Hybrid source-channel coding with bandwidth expansion for speech data

Minh Quang Nguyen, Hang Nguyen, Eric Renault, Yusheng Ji

Hybrid Source-Channel Coding with Bandwidth Expansion for Speech Data

Minh-Quang Nguyen†, Hang Nguyen†, Éric Renault†, Yusheng Ji⋆
†SAMOVAR, Télécom SudParis, CNRS, Université Paris-Saclay, 9 rue Charles Fourier - 91011 Evry Cedex
⋆National Institute of Informatics, Tokyo, Japan
Email: {minh_quang.nguyen, hang.nguyen, eric.renault}@telecom-sudparis.eu, kei@nii.ac.jp

***Abstract*—Hybrid digital-analog (HDA) architectures have been widely developed for efficient digital transmission of analog speech, audio or video data. By considering the advantage of both digital and analog components, HDA systems gain better performances than purely analog and digital schemes in a wide range of channel conditions. However, HDA systems described in previous works are mostly designed for continuous-valued sources. In this paper, we address the problem of transmission of discrete sources over noisy channels. In particular, our work focuses on digital speech data in PCM format. We proposed two analog schemes, linear mapping, and non-linear mappings. The linear analog mapping employs an equal error protection scheme while the non-linear mapping takes into account the heterogeneous nature of error values to provide better protection to important values. The experiment results show that our HDA systems provide a better performance on a wide range of channel qualities in comparison with traditional purely digital systems.**

***Index Terms*—joint source channel coding, Shannon theory, information theory, speech, unequal error protection.**

I. Introduction

We consider the problem of a communication system for transmitting discrete sources over a discrete-time memoryless Gaussian channel. In particular, we focus on a system in a point-to-point setting that conveys the digital form of speech data. One of the most importance issues in such systems is to assure a high quality of speech while maintaining the bandwidth used.

Motivated by Shannon's theory of the joint source and channel coding, common digital communication systems usually employ a separate design of optimal source codes and channel codes. Moreover, in some essential applications, such as multimedia transmission, parts of the information are considered to have different sensitivities to errors and various contributions to the overall performance. By taking into account the heterogeneous nature of information, unequal error protection (UEP) technique was proposed to provide different coding redundancy to the information parts according to their importance as well as sensitivities to errors. In published literature, mainly two different strategies of UEP have been considered: Bit-wise UEP and message-wise UEP [1]. In bitwise strategy, information bits are partitioned into different subsets and provided different error protection levels. On the other hand, message-wise UEP provides better protection to several important subsets of messages.

However, a fundamental disadvantage associated with a digital system with source-channel code design is the "leveling-off" effect [2] in which the system performance remains constant even in the case of infinitive channel signal-to-noise ratio (CSNR). The hybrid digital-analog (HDA) systems, on the other hand, do not suffer from this effect thanks to the advantages of their analog components. In an HDA system, one can use a simple concatenation of source and channel code at the digital part to improve the transmission performance against strict channel conditions, while employing an analog mapping in the analog components to achieve a graceful improvement at the high range of CSNRs. Various approaches in HDA architectures have been presented in the literature. In the point-to-point setting, Mittal and Phamdo [3] proposed three HDA systems which provide better performances than the purely digital systems in terms of achievable distortion regions. On the basis of these schemes, Skoglund et al. [4] proposed linear and non-linear coding schemes for the analog components, which can accommodate all bandwidth ratios. For speech coding, works in [5], [6], [7] applied HDA transmission designs to gain a superior improvement for several ranges of CSNR. Matthias et al. [8], [9] designed different versions of HDA transmission systems with pulse-code modulation and Adaptive DPCM. However, their proposed linear mapping and non-linear mapping at the analog part, which uses an Archimedes spiral [10], only works with continuous-amplitude variables. A little attention has been paid to apply HDA scheme for the discrete sources.

