On the Performance of Pilot-Free Semantic Communication over Time-Varying OFDM Channels

Jihun Park and Yo-Seb Jeon
Department of Electrical Engineering, POSTECH, Pohang, Gyeongbuk 37673, Republic of Korea
Email: {jihun.park, yoseb.jeon}@postech.ac.kr

***Abstract*—This paper investigates pilot-free semantic communication over time-varying orthogonal frequency division multiplexing (OFDM) channels. By eliminating pilots, the framework improves bandwidth efficiency and leverages end-to-end learning to implicitly adapt to fading dynamics. Performance evaluation shows that the system achieves strong reconstruction quality when training and deployment channels are matched, but suffers significant degradation under mismatched conditions. In particular, models trained on more challenging fading environments generalize better and maintain higher robustness across diverse scenarios. These findings highlight both the potential and vulnerability of pilot-free semantic communication and underscore the importance of robust training strategies for reliable deployment.**

I. Introduction

Semantic communication aims to reconstruct the source data in a task-relevant manner, in contrast to conventional systems that are based on exact bit-level recovery. From the perspective of joint source–channel coding (JSCC), this paradigm is particularly advantageous in short block length regimes, where traditional separation-based designs suffer from coding overhead and error propagation [1]–[3]. By jointly optimizing the entire communication chain, semantic communication can achieve higher efficiency and robustness, maintaining meaningful reconstruction quality even when reliable bit-wise recovery is infeasible. The dominant physical layer technology, orthogonal frequency division multiplexing (OFDM), combats the frequency-selective nature of wireless channels but traditionally relies on pilot symbols for channel estimation. This pilot-assisted approach, however, faces a fundamental trade-off in mobile environments. High user mobility induces Doppler shifts, causing the channel to be time-varying. To track these rapid changes, a higher density of pilots is required, which significantly reduces spectral efficiency. Conversely, using fewer pilots can lead to performance loss from channel aging, where the channel estimate becomes outdated before it can be applied to the data symbols.

In this context, pilot-free semantic communication has the potential to offer significant advantages. By eliminating the need for dedicated pilot symbols, such systems can reduce signaling overhead and improve spectral efficiency [2], while leveraging end-to-end learning to implicitly adapt to channel variations. This philosophy aligns naturally with neural network–based transceivers, which treat the channel as an integrated, though unknown, component of the communication system. Despite these advantages, a critical challenge remains: robustness to channel mismatch. When the propagation environment during deployment differs from that used for training—for example, in terms of power delay profile (PDP) or user velocity—the learned transceiver may fail to generalize, leading to significant performance degradation. While the potential of pilot-free systems has been recognized, their reliability in time-varying OFDM fading channels has not been systematically explored. Motivated by this gap, this paper investigates pilot-free semantic communication under standardized 3GPP fading models [4]. We train end-to-end transceivers under specific channel conditions and evaluate their reconstruction quality across matched and mismatched environments. The results reveal that while pilot-free semantic communication achieves strong performance under matched conditions, it suffers from pronounced degradation under channel profile mismatch. These findings highlight the dual nature of pilot-free systems—high efficiency but vulnerability to mismatch—and underscore the need for robust training strategies to ensure reliable deployment in dynamic wireless environments.

II. System Model and End-to-End Training

We consider a single-input single-output OFDM-based semantic communication system that transmits source image over a frequency-selective fading channel without explicit pilot symbols or channel estimation. Let $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ denote the source image. A learnable encoder $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ produces a latent tensor

$$\mathbf{Z} = \mathcal{E}_{\boldsymbol{\theta}}(\mathbf{X}) \in \mathbb{R}^{C_\ell\times H_\ell\times W_\ell}. \tag{1}$$

