Audio Engineering Society

Convention Paper 10491

Presented at the 150th Convention
2021 May 25–28, Online

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

LC3 and LC3plus: The new audio transmission standards for wireless communication

Markus Schnell[1], Emmanuel Ravelli[1], Jan Büthe[1], Maximilian Schlegel[1], Adrian Tomasek[1], Alexander Tschekalinskij[1], Jonas Svedberg[2], and Martin Sehlstedt[2]

[1]*Fraunhofer IIS, Germany*
[2]*Telefonaktiebolaget LM Ericsson, Sweden*

Correspondence should be addressed to Markus Schnell (markus.schnell@iis.fraunhofer.de)

ABSTRACT

The new Low Complexity Communication Codec (LC3) and its sibling Low Complexity Communication Codec Plus (LC3plus) were developed to solve essential shortcomings present in today's short-range wireless communication platforms such as Bluetooth and Digital Enhanced Cordless Telecommunications (DECT). The codec operation modes range from medium bit rates for optimal voice transmission to high bit rates for high-resolution music streaming services. Furthermore, the codecs operate at low latency, low computational complexity, and a low memory footprint.

1 Introduction

More than one billion wireless audio devices are sold every year. These include loudspeakers, headsets, headphones, true wireless earbuds, smart speakers, hearing aids, cordless phones, and many more hearables. The leading standards of wireless short-range communication are Bluetooth and ETSI DECT, which recently received a significant technology upgrade to fulfill today's requirements. One major component is the underlying audio codec, the Low Complexity Communication Codec LC3, and its sibling LC3plus. This paper aims to explain the technical details of these new audio codecs.

2 Requirements

The voice communication over mobile networks got already significantly improved by the introduction of the 3GPP EVS codec. LC3 and LC3plus can be seen as the counterparts of EVS, extending the high quality voice link to wireless accessories.

The use cases of LC3 and LC3plus include not only voice calls, but all kind of audio applications. The codec design was constrained by three major criteria: audio quality, latency, and complexity.

The codec was required to provide wideband (WB) voice service quality level at medium bit rates (32 kbps per channel at 16 kHz sampling rate), and to scale up to transparent music streaming quality at high rates (124 kbps per channel at 48 kHz sampling rate). In general, the goal was to reduce the required bit rate by roughly 50 % compared to the legacy codecs. The bit-rate reduction is one key component to reduce radio emission power or gain radio capacity. Specific stereo coding tools have been excluded to allow transmission of individual audio channels for true wireless hearables.

As LC3 and LC3plus are considered transcoders to enable audio accessories, the additional delay needed to be kept to a minimum. This allows forwarding of phone calls without disturbance or TV sound distribution without breaking the lip-sync criteria between video and audio. With LC3plus, even extremely low-delay cases for gaming headsets can be addressed. The latency of the codec is determined by the frame duration and the algorithmic delay, as outlined in Table 1.

Codec and Standard	Frame duration [ms]	Alg. delay [ms]	Total latency [ms]
Bluetooth LC3	10, 7.5[1]	2.5, 4	12.5, 11.5
ETSI LC3plus	10, 5, 2.5	2.5	12.5, 7.5, 5

[1] Added due to backwards compatibility reasons to Bluetooth Classic

Table 1: Latency of LC3 and LC3plus

Another primary goal was to connect hearing aids directly to the Bluetooth world. As hearing aids are highly constrained by size and computational power, low algorithmic complexity, low memory footprint, and small code size were targeted right from the start. As a result, LC3 and LC3plus require less than 10 WMOPS for encoding and decoding a WB voice call with a memory footprint well below 100 kByte.

Finally, it was essential to have one common design which can serve all the different needs spanning from hearing aids up to HiFi headphones.

3 Technical Solution

3.1 Codec Overview

LC3 and LC3plus are both an evolution of the modified discrete cosine transform (MDCT) based transform coded excitation (TCX) coder, which is part of the EVS codec [1, 2]. However, significant technical modifications had to be applied to gain efficiency as well as reduce delay and complexity. The improvements generally exploit the fact that the intended use cases for LC3 and LC3plus target a higher bit-rate range. This allowed replacing some components of the EVS codec with less complex solutions. An overview of the processing blocks are provided in Figures 1 and 2. In the following, all processing modules are explained in detail.