In this paper, we target a point-to-point communication system that conveys digital speech data over Gaussian channel on the basis of HDA architecture. We observe that the gain obtained by exploiting the advantage of analog components on the high range of CSNR is appreciable. Particularly, by an appropriate analog coding, one can transfer the digital compressed data by the digital part while simultaneously transferring the "error values", which contain the gap between the original speech data and the outputs after performing lossy compressing/decompressing operations. Moreover, the error values with higher frequencies have higher importance level than others. Motivated by this observation, our idea is to transfer digital speech data by the HDA architecture with an appropriate design of the analog mapping. Our contributions in this paper can be summarized as follows:

- We apply HDA architecture for speech data transmission, in which the concrete source values are drawn from digital speech data.
- We employ a linear mapping scheme for discrete error

Minh-Quang Nguyen is also affiliated with Hanoi University of Science and Technology in a cooperated program.

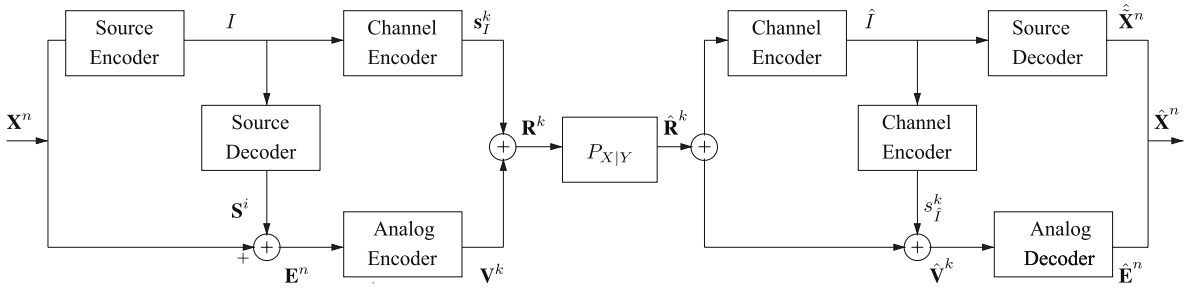


Figure 1: State-of-the-art HDA architecture.

values, which is a extension of the the analog mapping described in [4]. We also propose a non-linear mapping scheme which employs the message-wise UEP technique to improve the effect of analog components.

- We conduct computer simulations to demonstrate how well our system works.

The rest of the paper is organized as follows. After introducing a state-of-the-art HDA systems in Section II-A, we present our HDA model for digital speech data in Section II-B. In Section III we describe our proposed linear and non-linear mappings of the analog part and then subsequently show the experimental results in Section IV. Finally, we conclude this paper in Section V.

II. Hybrid digital-analog source and channel coding architecture

A. State-of-the-art HDA system

We consider a HDA communication system for analog-valued sources over discrete memoryless channels described in [4], [11]. Fig. 2 depicts a general version of HDA system, which contains a combination of a digital part and an analog part. We describe how the system works in the following.

1) Transmiter: At the transmitter, n samples $\mathbf{X}_i \in \mathbb{R}$ from a source are grouped into blocks $\mathbf{X}^n$. The source encoder, which contains a codebook including N vectors $\mathbf{S}_0, .., \mathbf{S}_N$, takes $\mathbf{X}^n$ and maps it to one of the codewords in its codebook. Each codeword $\mathbf{S}_i, 0 \leq i \leq N$, is a group of n real samples. Note that the source encoder and decoder agree upon the list of codewords in the codebook prior to the transmission. The output of the source encoder, the index I of a codeword, is then fed to the channel encoder which produces a channel symbols s_I^k , where $k > n$. The index I is also fed to a source decoder, which outputs the corresponding codeword S_I. At the analog part, an error vector E^n is calculated by subtracting S_I from $\mathbf{X}^n$. The vector $\mathbf{E}^n$ is further sent to a analog encoder which maps $\mathbf{E}^n$ to a k-dimensional vector $\mathbf{V}^k = (\mathbf{v}_1, .., \mathbf{v}_k)$, where $\mathbf{v}_i \in R, 1 \leq i \leq k$. The output s_I^k of the channel encoder and $\mathbf{V}^k$ are superposed and the result vector, $\mathbf{R}^k$, is sent to the receiver over a noisy channel.