Assuming the number of OFDM subcarrier $N_{sc} = H_\ell W_\ell$, the latent tensor is reshaped into a matrix $\mathbf{Z}_{\text{reshaped}} \in \mathbb{R}^{C_\ell\times N_{sc}}$. This matrix is subsequently partitioned into two halves, $\mathbf{Z}_{\text{real}} \in \mathbb{R}^{C_\ell/2\times N_{sc}}$ and $\mathbf{Z}_{\text{imag}} \in \mathbb{R}^{C_\ell/2\times N_{sc}}$, which are then used to construct the complex-valued channel input symbol matrix:

$$\mathbf{S} = \mathbf{Z}_{\text{real}} + j\mathbf{Z}_{\text{imag}} \in \mathbb{C}^{C_\ell/2\times N_{sc}}. \tag{2}$$

To maintain a consistent transmit power, each of the $C_\ell/2$ symbol vectors, denoted by $\mathbf{s}_t \in \mathbb{C}^{N_{\text{sc}}}$, is normalized such that its total power equals N_{sc}, i.e., $\bar{\mathbf{s}}_t = \sqrt{N_{\text{sc}}}\,\mathbf{s}_t/\|\mathbf{s}_t\|_2$.

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00453301) and in part by a grant of Korean ARPA-H Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (No. RS-2025-25454431).

We consider a time-varying, frequency-selective wireless channel. The channel is characterized by a time-varying channel impulse response (CIR) vector, $\mathbf{h}[t] = [h_0[t], h_1[t], \ldots, h_{L-1}[t]]^{\mathsf{T}} \in \mathbb{C}^L$, where t is the discrete time index corresponding to the OFDM symbol and L is the number of channel taps. At any given time t, each tap $h_\ell[t]$ is modeled as an independent complex Gaussian random variable with $h_\ell[t] \sim \mathcal{CN}(0, \sigma_\ell^2)$. The tap variances $\{\sigma_\ell^2\}$ are determined by a given power delay profile, which is normalized such that the total average power is unity, i.e., $\sum_{\ell=0}^{L-1} \sigma_\ell^2 = 1$. The temporal evolution of the CIR follows a first-order autoregressive model, which captures the channel's time correlation:

$$\mathbf{h}[t+1] = \rho\, \mathbf{h}[t] + \sqrt{1-\rho^2}\, \mathbf{w}[t], \tag{3}$$

where $\mathbf{w}[t] \in \mathbb{C}^L$ is a complex Gaussian innovation vector with the same statistics as $\mathbf{h}[t]$. The time correlation coefficient, ρ, is derived from the Jakes model, reflecting the impact of mobility:

$$\rho = J_0(2\pi f_d T_{\text{sym}}), \tag{4}$$

where $J_0(\cdot)$ is the zeroth-order Bessel function, T_{sym} is the OFDM symbol duration, and f_d is the maximum Doppler shift determined by the carrier frequency and receiver velocity. For each OFDM symbol at time t, the channel frequency response (CFR) on the k-th subcarrier, $h_f[t,k]$, is obtained via the discrete Fourier transform (DFT) of the corresponding CIR, $\mathbf{h}[t]$:

$$h_f[t,k] = \frac{1}{\sqrt{N_{\text{sc}}}} \sum_{\ell=0}^{L-1} h_\ell[t]\, e^{-j\frac{2\pi k \ell}{N_{\text{sc}}}}, \quad n = 0, \ldots, N_{\text{sc}}-1. \tag{5}$$

The received signal for the t-th symbol on subcarrier k is now modeled as:

