into multiple PDUs is required only when its size is bigger than maxPDU. Since the data rate of the LL must be at least equal to the data rate of the upper layer [9], the BN, maxPDU and ISI of a BIG should be configured based on the maximum data generation rate from the upper layer (i.e. a single SDU of maxSDU bytes every SDI). This leads to the following formula $\mathrm{BN} = \lceil \mathrm{maxSDU} \,/\, \mathrm{maxPDU} \rceil \times (\mathrm{ISI} \,/\, \mathrm{SDI})$. The specification presents this as the minimum required BN [9] but any larger number would present an unnecessary usage of bandwidth so the remainder of the paper assumes that the BN is set to the outcome of the provided formula. Moreover, the evaluation in this paper is limited to the SDI being equal to the ISI. Consequently, one SDU generates BN fragments so the recombination process can use this information to track the current SDU being recombined. Based on the provided fragmentation mechanism, PDU losses can either be handled via remaining retransmissions or will, in case all SEs for a single PDU are lost, lead to an erroneous SDU.

2.3. Broadcast Audio

BIGs can be used for any use case that requires observers to act on received data in a synchronized manner, including wireless audio. This paper provides a performance evaluation of wireless audio transmission under diverse configurations of a BIG. Audio quality and latency are two important metrics to consider when assessing the performance of a wireless audio stream [1]. Audio transmission on unpredictable wireless media can experience frame loss, so the inclusion of retransmissions on a BIG is a logical next step to ensure that audio quality at the observer remains above an acceptable threshold. However, more redundancy leads to a higher latency, so a trade-off presents itself. Therefore, the size of the audio that needs to be transmitted (and potentially retransmitted) should be limited, for which codecs can be used. Audio codecs encode digital audio at the broadcaster before transmission and decode it at the observers after reception. The encoding mechanism at the broadcaster requires a certain amount of subsequent samples to have sufficient information to perform compression. Generally, a larger number of samples can lead to more efficient compression, since more information is available, but will introduce a higher encoding latency. The set of subsequent samples used to generate a single encoded data block is called a frame and the number of samples it is based on can be expressed as a duration, called the frame interval (FI). The FI presents a similar trade-off between latency (shorter frames lead to a lower latency) and audio quality (larger frames can be encoded more efficiently). Decoding at an observer should include a PLC mechanism in order to fill in potential gaps in the decoded audio. The evaluation in this paper will look at two recently introduced codecs in the context of BLE: LC3 and LC3 Plus. Both codecs support several sampling frequencies and number of bits per sample. The encoding procedure is built upon the Modified Discrete Cosine Transform (MDCT) mechanism, that converts N audio input time domain samples into N spectral coefficients. Subsequently, the MDCT data is transmitted over the air and the decoding procedure transforms the spectral coefficients back to a time domain signal [4]. The performance evaluation in this paper is limited to speech audio that is sampled at 16 kHz and represented using 16 bits per sample. LC3 is optimized for an FI of 10 ms, which is based on the frame length sweet spot for good audio quality at a reasonable latency, as determined by the industry [1]. The number of encoded bytes in one frame can be between 20 & 400 (i.e. nbytes parameter) per audio channel [4]. Consequently, this leads to supported EBRs between 16 and 320 kbps. In order to cope with the legacy intervals of Classic Bluetooth, LC3 can also operate at a 7.5 ms FI but this is out of scope for the current paper. Both LC3 and LC3 Plus provide PLC techniques. LC3 provides an example implementation of a PLC algorithm for lost frames, based on repeating the MDCT data associated with the last correctly received frame and performing pseudo-random sign scrambling on the spectral coefficients. The LC3 Plus codec provides additional functionality on top of LC3. It employs a more advanced PLC mechanism that is based on three techniques. The codec itself detects the technique to use for all lost frames in a single burst of losses based on characteristics in the last correctly received frame(s) [11]. The evaluation in the paper assumes that one SDU in a BIG corresponds to a single audio frame, which is default behavior [1]. Consequently, the SDI is equal to the FI and maxSDU is equal to the nbytes parameter.

3. Related work

Previous work has studied the impact of external interference on BLE traffic, for different technologies operating in the 2.4 GHz Industrial, Scientific, and Medical (ISM) band [3], and within the context of various environments [12]. Parameters of a BLE connection are shown to require continuous optimization in order to cope with dynamically changing environmental conditions [3]. Previous work that examines the impact of external interference on BIGs, is limited. The work in [13] presents an alternative approach to a fixed number of PDU retransmissions in a BIG event. It proposes to replace the passive n retransmissions with a per-SE retransmission decision scheme based on observers transmitting a non-acknowledgment (nack) back to the broadcaster in case the expected PDU was not received. The concept is interesting but still presents some challenges. Providing nacks back to the broadcaster is often not feasible. Observers will typically not be mains-powered and have less transmission power available so it will often not be possible to even reach the broadcaster's range. Moreover, the data rate of the uplink traffic will be limited by the downlink nack slots. The work in [11] presents a technical overview of LC3 and LC3 Plus and includes a comparison with other codecs for both speech audio and music streaming use cases. The evaluation is based on perceptual scores from ITU-T P.800 ACR and ITU-R BS.1116-3 experiments, and the impact of packet loss is examined as well. However, the paper does not examine the codecs in the context of a BIG, including the associated impact of retransmissions and channel hopping. Previous work in the Bluetooth-based wireless audio domain consists of a performance evaluation of standardized high quality audio transmission on top of Classic Bluetooth [14] and the development of a non-standardized approach to set up speech audio transmissions on top of BLE [15]. The work presented in the following sections contributes to this research domain by analyzing the performance of a standardized broadcast oriented wireless audio stream based on various packet loss scenarios and stream configurations. Instrumental metrics are used to quantify the audio quality and intelligibility of the evaluated speech audio.

4. Design and Implementation of the BIG Model

One of the main contributions of this paper is the design and implementation of an unframed BIG model. It is able to mimic a broadcaster that is packetizing outgoing SDUs into unframed PDUs within periodic BIG events, perform frequency channel selection for each SE according to the CSA #2 and mimic any synchronized observer that recombines incoming PDUs into a stream of SDUs. The main aspects of the model are depicted in Figure 2 and its design is based entirely on the theory described in Section 2. The incoming stream of SDUs at the observer can contain gaps due to one or more BLE packet loss techniques being applied on the SEs, since PDU losses can cause recombination of one or more SDUs to fail. The BLE packet loss technique evaluated in this paper, is based on a PER probability per channel, that remains constant for the entire BIG simulation. It can be used to probabilistically assess, per SE, whether receiving the PDU on the current channel f is successful. Therefore, event S_f represents a successful reception and $P(S_f) = 1 - PER_f$. For every SE, the model takes a sample from a uniform distribution between 0 and 1, and compares it to the threshold defined by the PER probability of the current channel to determine whether the PDU is lost or not. PLC strategies can then be used to fill the gaps created by lost SDUs.

