SYSTEM-COMPATIBLE ROBUSTNESS IMPROVEMENT FOR NEW GENERATION DECT DECODERS BY G.722 SOFT-DECISION DECODING

Domingo López-Oller†*, Sai Han*‡*, Angel M. Gomez*†*, José Luis Pérez-Córdoba*†*, and Tim Fingscheidt*‡

† Department of Signal Theory, Telematics and Communications, University of Granada, Spain
Email: {domingolopez,amgg,jlpc}@ugr.es
‡ Institute for Communications Technology, Technische Universität Braunschweig
Schleinitzstr. 22, 38106 Braunschweig, Germany
Email: {s.han,t.fingscheidt}@tu-bs.de

ABSTRACT

The ITU-T Recommendation G.722 about subband adaptive differential pulse code modulation (SB-ADPCM) is the mandatory wideband speech codec in the new generation digital enhanced cordless telephony (NG-DECT). Although in ADPCM the difference signal instead of the original signal is quantized and adaptive prediction is employed, redundancy is yet observed within the quantized samples. In this paper we apply a soft-decision speech decoding technique which exploits this redundancy in terms of a priori knowledge and the channel reliability information to NG-DECT. In that way, we propose a novel scheme in a standard-compliant fashion which improves the robustness of the decoder. The performance of our proposal is evaluated in terms of speech quality and a noticeable improvement over the standard codec and its own packet loss concealment algorithm is observed.

Index Terms— Error concealment, speech decoding, ADPCM, soft-decision decoding, NG-DECT

1. INTRODUCTION

In wireless communications, the digital enhanced cordless telephony (DECT) [1], and now the new generation of enhanced cordless telephony (NG-DECT or HD Sound) [2], have dominated the market because they allow for wireless access to a wide variety of services of any public or private network [3]. In traditional telephony, narrowband speech communications, with bandwidth being 300 to 3400 Hz, are widely used in DECT systems where the ITU-T Recommendation G.726 codec [4] is chosen as the mandatory narrowband codec. However, better speech quality can be achieved by extending the bandwidth to the range of 50-7000 Hz (wideband). As a consequence, the ITU-T G.722 wideband codec [5] is mandatory for NG-DECT because of its low delay, low complexity, and high quality speech at 64 kbps.

This work has been supported by an FPI grant from the Spanish Ministry of Education and by the MEC/FEDER Project TEC2013-46690-P.

Although the adaptive differential pulse code modulation (ADPCM) quantizes each sample in a scalar fashion, thereby being somewhat insensitive to channel errors, the wireless transmission errors can degrade the speech quality due to error propagation. This error propagation is caused by a corrupted quantized sample that modifies some parameters in the decoder and causes a desynchronization with the encoder. As a result, an error in one sample also affects the next correctly received samples. In the case of G.722, this error propagation can affect one or both subbands (SB-ADPCM) during the ADPCM decoding.

In traditional decoding or hard-decision decoding, the decoder operates on the received bits from the channel or with the application of the G.722 packet loss concealment (PLC) algorithm [6]. Instead of only receiving the bits, the soft-decision decoding scheme [7] can estimate the parameters in the decoder in order to minimize the error propagation from the channel reliability information [8, 9]. This technique has been successfully applied over different works such as G.726 ADPCM decoding [10], A-law PCM and GSM Full-rate speech coding [7, 11], high-quality PCM audio [12], and AMR-WB codec [13]. The main advantage of this proposal is that we apply the soft-decision decoding technique in a standard-compliant way, since in case of an error-free channel, this technique preserves the original scalar quantized sample.

The rest of paper is structured as follows: Section 2 describes the G.722 codec and the changes which are included by the soft-decision technique in the decoding stage. Section 3 presents our experimental framework and results. Finally, in Section 4 conclusions are drawn.