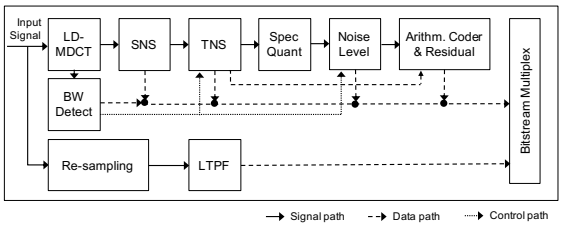


Fig. 1: Encoder overview

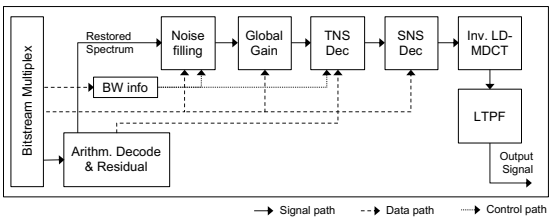


Fig. 2: Decoder overview

3.2 Low-Delay MDCT

MDCT is a widespread time/frequency transformation for perceptual audio coding. While the modulation is identical in most audio codec standards, the window shapes differ significantly. The AAC codec started with symmetric windows, AAC-ELD [3] introduced asymmetric low-delay windows, and CELT [4] implemented the low delay aspect with symmetric windows. A low-delay window means a window exhibiting $Z > 0$ leading zeros in this context. Such windows reduce the algorithmic delay of the MDCT from N to $N - Z$, where N denotes the frame length in samples. The MDCT is the only contribution to the algorithmic delay of LC3 and LC3plus.

LC3 and LC3plus utilize an asymmetric window shape to maximize the number of non-zero window coefficients and thus spectral channel separation capability. An analysis of the AAC-ELD window in [5] revealed that it is beneficial to add the temporal shape of the quantization error introduced in the spectral domain as a design criterion. This aspect was taken into account for the ALDO window used in the EVS codec. However, the design algorithm has rather limited freedom for optimization. The construction of the LC3 and LC3plus windows improves on that by using a mathematical optimization algorithm as described in [6] which takes both the frequency response and the temporal shape of the quantization noise into account.

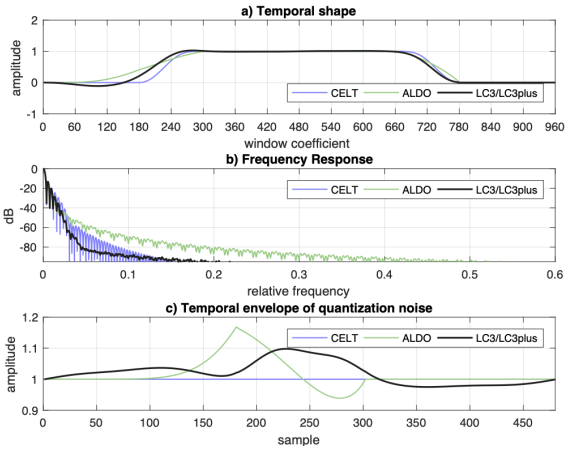


Fig. 3: Comparison of MDCT window shapes of CELT, ALDO and LC3/LC3plus regarding temporal shape, frequency response and temporal envelope of quantization noise

Figure 3 compares the MDCT window shapes of CELT, ALDO, and LC3/LC3plus for N=960 and Z=180. All windows induce the same algorithmic delay, see a). Regarding the frequency response shown in b), the LC3 and LC3plus window shows a higher attenuation than the CELT window over the complete frequency range. The ALDO window exhibits a slightly stronger attenuation of the first side lobes but with a considerably worse far-field attenuation than CELT and LC3/LC3plus. For LC3/LC3plus, a fast attenuation towards the 16-bit noise level (-96 dB) was of high priority to maximize the energy compaction gain for single tones. Figure 3-c) plots the temporal shape of the quantization error for the different windows. Here, the CELT window shows a perfectly flat shape due to its symmetry, and the quantization noise shaping of the LC3 and LC3plus windows is considerably weaker than that of the ALDO window.

3.3 Bandwidth Detector

For optimal support of phone calls, the bandwidth detector tests for the typical telephony bandwidths NB, WB, SWB, and FB. Additionally, SSWB was introduced, which refers to 12 kHz audio bandwidth. Whenever a signal is bandwidth-limited, e.g., due to an NB legacy headset, the bandwidth detector controls the Temporal Noise Shaping (TNS) and noise filling tool accordingly to avoid any blind extrapolation of the signal to the upper spectrum.

3.4 Spectral Noise Shaping

Spectral noise shaping (SNS) is an essential tool in any transform-based audio codec. It shapes the quantization noise in the frequency domain such that it is minimally perceived by the human ear, maximizing the perceptual quality of the decoded output. In EVS, spectral noise shaping is performed with an LPC-based perceptual filter, the same perceptual filter used in recent ACELP-based speech codecs (e.g., AMR-WB [7, 8]). However, the LPC estimation (autocorrelation, Levinson-Durbin), LPC quantization (LPC to LSF conversion, vector quantization), and LPC frequency response computation are costly operations. In LC3 and LC3plus, the LPC-based perceptual filter is replaced by 16 scale factors that mimic the LPC filter's frequency response but are entirely computed in the frequency domain to minimize the computational complexity.