Note that the channel symbol s_I^k must satisfy $\mathbb{E}\|\mathbf{s}_I^k\|^2 \leq k(1-t)P$, where P is the total input power per channel use constraint and $0 \leq t \leq 1$. Moreover, the output of linear encoder $\mathbf{V}^k$ must satisfy the power constraint $\mathbb{E}\|\mathbf{V}^k\|^2 \leq k \cdot t \cdot P$.

2) Receiver: At the receiver, upon receiving a channel output $\hat{\mathbf{R}}^k$, a channel decoder takes $\hat{\mathbf{R}}^k$ and outputs an estimate index $\hat{I}$ of the index I at the transmitter. The index $\hat{I}$ is then fed to a source decoder to produce a vector $\hat{\hat{\mathbf{X}}}^n = \hat{\mathbf{S}}_{\hat{I}}$, which is an estimate of $\mathbf{X}^n$ at the digital part. Simultaneously, the index $\hat{I}$ is also sent to a channel decoder which outputs a channel symbol $\mathbf{s}_{\hat{I}}^k$. Now $\mathbf{s}_{\hat{I}}^k$ is subtracted from $\hat{\mathbf{R}}^k$ to form a vector $\hat{\mathbf{V}}^k = (\hat{\mathbf{v}}_1, .., \hat{\mathbf{v}}_k)$, a estimation of the vector $\mathbf{V}^k$ at the transmitter. A analog decoder maps $\hat{\mathbf{V}}^k$ to an estimate $\hat{\mathbf{E}}^n$ of the error vector $\mathbf{E}^n$. Finally, the estimate of $\mathbf{X}^n$ by the HDA system is procedure by adding $\hat{\mathbf{E}}^n$ to $\hat{\hat{\mathbf{X}}}^n$.

B. HDA system for speech data

The main idea of our paper is to extend the state-of-the-art HDA system and apply to discrete-valued sources transmission. In our proposed system, which is depicted by Fig. 2, the digital segment works in the same way with the system shown in Fig. 1. In particular, we employ an Adaptive Multirate (AMR) compression, a turbo code and a binary phase-shift keying modulation (BPSK) module at the digital part. In the analog part, we investigate two types of mapping: linear mapping and non-linear mapping to produce the analog-valued error vectors. In the following, we describe how the proposed system works in detail.

1) Transmiter: In our system, the source is drawn from the raw digital speech signal in PCM format with the sample depth is 16 bits per sample. Unlike the continuous-amplitude source in state-of-the-art system, samples from the source are integer values $x, -2^{15} \leq x \leq 2^{15} - 1$. An array of Q consecutive values of x are grouped into a block $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_Q)$ and is fed to an AMR encoder. The output of the AMR encoder is n-bit compressed frames, where the frame size n depends on the codec modes used in the AMR encoder. The channel encoder is turbo encoder of rate r. We assume that the channel encoder is able to fix any error appeared in the digital part.

The output of turbo code (in M bit length) is then fed to a BPSK module, which outputs a symbol $\mathbf{s}_k \in \{1, -1\}^M$. The rate r is chosen to satisfy: $M = n/r \geq l \cdot P, l \in N, l > 1$.