2. IMPROVED NG-DECT DECODER

2.1. G.722 Standard Encoding Operation

The block diagram of the G.722 SB-ADPCM encoder [5] is depicted in Fig. 1. As can be observed, the frame with 16 kHz

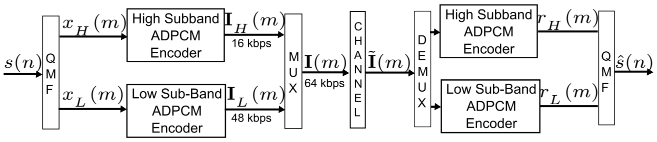

Fig. 1: High-level overview of the G.722 encoding and decoding operation.

sampling rate and composed by N samples ($s(n)$) is split into two components by a quadrature mirror filter (QMF). As a result, a lower subband vector (0-4000 Hz) $\mathbf{x}_L$ and a higher subband vector (4000-8000 Hz) $\mathbf{x}_H$, both with half the size of the frame length ($m \in \{0, 1, ..., , \frac{N}{2}-1\}$), are obtained. Then, an 8 bit combination $\mathbf{I}(m)$ is obtained for each index m from $x_L(m)$ and $x_H(m)$ and transmitted to the channel. The bit allocations are according to 64 kbps mode which is mandatory in NG-DECT. The ADPCM encoder process is described in [5] for both subbands.

2.2. G.722 Standard Decoding

Once the eight bits for each sample index m are received ($\tilde{\mathbf{I}}(m)$), the G.722 decoder [5] performs the reverse process of the encoder as can be observed in Fig. 2. Hence, the reconstructed speech samples $\hat{s}(n)$ and $\hat{s}(n+1)$ are obtained as:

$$\begin{aligned} \hat{s}(n) &= 2\sum_{i=0}^{11}(h_{(2i)} \cdot xd(i)) \\ \hat{s}(n+1) &= 2\sum_{i=0}^{11}(h_{(2i+1)} \cdot xs(i)), \end{aligned} \tag{1}$$

where $h_{(2i)}$ and $h_{(2i+1)}$ are a coefficient of the QMF filter, $xd(i) = r_L(m-i) - r_H(m-i)$, $xs(i) = r_L(m-i) + r_H(m-i)$, and $r_L(m)$ and $r_H(m)$ are the reconstructed signals for each subband.

Ideally, everything is updated in synchrony between encoder and decoder. However, in error-prone conditions, if the received bit combination $\tilde{\mathbf{I}}(m)$ is different to the original transmitted $\mathbf{I}(m)$, one or both reconstructed signals ($r_L(m)$ and $r_H(m)$) could be affected.

As a consequence of this change, we can see in Fig. 2 that the quantized differential signals ($d_{Lt}(m)$, $d_L(m)$ and $d_H(m)$) and the adaptive scale factors ($\Delta_L(m)$ and $\Delta_H(m)$) depend on the received bits $\tilde{\mathbf{I}}_H(m)$ and $\tilde{\mathbf{I}}_L(m)$ for each subband, obtained from $\tilde{\mathbf{I}}(m)$, so these parameters can be modified and this could generate an error propagation in one or both subbands. These parameters are defined in the G.722 standard [5] as:

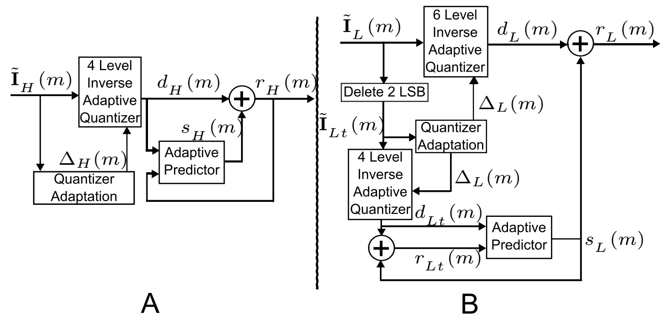

Fig. 2: ADPCM decoder scheme of both subbands in G.722 codec (A) high subband, (B) low subband

- Quantized differential signals:

$$\begin{aligned} d_L(m) &= Q6^{-1}[\tilde{\mathbf{I}}_L(m)] \cdot \Delta_L(m) \cdot \mathrm{sgn}(\tilde{\mathbf{I}}_L(m)), \\ d_{Lt}(m) &= Q4^{-1}[\tilde{\mathbf{I}}_{Lt}(m)] \cdot \Delta_L(m) \cdot \mathrm{sgn}(\tilde{\mathbf{I}}_{Lt}(m)), \\ d_H(m) &= Q2^{-1}[\tilde{\mathbf{I}}_H(m)] \cdot \Delta_H(m) \cdot \mathrm{sgn}(\tilde{\mathbf{I}}_H(m)). \end{aligned} \tag{2}$$

