Zero Perception Error Rate-Distortion for Semantic Communication

Jinho Choi*, Hyelin Nam†, Jihong Park‡, Seung-Woo Ko§, and Seong-Lyun Kim†
*School of Electrical and Mechanical Engineering, The University of Adelaide, SA 5005, Australia
Email: jinho.choi@adelaide.edu.au
†School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Korea
Emails: hylennam@yonsei.ac.kr, slkim@yonsei.ac.kr
‡ISTD Pillar, Singapore University of Technology and Design, Singapore 487372, Singapore
Email: jihong_park@sutd.edu.sg
§Department of Smart Mobility Engineering, Inha University, Incheon 21999, Korea
Email: swko@inha.ac.kr

***Abstract*—In this paper, we explore the rate-distortion theory for semantic communication with zero perception error (ZEPE). By incorporating a ZEPE constraint, the decoder can generate a reconstructed signal that maintains ZEPE while satisfying a given distortion constraint. This approach can be viewed as a special case of the conventional rate-distortion framework. We derive the optimal transition probability and extend the ZEPE rate-distortion framework to include inferred latent variables, which can capture the semantic meaning of the observed signal. For practical implementations, we discuss the use of generative models such as variational autoencoders (VAEs).**

***Index Terms*—Rate-Distortion Theory; Semantic Communication; Zero Perception Error**

I. Introduction

Recent advancements in deep learning have enabled semantic communication [1], [2], which focuses on transmitting the meaning or semantics extracted from source information to maximize effectiveness in a given task [3], [4]. In this context, deep joint source-channel coding has also been considered to enable end-to-end optimization of the encoding and decoding processes, aiming to preserve task-relevant semantics rather than achieving precise reconstruction [5]. While various measures can be used to capture the effectiveness of conveying the semantics of a given source, it is essential to minimize the perception error in the received information. To better understand the theoretical foundations of semantic communication and its connection to classical communication, it may be necessary to investigate how semantics and perception influence the tradeoff between rate and distortion. This can be done by extending classical rate-distortion (R-D) theory through the integration of the rate-distortion-perception (R-D-P) framework [6] and the indirect rate-distortion (I-R-D) framework [7].

In the R-D theory, the minimum required rate is derived to achieve a given distortion level [8], while the R-D-P tradeoff introduces an additional constraint to guarantee a target perception error [6], which increases the required rate. Semantic communication's aim of minimizing perception error is captured by the special case of zero perception error (ZEPE), also known as perfect perception [9]. The prior work in [6] has shown that, when distortion is measured using the mean squared error (MSE), achieving ZEPE incurs a cost equivalent to doubling the minimum achievable MSE distortion. However, since distortion is measured in the sample domain while perception is often defined over the distribution domain, generalizing this result and deriving an optimal solution for ZEPE R-D-P remains a challenging problem.

On the other hand, I-R-D extends R-D by introducing an unobserved state of the source [7]. This unobserved latent variable, which can be extracted using neural networks such as autoencoders [10], corresponds to the semantics in semantic communication. The prior work in [7] derives a closed-form solution only under the assumption of a linear relationship between the source and the latent variable. Generalizing this result and incorporating the missing ZEPE constraint into I-R-D is a non-trivial extension requiring further investigation.

In this work, we generalize R-D-P by deriving a closed-form relationship between the transmitting signal distribution P and the reconstructed signal distribution Q that satisfies the optimality of ZEPE R-D-P under a generic distortion measure. We further integrate this result with I-R-D and derive a closed-form solution for Q under the ZEPE constraint, given a generic latent variable. Leveraging this result with generative models [10] [11], we can provide practical insights into the design of generative semantic communication systems and hybrid semantic-classical communication architectures.

Notation: Let $\mathsf{D}(P;Q)$ be the Kullback–Leibler (KL) divergence between two distributions, P and Q, which is given by $\mathsf{D}(P;Q) = \sum_k p_k \log \frac{p_k}{q_k}$. The mutual information between $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathsf{I}(\mathbf{x};\mathbf{y})$, which is given by $\mathsf{I}(\mathbf{x};\mathbf{y}) = \sum_{\mathbf{x}} \sum_{\mathbf{y}} \Pr(\mathbf{x},\mathbf{y}) \log \frac{\Pr(\mathbf{x},\mathbf{y})}{\Pr(\mathbf{x})\Pr(\mathbf{y})} = \mathsf{D}(\Pr(\mathbf{x},\mathbf{y}); \Pr(\mathbf{x})\Pr(\mathbf{y}))$.

II. Background

A. Rate-Distortion Theory

Suppose that $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ represents a signal from a source, sampled from distribution $P(\mathbf{x})$, where $\mathcal{X}$ denotes the set of all signal vectors $\mathbf{x}$. Its reconstructed signal is represented by $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^n$, which is assumed to a sample from distribution

$Q(\mathbf{y})$. Here, $\mathcal{Y}$ denotes the set of all possible reconstructed signal vectors $\mathbf{y}$, and $Q(\mathbf{y})$ is the following marginal distribution: $Q(\mathbf{y}) = \sum_{\mathbf{x}} Q(\mathbf{y} \,|\, \mathbf{x}) P(\mathbf{x})$, where $Q(\mathbf{y} \,|\, \mathbf{x})$ is the conditional distribution. Denote by $d(\mathbf{x}, \mathbf{y})$ the distortion between $\mathbf{x}$ and $\mathbf{y}$, which is a function of $\mathbf{x}$ and $\mathbf{y}$, i.e., $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$, where $\mathbb{R}^+ = \{x : \; x \geq 0\}$. Then, the R-D relation is characterized by the following optimization problem:

$$\begin{aligned} R = \min\nolimits_{Q(\mathbf{y} \,|\, \mathbf{x})} \mathsf{I}(\mathbf{x}; \mathbf{y}) \\ \text{subject to (s.t.) } \mathbb{E}[d(\mathbf{x}, \mathbf{y})] \leq D, \end{aligned} \tag{1}$$

where $D \geq 0$ is a constant. The resulting pair (R, D) forms a R-D curve, i.e., the R-D function, denoted by $R(D)$, which is a convex function of D [8].

