Generating Underwater Acoustic Communication Channel Impulse Responses Using A Diffusion Model

Yongjie Zhuang – *Stony Brook University*

Dariush Kari – *University of Illinois Urbana-Champaign*

Zhengnan Li – *Northeastern University*

Milica Stojanovic – *Northeastern University*

Andrew C. Singer – *Stony Brook University*

Funded by ONR

When large numbers of channel realizations are required

Experimental data:

- **Expensive to collect**
- **Captures realistic channel characteristics**

Physics-based model:

- **Easy to simulate**
- **No standard models that accurately capture channel characteristics**

We need models that can generate channel responses capturing characteristics such as multipath, Doppler, and spatiotemporal variability

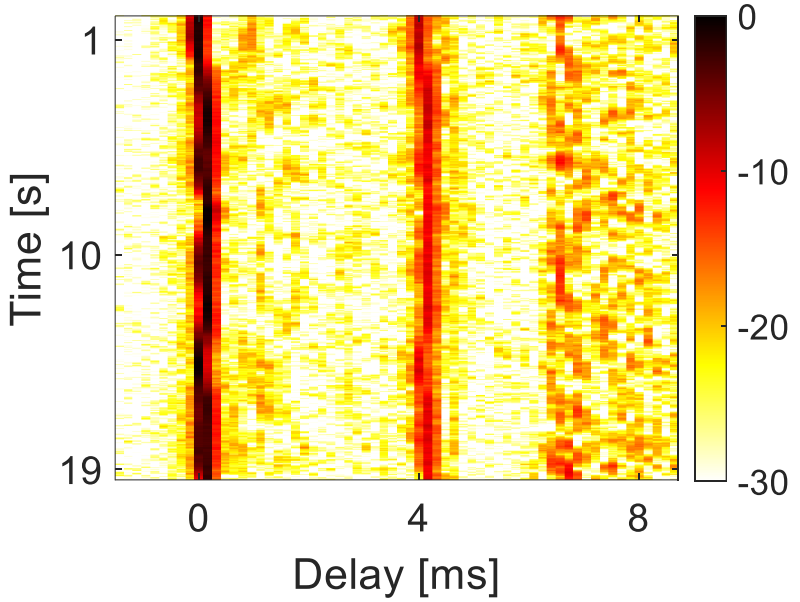
What do we have? → **Measured channel models collected over several decades**

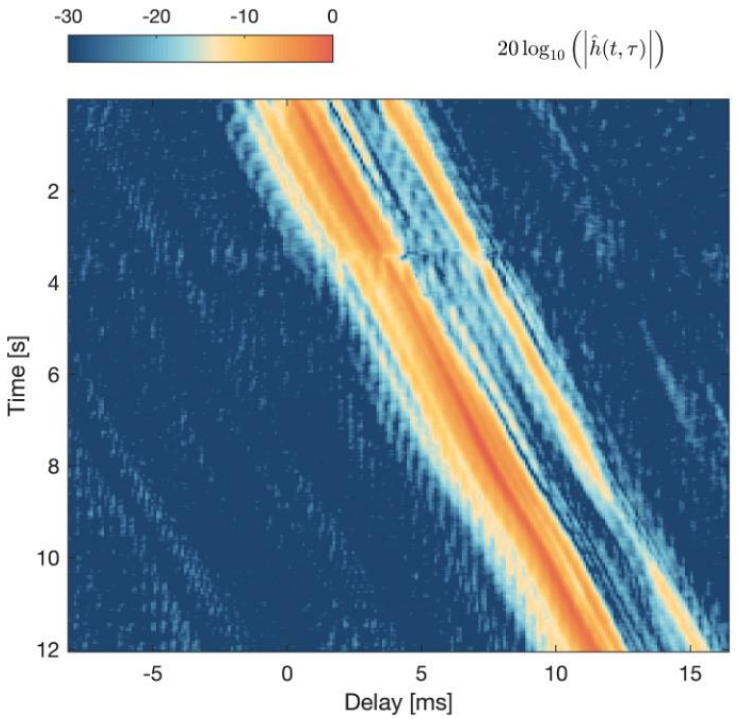
UWA-Channels:

A repository of standard channel models

This is a repository we are building to provide publicly available channel responses measured from a diverse collection of environments, using decades of field data:

- Kauai Acomms MURI 2011 (KAM'11)
- Mobile Acoustic Communication Experiment (MACE'10)
- Surface Processes and Acoustic Communications Experiment (SPACE'08)
- Rescheduled Acoustic Communications Experiment (RACE'08) ...