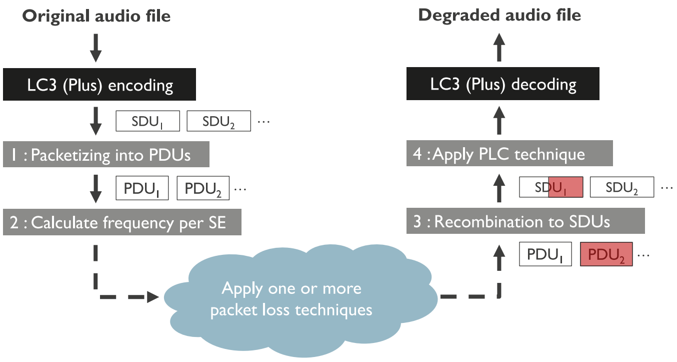


Figure 2: The unframed BIG model.

5. BIG Evaluation Setup

In order to assess the impact of various BLE packet loss scenarios on the audio quality and intelligibility of encoded speech audio, the model from Section 4 can be used. Each audio file is encoded using a combination of FI and EBR. The speech audio for the evaluation is taken from the PTDB-TUG dataset [16]. The complete dataset contains speech audio of 2342 distinct English sentences, recorded by at least one male and one female, at 48 kHz. 20 distinct speech audio files (10 per sex) are selected for the evaluation. To better represent mandatory speech audio use cases, all audio files are down-sampled to 16 kHz. The remainder of this section describes the evaluated codec parameters and the scoring metrics employed to quantify audio quality and intelligibility.

5.1. Codec Parameters

BIG observers must support 10 ms LC3 frames that are encoded at 16 kHz using a 40 byte SDU (i.e. 32 kbps EBR) [1]. This way, a BIG broadcaster knows that its audio streams can be decoded by any BIG observer. The first part of the evaluation examines the impact of various PDU loss rate (LR) scenarios and different numbers of SDU retransmissions in a default 32 kbps EBR-based BIG. The second part of the evaluation examines a 240 kbps

EBR-based BIG, which leads to an SDU size of 300 bytes and thus fragmentation into two PDUs of 150 bytes. Due to the larger SDU size and associated fragmentation, less bandwidth remains for SDU retransmissions in a single BIG event, compared to 32 kbps EBR-based BIGs.

5.2. Scoring Metrics

The ITU-T defines a five-degree Listening Quality (LQ) Objective (LQO) scale that can be used to grade the LQ of speech audio [17], also illustrated in Figure 2 of [18]. Perceptual scores from a sufficiently large enough group of participants will lead to a Mean Opinion Score (MOS). Instrumental metrics exist that analytically attempt to predict the MOS of an audio signal similar to how this would be done by a group of listeners. One such metric is Perceptual Evaluation of Speech Quality (PESQ) [19]. This is an intrusive metric and typically provides a MOS-LQO score between 1.0 and 4.5. It was developed to measure the audio quality in communications [18] and has been proven in the past to be applicable for codec evaluation [19]. The evaluation uses PESQ on the degraded audio signals to provide an initial understanding of the impact of frame loss and chosen PLC on the audio quality. Next to this, the Short-Time Objective Intelligibility (STOI) measure can be used to quantify the intelligibility of the degraded speech audio files [20]. It employs an intrusive approach to assess degraded speech audio and provides a scalar value between 0 and 1. The results in [20] have shown that any value above 0.8 can be considered to represent highly intelligible speech audio.

6. Evaluation of Shared PER Probability

The experiments evaluated in this section employ a shared PER probability for all BLE channels, in order to clearly show the impact of a specific PER on the speech audio. Therefore, there is no additional impact from the employed channel selection pattern based on the CSA #2. The evaluated shared PER probabilities can be structured into three LR categories: low (2.5% or 5%), middle (10% or 15%) and high (30% or 60%). Each combination of EBR, NSE and shared PER probability is evaluated 20 times per speech audio file. For each experiment, three distinct PLC methods are applied to the gaps in the incoming degraded encoded audio data at the observer. The first two methods replace all lost frames with either all zeroes (Zero Filling (ZF)) or the last correctly received frame (Repeat Previous (RP)). The third method is based on the PLC algorithm from the codec. The decoded audio signal of each PLC method, is given an audio quality and intelligibility score using PESQ and STOI respectively. For PESQ, the evaluation in the paper mainly focuses on the 2.0 and 3.0 MOS thresholds. Three associated zones are defined to categorize the PESQ scores: A (> 3.0), B (between 3.0 and 2.0) and C (< 2.0). The sample page [7] provides a demonstration of the relation between those zones and the actual listening experience of the associated audio signals. Speech audio in zone C experiences a huge drop in quality and the employed PLC technique can be deduced based on audible characteristics (robotic speech, many short silences, etc.). In Zone B, the employed PLC technique is no longer distinguishable but the drop in audio quality remains annoying. From Zone A, the drop in audio quality becomes less annoying.

6.1. Baseline Evaluation using LC3

The first set of experiments are performed using the default 32 kbps EBR. Each BIG configuration employs a BN of 1. The NSE parameter varies based on the number of retransmissions, which can go up to 14 [9]. Based on the following MPT of *(10 header bytes + 40 payload bytes) / 1M PHY = 400 μs*, this leads to consequent maximum BSD of *(15 - 1) * (400 μs + 150 μs) + 400 μs = 8.1 ms*. The largest control subevent currently defined by the specification is the Channel Map Update procedure and requires *(10 header bytes + 8 payload bytes) / 1M PHY = 144 μs*. This leads to a maximum BIG event duration of

*15 * (400 µs + 150 µs) + 144 µs = 8.4 ms*, which is still below the maximum allowed BIG event duration of 9.85 ms (10 ms ISI – T_IFS). Therefore, in theory, the current experiment set can employ a BIG configuration that supports any of the allowed retransmissions. However, the employed SUI and BCO in the calculations are minimum values and, in practice, the hardware implementing the BLE stack needs to combine the BIG schedule with other functionality, such as the advertisement of BIG availability to potential observers. Further research on the associated bandwidth limitations of potential combinations are not analyzed in this paper.

6.1.1. Evaluation of Audio Quality

Figure 3a illustrates the PESQ distribution for 2.5% shared PER probability. The boxplots indicate that low LRs already lead to an undesirable drop in audio quality when using no additional retransmissions. One additional retransmission ensures that most PESQ scores are in zone A. The notches of the boxplots for no additional retransmissions do not overlap, concluding that there is a 95% confidence that the differences are significant at the median level. The LC3 PLC is shown to perform worse compared to the other two concealment methods, of which RP performs best. This observation will be further analyzed in the next subsection. Figure 3b indicates that, for a 10% shared PER probability and no retransmissions, many scores are categorized in zone C. However, adding 1 retransmission is sufficient to ensure that almost all scores are at least in zone B. Moreover, from 2 retransmissions, all PESQ scores are shown to be in zone A. Figure 3c indicates that, for a 30% shared PER probability, more than 3 retransmissions are required to ensure that all PESQ scores are in zone A.