- Adaptive scale factors:

$$\begin{aligned} \Delta_L(m) &= 2^{[\triangledown_L(m)+2]} \cdot \Delta_{\min}, \\ \Delta_H(m) &= 2^{[\triangledown_H(m)]} \cdot \Delta_{\min}, \\ \text{with } \triangledown_R(m) &= \beta \triangledown_R(m-1) + W_R[\tilde{\mathbf{I}}_R(m-1)], \end{aligned} \tag{3}$$

where QM^{-1} is the M-level inverse adaptive quantizer, $\tilde{\mathbf{I}}_{Lt}(m)$ is $\tilde{\mathbf{I}}_L(m)$ truncated by two least significant bits (LSB), W_R is the corresponding logarithmic scaling factor multiplier defined in [5] for each subband ($R \in \{L, H\}$), $\mathrm{sgn}(\tilde{\mathbf{I}}_R(m))$ is the sign of the received bit combination $\tilde{\mathbf{I}}_R(m)$ and β and $\Delta_{\min}$ are constants.

Due to their dependency on the received bits $\tilde{\mathbf{I}}_L(m)$ and $\tilde{\mathbf{I}}_H(m)$, these parameters must be estimated in order to improve the robustness of the G.722 codec.

2.3. Novel G.722 Decoding by Soft-Decision Decoding

In order to apply the soft-decision decoding technique, the channel reliability information, which represents the mean bit error rate (BER) probability in a frame, is exploited in the decoding stage. As we can see in Fig. 3, the BER value and the received bits ($\tilde{\mathbf{I}}(m)$) are used to calculate the a posteriori probability (APP) $P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m))$ which represents the likelihood of a possible transmitted bit combination $\mathbf{I}^{(j)}$ given the received bits $\tilde{\mathbf{I}}(m)$, where j is the quantization index. This a posteriori probability will be used in the estimation of the quantized differential signals and the adaptive scale factors in both bands. The a posteriori probability is defined as [7]:

$$P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m)) = C \cdot P(\tilde{\mathbf{I}}(m)|\mathbf{I}^{(j)}) \cdot P(\mathbf{I}^{(j)}) \tag{4}$$

where C is always used to normalize the left hand side to 1, $P(\mathbf{I}^{(j)})$ is the a priori knowledge obtained from the training

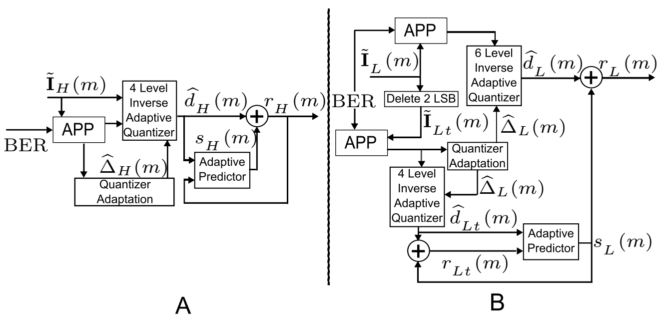

Fig. 3: Novel ADPCM scheme of both subbands in the G.722 codec with soft-decision decoding technique (A) high subband, (B) low subband.

database and $P(\tilde{\mathbf{I}}(m)|\mathbf{I}^{(j)})$ is the transition probability of any possible transmitted bit combination in $\mathbf{I}^{(j)}$ to the known received bit combination $\tilde{\mathbf{I}}(m)$ for each sample index m. In this work we use $P(\mathbf{I}^{(j)})$ as histogram knowledge because although this work can be extended to higher orders, some tests performed over G.726 in [7] did not show a big improvement compared to the histogram knowledge assumption, while the complexity increases for each order.