At the analog part, an AMR decoder is used to decode the compressed frame and produce a vector $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, .., \bar{\mathbf{x}}_Q)$, an estimate of $\mathbf{X}$. An error vector $\mathbf{E}$, which contains Q integers, is calculated by subtracting the output vector $\bar{X}$ of the AMR decoder from $\mathbf{X}$. To form the analog representation of $\mathbf{E}$, the analog encoder maps each integer element $\mathbf{e}_i \in \mathbf{E}$ to a l-tuple $(\mathbf{v}_{i \cdot l}, .., \mathbf{v}_{i \cdot l + l - 1}), \mathbf{v}_i \in \mathbb{R}, i = 1, .., l$. We group Q l-tuples, corresponded to Q elements of $\mathbf{E}$, into an analog-

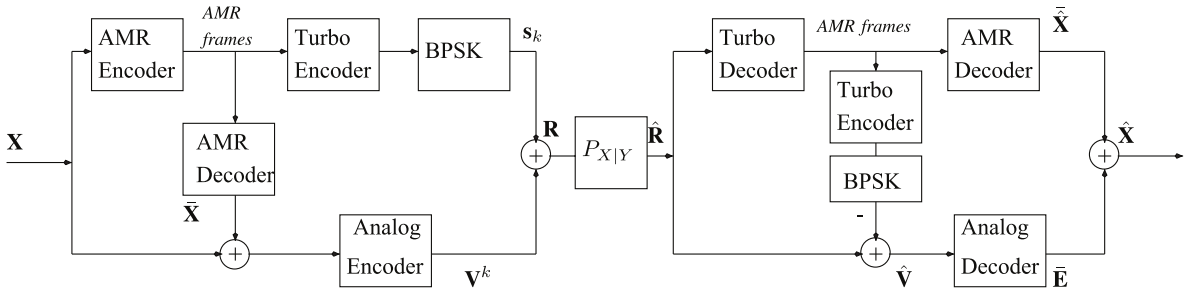


Figure 2: HDA scheme for speech data transmission.

valued representation $\mathbf{V} = (\mathbf{v}_1, .., \mathbf{v}_{Q \cdot l})$ of the error vector. Since the power in the digital segment is 1, then the vector V satisfies the power constraint:

$$\mathbb{E}\|\mathbf{V}\|^2 < k \cdot (1 - P).$$

The analog output V and digital output S are superimposed to produce the channel symbol $R = (r_1, .., r_M)$ before being transmitted over a discrete-time channel with zero mean additive white Gaussian noise W of variance σ_2 per component:

$$\mathbf{r}_j = \begin{cases} \mathbf{v}_j + \mathbf{s}_j & 0 \leq j \leq Q \cdot l, \\ 0 & \text{otherwise.} \end{cases}$$

2) Receiver: At the digital part, an AMR decoder and a turbo decoder are used to calculate a digital estimated version of $\mathbf{X}$. In particular, the received channel symbol $\hat{\mathbf{R}} = (\hat{\mathbf{r}}_1, .., \hat{\mathbf{r}}_M)$ is first fed to the turbo decoder. The output of the turbo decoder, an AMR frame in n-bit length, is sent to the AMR decoder, while simultaneously being encoded by a turbo encoder and mapped into BPSK symbols by the modulator. Outputs of the AMR decoder is an estimate version $\bar{\bar{\mathbf{X}}}$ of $\mathbf{X}$ at the digital part, where $\bar{\bar{\mathbf{X}}} = (\bar{\bar{\mathbf{x}}}_1, .., \bar{\bar{\mathbf{x}}}_Q)$.

At the analog part, an estimate of the error vector $\mathbf{V}$, $\hat{\mathbf{V}} = (\hat{\mathbf{v}}_1, .., \hat{\mathbf{v}}_M)$, is calculated by subtracting the output of the modulator from the received channel symbol $\hat{\mathbf{R}}$. The analog decoder maps each subset $(\hat{\mathbf{v}}_{i \cdot l}, .., \hat{\mathbf{v}}_{i \cdot l + l - 1}) \subset \hat{\mathbf{V}}$ into a discrete-valued error $\mathbf{e}_i$, which is the estimate of the element $\mathbf{e}_i$ of the error vector $\mathbf{E}$. The final estimation of $\mathbf{X}, \hat{\mathbf{X}} = (\hat{\mathbf{x}}_1, .., \hat{\mathbf{x}}_Q)$, is calculated by superimposing the vector $\hat{\mathbf{E}} = (\hat{\mathbf{e}}_1, .., \hat{\mathbf{e}}_Q)$ and the estimate version $\bar{\bar{\mathbf{X}}}$ which is produced by the AMR decoder:

$$\hat{x}_i = \begin{cases} \bar{\bar{\mathbf{x}}}_i + \hat{\mathbf{e}}_i & 0 \leq i \leq Q \cdot k, \\ \bar{\bar{\mathbf{x}}}_i & \text{otherwise.} \end{cases}$$