6.1.2. Evaluation of Intelligibility

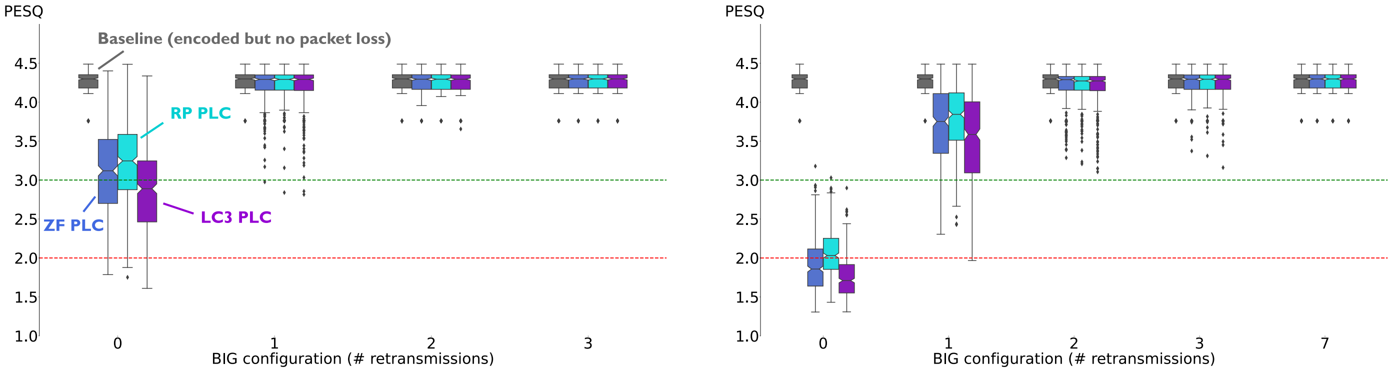
The results in the dataset [6] indicate that the shared PER probability needs to go beyond 30% for the STOI scores to significantly drop below the 0.8 threshold for all three PLC methods. Figure 3f illustrates that, for a shared PER probability of 60%, 2 retransmissions are required to ensure that at least all RP and LC3 PLC results yield a STOI score above 0.8, which, in practice, indicates high intelligibility. The ZF PLC results indicate less intelligible speech audio compared to the other two methods. Since it will fill the gaps with zeroes, the associated audio signals will contain many silence segments, and will thus result in much less intelligible speech audio.

6.1.3. Evaluation of Frame Loss

Table 1 illustrates the average SDU LR over all experiments, for a subset of the evaluated shared PER probabilities and BIG configurations. The experimental LRs are compared with the theoretical (Th) expected SDU LR, calculated based on the current BIG configuration and shared PER probability. Each SDU fits into a single PDU so the successful reception of an SDU depends on at least one of the NSE subevents in a single BIG event succeeding. Moreover, due to the shared PER probability between all BLE channels, each SE has an equal probability of succeeding. Consequently, the theoretical SDU LR becomes $\mathrm{PER}^{\mathrm{NSE}}$. Table 1 validates that each set of experiments mimics the expected theoretical SDU LR. Table 2 indicates the mean percentages of the contribution (contr.) of a certain loss burst (LB) length to the total number of LBs in an experiment, based on the shared PER probability and BIG configuration. Each LB represents a set of consecutive lost frames (CLFs) between two correctly received frames. High LR PER probabilities present a more noticeable contribution of larger LB lengths but indicate that, with an increasing number of retransmissions, the LB length of 1 becomes dominant again.

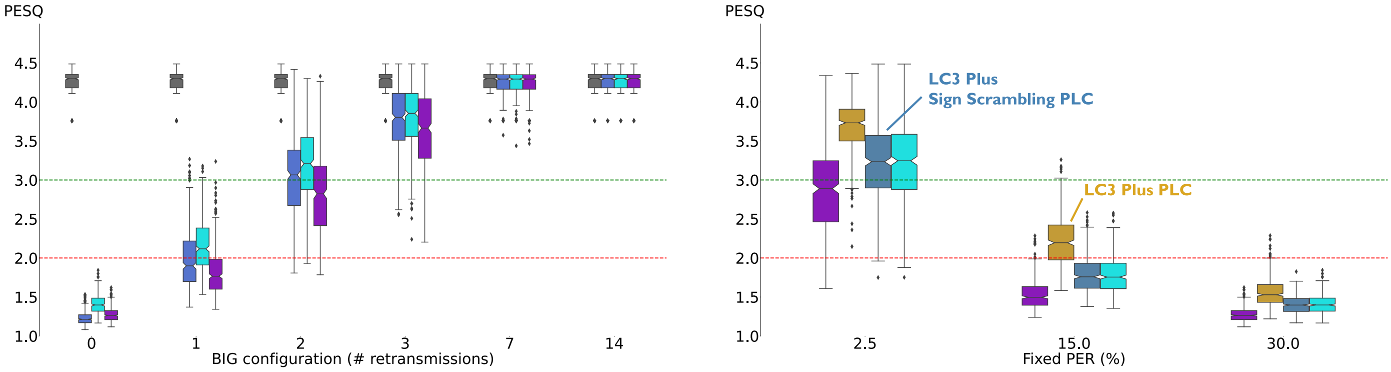
6.1.4. Relation between Frame Loss and Score Metrics

Figure 4a illustrates the PESQ score distribution of the LC3 PLC, for 20 distinct frame loss categories from 1% to 20%. Each boxplot contains the PESQ scores of the experiments where the SDU LR was more than the previous



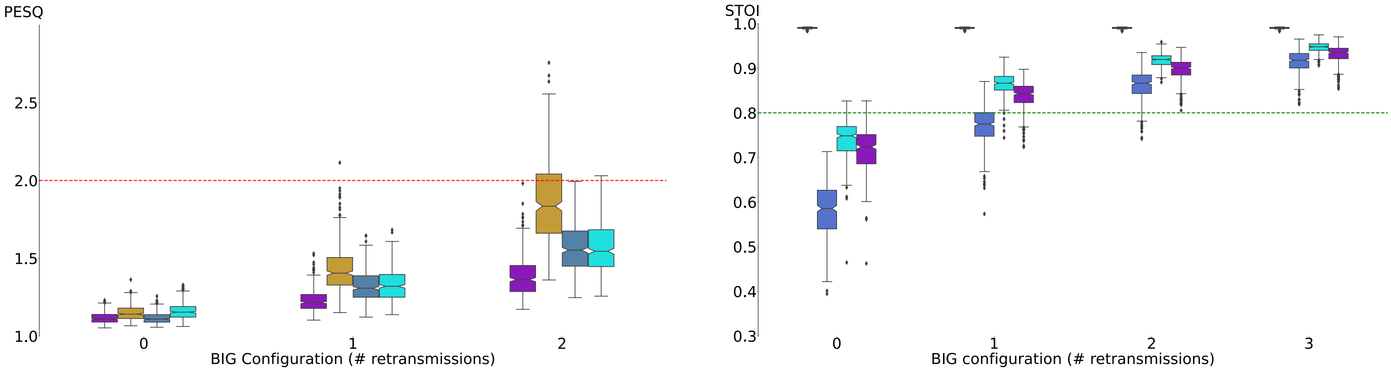
(a) PESQ for 2.5% PER.