$$y[t,k] = h_f[t,k]\, \bar{s}[t,k] + v[t,k], \tag{6}$$

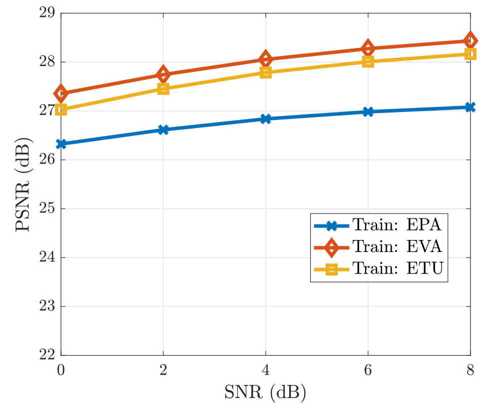
where $v[t,k] \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise at that time and subcarrier. By stacking $\{y[t,k]\}$ over all subcarriers and OFDM symbols, we obtain the frequency-domain receive matrix $\mathbf{Y} \in \mathbb{C}^{C_\ell/2 \times N_{\text{sc}}}$. Its real and imaginary parts are concatenated along the channel dimension to form $\widetilde{\mathbf{Z}} \in \mathbb{R}^{C_\ell \times N_{\text{sc}}}$, which is then reshaped to $\mathbb{R}^{C_\ell \times H_\ell \times W_\ell}$ and fed into the decoder $\mathcal{D}_{\boldsymbol{\phi}}(\cdot)$ for source reconstruction:

$$\widehat{\mathbf{X}} = \mathcal{D}_{\boldsymbol{\phi}}(\widetilde{\mathbf{Z}}_{\text{reshaped}}). \tag{7}$$

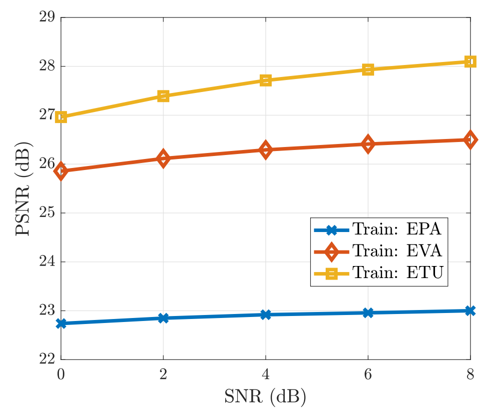
The parameters $\boldsymbol{\theta}$ of the encoder $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ and $\boldsymbol{\phi}$ of the decoder $\mathcal{D}_{\boldsymbol{\phi}}(\cdot)$ are jointly trained end-to-end to minimize the mean squared error (MSE) between the source image $\mathbf{X}$ and the reconstructed image $\widehat{\mathbf{X}}$. This objective is formulated as the loss function:

$$\mathcal{L}_{\text{MSE}} = \mathbb{E}_{\mathbf{X},\, \{\mathbf{h}[t]\},\, v}\left[\left\|\mathbf{X} - \widehat{\mathbf{X}}\right\|_2^2\right], \tag{8}$$

where the expectation is taken over the distributions of the source data, the channel realizations, and the additive noise. Minimizing this loss enables the system to learn a communication strategy that is implicitly robust to the channel dynamics.



(a) Test: EVA



(b) Test: ETU

Fig. 1. PSNR performance of pilot-free semantic communication under different SNR range. Models are trained on EPA, EVA, and ETU channels and tested on (a) EVA and (b) ETU channels.

III. Simulation Results

We conducted training and evaluation using the CIFAR-10 dataset, where each image has a resolution of $3 \times 32 \times 32$. The semantic encoder–decoder was implemented with a swin transformer backbone [5], which maps each image into a latent tensor of dimension $16 \times 8 \times 8$. The encoder is constructed with two hierarchical stages, each with depth 2, embedding dimensions $[32, 16]$, and 8 attention heads per stage, using a window size of 8 and an MLP expansion ratio of 4. The decoder mirrors this design with embedding dimensions $[16, 32]$ while maintaining the same depth, number of heads, and window size. This latent representation was subsequently reshaped into a complex-valued OFDM symbol matrix and transmitted over an OFDM system with $N_{\text{sc}} = 64$ subcarriers and 8 OFDM symbols.