To this end, the energy of the signals computed in 64 frequency bands from the MDCT coefficients. The bands are non-uniform and follow the perceptually-relevant Bark scale. Afterward, the energy values are smoothed, pre-emphasized and subsequently a noise floor at -40 dB is added. The resulting 64 values are then transformed into the logarithmic domain and downsampled by a factor of 4. The final 16 scale factors are then obtained by removing the mean value and scaling by a factor of 0.85.

The 16 scale factors are subsequently quantized with a fixed budget of 38 bits. The quantizer is explicitly designed for LC3 and LC3plus and targets low complexity and low memory footprint. This is achieved by a two-stage approach.

In the first stage, the lower and upper eight scale factors are quantized using an exhaustive codebook search. The codebooks each have 32 entries generated using the Linde-Buzo-Gray algorithm [9]. In the second stage, the first stage's quantization residual is further quantized using a pulse vector quantizer (PVQ), which allows for an efficient iterative codebook search. In particular, the codebooks do not need to be stored, which keeps the memory footprint small.

The second quantization stage utilizes a DCT-II transform. Four sub-modes are supported to cover all relevant shapes, which differ with respect to the PVQ layout and the number of quantization levels for a gain parameter. The sub-modes implement different trade-offs between shape resolution (more pulses) and dynamic

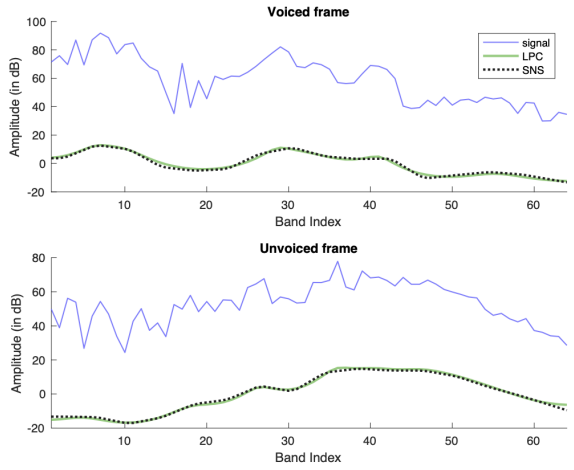


Fig. 4: Comparison of the frequency shapes of an LPC-based approach and the new SNS approach

range (more gain levels). Since the shape codebooks for different sub-modes overlap, a codebook search in all sub-modes can be implemented efficiently. The resulting shape pulse vector is encoded in a similar way as in [10], however the codebooks employed for LC3 and LC3plus have been optimized such that decoding can be done efficiently within 24-bit word indices.

The quantized scale factors are used as follows to shape the MDCT spectrum. The quantized 16 scale factors are upsampled by a factor of four and the resulting 64 factors are subsequently used to scale the MDCT spectrum in the corresponding 64 bands.

Figure 4 compares the spectral envelopes derived with the LPC-based approach and the new SNS approach for a voiced and an unvoiced frame from a german male speaker signal. As can be seen from the figure, the envelopes are approximately the same for both methods. Furthermore, it has been verified in additional listening tests that both methods deliver the same quality, despite the complexity of the new SNS approach being significantly lower.

3.5 Temporal Noise Shaping

Temporal Noise Shaping (TNS) effectively reduces the pre-echo artifact on signals containing sharp attacks such as castanets. It is also helpful for signals containing pseudo-stationary series of impulse-like signals such as speech. It is implemented similarly as in the EVS codec [1], with two modifications introduced for low bit rates. This was necessary as TNS was formerly applied at higher bit rates, and extending it to lower bit rates can sometimes lead to click- or noise-like artifacts.

The first modification concerns the position of TNS in the processing chain. For LC3 and LC3plus, it is applied on the flattened spectrum after SNS processing and not before, as in the EVS codec [1]. This reduces the risk of instabilities when TNS inverse filtering is performed on the sparse quantized spectrum. The second modification concerns the TNS filter itself, which is attenuated using LPC weighting when the prediction gain is low.

3.6 Attack Detector

Because of complexity reasons, the LC3 and LC3plus audio codecs do not incorporate switching of block lengths; pre-echoes have to be controlled by temporal noise shaping alone. To boost the performance of TNS, the attack detector controls the effectiveness of TNS by softening the SNS impact.

3.7 Spectral Quantization

The shaped MDCT spectrum after SNS and TNS is quantized using a uniform quantizer (with an additional dead-zone) whose step size is controlled by a unique global gain, denoted γ. Finding the optimal global gain is computationally the most expensive part of spectral quantization. The optimal value is defined as the smallest value for which the resulting quantized spectrum can still be encoded given the bit budget. In EVS-TCX [2], the global gain is calculated in an iterative way, also known as rate-loop. Starting from an initial value, the spectrum is quantized in every step of the iteration, and the number of required bits for encoding the quantized spectrum is estimated accurately. Based on a comparison to the given bit budget, the global gain may be adjusted in each step. This procedure may be carried out up to four times.