(b) PESQ for 10% PER.



(c) PESQ for 30% PER.

(d) PESQ for no retransmissions.

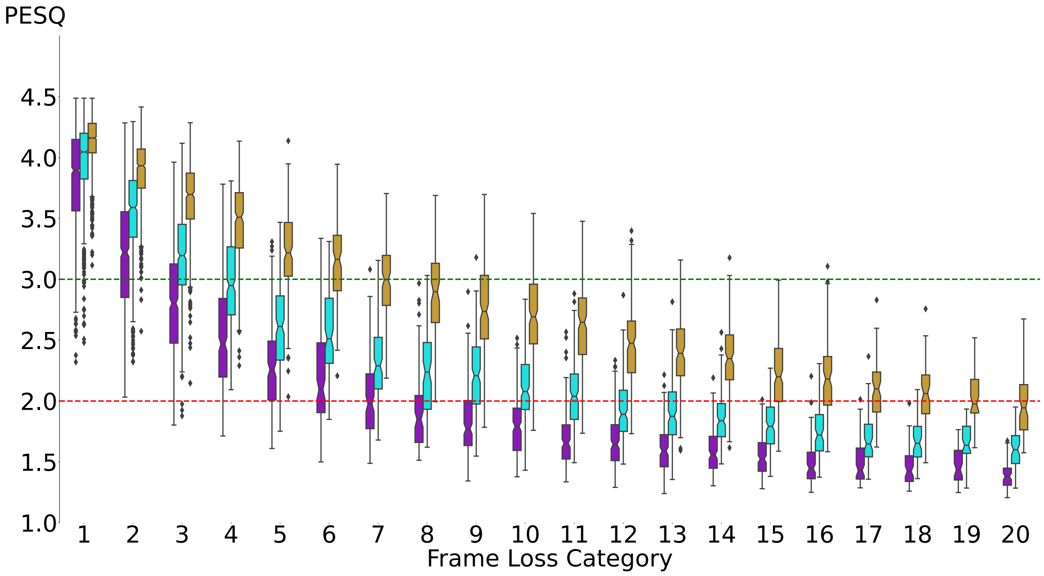


(e) PESQ for 60% PER.

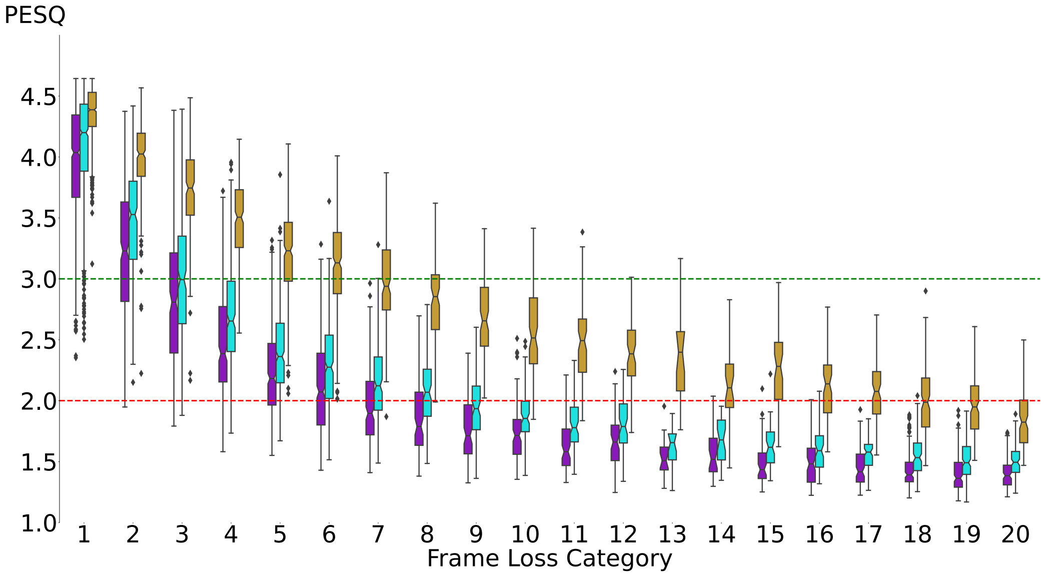
(f) STOI for 60% PER.

Figure 3: Evaluation of 32 EBR-based BIG configurations (the color scheme remains the same over all subfigures).

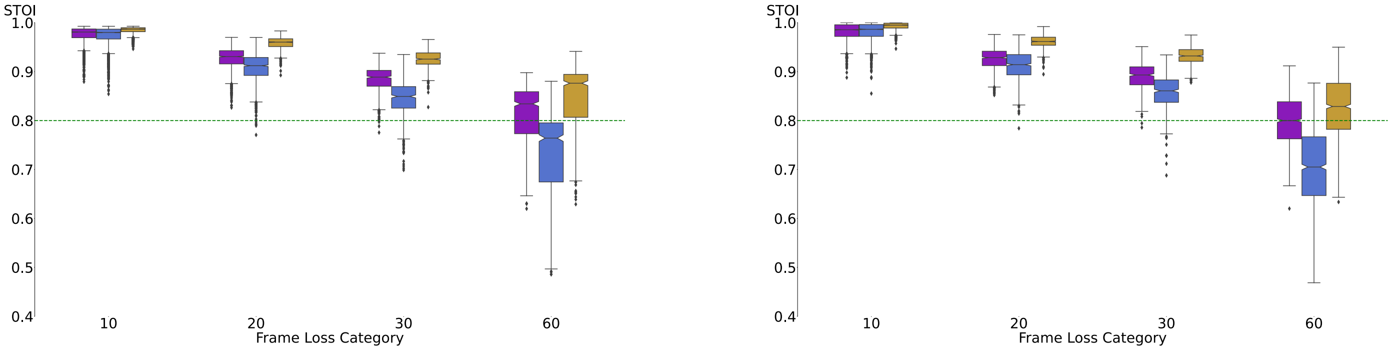
frame loss category (or more than 0% in case of frame loss category 1%) and less than or equal to the current frame loss category. The lowest frame loss categories experience the largest drop in performance for each additional loss percentage, indicated by the associated boxplots illustrating a difference in the PESQ scores that is significantly different at the median level. Upon getting closer to the zone C threshold, the additional loss percentages have less impact but the audio quality is not sufficient anyway, as can be heard on the sample page [7]. It is clear that frame loss percentages larger than 5% are undesirable due to more than 25% of the scores being in zone C. Moreover, a threshold of only 2% frame loss percentage is shown to be the maximum (integer) SDU LR that does not lead to scores ending up in zone C. Figure 4a also includes a similar PESQ score distribution for the RP PLC. Using this



(a) Relation PESQ & SDU LR (32 kbps).



(b) Relation PESQ & SDU LR (240 kbps).



(c) Relation STOI & SDU LR (32 kbps).

(d) Relation STOI & SDU LR (240 kbps).

Figure 4: Relation between the score metrics and frame loss for 32 and 240 kbps EBR-based BIG configurations (the color scheme is the same as Figure 3).