B. Rate-Distortion-Perception Tradeoff

In [6], the concept of the R-D tradeoff is expanded to incorporate considerations of perceptual quality in signal reconstruction. This extension addresses a fundamental limitation in conventional distortion measures, which typically quantify the difference between the original signal and its reconstruction using metrics such as $||\mathbf{x} - \mathbf{y}||^2$, which leads to the conventional MSE measure for the distortion. However, such measures often fail to accurately capture human perception of quality, as they may not align well with the characteristics of human sensory systems. Consequently, it is desirable to explore alternative measures that better capture perceptual aspects [12].

When the perceptual quality can be measured by distributions rather than samples themselves, the KL divergence is commonly used [9]. In this case, with both the distortion and perception constraints, the following optimization problem can be formulated [6]:

$$\begin{aligned} R = \min\nolimits_{Q(\mathbf{y} \,|\, \mathbf{x})} \mathsf{I}(\mathbf{x}; \mathbf{y}) \\ \text{s.t. } \mathsf{D}(P; Q) \leq \delta \text{ and } \mathbb{E}[d(\mathbf{x}, \mathbf{y})] \leq D, \end{aligned} \tag{2}$$

where $\delta \geq 0$. Clearly, $R \geq R(D)$ due to the additional perception constraint, $\mathsf{D}(P; Q) \leq \delta$. The resulting R-D-P function, $R(D, \delta)$, is convex in D and δ [6].

From (2), we can also consider an extreme case as follows:

$$\begin{aligned} R = \min\nolimits_{Q(\mathbf{y} \,|\, \mathbf{x})} \mathsf{I}(\mathbf{x}; \mathbf{y}) \\ \text{s.t. } \mathsf{D}(P; Q) \leq \delta, \end{aligned} \tag{3}$$

where the distortion constraint, $\mathbb{E}[d(\mathbf{x}, \mathbf{y})] \leq D$, is fully relaxed. In this case, it is possible to achieve $(R, \delta) = (\mathsf{I}(\mathbf{x}; \mathbf{y}), \mathsf{D}(P; Q)) = (0, 0)$ by setting $Q(\mathbf{y} \,|\, \mathbf{x}) = Q(\mathbf{y}) = P(\mathbf{y})$. That is, the decoder generates the output independently of the input, $\mathbf{x}$, while the output being a sample from distribution $P(\mathbf{y})$. As a result, we have $\mathsf{I}(\mathbf{x}; \mathbf{y}) = 0$.

For example, let $\mathbf{x}$ be an image of handwritten digits. If the perceptual constraint in (3) allows distinguishing digits, then any image of the same digit could suffice for recognition, even with $R = 0$. This illustrates that perfect perceptual quality can occur without transmitting information.

III. Zero Perception Error Rate-Distortion Function

In this section, we consider a modified R-D function with a constraint of $\mathsf{D}(P; Q) = 0$, which can provide ZEPE reconstructions by ensuring that the reconstructed signal $\mathbf{y} \sim Q$ perfectly preserves the perceptual characteristics of the original signal $\mathbf{x} \sim P$. Throughout the paper, we assume that $\mathcal{Y} = \mathcal{X}$. In addition, let $X = \mathbf{x}$ and $Y = \mathbf{y}$ for notational convenience.

Based on the standard R-D function characterized as an optimization problem in (1), we can consider the following modified optimization problem:

$$\begin{aligned} R = \min\nolimits_{Q(y \,|\, x)} \mathsf{I}(X; Y) \\ \text{s.t. } P(y) = \textstyle\sum_x Q(y \,|\, x) P(x), \forall y, \; \mathbb{E}[d(X, Y)] \leq D, \end{aligned} \tag{4}$$

The first constraint ensures ZEPE and is thus referred to as the ZEPE constraint, with the resulting R-D function called the ZEPE R-D function. This additional constraint leads to a higher rate compared to the conventional R-D function for the same distortion level. While the perception error is eliminated, traditional distortion metrics (e.g., MSE) may increase, requiring more bits to maintain the same level of conventional distortion. This reflects a fundamental trade-off: achieving perceptual fidelity under the ZEPE constraint may incur a rate penalty due to the stricter perceptual requirement.

Lemma 1: The problem in (4) is a convex optimization problem, and if the solution is given by (if it exists)

$$Q(y \,|\, x) \propto e^{-\lambda d(x,y) - \mu(y)}, \tag{5}$$

where λ and $\mu(y)$ are Lagrange multipliers.