Diffusion models are powerful tools for augmenting datasets

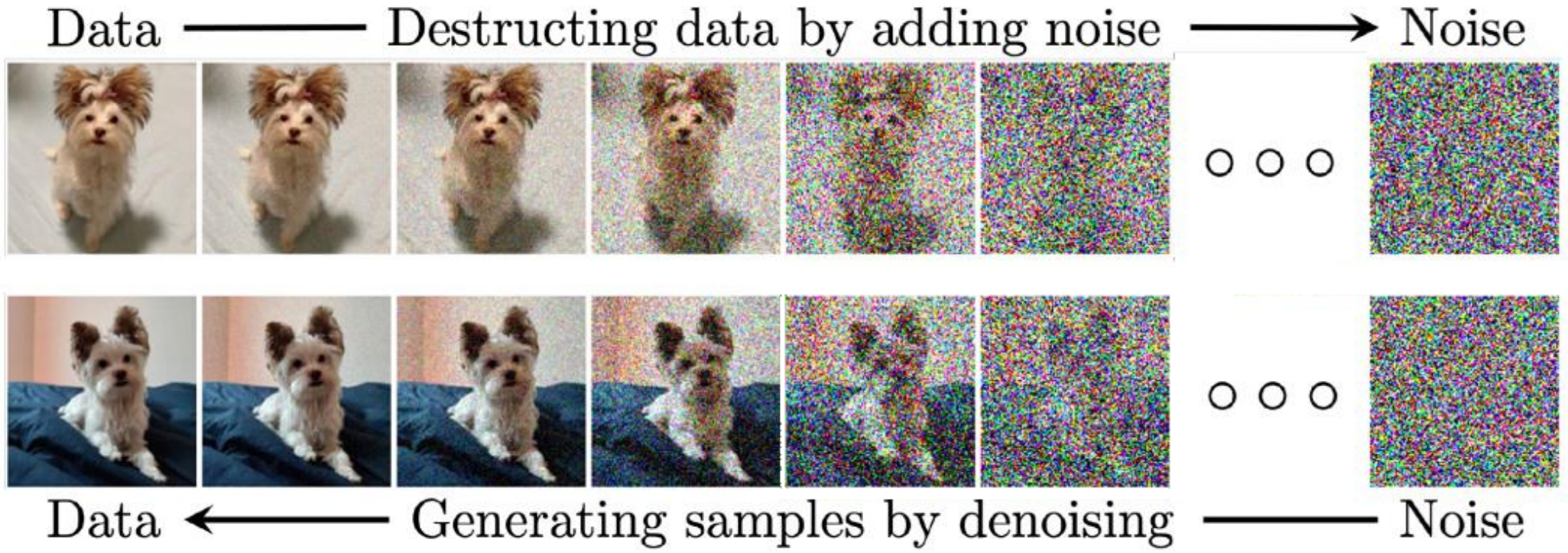


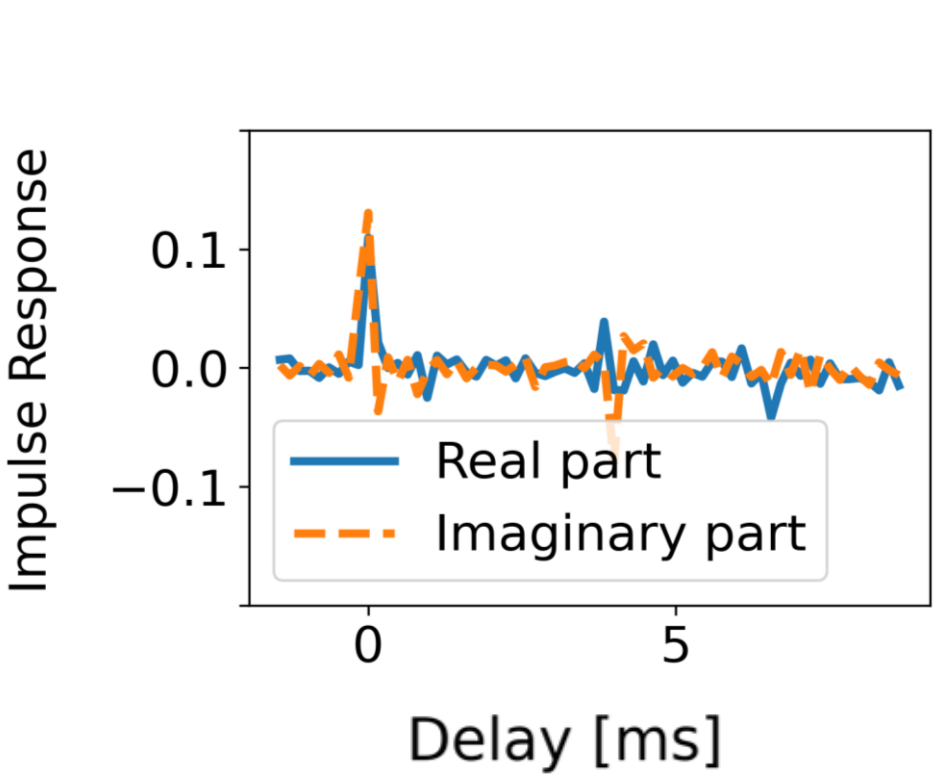
Figure cited from Yang, Ling, et al. "Diffusion