The accurate estimate of required number of bits for encoding a quantized spectrum is rather complex, so a less costly method is applied in LC3 and LC3plus. Instead of calculating the number of bits by trial encoding, it is estimated based on the power spectrum. To this end, the energy values $e(k) = \sum_{i=0}^{3} c(4k+i)^2$ are calculated for $k = 0, 1, ..., N/4-1$, where $c(i)$ denotes the i-th MDCT

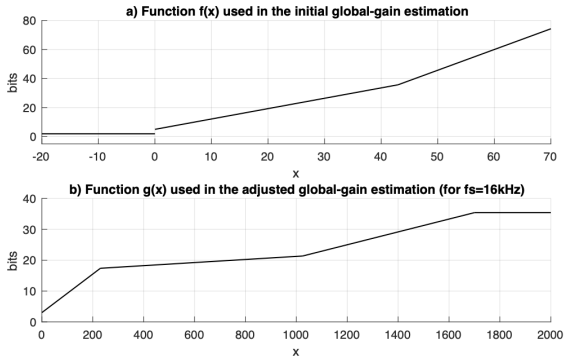


Fig. 5: Piecewise linear functions used in the global-gain estimation

coefficient and N denotes the frame length. The required bits for encoding the four coefficients collected in $e(k)$ is then estimated as $f(10\log_{10}(e(k)/\gamma^2)$, where f is a piece-wise linear function. Based on these estimates, the global gain is determined by a bisection algorithm. This procedure is very efficient since the values $10\log_{10}(e(k))$ can be re-used in every iteration. Once the global gain γ is determined, the spectrum is quantized, and the bit requirement is estimated accurately. If this estimate deviates too much from the targeted bit budget $nbits_{target}$, the global gain is adjusted once in a second stage. This is done based on a second piecewise linear function g. If the bit requirement $nbits$ is smaller than $nbits_{target} - g(nbits) + 2$, the global gain is decreased and if it is larger than $nbits_{target} + g(nbits)$, it is increased. The functions f and g are displayed in Figure 5. Tests showed that this low complexity scheme performs as well as the considerably more complex rate-loop in EVS-TCX.

3.8 Noise Filling

With decreasing bit rate, more and more spectral coefficients are quantized to zero. This introduces spectral holes at low bit rates, which are known to cause undesired artifacts, e.g., the well-known birdies artifacts. This can be mitigated by filling the spectral holes with pseudo-random noise. The energy of this noise is a parameter computed and transmitted by the encoder. The noise filling algorithm is very similar to the one used by MPEG-D USAC [11] and has lower complexity than the one used by 3GPP EVS.

3.9 Entropy Coding

Entropy coding is applied to both the quantized TNS coefficients and the quantized spectrum. The general scheme is similar to encoding in EVS-TCX, but a base-256 range coder replaces the arithmetic coder. This reduces the encoding complexity by a factor of three and the decoding complexity by a factor of two without significant loss in compression efficiency. As in EVS-TCX [2], the spectral coefficients are encoded in pairs using context-adaptive symbol frequencies. Furthermore, the context depends only on intra-frame data to keep the memory footprint small.

3.10 Residual Coding

Excess bits resulting from overestimating the bit budget for the range encoder are used to increase the precision of the non-zero quantized spectral coefficients. At most one bit per non-zero coefficient is spent, signaling whether the quantization error is positive or not. To take perceptual relevance into account, this is done from lowest to highest frequency until the excess bits are exhausted.

3.11 Long Term Post Filter

A long-term post filter (LTPF) is applied to reduce interharmonic coding noise in tonal frames at lower bit rates. This is a particularly valuable tool for enhancing speech at low bit rates. The LTPF is steered from the encoder, which provides a pitch lag estimate and an activation flag to the decoder. The LTPF is obtained at the decoder from the estimated pitch lag and a bit-rate dependent gain parameter and subsequently applied to the decoded signal as an enhancement tool.

4 LC3plus Extensions

4.1 Packet Loss Concealment

Packet loss is usually an unavoidable issue in wireless communication, from which both audio quality and speech intelligibility can suffer. To mitigate this problem, the LC3plus codec features a very efficient packet loss concealment (PLC) algorithm. It features three different methods for handling the wide range of signals targeted by LC3plus, such as speech and music.