Proof: Due to the ZEPE constraint, the mutual information becomes

$$\mathsf{I}(X; Y) = \mathsf{H}(Y) - \mathsf{H}(Y \,|\, X) = \mathsf{H}(X) - \mathsf{H}(Y \,|\, X). \tag{6}$$

Thus, minimizing $\mathsf{I}(X; Y)$ becomes equivalent to maximizing $\mathsf{H}(Y \,|\, X)$ or minimizing $\sum_{x,y} Q(y \,|\, x) P(x) \log Q(y \,|\, x)$. The resulting unconstrained optimization problem is given by

$$\begin{aligned} \mathcal{L}(Q, \lambda, \mu) = & \sum_{x,y} Q(y \,|\, x) P(x) \log Q(y \,|\, x) \\ & + \lambda \left(\sum_{x,y} d(x, y) Q(y \,|\, x) P(x) - D \right) \\ & + \sum_y \mu(y) \left(\sum_x Q(y \,|\, x) P(x) - P(y) \right), \end{aligned} \tag{7}$$

which shows that the objective function is a convex function of $Q(y \,|\, x)$ and constraints are linear. By differentiating $\mathcal{L}(Q, \lambda, \mu)$ and setting it to zero, we have (5), which completes the proof. ■

Lemma 2: With the ZEPE constraint, $Q(y \,|\, x)$ in (5) is now dependent only on λ as follows:

$$Q_\lambda(y \,|\, x) = v(y) e^{-\lambda d(x,y)}, \tag{8}$$

where

$$v(y) = \frac{P(y)}{\sum_x e^{-\lambda d(x,y)} P(x)} \geq 0. \tag{9}$$

Furthermore, $v(y)e^{-\lambda d(x,y)}$ is decreasing in λ.

Proof: In (5), $Q(y \,|\, x)$ can be rewritten as

$$Q(y \,|\, x) = v(y)e^{-\lambda d(x,y)} \tag{10}$$

where $v(y) \propto e^{-\mu(y)} \geq 0$ is a function of y. Due to the ZEPE constraint, we have

$$\begin{aligned} P(y) &= \sum_x Q(y \,|\, x)P(x) = \sum_x v(y)e^{-\lambda d(x,y)}P(x) \\ &= v(y) \sum_x e^{-\lambda d(x,y)} P(x), \end{aligned} \tag{11}$$

which leads to (9). We can readily show that $Q_\lambda(y \,|\, x)$ in (8) decreases with λ. This completes the proof. ■

Thanks to Lemma 2, the ZEPE R-D function can be found by optimizing λ only. To this end, the bi-section method can be used.

Example 1: Let $\mathcal{X} = \mathcal{Y} = \{1, \ldots, 6\}$, $P(x) = \frac{x}{21}$, $x \in \mathcal{X}$, and $d(x,y) = \frac{2}{5}|x-y|^2$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. In Fig. 1, we show two R-D functions: the standard (or conventional) R-D function and the ZEPE R-D function. It is clearly shown that the rate of the ZEPE R-D function is higher than that of the conventional one for the same distortion. Notably, the KL divergence of the ZEPE reconstruction is zero, as expected, whereas that of the conventional reconstruction can be arbitrarily high unless the distortion is sufficiently small. This indicates that the conventional reconstruction is not necessarily close to the original signal in terms of perceptual quality.

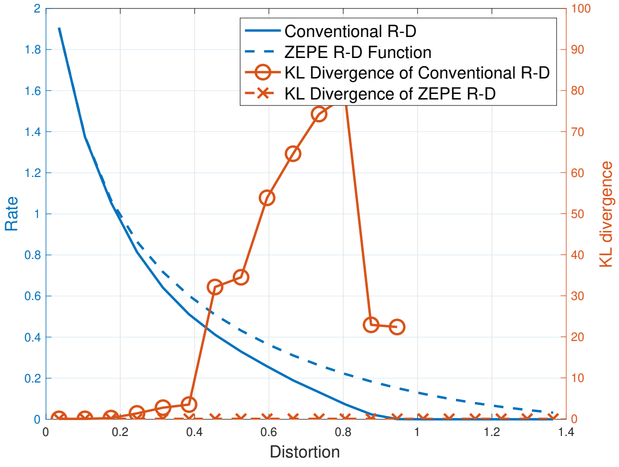


Fig. 1: Conventional and ZEPE R-D functions.

It is noteworthy that since the rate of the ZEPE R-D function is higher than that of the conventional one, it might be seen as less bandwidth-efficient. However, in an extreme case, the rate can be reduced to zero if actual distortion is not a concern, while still maintaining ZEPE. Thus, when considering the ZEPE R-D function, we may allow a larger distortion, as the reconstructed signal may preserve the same perceptual quality despite higher conventional distortion.

IV. Rate Distortion Function of Semantic Sources

A. Semantic Rate-Distortion

In [7], the signal source X is characterized as an observation that depends on a hidden or latent variable $Z \in \mathcal{Z}$, where $\mathcal{Z}$ represents the latent space. In other words, X follows the conditional probability distribution $P(X \,|\, Z)$ for a given Z. The pair of Z and X, i.e., (Z, X), is called a semantic source. It is assumed that X is available as an observation, while Z is unknown. In this sense, Z and X are also referred to as the intrinsic state variable and the extrinsic observation, respectively. An encoder's task is to estimate Z from the observation X and encode X for reconstruction at a decoder.