What remains to be demystified is the analog encoder and decoder at the analog part. We investigated two systems, which employ linear and non-linear analog mapping. The two systems are described in the following section.

III. The analog encoder/decoder

A. The linear mapping

In the first system, we employ a linear mapping at the analog part. Roughly speaking, each element of $\mathbf{E}$ is sent l times in the output vector of the linear analog encoder. In other words, an element $\mathbf{e}_i \in \mathbf{E}$ is sent by l equal elements of $\mathbf{V}$:

$$\mathbf{v}_{i \cdot l} = \mathbf{v}_{i \cdot l + 1} = .. = \mathbf{v}_{i \cdot l + l - 1} = a \cdot \mathbf{e}_i, \tag{1}$$

where a is a scaling constant that regulates the magnitude of the analog part. To calculate the value a, we use the method presented in [4]. In particular, we assign a power allocation coefficient $0 < \Delta < 1$ to the analog part. The scaling constant a is chosen to satisfy the following equation:

$$\Delta = \frac{Qa^2 E\|\mathbf{E}\|}{M + Qa^2 E\|\mathbf{E}\|} \tag{2}$$

At the receiver, upon receiving $(\hat{\mathbf{v}}_{i \cdot l}, .., \hat{\mathbf{v}}_{i \cdot l + l - 1})$, the linear decoder produces an estimation of each element e_i of the error vector. Assumming that σ is known at the receiver, we extend the method previously described by Skoglund et al. [4] for the case of l-tuple as follows:

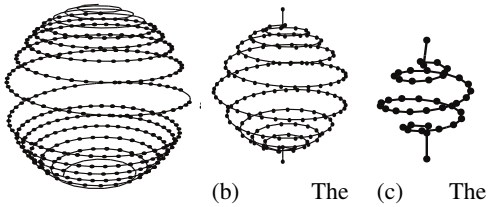
$$\hat{\mathbf{e}}_i = \frac{1}{l \cdot q} \cdot \frac{\frac{\Delta}{1-\Delta}}{\sigma^2 + \frac{\Delta}{1-\Delta}} \sum_{j=0}^{l} \hat{\mathbf{v}}_j. \tag{3}$$

B. The non-linear mapping

We propose a non-linear mapping for the analog part in the case of discrete-valued error vector and bandwidth expansion. For the proposed system where error vectors include discrete values, we observe that some values have higher frequencies than others. Therefore, the distortion of these values causes greater overall distortion. The observation gives us an idea to design a UEP scheme in the analog segment which protects "important" values in error vectors more efficient than the rest.

First, we choose K integers $\mathbf{e}_i, 1 < i < K$ which have highest frequencies from error vectors. Each value $\mathbf{e}_i$ is mapped to a point in l-dimensional space, corresponds to a l-tuple. We note that for amplitude continuous sources, Pal et al. [10] proved that one can decrease distortion by stretching the signal curve (i.e. the representation of signal in l-dimensional channel space). Considering a power-constrained channel and 2-dimensional space, they suggest using an Archimedes spiral to present signal values in 2-dimensional space. Motivated by this result, for l-dimensional space, we propose a signal curve which is the projections of an Archimedes spiral onto different hyper-spheres. The projections are connected at their poles. At the signal curve, we choose K points which divide the curve into $K + 1$ equal segments, each point corresponds to one in K error values $\mathbf{e}_i$.

For example, in 3-dimensional space, the curve is constructed by connecting projections of an Archimedes spiral onto different spheres. Fig. 3 illustrates a signal curve which locates on three different spheres whereby the connection between projections is at their poles.