4.1.1 PLC Methods

MDCT frame repetition with sign scrambling (MDCT-FRSS) is a method best suited for noise-like signals without any dominant harmonic structure. The missing MDCT data is replaced by the spectrum of the last decodable frame, and the sign of some MDCT coefficients is flipped in each lost frame based on a pseudo-random number generator. Furthermore, a decaying fade-out factor is applied to every substituted spectrum. The probability of the sign flips and the fade-out factor depends on the signal characteristic and the number of preceding lost frames.

Time Domain Concealment (TDC) is best suited for monophonic signals with a periodic structure, such as voiced speech. It uses a prediction filter based on the most recent decodable frame. The prediction error signal, i.e., the filtered signal, is extrapolated using a mixture of a periodic continuation, using the last decodable frame's pitch lag, and a generated pseudo-random noise signal. The substitute signal is then obtained by inverse filtering of the extrapolated prediction error signal. Furthermore, this extrapolated signal is faded out over a signal-dependent period of time.

The Phase Error Concealment Unit (PhaseECU) is an adapted and updated version of the sinusoidal phase evolution method [12] used for the HQ MDCT frame concealment in EVS and is best suited to conceal packet losses for complex signals exhibiting both harmonic structure and noise-like components. It operates in the STFT domain, where the dominant frequencies are estimated and extrapolated, maintaining the correct phase evolution. The modifications compared to EVS consist of using a shorter 16 ms prototype window in combination with an adapted version of complex domain interpolation [13] to maintain a high resolution for the sinusoid frequency estimation step. The PhaseECU time-domain windowing was updated with MDCT frame completion before the TDA/ITDA steps, using a copy and overlap-add method of the previously decoded or reconstructed signal and the current PhaseECU reconstructed frame. These combined steps improve tonal handling by avoiding time-domain discontinuities. Furthermore, the complexity at the first lost frame has been reduced by updating the sub-band transient detector spectral estimation method to operate in the MDCT domain.

4.1.2 PLC Method Selection

The selection of a concealment method is based on a set of signal features calculated on the most recent decodable frame(s) and remains unchanged independent of the number of lost packets. One feature is periodicity, which is indicated by the pitch value in the LTPF data. If not present, MDCT-FRSS is selected. Otherwise, either TDC or PhaseECU is selected based on the spectral centroid and the temporal predictability, which is the calculated auto-correlation coefficient at the pitch lag. TDC is preferred over PhaseECU when the spectral centroid is low and the temporal predictability is high.

4.2 Forward Error Correction

4.2.1 Channel Coder

The LC3plus codec features a channel coder designed for the DECT environment supporting gross bit rates from 40 to 400 bytes per frame, which corresponds to 32 to 320 kbps for a frame duration of 10ms. It offers four error protection modes (EP modes), where EP mode 1 provides robust error detection, and EP modes 2 to 4 provide an increasing amount of error correction capability at the expense of decreasing data rates.

In an earlier study [14], the optimal choice of BCH codes at 32 kbps has been studied extensively. However, to save complexity and provide good scalability, the LC3plus channel coder is based on Reed-Solomon codes (RS codes) over $\mathrm{GF}(16)$ with error correction capabilities of one, two, or three 4-bit symbols. The data is split and encoded in many truncated RS codes to encode different frame sizes, which are subsequently interleaved to improve the correctability of clustered transmission errors. This scheme provides an equivalent performance on the DECT error patterns studied in [14] with a complexity that is three to four times lower than for BCH codes.

EP modes can be selected frame-by-frame to adapt to varying channel conditions. This requires the transmission of the EP mode for every frame. An implicit signaling scheme is employed to avoid the inefficiency of transmitting and protecting the EP mode separately. It is based on affine-linear mode-dependent transformations of the RS codes at the encoder, aiming for maximal separation of the transformed codes and simultaneous trial decoding at the decoder, using the chosen RS codes' subcode structure.

4.2.2 Redundancy Frames

For channel conditions without bit errors but with frame losses, LC3plus also features a redundancy mode similar to the channel-aware mode of EVS. In this mode, each frame is encoded twice, once as a main frame and once as a helper frame, where the helper frame usually has a lower bit budget and a reduced bandwidth. The helper frame is transmitted timewise shifted compared to the main frame, meaning that the helper frame is available to steer partial concealment in case of packet loss.

4.3 High-Resolution Mode

The LC3plus codec offers a high-resolution audio mode dedicated to operate at high sampling frequencies of 48 and 96 kHz and high precision of 24 bits per sample. Higher accuracy is achieved by a different set of low-delay MDCT windows exhibiting a stronger stop-band attenuation towards the 24-bit noise level of −144 dB. This enables energy compaction to only a small number of non-zero spectral coefficients for audio components with 24-bit resolution. Furthermore, the quantizer's dynamic range is extended to full 24-bit resolution, and residual coding is reconditioned to allow spending more than one bit on each non-zero quantized coefficient. Results of AudioPrecision THD+N measurements are given in section 6.3.