Table 1: SDU LR for 32 kbps EBR.

NSE	PER probability	Th SDU LR	Mean SDU LR (StDev)
1	5%	5.00%	4.92 (1.24)%
	10%	10.00%	10.01 (1.82)%
	30%	30.00%	29.97 (2.73)%
2	5%	0.25%	0.27 (0.30)%
	10%	1.00%	0.98 (0.60)%
	30%	9.00%	8.94 (1.65)%
3	5%	0.01%	0.01 (0.07)%
	10%	0.10%	0.10 (0.19)%
	30%	2.70%	2.62 (1.01)%

Table 2: Contribution per LB length for 32 kbps EBR.

NSE	PER probability	LB length	Mean Contr. (StDev)%
1	10%	1	89.88 (6.30)%
		2	9.13 (5.96)%
		≥ 3	0.98 (1.90)%
	30%	1	70.56 (5.89)%
		2	20.52 (5.26)%
		≥ 3	8.91 (3.95)%
	60%	1	40.06 (5.45)%
		2	23.86 (5.02)%
		≥ 3	36.08 (5.52)%
2	30%	1	91.20 (5.84)%
		2	7.89 (5.56)%
		≥ 3	0.92 (1.94)%
	60%	1	64.02 (5.87)%
		2	23.07 (5.11)%
		≥ 3	12.90 (4.28)%

method, no PESQ scores for SDU LRs less than or equal to 4%, go into zone C. Moreover, the PESQ scores are divided between zone B and C for much longer, and are only fully in zone C starting from frame loss category 19%. Figure 4c presents an illustration of the relation between STOI scores of the ZF & LC3 PLC and distinct frame loss categories. Most audio signals for LC3 PLC experiencing an SDU LR lower than or equal to 30% are shown to be still intelligible. The results for ZF PLC present a lower SDU LR threshold for intelligibility of the associated audio signals, due to the increased number of (consecutive) short silences in the signal.

6.2. Comparing LC3 and LC3 Plus

The experiments from Subsection 6.1 are repeated using LC3 Plus. The evaluated BIG configurations and shared PER probabilities are the same and each LC3 Plus experiment mimics the frame loss pattern of its LC3 counterpart. This way, the impact of employing the more well-thought-out approach to PLC that is present in LC3 Plus, can be directly compared to the simplistic LC3 PLC method as well as to the other two evaluated concealment methods, applied to the same frames being lost on the speech audio signals. The LC3 PLC method can be found in Appendix B.3 of the specification [4]. In case of a lost frame, the MDCT data associated with the last good frame is repeated. Next, the PLC algorithm applies pseudo-random sign scrambling on the spectral coefficients in the frame. The employed sign per spectral coefficient is determined based on a recalculation of an *unsigned* 16-bit PLC seed being lower than a fixed threshold of 32768 or not. Moreover, an increased number of CLFs in a single burst lead to the substituted spectrum fading out to zero faster [4]. The employed sign per spectral coefficient decision in LC3 PLC does not consider any additional characteristic of the last good frame, which can indicate whether sign scrambling could be beneficial or detrimental here, and thus leads to an unreliable performance. The LC3 Plus PLC method includes three distinct PLC techniques and at the start of each LB, the codec selects the most applicable technique for the lost frames in the burst, based on a set of signal features from the last good frame(s). One of the techniques is also based on sign scrambling but now takes into account an additional characteristic of the last good frame, which indicates, based on a pitch presence detection scheme in the codec [8], whether the frame was voiced or not. The associated PLC algorithm can be found in section 5.6.3.2 of the LC3 Plus specification [8]. The sign is not changed in

Table 3: Additional thresholds for when a pitch is present in the last good frame (10 ms FI and 60 ms PTE)

Number of CLFs	Additional threshold
< 3	-32768
3	-24576
4	-16384
5	-8192
> 5	0

case the current *signed* 16-bit PLC seed is a non-negative number or, it is not lower than an additional threshold and there is a pitch present in the current frame. Thereby, in case a pitch is present in the last correctly received frame, the probability of a negated spectral coefficient is potentially much lower compared to the LC3 PLC, depending on the current value of the additional threshold. The calculation of the additional threshold for the current lost frame employs the FI, the current number of CLFs and an additional parameter PLC4_TRANSIT_END_IN_MS (PTE). The combination of FI and PTE controls how many intermediate thresholds can be calculated between the lowest threshold and a threshold of 0, depending on the current number of CLFs. For the experiments performed in this paper, PTE was set to 60 ms (which is the default value) and FI is 10 ms, which leads to the distinct thresholds presented in Table 3. Consequently, for a short number of CLFs (less than 3) and in case the last good frame had a pitch present, the current PLC seed should be lower than a threshold of -32768 for the sign of the spectral coefficient to be altered. But a signed 16-bit PLC seed can never be lower than -32768. Therefore, in case of predominantly short LB lengths, the behavior of the sign scrambling method in LC3 Plus PLC is expected to represent a more comparable performance to the RP PLC from Subsection 6.1. Figure 3d illustrates PESQ score distributions for the encoded speech audio subjected to either 2.5%, 15% or 30% shared PER probabilities and all three employing a BIG configuration with no additional retransmissions. It presents a comparison between the various PLC methods, including RP PLC and LC3 Plus PLC where the sign scrambling approach is enforced for every lost frame. It validates that the sign scrambling method in LC3 Plus PLC outperforms LC3 PLC and is more comparable to the performance observed for the RP PLC. Table 3 illustrates that 3 or more CLFs are required in order to have a non-zero probability by which the sign of a spectral coefficient is altered during the PLC algorithm, in case the last good frame was detected as voiced. Table 2 shows that, for 30% shared PER probability and no retransmissions, almost 10% of the LB lengths are at least equal to 3. However, regardless of the associated increased probability of negated spectral coefficients in the PLC algorithm, the distributions in Figure 3d illustrate that the results for enforced sign scrambling in LC3 Plus PLC are still comparable to the RP PLC. For 60% shared PER probability and no retransmissions, Table 2 shows that almost 40% of all LB lengths are at least equal to 3. For this combination, Figure 3e illustrates a clear difference between the results for RP and enforced sign scrambling in LC3 Plus PLC, and also indicates that the latter is now more comparable to the LC3 PLC distribution instead (due to the increased probability of sign scrambling). Adding 1 retransmission reinstates the more comparable performance to RP PLC, which is validated by Table 2 showing a decrease in the contribution of LB lengths at least equal to 3, to only a bit more than 10% for 60% shared PER probability. In general, the sign scrambling based PLC strategy in LC3 Plus is presented as best suited for noise-like signals without any dominant harmonic structure [11]. This property is reflected by the PLC technique selection mechanism in LC3 Plus [8], which selects between the sign scrambling based technique or two other techniques, based on a pitch related characteristic of the signal. The other two techniques are presented as more suited for either monophonic signals with a periodic structure (e.g. voiced speech) or complex signals that exhibit both harmonic structure and noise-like components [11]. Therefore, employing unaltered LC3 Plus PLC (without enforced sign scrambling) on the speech audio evaluated in this paper, should exhibit an even higher audio quality preservation. This is reflected by Figure 3d and Figure 3e. Figure 4a includes the PESQ score distribution of the LC3 Plus PLC. Compared to the results for LC3 and RP PLC, the SDU LR threshold before scores are in zone C, has increased to 8%. Moreover, at 20% frame loss, the median of the boxplot is only slightly below the zone C threshold. Figure 4c illustrates that the intelligibility gain of using LC3 Plus PLC is more marginal since only an additional portion of the 60% shared PER probability results yield a STOI score above the 0.8 threshold and the other frame loss categories have a bit more margin on top of the minimum threshold. The sample page in [7] presents an audible comparison between the different PLC techniques. The speech audio from RP PLC in Zone C sounds very robotic. Applying pseudo-random sign scrambling can counter this effect, but the samples for LC3 PLC illustrate that an annoying addition of vocoding artefacts is present instead. The sign scrambling algorithm in LC3 Plus PLC presents a decreased probability of sign scrambling for lost frames where the last good frame was detected as voiced, especially for short LB lengths, thereby reducing the vocoding artefacts for dropped frames. The associated instrumental PESQ scores in [7] are shown to be similar to their RP PLC counterparts and the audio samples illustrate an audible similarity too. Lastly, in comparison, listening to the audio samples for LC3 Plus PLC indicates both a reduction of vocoding artefacts and robotic speech. Based on the PESQ score comparison, this audio quality increase can be expected for default LC3 Plus PLC.