(a) The outermost sphere. (b) The intermediate sphere. (c) The innermost sphere.

Figure 3: Illustration of the signal curve located on different spheres.

IV. Simulation and Results

We have conducted some experiments to demonstrate how well our proposed models can transfer digital speech data over discrete channels, under two scenarios: linear mapping and non-linear mapping. In the following, we first describe our simulation in section V-A and then use this simulation to compare our model with the existing scheme in section V-B.

A. Simulation description

Our simulation is based on speech data with the Adaptive Multirate (AMR) audio compression as the source code and the turbo code as the channel code. The AMR encoder compresses $P = 380$ integer values from digital raw speech data to obtain an AMR frame. Here we use the codec mode 6.7 kbits/s, i.e., the AMR frames are in 144-bit form ($L = 144$). The turbo code with rate $r = 1/3$ is used, with generator $(37, 21)$ and with a random interleaver. Consequently, outputs of the digital part are 432-dimensional vectors.

In the linear analog segment, each error vector element is sent $k = 3$ times by V. Precisely, we use a 3-tuple of three identical analog values to present an error value which is calculated by the formula (1) and (2). The estimate values of error is calculated by the formula (3), again, under the assumption that σ_2 is known at the receiver.

In non-linear analog mapping, we connect the projections of Archimedes spirals on four spheres to produce the signal curve. Then we present highest frequent values of error vectors by the coordinates of points located at the curve. In particular, we use four signal curves, which are the projections of an Archimedes spiral onto eight hemispheres (corresponds four spheres) as in the following (the variables are x, y and z):

$$x = c \cdot \phi \cdot cos\phi,$$
$$y = c \cdot \phi \cdot sin\phi,$$
$$z = \pm\sqrt{r_s - c^2 \cdot \phi^2},$$

where c is a constant, $r_s, s = 1, 2, 3$ and 4 are the radius of spheres. We choose $c = 0.065$ and assign a constant Δ to the analog components, r_s are trained to satisfy:

$$\Delta = \frac{QE\|\mathbf{E}\|}{M + QE\|\mathbf{E}\|}.$$

B. Simulation results

In Figs. 4 and 5, we show performance results in terms of the Perceptual Evaluation of Speech Quality Mean Opinion Score (MOS) [12] and the mean square error

$$MSE = 10 \cdot \log_{10} E\|\hat{\mathbf{X}} - \mathbf{X}\|^2$$

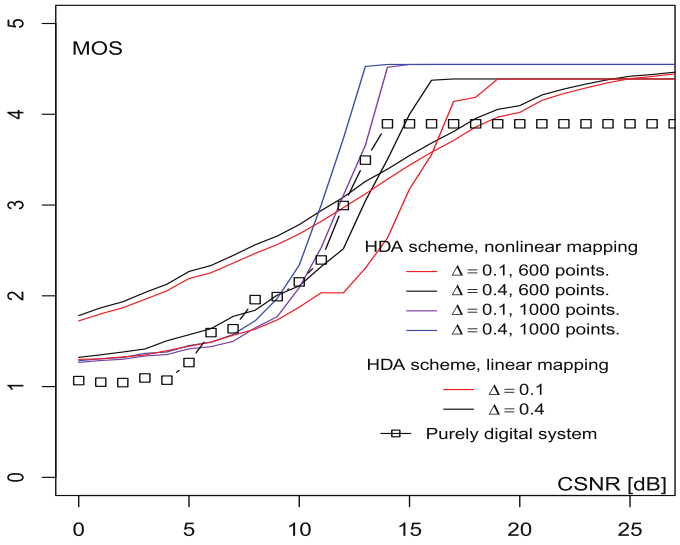


Figure 4: Performance in terms of MOS.