Assuming a memoryless channel, the transition probability $P(\tilde{\mathbf{I}}(m)|\mathbf{I}^{(j)})$ for each quantization index j can be defined as:

$$P(\tilde{\mathbf{I}}(m)|\mathbf{I}^{(j)}) = \prod_{b=0}^{7} P(\tilde{I}_{(b)}(m)|I_{(b)}^{(j)}) \quad (5)$$

where $\tilde{I}_{(b)}$ and $I_{(b)}^{(j)}$ are the corresponding bits in the bit combination of $\tilde{\mathbf{I}}(m)$ and $\mathbf{I}^{(j)}$, respectively, and the conditional bit probability $P(\tilde{I}_{(b)}(m)|I_{(b)}^{(j)})$ is defined as [7]:

$$P(\tilde{I}_{(b)}(m)|I_{(b)}^{(j)}) = \begin{cases} 1-\mathrm{BER}, & \text{if } \tilde{I}_{(b)}(m) = I_{(b)}^{(j)}, \\ \mathrm{BER}, & \text{otherwise,} \end{cases} \quad (6)$$

where the BER value is fixed for all the m quantized samples in a frame.

Once the a posteriori probability (5) is calculated, the scale factor must be estimated from (3) by minimum mean squared error (MMSE) as [7]:

$$\widehat{\Delta}_R(m) = \sum_{j=0}^{2^M-1} \left(\Delta_R(m)^{(j)} P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m-1)\right) = \\ 2^K \Delta_{\min} \sum_{j=0}^{2^M-1} \left(\left(2^{\beta \nabla_R(m-1) + W_R[\mathbf{I}^{(j)}]} \right) P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m-1)) \right), \quad (7)$$

where $R \in \{L, H\}$ represents each subband, K is the constant which appears added in $\Delta_L(m)$ in (3), $\mathbf{I}^{(j)}$ is the bit combination according to the index j, $W_R[\mathbf{I}^{(j)}]$ is the logarithmic scale factor and $M \in \{2, 4\}$ is the size of tables for high and low subband respectively [5].

Then, the quantized difference signal is also given by MMSE from (2) but taking into account the previous estimated scale factor as:

$$\widehat{d}_L(m) = \\ \left(\sum_{j=0}^{2^6-1} (Q6^{-1}[\mathbf{I}^{(j)}] \cdot \mathrm{sgn}(\mathbf{I}^{(j)}) \cdot P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m)) \right) \widehat{\Delta}_L(m); \\ \widehat{d}_{Lt}(m) = \\ \left(\sum_{j=0}^{2^4-1} (Q4^{-1}[\mathbf{I}^{(j)}] \cdot \mathrm{sgn}(\mathbf{I}^{(j)}) \cdot P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m)) \right) \widehat{\Delta}_L(m); \\ \widehat{d}_H(m) = \\ \left(\sum_{j=0}^{2^2-1} (Q2^{-1}[\mathbf{I}^{(j)}] \cdot \mathrm{sgn}(\mathbf{I}^{(j)}) \cdot P(\mathbf{I}^{(j)}|\tilde{\mathbf{I}}(m)) \right) \widehat{\Delta}_H(m). \quad (8)$$

As a result, these estimates are used to compute the novel reconstructed signals $r_L(m)$ and $r_H(m)$. While hard-decision decoding could lead to an error propagation due to the bit changes, soft-decision decoding provides a reconstructed signal which minimizes this effect because of the MMSE estimation.

3. FRAMEWORK AND RESULTS

In order to evaluate our proposals, we have considered the NG-DECT operating in the frequency band 1880-1980 MHz where the frame length is 10 ms and the G.722 codec is working on the 64 kbps mode. An objective test is performed by the ITU wideband extension perceptual evaluation of speech quality (WB PESQ) algorithm [14] over a testing set from the NTT database [15]. The NTT database covers 21 languages from all over the world, where 15 languages are used as training set, excluding American English, British English, Chinese, French, Spanish and German for testing, in order to obtain the a priori knowledge. The performance is tested over the training set where each language is composed by 96 sentences, including 4 male and 4 female speakers, with each sentence being 8 seconds long and sampled at 16 kHz.