models: A comprehensive survey of methods and applications." *ACM Computing Surveys* 56.4 (2023): 1-39.

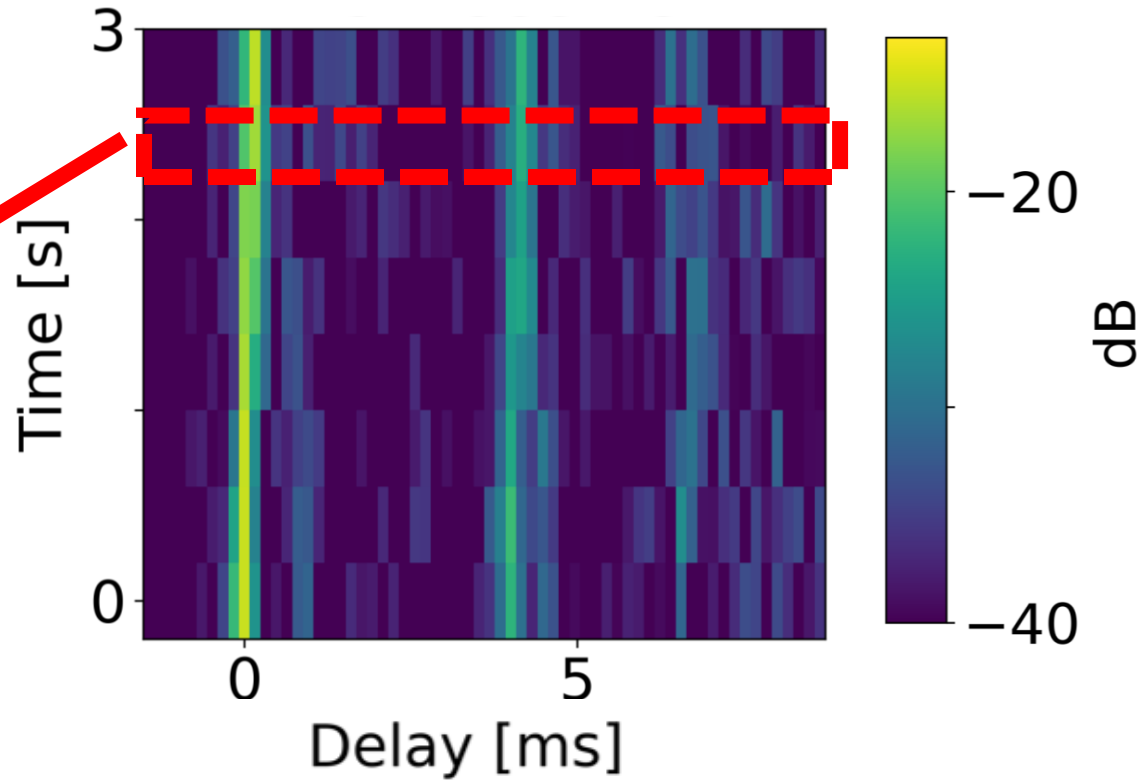
Using the KAM'11 dataset to train a diffusion model