5 Standardization

LC3 is specified as the mandatory codec of Bluetooth LE audio. LC3plus is standardized as ETSI TS 103 634 and has been adopted to the latest DECT standard.

6 Audio Quality Evaluation

In the following, an excerpt of the extensive evaluation in the standardization course is presented and discussed.

6.1 Voice Quality

One key use case for LC3 and LC3plus is the forwarding of phone calls to Bluetooth or DECT devices. For the evaluation of WB and SWB calls, two ITU-T P.800 ACR experiments have been conducted to compare LC3 and LC3plus to the legacy codecs (m)SBC and G.722, as well as to OPUS as an additional reference.

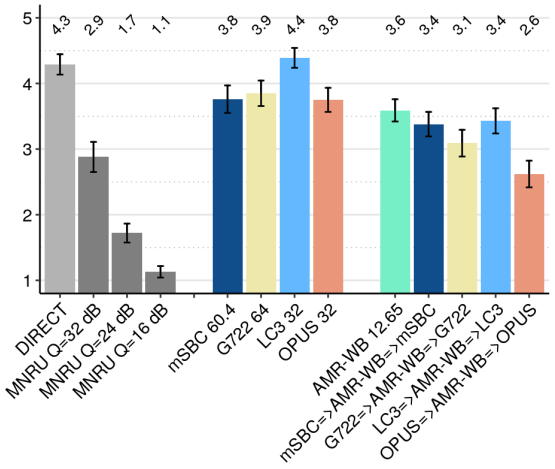


Fig. 6: P.800 ACR results for WB clean speech signals. Mean scores and 95 % confidence intervals of 25 subjects. Codec configurations identical in transcoding and direct case.

The exact codec configurations are listed in Table 2. Besides the intrinsic audio quality, the transcoding performance in conjunction with the most relevant mobile codecs, i.e., AMR-WB or EVS, has been investigated. Here, the worst-case scenario is considered meaning both participants use a wireless accessory leading to two transcoding steps. Figure 6 shows the WB results, while the results for SWB are displayed in Figure 7.

Codec	Bandwidth	Bit rate [kbps]	PLC
LC3 and LC3plus[1]	WB	32	std./adv.
G.722	WB	64	App. IV
mSBC	WB	60.4	BT HFP
OPUS[1 2]	WB	32	V1.3.0
LC3 and LC3plus	SWB	64	std./adv.
OPUS[1 2 3]	SWB	64	V1.3.0
SBC	SWB	128	-

[1] 10 ms frame duration
[2] restricted-lowdelay configuration
[3] resampled to 48 kHz sampling rate

Table 2: Configurations for WB and SWB voice

Regarding the intrinsic WB quality, LC3 and LC3plus outperform G.722 and mSBC, both operating at twice the bit rate of LC3 and LC3plus, and they outperform

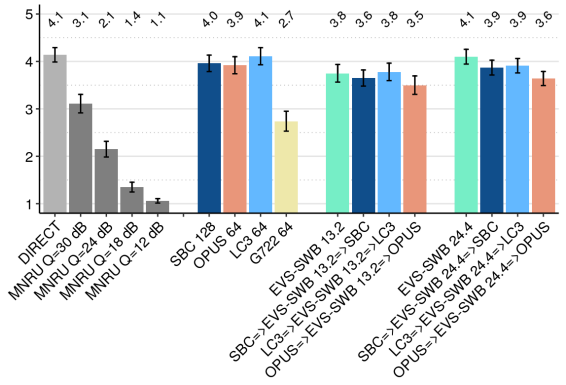


Fig. 7: P.800 ACR results for SWB clean speech signals. Mean scores and 95 % confidence intervals of 25 subjects. Codec configurations identical in transcoding and direct case.

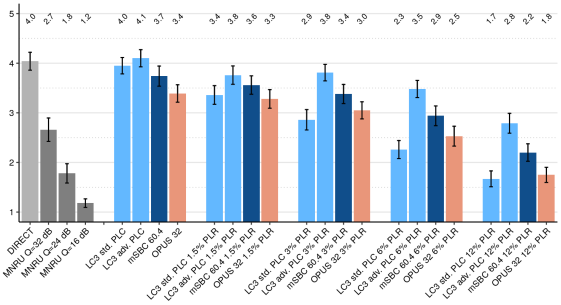


Fig. 8: P.800 ACR results for WB packet loss conditions. Mean scores and 95 % confidence intervals of 25 subjects. PLR: Packet Loss Rate.