Let $\hat{Z} \in \hat{\mathcal{Z}}$ be the estimate of Z, where $\hat{\mathcal{Z}}$ is the space of all possible $\hat{Z}$. Then, the outputs of the decoder are Y and $\hat{Z}$. Thus, the following optimization problem is considered [7]:

$$\begin{aligned} &R = \min\nolimits_{Q(\hat{Z},Y \,|\, X)} \mathsf{I}(X;Y) \\ &\text{subject to } \mathbb{E}[d_\text{s}(Z,\hat{Z})] \leq D_\text{s} \text{ and } \mathbb{E}[d(X,Y)] \leq D, \end{aligned} \tag{12}$$

where $D_\text{s} \geq 0$ is a constant and $d_\text{s} : \mathcal{Z} \times \hat{\mathcal{Z}} \to \mathbb{R}^+$ is the distortion function between Z and $\hat{Z}$, and $Q(\hat{Z}, Y \,|\, X)$ is the joint conditional distribution of $\hat{Z}$ and Y for given X. In [7], it is shown that

$$\mathbb{E}[d_\text{s}(Z,\hat{Z})] = \mathbb{E}[\hat{d}_\text{s}(X,\hat{Z})] \tag{13}$$

where

$$\hat{d}_\text{s}(X,\hat{Z}) = \mathbb{E}[d_\text{s}(Z,\hat{Z}) \,|\, X] = \sum_{z \in \mathcal{Z}} \pi(z \,|\, x) d_\text{s}(z,\hat{Z}). \tag{14}$$

Here, $\pi(z \,|\, x)$ is the conditional distribution of z for given $X = x$, i.e.,

$$\pi(z \,|\, x) = \frac{P(x|z)\pi(z)}{\sum_{z \in \mathcal{Z}} P(x|z)\pi(z)}, \tag{15}$$

where $\pi(z)$ is the distribution of Z. Using the Markov chain, $Z \to X \to (Y, \hat{Z})$, we can also readily shown that

$$\begin{aligned} \mathbb{E}[d_\text{s}(Z,\hat{Z})] &= \sum_{z,\hat{z}} d_\text{s}(z,\hat{z}) \sum_{x,y} Q(\hat{z}, y \,|\, x)P(x \,|\, z)\pi(z) \\ &= \sum_{z,\hat{z}} d_\text{s}(z,\hat{z}) Q(\hat{z} \,|\, z)\pi(z), \end{aligned} \tag{16}$$

where $Q(\hat{z} \,|\, z) = \sum_{x,y} Q(\hat{z}, y \,|\, x)P(x \,|\, z)$. Then, it can be shown that $R(D, D_\text{s})$, which is the solution of (12), is convex in D and D_s.

B. ZEPE Rate-Distortion with Latent Variables

While the distortion of the latent variable has been studied in [7] for semantic communication, we take a different approach by incorporating ZEPE, which will be discussed in this subsection.

With the introduction of the latent variable Z, the signal transmission process can be understood through the following Markov chain:

$$Z \to X \to Y. \tag{17}$$

In general, in the context of semantic communication, we assume that

$$|\mathcal{Z}| \ll |\mathcal{X}|, \tag{18}$$

i.e., the number of possible latent representations in $\mathcal{Z}$ is significantly smaller than the number of possible observed signals in $\mathcal{X}$. This implies that semantic communication aims to achieve efficient compression by mapping a large input space to a much smaller latent space while preserving the essential meaning of the information. Thus, the sender (or encoder) can infer the latent variable Z as in [7] and transmit it instead of X to the receiver (or decoder). The receiver can generate a sample Y', which requires a lower transmission rate bounded as follow: $R_{\rm sc} = \mathsf{H}(Z) \le \log_2 |\mathcal{Z}|$, while the associated Markov chain is given by

$$Z \to Y'. \tag{19}$$

More importantly, when aiming to achieve ZEPE transmissions, we need to ensure that $\mathsf{D}(P(X \,|\, Z); Q(Y' \,|\, Z)) = 0$, which can be satisfied if the following condition holds:

$$Q(y' \,|\, z) = P(y' \,|\, z), \forall (y,z) \in \mathcal{Y} \times \mathcal{Z}. \tag{20}$$

In other words, with the latent variable, Z, the receiver is expected to generate a signal, $\hat{Y}$, from $P(y' \,|\, z)$ with ZEPE.

The above ZEPE approach can be generalized to take into account the distortion. To this end, we can formulate the following optimization problem for the ZEPE semantic R-D function:

$$R_{\rm sc} = \min_{Q(y \,|\, x,z)} \mathsf{I}(X;Y \,|\, Z)$$

$$\text{s.t.} \quad \left\{ \begin{array}{l} P(y \,|\, z) = \sum_x Q(y \,|\, x,z) P(x \,|\, z), \forall (y,z) \in \mathcal{Y} \times \mathcal{Z}, \\ \mathbb{E}[d(X,Y) \,|\, Z = z] \le D(z), \ z \in \mathcal{Z}, \end{array} \right. \tag{21}$$

where $D(z)$ represents the distortion threshold for a given latent variable $Z = z$. In (21), the first constraint is to ensure ZEPE, i.e., $\mathsf{D}(Q(Y \,|\, Z); P(Y \,|\, Z)) = 0$. It is noteworthy that $R_{\rm sc}$ in (21) is the rate that does not include the rate to transmit the latent variable, Z. Thus, the actual rate becomes $R_{\rm sc} + \mathsf{H}(Z)$.