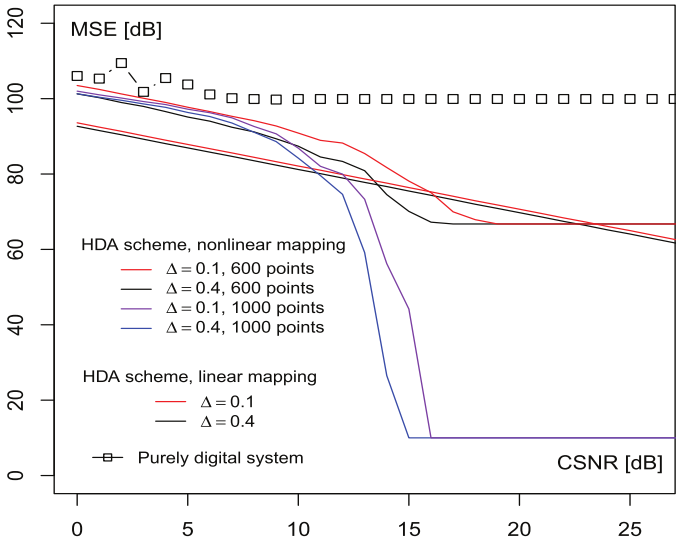


Figure 5: Performance in terms of MSE.

for the following systems:

- Two HDA systems with linear analog mapping. They are evaluated with the power level fractions Δ = 0.1 and 0.4.
- Two HDA systems with non-linear analog mapping which presents 600 error values by 600 points on the signal curve. The power level fractions are $\Delta = 0.1$ and 0.4,
- Two other HDA systems with non-linear analog mapping. They use 1000 points located on the signal curve to present 1000 error values and the power level fractions are also $\Delta = 0.1$ and 0.4,
- A UEP system, which employs source code (AMR compression) at the codec mode 6.7 kbits/s, a turbo code of rate $r_h = 1/5$ for the AMR frame header and a turbo code of rate $r_c = 8/23$ for classes A and B bits.

As can be seen in Fig. 4, HDA systems outperform the UEP purely digital system for the wide range of CSNRs in terms of MOS. In particular, HDA systems with linear analog mapping perform better than the digital system in the ranges of CSNRs from 0 dB to 12 dB and 17 dB to 25 dB. Moreover, systems

equipped with a non-linear mapping and 1000 points on the signal curve also achieve better MOS than the digital system in the low range (from 0 dB to 5 dB) and high range (from 10 dB to 25 dB) of CSNRs. Similarly, in Fig. 5, we can see a significant improvement of HDA systems over the UEP system in terms of MSE. For example, in the case of linear analog mapping and $\Delta = 0.4$, the HDA system outperforms the UEP system by 6 dB and 12 dB in terms of MSE, at CSNR = 0 dB and 3 dB, respectively. The improvement is more significant on the high range of CSNRs, with the minimum gap is 22.5 dB for the range of CSNRs from 15 dB to 25 dB. This demonstrates the effect of analog parts on speech transmission.

We remark that the systems with non-linear mapping suffer from the "leveling-off effect", i.e., the performance remains constant even when CSNR approaches infinity. For example, the performance of the systems with non-linear mapping, 600 points on the signal curve and Δ = 0.1 and 0.4, remain at MOS = 4.389 for all CNRSs greater than 14 dB and 17 dB, respectively. Similarly, in terms of MSE, they remain at MSE = 65.8 dB for all CSNRs greater than 15 dB and 18 dB. The reason is that one only can transfer the approximated values of error vectors at the transmitter. Hence, there is no reason to expect that the original error values can be recovered at the analog part even in the case of noiseless channels. In contrast, the linear mapping systems does not experience the leveling-off effect. In another words, MOS in the system with linear mapping increases alongside the whole range of CSNRs. This is due to the nature of linear analog mappings, in which we can recover every error values at good channel conditions.