OPUS operating at the same bit rate. For transcoding with AMR-WB at 12.65 kbps, LC3 and LC3plus are not worse than mSBC and significantly better than G.722 and OPUS. Regarding the SWB quality, LC3 and LC3plus are on par with SBC operating at twice the bit rate of LC3 and LC3plus, and OPUS operating at the same bit rate. For transcoding with EVS 13.2 kbps and 24.4 kbps, LC3 and LC3plus are not worse than any reference condition. These experiments confirm that LC3 and LC3plus are able to reduce the bit rate by roughly 50 % compared to the legacy codecs G.722 and (m)SBC.

Another essential factor for phone calls' quality is the robustness against bit errors and frame losses due to transmission errors. This has been evaluated in two additional ACR WB experiments that have been conducted to assess the quality over distorted Bluetooth (packet loss) and DECT (bit errors and packet loss) channels. In the Bluetooth context, LC3 utilizing Bluetooth's standard example concealment is compared to mSBC, OPUS, and LC3 using an advanced PLC algorithm based on MDCT-FRSS and TDC. As shown in Figure 8, the LC3 utilizing the advanced PLC performs significantly better than any reference solution.

For verifying the robustness over error-prone DECT channels, LC3plus was compared to the legacy codec G.722. To simulate typical channel conditions, the profiles DP 0 to DP 3 - corresponding to RSSI values of 80, 56, 48, and 40 dB as described in [14] - have been applied to the bitstreams. DP 0 contains no errors, while

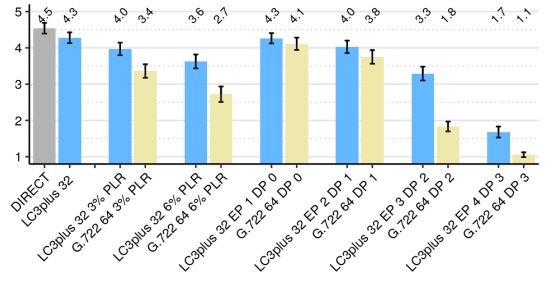


Fig. 9: P.800 ACR results for error-prone WB speech signals. Mean scores and 95 % confidence intervals of 26 subjects. PLR: Packet Loss Rate. EP: error protection mode. DP: DECT bit error pattern.

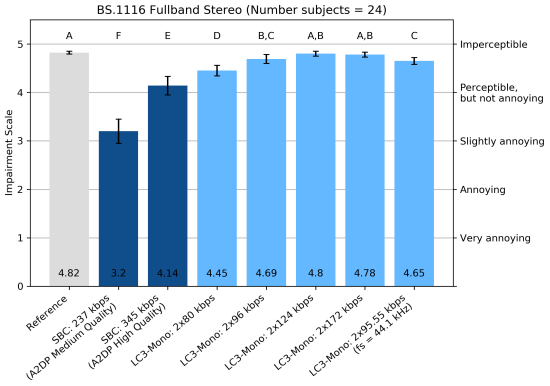


Fig. 10: Mean scores, 95 % confidence intervals and rank-grouping over all items in the BS.1116-3 stereo test. Rank-grouping as result of pairwise significance test per condition, based on a parametric multiple comparison two-sided t-test with non-pooled standard deviation estimates, and the "Holm" correction for multiple comparisons. Conditions with the same letter show no statistically significant difference. LC3 operates in dualmono with 2x the bit rates for each channel and 10 ms frame duration.

DP 1 to DP 3 exhibits an increasing number of bit errors and frame losses. Additionally, the codec's PLC is compared for 3 % and 6 % packet loss rates (PLR). According to the results in Figure 9, LC3plus outperforms G.722, especially for the error profiles with higher bit error rates (2 and 3). Furthermore, LC3plus performs significantly better than G.722 under packet loss conditions. This result shows that LC3plus is able to double the capacity for DECT calls while enabling more robust connections.

6.2 Music Streaming

For the Bluetooth LE Audio development, an independent ITU-R BS.1116-3 [15] experiment with 24 expert listeners was conducted by SenseLab, the test lab department of FORCE Technology, in order to verify the music streaming use case. The legacy Bluetooth codec SBC served as reference condition, operating at medium and high-quality stereo configuration according to the A2DP [16]. Six critical items have been selected from the EBU SQAM CD: castanets, bass clarinet, ABBA, Mozart, and triangle. Figure 10 presents the results in the form of z-normalized mean scores.

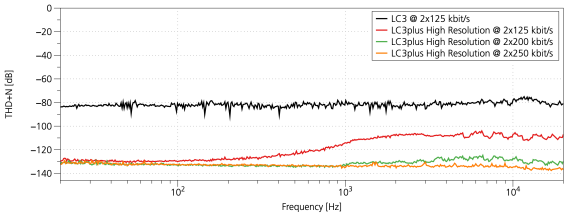


Fig. 11: THD+N measured with AudioPrecision APx500

LC3 at 2x80 kbps performs significantly better than SBC at 345 kbps. Starting from 2x124, LC3 shows no statistically significant difference to the uncoded original. Therefore, LC3 can provide a higher level of quality than SBC and enables a bit rate reduction compared to legacy codecs by roughly 50 %. Note that LC3 and LC3plus are identical for this scenario.