Lemma 3: With the ZEPE constraint, for given $\lambda(z)$, which are the Lagrange multipliers for the distortion constraints, the optimal transition probability in (22) is given by

$$Q_\lambda(y \,|\, x,z) = v(y,z) e^{-\lambda(z) d(x,y)}, \tag{22}$$

where

$$v(y,z) = \frac{P(y \,|\, z)}{\sum_x e^{-\lambda(z) d(x,y)} P(x \,|\, z)} \ge 0. \tag{23}$$

Then, for each $z \in \mathcal{Z}$, $\lambda(z)$ can be determined to satisfy the second constraint in (22) for a given $D(z)$.

Proof: It can be shown that

$$\begin{aligned} \mathsf{I}(X;Y \,|\, Z) &= \mathsf{H}(Y \,|\, Z) - \mathsf{H}(Y \,|\, Y,Z) \\ &= \mathsf{H}(X \,|\, Z) - \mathsf{H}(Y \,|\, X,Z), \end{aligned} \tag{24}$$

where the last equality is due to the ZEPE constraint. Thus, we need to minimize

$$-\mathsf{H}(Y \,|\, X,Z) = \sum_{x,y,z} Q(y \,|\, x,z) P(x,z) \log Q(y \,|\, x,z).$$

Then, we can consider the following objective function for an unconstrained optimization problem:

$$\begin{aligned} \mathcal{L}(Q,\lambda,\mu) &= \sum_{x,y,z} Q(y \,|\, x,z) P(x,z) \log Q(y \,|\, x,z) \\ &+ \sum_z \lambda(z) \left(\sum_{x,y} d(x,y) Q(y \,|\, x,z) P(x,z) - D(z) \right) \\ &+ \sum_{y,z} \mu(y,z) \left(\sum_x Q(y \,|\, x,z) P(x \,|\, z) - P(y \,|\, z) \right), \end{aligned} \tag{25}$$

where the last term can be modified as follows:

$$\sum_{y,z} \frac{\mu(y,z)}{\pi(z)} \left(\sum_x Q(y \,|\, x,z) P(x,z) - P(y,z) \right).$$

Then, letting $\tilde{\mu}(y,z) = \frac{\mu(y,z)}{\pi(z)}$, thanks to Lemma 1, we have

$$Q(y \,|\, x,z) \propto e^{-\lambda(z) d(x,y) - \tilde{\mu}(y,z)}. \tag{26}$$

From Lemma 2, it can also be shown that $Q(y \,|\, x,z)$ becomes $Q_\lambda(y \,|\, x,z)$ in (22) for a given λ. ■

In the distortion-constrained ZEPE approach, the latent variable Z can be seen as side information that aids in generating the reconstructed signal for ZEPE. Additionally, the distortion thresholds $D(z)$ can be adjusted to produce reconstructed signals that more closely resemble X, at the cost of an increased number of bits, quantified by $\mathsf{I}(Y;X|Z)$.

C. Relation to Other Approaches

It is noteworthy that the semantic rate-distortion framework is related to the rate-perception-classification (R-P-C) framework for task-oriented communication [13]. In particular, the R-P-C framework imposes an additional constraint on the classification in (3), as follows:

$$\begin{array}{rl} R = & \min_{Q(y \,|\, x)} \mathsf{I}(X;Y) \\ \text{s.t.} & \left\{ \begin{array}{ll} \mathsf{D}(P;Q) & \le \delta \\ \mathsf{H}(S \,|\, Y) & \le C, \end{array} \right. \end{array} \tag{27}$$

where S is the classification or label variable, which is not directly observable, but can be inferred from X, and $C \ge 0$. Thus, the second constraint in (27) ensures that the received signal Y retains sufficient information to identify the label S with an uncertainty no greater than C. This reflects the task-oriented nature of the R-P-C framework, where the reconstruction is not only judged by fidelity but also by its utility in downstream tasks such as classification.

However, the main difference from the R-P-C framework is that S is considered the semantic variable in the semantic rate-distortion framework (i.e., $S = Z$), which is extracted at the semantic encoder and transmitted to the receiver. In contrast to R-P-C, where the decoder must infer S from the received representation $\mathbf{y}$, the semantic rate-distortion framework treats S as an explicit part of the communication process.

Another approach related to the semantic rate-distortion framework is optimal transport [14], which has also been extensively studied in the context of quantization and lossy compression [15] [16] [17]. In particular, [16] considers the following problem under the Markov chain $X \to Z \to Y$:

$$\begin{aligned} D(P(x), Q(y), R) = \min\nolimits_{Q_{X,Z,Y} \in \mathcal{Q}} \mathbb{E}[d(X,Y)] \\ \text{s.t. } \mathsf{H}(Z) \le R, \end{aligned} \tag{28}$$

where $\mathcal{Q}$ denotes the set of joint distributions defined as

$$\begin{aligned} \mathcal{Q} = \{Q: \ Q(x,z,y) = P(x)P(z \,|\, x)Q(y \,|\, z), \\ \sum_{x,z} Q(x,z,y) = Q(y)\}. \end{aligned}$$

In (28), $P(z \,|\, x)$ can be assumed to be given, acting as a stochastic semantic encoder, where Z represents the inferred semantic meaning of X, subject to the rate constraint, $\mathsf{H}(Z) \le R$. This constraint ensures that the semantic representation Z remains compressible within the target rate budget, while still preserving sufficient information for accurate reconstruction of Y. The decoder then maps Z to Y through a learned or optimized conditional distribution $Q(y \,|\, z)$, aiming to minimize the expected distortion between the original signal X and its reconstruction Y.