6.3. Impact of SDU Fragmentation

The next part of the evaluation examines the impact of SDU fragmentation on the audio quality and intelligibility of speech audio by means of employing a 240 kbps EBR. Due to the fact that a single SDU requires two larger PDUs, less retransmissions are possible in a single BIG event. The associated MPT is *(10 header bytes + 150 payload bytes) / 1M PHY = 1.28 ms* and, for 2 additional retransmissions, the BSD is *(6 - 1) * (1.28 ms + 150 µs) + 1.28 ms = 8.43 ms.* Including the Channel Map Update procedure, leads to a maximum BIG event duration of *6 * (1.28 ms + 150 µs) + 144 µs = 8.724 ms*, which is below the maximum allowed BIG event duration of 9.85 ms. However, it is clear that adding an additional SDU retransmission would surpass that threshold. The fragmentation of an SDU into two halves, requires the reception of both fragments at an observer in order to correctly reconstruct the SDU. The associated theoretical probability of successful reception of an SDU depends on the correct reception of each PDU individually (1 - $(\mathrm{PER})^{\mathrm{GC}}$, for each PDU). In case one or more PDUs are lost, the SDU is also considered lost. Therefore, the theoretical SDU LR can be formulated as $1 - (1 - (\mathrm{PER})^{\mathrm{GC}})^{\mathrm{BN}}$. Table 4 validates that each set of experiments mimics the expected theoretical SDU LR. It also indicates that, for the same shared PER probability, higher SDU LRs are expected compared to the 32 kbps EBR-based results presented in Table 1, and that more retransmissions are required to ensure that the SDU LRs for higher shared PER probabilities are below a certain threshold. Figure 4b and Figure 4d illustrate a similar relation between frame loss rate and score metrics as Figure 4a and Figure 4c, but now for the 240 kbps EBR results, which use 300 bytes frames. The 1% frame loss category in Figure 4b illustrates a clear difference compared to the same category in Figure 4a, which can be related to the expected gain in encoding efficiency by employing the increased EBR. However, from the 2% frame loss category, the difference between the two EBR score distributions decreases and not every frame loss category in the 240 kbps EBR distribution is shown to exhibit a better PESQ score preservation compared to its 32 kbps

Table 4: SDU LR for 240 kbps EBR.

NSE	PDU LR	Th SDU LR	Mean SDU LR (StDev)
2	5%	9.75%	9.58 (1.64)%
	10%	19.00%	19.18 (2.25)%
	30%	51.00%	50.99 (3.00)%
4	5%	0.50%	0.48 (0.39)%
	10%	1.99%	2.01 (0.81)%
	30%	17.19%	17.11 (2.24)%
6	5%	0.02%	0.02 (0.09)%
	10%	0.20%	0.19 (0.25)%
	30%	5.33%	5.33 (1.22)%

EBR counterpart. Similarly, for the STOI metric, Figure 4d is shown to experience small differences compared to Figure 4c and some 240 kbps EBR score distributions are shown to behave worse compared to their 32 kbps EBR counterpart, for the same frame loss category. Comparing the zone C thresholds for PESQ, indicates that Figure 4b contains a similar threshold for LC3 PLC (2%), a small increase for LC3 Plus PLC (9%) and a small decrease for RP PLC (3%). Moreover, for LC3 and RP PLC, the frame loss category associated with the provided threshold already observes a small portion of score metrics in zone C. The increase in audio quality preservation is still clearly observable from the zone C thresholds for the different PLC techniques. However, further studies are required into the observed threshold variability and the same frame loss category not always exhibiting similar or better audio quality and intelligibility preservation for 240 kbps EBR, compared to its 32 kbps EBR counterpart. This way, it can be determined whether these observations can be mainly related to insufficient size and variety for each EBR dataset or other aspects are involved. Comparing the observed zone C thresholds for 240 kbps EBR against the SDU LRs in Table 4, indicates that the maximum number of retransmissions is insufficient to ensure that the shared PER probability of 30% leads to an SDU LR that is lower than or equal to 2%. Therefore, LC3 PLC will not be able to guarantee PESQ scores above zone C for any number of retransmissions. For LC3 Plus, only the maximum number of retransmissions for a 30% shared PER probability, lead to an SDU LR that is below the zone C threshold of 9%. Moreover, comparing the SDU LRs with Figure 4d, indicates that using no retransmissions for 30% shared PER probability will potentially lead to STOI scores lower than 0.8 when employing either LC3 or LC3 Plus PLC.

7. Evaluation of Variable PER Probability

The evaluation in Section 6 limits its experiments to a shared PER probability for all BLE channels and does not fully exploit the capabilities of the BIG model. This section examines the impact of a variable PER probability on each BLE channel and how the combination of CSA #2 and one or more retransmissions in a BIG event can limit frame loss. The variable PER probabilities are based on observations in real-life environments, in particular related to the BLE channels that overlap with the three non-overlapping 802.11 channels (1, 6 and 11) in the 2.4 GHz ISM band. The work in [12] presents the results of a spectrum survey performed in various real-life environments, for several hours per environment. It can quantify per-channel availability probabilities for a BLE connection operating in such environments, depending on its sensitivity to the underlying interference (i.e. interference sensitivity threshold (ITH), in dBm). The ITH is related to the expected signal power of a received BLE signal at the physical position where the spectrum survey is conducted. The spectrum survey collects several power measurements for each BLE channel over a period of time. Based on the number of spectrum entries for which the measured power value of the BLE channel does not surpass the ITH, divided by the total number of spectrum entries for that channel, a per-channel availability percentage can be calculated. The evaluation in this paper employs these values to quantify a PER probability for each channel (100% - per-channel availability percentage). In theory, success-