For the NG-DECT channel model, in this paper a frequency-nonselective Rayleigh fading channel model with perfect frame synchronization and 2-path selection diversity is used [11]. For a given ratio of signal energy per bit to noise power spectral density (E_b/N_0) and an user speed, a fading factor α is obtained for each frame and is considered as a constant for all the samples in the frame. Thus, assuming a binary frequency shift keying (BFSK) modulation, the BER value for each frame is obtained as [16]:

$$\mathrm{BER} = \frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\alpha^2 \frac{E_b}{2N_0}} \right). \quad (9)$$

Table 1 and Fig. 4 show the WB-PESQ average results of different soft-decision decoding proposals in comparison

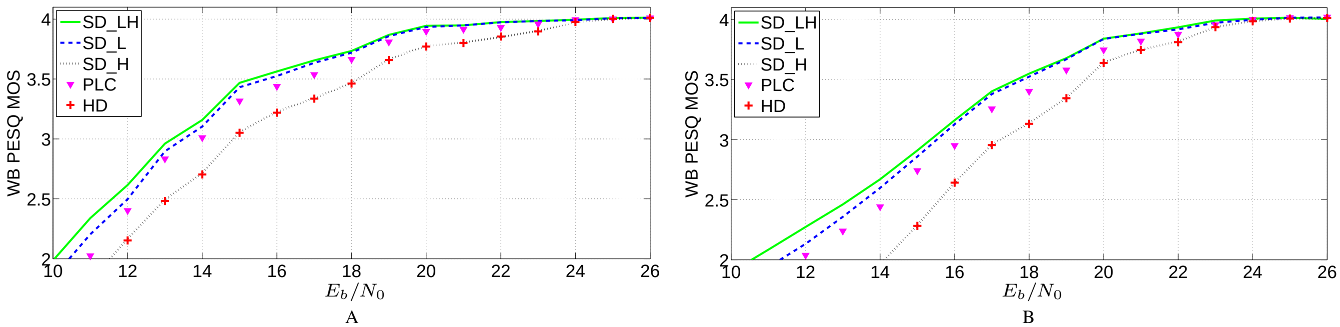

Fig. 4: WB-PESQ results for an NG-DECT speech transmission with user speeds (A) 0.3 m/s and (B) 3 m/s for our SD approaches, the G.722 PLC algorithm, and hard-decision (HD) decoding over different E_b/N_0 values.

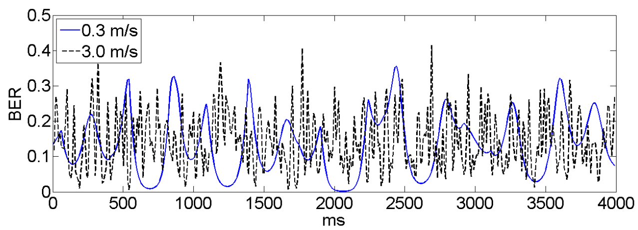

Fig. 5: BER output values in 4000 ms for both user speeds of 0.3 m/s and 3 m/s, respectively, when $E_b/N_0 = 0$ dB.

		E_b/N_0(dB)					
	speed	0	5	10	15	20	25
HD	0.3	1.06	1.12	1.55	3.05	3.77	4.01
	3	1.05	1.07	1.28	2.28	3.64	4.01
PLC	0.3	1.23	1.14	1.73	3.32	3.90	4.02
	3	1.10	1.18	1.68	2.74	3.75	4.02
SD_H	0.3	1.08	1.13	1.56	3.06	3.78	4.01
	3	1.05	1.08	1.28	2.29	3.65	4.01
SD_L	0.3	1.02	1.12	1.84	3.43	3.93	4.02
	3	1.02	1.09	1.74	2.86	3.84	4.02
SD_LH	0.3	1.05	1.23	1.72	3.47	3.94	4.02
	3	1.03	1.19	1.90	2.91	3.84	4.02

Table 1: WB-PESQ results of our soft-decision decoding techniques, G.722 PLC algorithm and hard-decision decoding under different E_b/N_0 values and user speeds.

with the G.722 PLC algorithm (PLC) and hard-decision decoding (HD) under several user speeds (0.3 m/s and 3 m/s) and E_b/N_0 values. In order to apply the G.722 PLC, we assume a BER value bigger than 10% as a reasonable point to consider the frame as lost [17]. The soft-decision decoding tests are the following: SD_LH represents the soft-decision decoding technique applied to both (high and low) subbands, SD_L represents the soft-decision decoding technique applied to the low subband and SD_H represents the soft-decision decoding technique applied to the high subband. It must be noted that as the user speed is higher, the fading effect is bigger and more consecutive frames are affected by a high BER value, thus, the predictor adaptation may not be able to recover properly and the perceptual quality decreases (see the HD results in Fig. 5).