Time-varying baseband channel impulse responses are obtained:

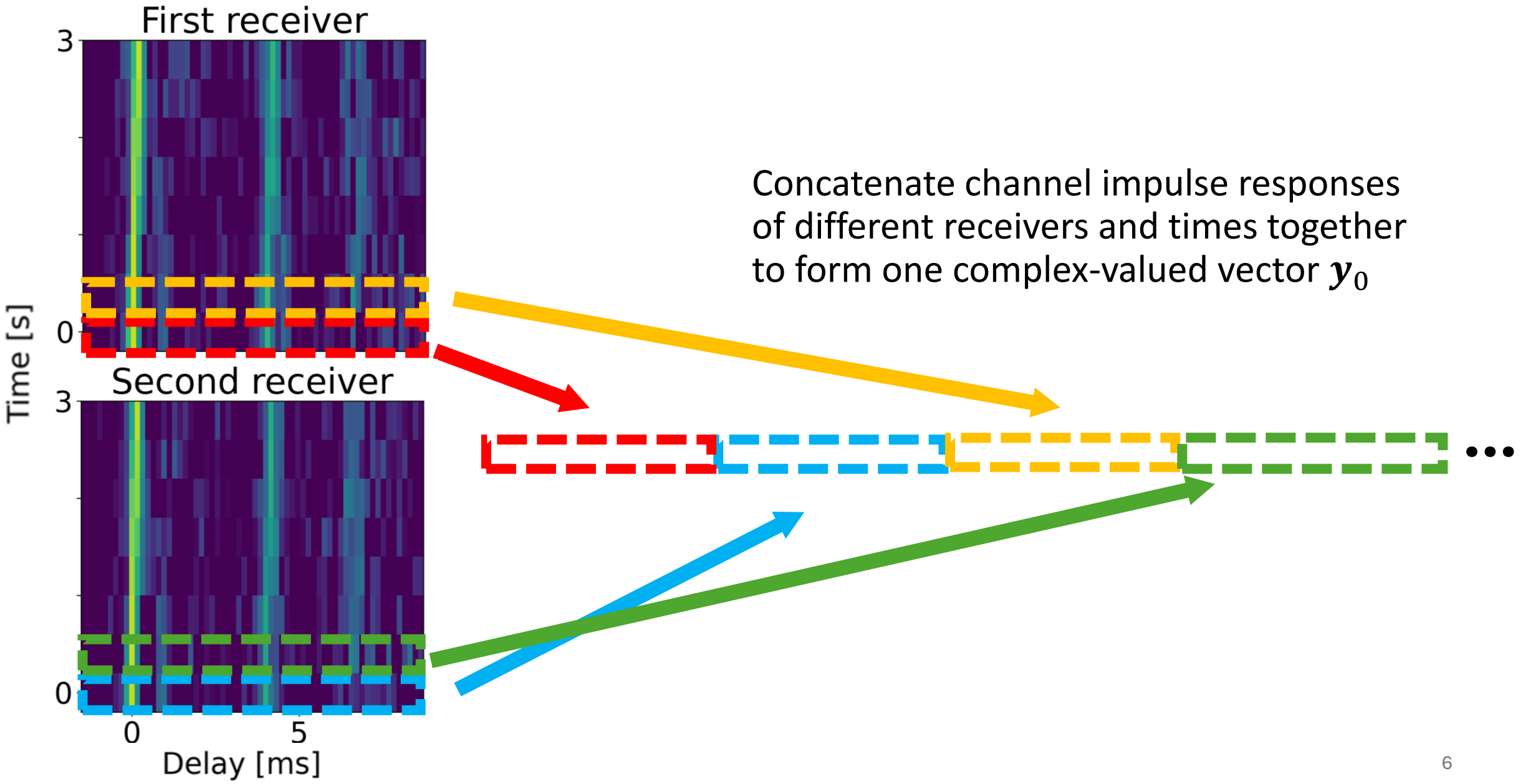
- Two hydrophones and 3.2 s time span are considered for spatiotemporal variability
- Sampling rate: 6.25 kHz
- Filter length 64-point (10.24 ms)