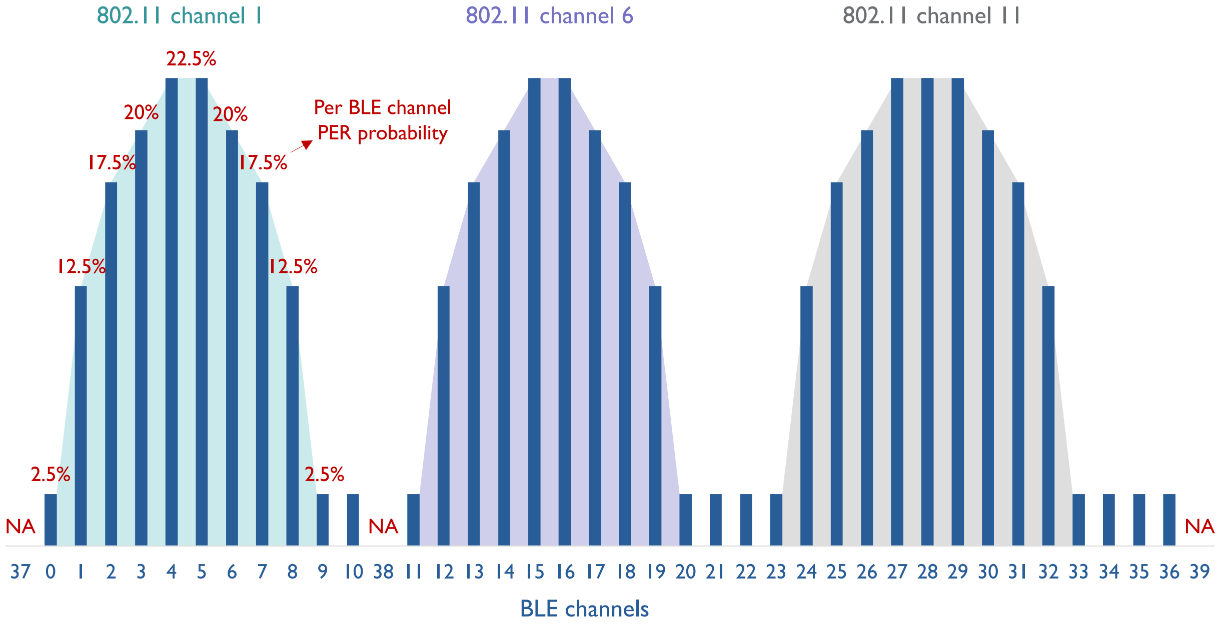


Figure 5: PER probability per BLE channel (-75 dBm ITH) [12].

ful decoding of a BLE signal at a receiver depends on the relation between the signal-to-interference-plus-noise ratio (SINR) and the receiver sensitivity (depending on the BLE PHY). But the goal of this section is to illustrate that the model can test out any measured set of per-channel PER probabilities, thus the underlying evaluation assumes that measured power surpassing the ITH in [12] will always lead to unsuccessful decoding of the BLE signal at the receiver. Figure 5 illustrates the PER probabilities that can be quantified for BLE channels overlapping with 802.11 channel 1, roughly based on a spectrum survey conducted in a university food court [12]. The indicated PER probabilities are based on an ITH of -75 dBm. The evaluation in this section employs the PER probabilities for three distinct ITHs: -75, -85 and -95 dBm. Based on the results in [12], it can be assumed that the PER probabilities for the -85 and -95 dBm ITHs are respectively 10% and 20% higher than the results for -75 dBm ITH indicated in Figure 5. In general, the transmission signal power on an 802.11 channel should remain below a specified spectrum mask, which decreases closer to either ends of the channel. Therefore, the BLE signal power of the overlapping BLE channels closer to either ends of the 802.11 channel, needs to be even lower in order for the Wi-Fi interference to still have a detrimental impact on the successful reception of a BLE signal. Thus, in a realistic environment, it can be expected that the PER probabilities for BLE channels closer to the left or right border of the 802.11 channel, are lower compared to the PER probabilities of the BLE channels closer to the center frequency of the 802.11 channel. This behavior is clearly present in Figure 5. Consequently, the evaluation in this section assumes that the PER probabilities of the outer overlapping BLE channels 0 and 9 are actually no longer impacted by the interference on the 802.11 channel and are thus solely related to interference from other technologies. In order to provide every remaining non-overlapping BLE channel with a realistic PER probability as well, the evaluation in this section assumes that the remaining non-overlapping BLE channels have the same PER probability as channel 0 and 9. Moreover, the employed PER probabilities for BLE channels overlapping with 802.11 channel 1 can be mapped on 802.11 channels 6 and 11 too, as illustrated in Figure 5. This way, the evaluation can be extended towards an environment where there is similar Wi-Fi interference on two or three 802.11 channels simultaneously. The 802.11 channel 11 overlaps with one additional BLE channel compared to the other two 802.11 channels. The evaluation assumes that, in that case, three (instead of two) BLE channels closest to the center frequency experience the highest PER probability. The evaluation examines the impact of each probabilistic combination of ITH and number of active 802.11 channels on the 32 kbps EBR audio files. It uses the BIG configurations from Subsection 6.1 and 20 distinct experiments are performed per BIG configuration and probabilistic combination on each of the 20 speech audio files. Each experiment employs a randomly generated bigEventCounter between 0 and 30000 for its first BIG event, emulating that speech audio is started within the first 5 minutes after the BIG is set up (i.e. 5 minutes / 10 ms FI = 30000), and also employs a BIS access address based on a randomly generated SAA. This way, the two input parameters to the CSA #2 are randomly generated for each experiment. Table 5 illustrates a subset of the results for the average SDU LR over all experiments, per probabilistic combination. It can be seen that, without any retransmissions, frame LR is always more than 5%. Based on Figure 4a, this results in at least 25% of the PESQ scores being in zone C for LC3 PLC. Additional overlapping 802.11 channels that experience the same range of PER probabilities are shown to provide less impact on the SDU LR compared to a decreased ITH for an unchanged number of overlapping channels. The dataset in [6] illustrates that the worst case probabilistic combination of -95 dBm ITH and all three 802.11 channels being actively in use, requires 2 or 3 additional retransmissions to be below the 8 or 2% SDU LR thresholds of respectively LC3 Plus or LC3 PLC from Figure 4a (i.e. 3.66% and 1.22% SDU LR respectively). In general, the SDU LRs exhibit a large variability, depending on the current interference environment and employed BIG configuration. Each LR category lower than or equal to 30%, evaluated in Section 6, is relatable to one or more SDU LR(s) in Table 5, which further validates the relevance of examining these different LR categories. The maximum SDU LR from Table 5 is around 30% but ITHs lower than -95 dBm in [12] present even higher PER probabilities for overlapping BLE channels, close to or surpassing 60%. Therefore, studying the edge case of a 60% LR category in Section 6 is relevant as well.

Table 5: SDU LR for overlap with Wi-Fi.