Note that in (28), the marginal distributions, $P(x)$ and $Q(y)$, are assumed to be given. In general, $P(x)$ represents the distribution of noise-corrupted or degraded images, while $Q(y)$ denotes the distribution of clean or high-quality reconstructions. Thus, through the formulation in (28), the goal is to compress and restore a degraded image, where the intermediate semantic representation Z enables meaningful reconstruction under a rate constraint.

On the other hand, in the semantic rate-distortion framework, X represents an original image, which is not necessarily degraded or noisy. The focus is not on reconstruction but on capturing and transmitting the semantic content of the input with minimal rate and acceptable distortion. In this setting, the semantic variable Z is inferred from X and directly communicated to the receiver, who reconstructs a meaningful output Y based on Z.

V. Generative Models for Implementation

As in [7], the latent variable, Z, in the context of semantic communication can be seen as the intention or meaning of a given observation, X. This is also considered for modeling the LLM [18], where X is seen as a message that is associated with an intention or meaning. Thus, it is essential to infer Z from X at the encoder, which is also essential for ZEPE as shown earlier. In this section, we consider two approaches for ZEPE R-D framework based on generative models. To discuss practical approaches, we now consider the case where the signal or observation is a vector $\mathbf{x} \in \mathbb{R}^n$, as defined earlier.

A. ZEPE via VAE

In this subsection, we demonstrate that the VAE [10] for lossy compression can be used for ZEPE, which has also been used for joint source-channel coding [19]. In particular, we consider the system model shown in Fig. 2, where the signal, $\mathbf{x}$, becomes the input to the semantic encoder that infers the latent variable, $\mathbf{z}$. The latent variable is transmitted to the decoder, which generates a reconstructed signal, denoted by $\hat{\mathbf{x}}$. There are two key assumptions. First, the meaning or intention of $\mathbf{x}$, denoted by $\mathbf{w}$, can be perfectly inferred by the encoder, ensuring that $\mathbf{w} \Leftrightarrow \mathbf{z}$, i.e., $\mathbf{z}$ and $\mathbf{w}$ have a bijective relationship. Second, the transmission is error-free, meaning that $\hat{\mathbf{z}} = \mathbf{z}$. Consequently, we assume that $\mathbf{w} = \mathbf{z} = \hat{\mathbf{z}}$.

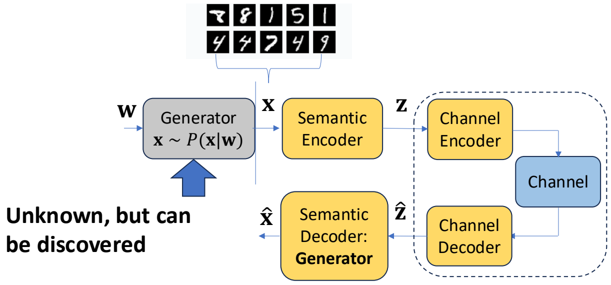


Fig. 2: A diagram for semantic communication with ZEPE.

When the VAE is used, the encoder functions as the semantic encoder to infer $\mathbf{z}$, while the decoder acts as a generator to produce $\hat{\mathbf{x}}$ from $\mathbf{z}$. Based on this, the input-output relation of the VAE can be expressed as the following transition probability or conditional distribution involving the encoder and decoder distributions:

$$Q(\mathbf{y} \,|\, \mathbf{x}) = \sum_{\mathbf{z}} \mathcal{G}(\mathbf{y} \,|\, \mathbf{z}) \xi(\mathbf{z} \,|\, \mathbf{x}), \tag{29}$$

where $\xi(\mathbf{z} \,|\, \mathbf{x})$ represents the encoder distribution, and $\mathcal{G}(\mathbf{y} \,|\, \mathbf{z})$ represents the decoder distribution.

Suppose that the input space, $\mathcal{X}$, is given by $\mathcal{X} = \bigcup_{m=1}^{M} \mathcal{X}_m$, where $\mathcal{X}_m$ is the subset associated with the intention $Z = m$. When $\mathbf{x} \in \mathcal{X}_m$, suppose that the following relation holds:

$$\xi(\mathbf{z} \,|\, \mathbf{x}) = \delta(\mathbf{z} - \mathbf{z}_m), \tag{30}$$

where $\mathbf{z}_m$ is the latent vector associated with $Z = m$. For any unambiguous model, the encoder can uniquely identify the intention [18]. In this case, we can assume that $\mathcal{X}_m \cap \mathcal{X}_l = \emptyset$ for $l \ne m$. Thus, from (29), we have

$$Q(\mathbf{y} \,|\, \mathbf{x}) = \mathcal{G}(\mathbf{y} \,|\, \mathbf{z}_m) \text{ for } \mathbf{x} \in \mathcal{X}_m. \tag{31}$$

From this, it can be shown that

$$\begin{aligned} Q(\mathbf{y}) &= \sum_{\mathbf{x}} Q(\mathbf{y}|\mathbf{x}) P(\mathbf{x}) \\ &= \sum_{m} \mathcal{G}(\mathbf{y}|\mathbf{z}_m) \Pr(\mathbf{x} \in \mathcal{X}_m). \end{aligned} \tag{32}$$

Note that when there is no ambiguity in the model, we can assume that $\pi_m(\mathbf{z}) = \delta(\mathbf{z} - \mathbf{z}_m)$ in (30). Thus, $P(\mathbf{x}) =$