We also remark that in non-linear analog mapping, systems with 1000 points on the signal curve performs better than systems with 600-point signal curve over the high range of CSNRs. The 1000-point signal curves also achieve greater MOS thresholds than that of 600-point signal curves. The reason for this better performance is that systems equipped with 1000-point representation can transfer error values more accurately than the 600-point signal curve. On the high range of CSNRs where the noise is weak, the receivers of these systems can recover the original value with small distortion. Hence, one can achieves a better reproduction fidelity with 1000-point-curve systems than systems implemented with 600 points on their signal curve.

It can be observed from Figs. 4 and 5 that the performances of non-linear analog mapping are better than that of linear analog mapping over the high range of CSNRs. For example, in terms of MOS, the non-linear 600-point analog mapping provides a greater MOS than the linear analog mapping for the range of CSNRs from 14 dB to 24 dB in the system with $\Delta = 0.4$. Moreover, the improvement is more significant in terms of MSE. At the range of CSNRs from 17 dB to 25 dB, the non-linear 1000-point analog mapping outperforms the linear mapping by at least 55.1 dB and 55.3 dB for $\Delta = 0.4$ and 0.1, respectively. This demonstrates the advantage of our UEP scheme implemented at the analog segment. We also observe that the linear mappings perform better than the nonlinear mappings over the low range of CSNRs. Namely, the system equipped with linear mapping and $\Delta = 0.4$ achieves better results than systems with non-linear mapping for CSNRs from 0 dB to 10 dB and the minimum gap in terms of MOS and MSE is 0.45 and 8.45 dB, respectively. The reason is that the strong noise in bad channel conditions may takes a point at a particular part to another part of the curve which is far from the original point at the signal curve. This leads to worse performances of the non-linear mappings in comparison with linear mappings.

Finally, it is worth noting that the systems with greater values of Δ provide better performance than the systems with low values of Δ. This is because that the high values of Δ help the transmitter to distinguish original values of error vectors better than low Δs, while suffering less distortion by the noisy channel.

V. Conclusion

In this paper, we have proposed a HDA joint source channel coding for the reliable communication and reproduction of digital raw speech data over AWGN channel. The analog segments employs the unequal error protection by taking into account the unequal distribution of error vectors to provide a better protection to important values. Numerical results verify the better performance of the proposed scheme compared to traditional schemes.

References

[1] S. Borade, B. Nakiboglu, and L. Zheng, "Unequal error protection: some fundamental limits," in *IN PROC. OF THE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY*. Citeseer, 2008.

[2] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[3] U. Mittal and N. Phamdo, "Hybrid digital-analog (hda) joint source-channel codes for broadcasting and robust communications," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1082–1102, 2002.

[4] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital–analog source–channel coding for bandwidth compression/expansion," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3757–3763, 2006.

[5] T. Miki, C.-E. Sundberg, and N. Seshadri, "Pseudo-analog speech transmission in mobile radio communication systems," *IEEE transactions on vehicular technology*, vol. 42, no. 1, pp. 69–77, 1993.

[6] N. Phamdo and U. Mittal, "A joint source-channel speech coder using hybrid digital-analog (hda) modulation," *IEEE transactions on speech and audio processing*, vol. 10, no. 4, pp. 222–231, 2002.

[7] C. Hoelper and P. Vary, "Bandwidth-efficient mixed pseudo analogue-digital speech and audio transmission," in *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*. IEEE, 2006, pp. 141–145.

[8] M. Rüngeler, J. Bunte, and P. Vary, "Design and evaluation of hybrid digital-analog transmission outperforming purely digital concepts," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3983–3996, 2014.

[9] M. Rüngeler, F. Kleifgen, and P. Vary, "Wideband speech coding with hybrid digital-analog transmission," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 784–788.

[10] P. A. Floor and T. A. Ramstad, "Optimality of dimension expanding shannon-kotel'nikov mappings," in *Information Theory Workshop, 2007. ITW'07. IEEE*. IEEE, 2007, pp. 289–294.

[11] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1509–1523, 2015.

[12] "P.862: Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codes. [online]. available: http://www.itu.int/rec/t-rec-p.862/en."