NSE	802.11 channel(s)	ITH	Mean SDU LR (StDev)%
1	1	-75 dBm	5.88 (1.30)%
		-85 dBm	15.91 (2.16)%
		-95 dBm	25.98 (2.79)%
	1 & 6	-75 dBm	9.27 (1.64)%
		-85 dBm	19.30 (2.27)%
		-95 dBm	29.28 (2.66)%
	1, 6 & 11	-75 dBm	13.33 (2.14)%
		-85 dBm	23.27 (2.68)%
		-95 dBm	32.88 (2.71)%
2	1	-75 dBm	0.21 (0.27)%
		-85 dBm	2.42 (0.91)%
		-95 dBm	6.68 (1.55)%
	1 & 6	-75 dBm	0.72 (0.53)%
		-85 dBm	3.56 (1.10)%
		-95 dBm	8.43 (1.65)%
	1, 6 & 11	-75 dBm	1.82 (0.76)%
		-85 dBm	5.56 (1.34)%
		-95 dBm	11.17 (1.72)%

8. Conclusion and Future Work

The BIG model is shown to be a useful tool to assess the impact of a diverse number of retransmissions on the audio quality and intelligibility of encoded speech audio, subjected to various PER probabilities, either shared or variable per BLE channel. The evaluation illustrates that the LC3 PLC method has limited potential to deal with frame loss, even being surpassed in performance by simply repeating the last correctly received audio frame. The more advanced PLC algorithms in LC3 Plus are shown to provide an increased robustness against frame loss for the audio quality and, for both codec PLC methods, the encoded speech audio is shown to remain intelligible up until high LRs. Future evaluation of the model could examine other methods to emulate environmental conditions on BLE channels. Moreover, a comparison of the model-based experiments against the performance evaluation of a solution implemented on real hardware, can provide additional validation of the model itself. Lastly, the current implementation can also be extended with additional functionality (multiple BISes, PTO-based pre-transmissions, etc.) in order to evaluate other BIG implementations (including other PHYs and limited channel maps).

Acknowledgments

The authors would like to thank Loubna Kadri for her assistance during the evaluation. The paper results from the imec.icon project BLE2AV, financed by imec and Flanders Innovation & Entrepreneurship.

References

[1] N. Hunn, Introducing Bluetooth LE Audio, Independently published (January 1, 2022) (Jan 2022).

[2] Bluetooth SIG, Auracast (2023. [Online]).
URL `https://www.bluetooth.com/auracast/`

[3] M. Spörk, J. Classen, C. A. Boano, M. Hollick, K. Römer, Improving the Reliability of Bluetooth Low Energy Connections, in: Proceedings of the 2020 International Conference on Embedded Wireless Systems and Networks on Proceedings of the 2020 International Conference on Embedded Wireless Systems and Networks, EWSN '20, Junction Publishing, USA, 2020, p. 144–155.

[4] Bluetooth SIG, Low Complexity Communication Codec, `https://www.bluetooth.com/specifications/specs/low-complexity-communication-codec-1-0` (Sep 2020. [Online]).

[5] M. Afaneh, A Technical Overview of LC3, `https://www.bluetooth.com/blog/a-technical-overview-of-lc3/` (Nov 2020. [Online]).

[6] M. Baert, BIG Model executable and dataset, `https://gitlab.ilabt.imec.be/mrobaert/multtransbigmodeldata` (March 2023. [Online]).

[7] M. Baert, Audio Samples, `https://imec-idlab.github.io/MultTransSamplePage/` (Feb 2023. [Online]).

[8] Technical Committee Digital Enhanced Cordless Telecommunications, Low complexity communication codec plus, `https://www.iis.fraunhofer.de/en/ff/amm/communication/lc3.html` (2021. [Online]).

[9] Bluetooth SIG, Core Specification 5.2, `https://www.bluetooth.com/specifications/specs/core-specification-5-2/` (Dec 2019. [Online]).

[10] B.-Z. Pang, T. Claeys, D. Pissoort, H. Hallez, J. Boydens, Comparative Study on AFH Techniques in Different Interference Environments, in: 2019 IEEE XXVIII International Scientific Conference Electronics (ET), 2019, pp. 1–4. doi:10.1109/ET.2019.8878594.

[11] M. Schnell, E. Ravelli, J. Büthe, M. Schlegel, A. Tomasek, A. Tschekalinskij, J. Svedberg, M. Sehlstedt, LC3 and LC3plus: the New Audio Transmission Standards for Wireless Communication, Journal of the Audio Engineering Society (May 2021).

[12] M. Omar Al Kalaa, W. Balid, N. Bitar, H. H. Refai, Evaluating Bluetooth Low Energy in realistic wireless environments, in: 2016 IEEE Wireless Communications and Networking Conference, 2016, pp. 1–6. doi:10.1109/WCNC.2016.7564809.

[13] J. Lee, Broadcast Audio Transmission for Bluetooth LE on an Interfered ISM Band, IEEE Internet of Things Journal 6 (4) (2019) 6140–6150. doi:10.1109/JIOT.2018.2875773.

[14] S. Zeadally, A. Kumar, Design, implementation, and evaluation of the audio/video distribution transport protocol (avdtp) for high quality audio support over bluetooth, Computer Communications 28 (2) (2005) 215–223. doi:https://doi.org/10.1016/j.comcom.2004.09.007.
URL `https://www.sciencedirect.com/science/article/pii/S0140366404003408`

[15] M. Gentili, R. Sannino, M. Petracca, Bluevoice: Voice communications over bluetooth low energy in the internet of things scenario, Computer Communications 89-90 (2016) 51–59, internet of Things Research challenges and Solutions. doi:https://doi.org/10.1016/j.comcom.2016.03.004.
URL `https://www.sciencedirect.com/science/article/pii/S0140366416300627`

[16] G. Pirker, M. Wohlmayr, S. Petrik, F. Pernkopf, A Pitch Tracking Corpus with Evaluation on Multipitch Tracking Scenario., in: Interspeech - International Conference on Spoken Language Processing, 2011, pp. 1509–1512. doi:10.21437/Interspeech.2011-317.

[17] International Telecommunication Union, ITU-T Rec. P. 800. methods for subjective determination of transmission quality, `https://www.itu.int/rec/T-REC-P.800-199608-I` (Aug 1996. [Online]).

[18] D. Vasicek, M. Mikulec, E. Gresak, F. Rezac, E. Chromy, Mobile Probe for Cellular Network Coverage and Quality Measurement, in: J. Gottvald, P. Praus (Eds.), Proceedings of the 3rd Czech-China Scientific Conference 2017, IntechOpen, Rijeka, 2017, Ch. 6. doi:10.5772/intechopen.71105.
URL `https://doi.org/10.5772/intechopen.71105`

[19] International Telecommunication Union, ITU-T Rec. P. 862. perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs., `https://www.itu.int/rec/T-REC-P.862` (Feb 2001. [Online]).

[20] C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen, An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech, IEEE Transactions on Audio, Speech, and Language Processing 19 (7) (2011) 2125–2136. doi:10.1109/TASL.2011.2114881.