As can be observed in Table 1 and Fig. 4, the PLC achieves a noticeable improvement over the hard-decision decoding while the proposal SD_H barely improves the scores of hard-decision decoding as only High component is estimated (only 2 bits of the 8 bits of $\mathbf{I}(m)$). However, when it is applied to estimate the lowband (SD_L) and both subbands (SD_LH), our proposal provides a noticeable improvement over the hard-decision decoding and its own PLC algorithm for both user speeds over most of the E_b/N_0 values as we estimate more bits of $\mathbf{I}(m)$.

Furthermore, we can see in Fig. 4 that the proposed SD_LH achieves better results than those of the G.722 PLC algorithm for E_b/N_0 above 10dB in the different user speeds. This is due to the fact that as the value of E_b/N_0 is higher, the BER values are lower and better estimates are obtained by the soft-decision technique. However, as can be seen in Table 1, below 5dB, the PLC algorithm achieves better results because we assume the same BER value for all the samples in the frame. Thus, if a bit combination is not modified but the BER value is high, the soft-decision estimation could be poor. This can be seen in Table 1 at E_b/N_0 of 0 dB, where the SD_H score is bigger than that of both SD_L and SD_LH.

4. CONCLUSIONS

In this paper, we have applied a soft-decision speech decoding technique to the G.722 codec under the NG-DECT specifications. Our proposal estimates the quantized differential signals and the adaptive scale factor for each subband ADPCM process as a way to minimize the error propagation caused by a modification in the quantized received sample.

The objective test has shown the suitability of our proposal in different E_b/N_0 conditions, where the soft-decoding technique applied to both bands achieves better scores than the G.722 PLC algorithm and the hard-decision decoding. Thus, our proposal shows a significant robustness improvement in a standard-compatible way.

5. REFERENCES

[1] "ETSI EN 300 175, Digital Enhanced Cordless Telecommunications Common Interface," ETSI, 1992.

[2] "ETSI TS 102 527, Digital Enhanced Cordless Telecommunications New Generation DECT," ETSI, 2007.

[3] "http://www.etsi.org/technologies-clusters/technologies/dect," ETSI, 2014.

[4] "ITU-T Recommendation G.726, 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)," ITU-T, Aug. 1990.

[5] "ITU-T Recommendation G.722, 7 kHz Audio-Coding Within 64 kbit/s," ITU-T, Nov. 1988.

[6] "ITU-T Recommendation G.722 Appendix III, A High-Quality Packet Loss Concealment Algorithm for G.722," ITU-T, Nov. 2006.

[7] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, pp. 240–251, Mar. 2001.

[8] J. Hagenauer and P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications," in *Proc. of GLOBECOM*, Dallas, TX, USA, Nov. 1989, pp. 1680–1686.

[9] J. Hagenauer, "Source-Controlled Channel Decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sep. 1995.

[10] T. Fingscheidt, "Graceful Degradation in ADPCM Speech Transmission," in *Proc. of DAGA*, Aachen, Germany, Mar. 2003, pp. 748–749.

[11] T. Fingscheidt and O. Scheufen, "Robust GSM Speech Decoding Using the Channel Decoder's Soft Output," in *Proc. of EUROSPEECH*, Rhodos, Greece, 1997, pp. 1315–1318.

[12] F. Pflug and T. Fingscheidt, "Delayless Soft-Decision Decoding of High-Quality Audio Transmitted Over AWGN Channels," in *Proc. of ICASSP 2011*, Prague, Czech Republic, May. 2011, pp. 489–492.

[13] S. Han, F. Pflug, and T. Fingscheidt, "Improved AMR Wideband Error Concealment for Mobile Communications," in *Proc. of EUSIPCO*, Marrakech, Morocco, Sep. 2013, pp. 1–5.

[14] "ITU-T Recommendation P.862.2, Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs," ITU-T, Nov. 2007.

[15] "Multi-Lingual Speech Database for Telephonometry," NTT-AT, 1994.

[16] J.G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1989.

[17] C. Lamblin and H. Taddei, "Terms of Reference (ToR) and Time Schedule for the G.722 Packet Loss Concealment (G.722 PLC) standardisation," *ITU-T WP3/16 Doc. AC-06-14*, Jun. 2006.