$\sum_{m=1}^{M} P(\mathbf{x}\,|\,\mathbf{z}_m)\pi_m$, where $\pi_m = \Pr(Z = m)$. As a result, in (32), if

$$\mathcal{G}(\mathbf{x}\,|\,\mathbf{z}_m) \leftarrow P(\mathbf{x}\,|\,\mathbf{z}_m) \tag{33}$$

$$\Pr(\mathbf{x} \in \mathcal{X}_m) \leftarrow \pi_m, \tag{34}$$

we have $Q(\mathbf{x}) = P(\mathbf{x})$, which demonstrates that the VAE is capable of ZEPE. Note that (33) is equivalent to the ZEPE constraint in the ZEPE R-D framework, as $\mathcal{G}(\mathbf{x}\,|\,\mathbf{z}_m) = Q(\mathbf{y}\,|\,\mathbf{z}_m)$.

B. A Hybrid Approach

The semantic rate-distortion framework states that satisfying the ZEPE constraint requires a higher rate to achieve the same distortion (see Fig. 1), which can be achieved by transmitting additional bits. However, this method depends on detailed statistical knowledge of the signal, which may not be practical or adaptable to specific samples. Furthermore, in practice, the encoder and decoder of semantic communication are commonly implemented as a neural network pre-trained over a large dataset, and thus cannot flexibly adjust their input and output dimensions for each sample. To overcome these challenges, we introduce a hybrid approach, where a semantic communciation link is employed to meet the ZEPE constraint (or achieve sufficiently low perception error), while a classical communication link is used to further reduce distortion, conditioned on the outcome of the semantic communication link.

Precisely, the encoder is able to extract the latent vector, $\mathbf{z}$, which is transmitted to the decoder. Subsequently, $\mathbf{y}$ is sampled from $Q(\mathbf{y}\,|\,\mathbf{z}) = P(\mathbf{y}\,|\,\mathbf{z})$, which is a generated signal with ZEPE. To further minimize distortion, based on the null space decomposition [20], the encoder may transmit additional bits or signals. To determine what additional signals are needed, we express $\mathbf{x}$ as follows:

$$\mathbf{x} = \mathbf{A}^\dagger \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{x}, \tag{35}$$

where $\mathbf{A} \in \mathbb{R}^{d \times n}$ represents a projection matrix that maps the signal vector $\mathbf{x}$ onto a lower-dimensional space (of dimension $d \ll n$). Here, $\mathbf{A}^\dagger$ denotes the pseudo-inverse of $\mathbf{A}$. It is now assumed that $\mathbf{x}$ in the second term on the right-hand side (RHS) is replaced with a generated signal at the decoder, $\hat{\mathbf{x}} \sim Q(\mathbf{y}\,|\,\mathbf{z})$. Then, the reconstructed signal becomes

$$\tilde{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\hat{\mathbf{x}}, \tag{36}$$

where the projected signal, $\mathbf{A}\mathbf{x} \in \mathbb{R}^d$, is additionally transmitted from the encoder to the decoder. The resulting approach is referred to as the hybrid approach, since the encoder sends $\mathbf{z}$ (for ZEPE) as well as $\mathbf{A}\mathbf{x}$ (for low distortion).

It is noteworthy that the projection matrix, $\mathbf{A}$, can be optimized to minimize the distortion, i.e.,

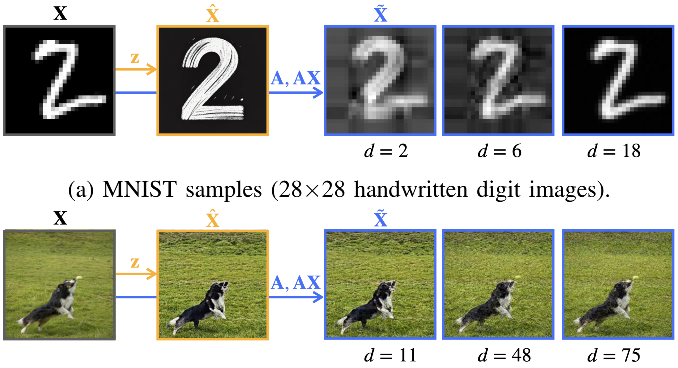
$$\hat{\mathbf{A}} = \underset{\mathbf{A}}{\operatorname{argmin}}\, \mathbb{E}[d(\mathbf{x}, \tilde{\mathbf{x}})]. \tag{37}$$

If the MSE is used, it can be shown that

$$\begin{aligned} \mathbb{E}[d(\mathbf{x}, \tilde{\mathbf{x}})] &= \mathbb{E}[\mathbb{E}[||\mathbf{x} - \tilde{\mathbf{x}}||^2\,|\,\mathbf{z}]] \\ &= \mathbb{E}[\mathbb{E}[||\mathbf{x} - \mathbf{A}^\dagger \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\hat{\mathbf{x}}||^2\,|\,\mathbf{z}]] \\ &= (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbb{E}[\mathbb{E}[||\mathbf{x} - \hat{\mathbf{x}}||^2\,|\,\mathbf{z}]], \end{aligned} \tag{38}$$

where $\hat{\mathbf{x}} \sim Q(\mathbf{y}\,|\,\mathbf{z})$. It is also possible to separately optimize $\mathbf{A}$ for each $\mathbf{z} \in \mathcal{Z}$.