6.3 LC3plus High-Resolution

The LC3plus high-resolution algorithm's precision was verified with AudioPrecision APx500 over a full digital 24-bit codec chain using the floating-point implementation of LC3plus. To complete the picture, LC3, as standardized in Bluetooth, was taken as a reference. Figure 11 shows that LC3 saturates at a distortion level around -80 dB while LC3plus provides even lower distortions down to -130 dB over the complete spectrum.

7 Acknowledgement

The authors would like to thank FORCE Technology - SenseLab for administrating, analysing and reporting the BS-1116 test as well as the SIG Bluetooth for sharing the data.

8 Conclusion

This paper provides the technical background of LC3 and LC3plus, the new audio transmission standards for Bluetooth and DECT. The flexible design of these codecs serves all needs of current audio applications, such as call forwarding, music streaming, streaming of high-resolution audio content, wireless gaming headphones, wireless microphones, and many more, making it a powerful general-purpose tool for medium to high bit rate wireless audio transmissions.

References

[1] 3GPP, "Codec for Enhanced Voice Services (EVS); Detailed algorithmic description," Technical Specification (TS) 26.445, 3rd Generation Partnership Project (3GPP), 2014, version 16.0.0.

[2] Fuchs, G., Helmrich, C. R., Marković, G., Neusinger, M., Ravelli, E., and Moriya, T., "Low delay LPC and MDCT-based audio coding in the EVS codec," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5723–5727, 2015.

[3] Schnell, M., Geiger, R., Schmidt, M., Multrus, M., Mellar, M., Herre, J., and Schuller, G., "Low Delay Filterbanks for Enhanced Low Delay Audio Coding," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 235–238, 2007, ISSN 1947-1629, doi: 10.1109/ASPAA.2007.4392985.

[4] Valin, J., Terriberry, T. B., Montgomery, C., and Maxwell, G., "A High-Quality Speech and Audio Codec With Less Than 10-ms Delay," *IEEE Transactions on Audio, Speech, and Language Processing*, 18(1), pp. 58–67, 2010, ISSN 1558-7924, doi:10.1109/TASL.2009.2023186.

[5] Helmrich, C. R., Marković, G., and Edler, B., "Improved low-delay MDCT-based coding of both stationary and transient audio signals," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6954–6958, 2014, ISSN 2379-190X, doi: 10.1109/ICASSP.2014.6854948.

[6] Schuller, G. D. T. and Smith, M. J. T., "New framework for modulated perfect reconstruction filter banks," *IEEE Transactions on Signal Processing*, 44(8), pp. 1941–1954, 1996, ISSN 1941-0476, doi:10.1109/78.533715.

[7] 3GPP TS 26.171, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description," Technical Specification (TS), 3GPP, 2005, Version 5.0.0.

[8] Bessette, B., Salami, R., Lefebvre, R., Jelinek, M., Rotola-Pukkila, J., Vainio, J., Mikkola, H., and Jarvinen, K., "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, 10(8), pp. 620–636, 2002, doi:10.1109/TSA.2002.804299.

[9] Linde, Y., Buzo, A., and Gray, R., "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, 28(1), pp. 84–95, 1980.

[10] Svedberg, J., Grancharov, V., Sverrisson, S., Norvell, E., Toftgård, T., Pobloth, H., and Bruhn, S., "MDCT audio coding with pulse vector quantizers," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5937–5941, 2015.

[11] ISO/IEC 23003-3:2012, "Information technology — MPEG audio technologies — Part 3: Unified speech and audio coding," Standard, ISO/IEC, 2012.

[12] Bruhn, S., Norvell, E., Svedberg, J., and Sverrisson, S., "A novel sinusoidal approach to audio signal frame loss concealment and its application in the new evs codec standard," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5142–5146, 2015.

[13] Jacobsen, E. and Kootsooks, P., "Fast, Accurate Frequency Estimators," *IEEE Signal Processing Magazine*, 24(3), pp. 123–125, 2007.

[14] ETSI, "Digital Enhanced Cordless Telecommunications (DECT); Study of Super Wideband Codec in DECT for narrowband, wideband and super-wideband audio communication including options of low delay audio connections ($\leq$ 10 ms framing) ," Technical report (tr), ETSI, 2018.

[15] Recommendation ITU-R BS.1116-3, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, International Telecommunication Union (ITU), 2015.

[16] A2DP, "Advanced Audio Distribution," Standard, Bluetooth, 2019.