The network was trained for 200 epochs using the Adam optimizer with a learning rate of 1×10^{-3} and a batch size of 16. During training, the propagation channel was simulated as a frequency-selective, time-varying fading process following a 3GPP PDP. Specifically, we considered three standardized profiles: Extended Pedestrian A (EPA), Extended Vehicular A

(EVA), and Extended Typical Urban (ETU) [4]. The temporal correlation of the channel taps was modeled according to Jakes' model with Doppler spread determined by the carrier frequency $f_c = 3.5$ GHz and receiver velocity (EPA: 5 km/h, EVA: 120 km/h, ETU: 300 km/h). We use the peak signal-to-noise ratio (PSNR) as the evaluation metric for reconstruction quality, defined as

$$\mathrm{PSNR} \triangleq 10 \log_{10} \frac{\mathrm{MAX}^2}{\mathrm{MSE}}, \tag{9}$$

where MAX is the maximum pixel value (255 for 8-bit images). The signal-to-noise ratio (SNR) quantifies the relative power of the transmitted signal to the noise, defined as

$$\mathrm{SNR} \triangleq 10 \log_{10} \frac{N_{\mathrm{sc}} \|\bar{\mathbf{s}}_t\|^2 \sum_{\ell=0}^{L-1} \sigma_\ell^2}{\sigma^2}. \tag{10}$$

To evaluate both the effectiveness of pilot-free semantic communication and its robustness against channel profile mismatch, we trained three independent models at an SNR of 0 dB under EPA, EVA, and ETU channels, respectively, and tested them on two deployment channels (EVA and ETU). Figure 1 shows the average PSNR (dB) over 0–8 dB. When training and testing channels are matched, the pilot-free models demonstrate strong reconstruction quality even with short block length, limited bandwidth, and no explicit CSI, highlighting the effectiveness of end-to-end learning under consistent fading conditions. Under mismatched conditions, however, clear differences emerge: the EPA-trained model exhibits the most severe degradation on both EVA and ETU, indicating poor generalization beyond its training environment. The EVA-trained model performs well on EVA and maintains reasonable performance on ETU, though with a noticeable drop. By contrast, the ETU-trained model demonstrates the strongest robustness, achieving the best performance on ETU and maintaining more robust performance than the EPA-trained model on EVA. These results emphasize that pilot-free semantic systems are vulnerable to channel profile mismatch and that robust training strategies—potentially incorporating diverse or joint channel profiles—are critical for reliable deployment.

IV. Conclusion

This paper investigated pilot-free semantic communication over time-varying OFDM channels. By eliminating dedicated pilot symbols, the proposed framework improves bandwidth efficiency while relying on end-to-end learning to implicitly adapt to channel variations. Performance was evaluated under standardized 3GPP fading models (EPA, EVA, ETU). The results show that, under matched training and testing conditions, pilot-free semantic communication achieves strong reconstruction quality even with short block length, limited bandwidth, and no explicit CSI. However, the system is sensitive to channel profile mismatch: the EPA-trained model exhibits the most severe degradation when tested on different deployment channels, whereas the ETU-trained model demonstrates the strongest robustness. These findings highlight both the potential and the vulnerability of pilot-free semantic communication. Future work will focus on robust training strategies that incorporate diverse or mixed channel profiles, as well as adaptive transceivers capable of generalizing to unseen propagation environments. Promising directions include mixture-of-experts–based architectures that dynamically activate specialized sub-models for different fading conditions, and meta-learning techniques that enable rapid adaptation to previously unseen channels.

References

[1] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, May 2019.

[2] Z. Cao, H. Zhang, L. Liang, and H. Wang, "Pilot-free semantic communication systems for frequency-selective fading channels," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2023, pp. 1860–1865.

[3] J. Park, H. Kim, J. Shin, Y. Oh, and Y.-S. Jeon, "End-to-end training and adaptive transmission for OFDM-based semantic communication," *ICT Express*, 2025.

[4] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Relay radio transmission and reception," 3rd Generation Partnership Project (3GPP), TS TS 36.133 V10.4.0, Apr. 2015.

[5] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10 012–10 022.