Fig. 3a shows an example of a handwritten digit image. By transmitting the latent variable corresponding to digit 2, the decoder is able to generate an image of digit 2, which is $\hat{\mathbf{x}}$. While both the original and generated images (i.e., $\mathbf{x}$ and $\hat{\mathbf{x}}$, respectively) convey the same semantic meaning - that they represent the digit 2 - the unique characteristics of the original image are not effectively preserved in $\hat{\mathbf{x}}$. However, by transmitting additional bits for $\mathbf{A}\mathbf{x}$, the reconstructed image can better capture the finer details and structure of the original. Furthermore, increasing the dimension of the projected space, d, allows for a richer representation, further improving the fidelity of the reconstructed image. Similar observations hold for higher-resolution images, as shown in Fig. 3b. A more detailed approach and simulation results can be found in [21], where a hybrid semantic-classical coding scheme is proposed to balance semantic preservation and perceptual quality.



(a) MNIST samples (28×28 handwritten digit images).

(b) Flickr 8k (256×256 images)

Fig. 3: Hybrid transmission experiments under (a) the MNIST and (b) Flickr 8k images.

VI. Conclusion

In this paper, a semantic R-D theory was studied with a ZEPE constraint so that the reconstructed signals should have no perceptual difference from the original signals while allowing for a certain level of distortion in terms of traditional metrics. By incorporating this constraint, we derived the optimal transition probability and analyzed the trade-off between rate and distortion under the ZEPE condition. Furthermore, we extended the framework to include inferred latent variables, which capture the semantic meaning of the original signals. For practical implementation, we discussed the use of generative models such as VAEs to achieve effective semantic reconstruction.

There are several important issues to be addressed in future work. For instance, in the hybrid approach, a linear projection was employed to send an additional signal in order to reduce the distortion. However, this approach does not guarantee that the resulting signal will have ZEPE. To mitigate this issue, we can explore the use of machine learning models, such as generative networks, that can learn to reconstruct signals with ZEPE while minimizing distortion.

Acknowledgment

This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ITRC (Information Technology Research Center) grant funded by the Korea government (MSIT) (IITP-2025-RS-2023-00259991, 25%), in part by IITP grant funded by the MSIT (No. 2021-0-00347, 6G Post-MAC (POsitioning- & Spectrum-aware intelligenT MAC for Computing & Communication Convergence), 25%), in part by the National Research Foundation of Korea (NRF) grant funded by the MSIT (RS-2025-02217000, 25%), and in part by SUTD Kickstarter Initiative (SKI 2021_06_08, 25%).

References

[1] W. Weaver, "Recent contributions to the mathematical theory of communication," *ETC: A Review of General Semantics*, vol. 10, no. 4, pp. 261–281, 1953.

[2] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *2011 IEEE Network Science Workshop*, pp. 110–117, 2011.

[3] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.

[4] P. Yi, Y. Cao, X. Kang, and Y.-C. Liang, "Deep learning-empowered semantic communication systems with a shared knowledge base," *IEEE Transactions on Wireless Communications*, vol. 23, no. 6, pp. 6174–6187, 2024.

[5] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gündüz, "Deep joint source-channel coding for semantic communications," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 42–48, 2023.

[6] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 675–685, PMLR, 09–15 Jun 2019.

[7] J. Liu, S. Shao, W. Zhang, and H. V. Poor, "An indirect rate-distortion characterization for semantic sources: General model and the case of Gaussian observation," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 5946–5959, 2022.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. NJ: John Wiley, second ed., 2006.

[9] X. Niu, B. Bai, N. Guo, W. Zhang, and W. Han, "Rate–distortion–perception trade-off in information theory, generative models, and intelligent communications," *Entropy*, vol. 27, no. 4, 2025.

[10] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[11] E. Grassucci, S. Barbarossa, and D. Comminiello, "Generative semantic communication: Diffusion models beyond bit recovery." arXiv:2306.04321v1, 2023.

[12] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.

[13] Y. Wang, Y. Wu, S. Ma, and Y.-J. A. Zhang, "Task-oriented lossy compression with data, perception, and classification constraints," 2025.

[14] C. Villani, *Optimal Transport*. Grundlehren der mathematischen Wissenschaften, Berlin, Germany: Springer, Dec. 2009.

[15] R. M. Gray, "Transportation distance, Shannon information, and source coding," in *GRETSI Symposium on Signal and Image Processing*, 2013.

[16] H. Liu, G. Zhang, J. Chen, and A. Khisti, "Cross-domain lossy compression as entropy constrained optimal transport," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 3, pp. 513–527, 2022.

[17] L. Xie, L. Li, J. Chen, and Z. Zhang, "Output-constrained lossy source coding with application to rate-distortion-perception theory," *IEEE Transactions on Communications*, vol. 73, no. 3, pp. 1801–1815, 2025.

[18] H. Jiang, "A latent space theory for emergent abilities in large language models." arXiv:2304.09960v3, 2023.

[19] M. Nemati, J. Park, and J. Choi, "VQ-VAE empowered wireless communication for joint source-channel coding and beyond," in *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, pp. 3155–3160, 2023.

[20] J. Schwab, S. Antholzer, and M. Haltmeier, "Deep null space learning for inverse problems: convergence analysis and rates," *Inverse Problems*, vol. 35, p. 025008, jan 2019.

[21] H. Nam, J. Park, J. Choi, and S.-L. Kim, "Hybrid semantic-complementary transmission for high-fidelity image reconstruction." arxiv:2507.17196, 2025, 2025.