One example of measured response



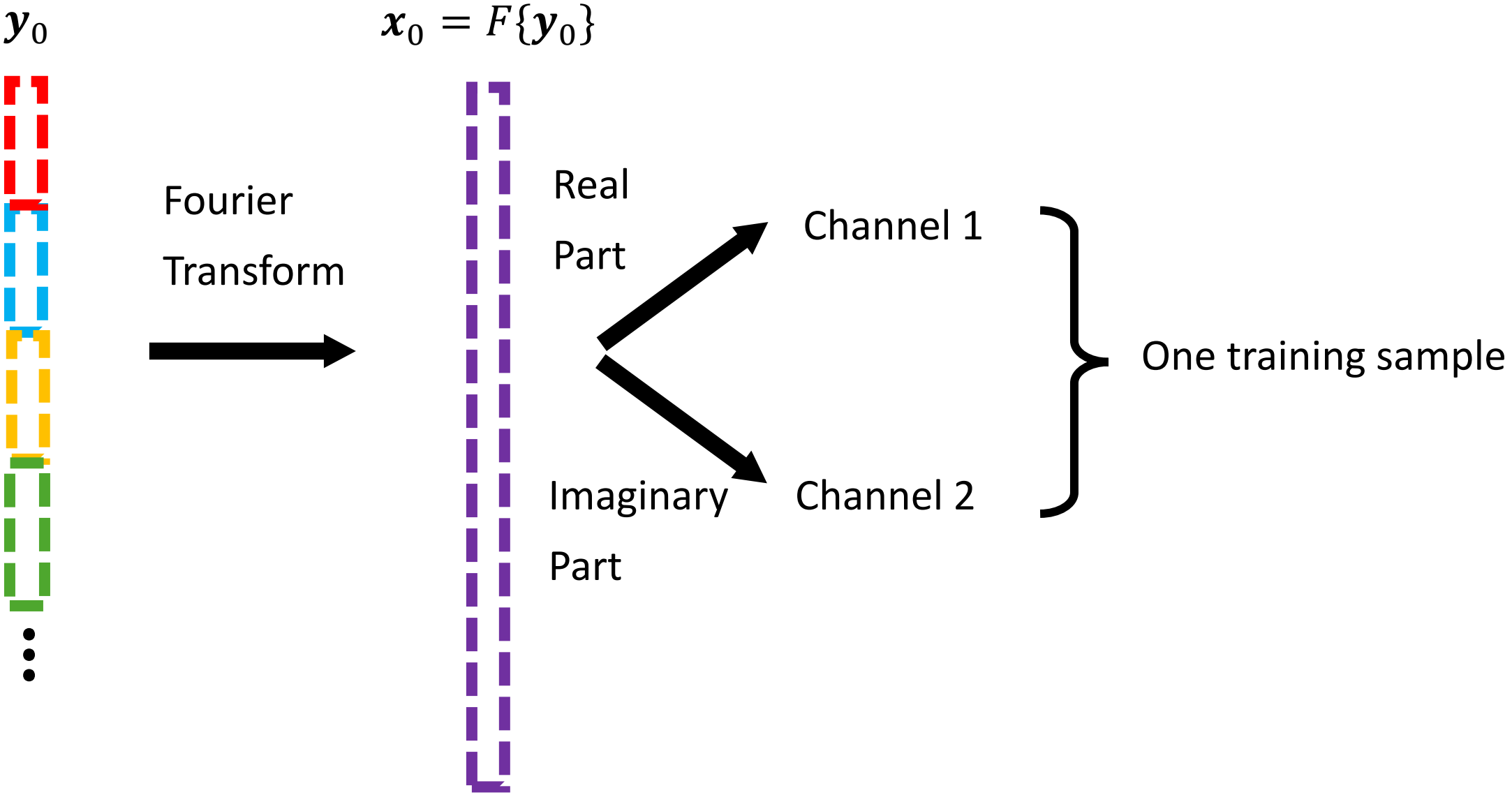


Construct one training sample

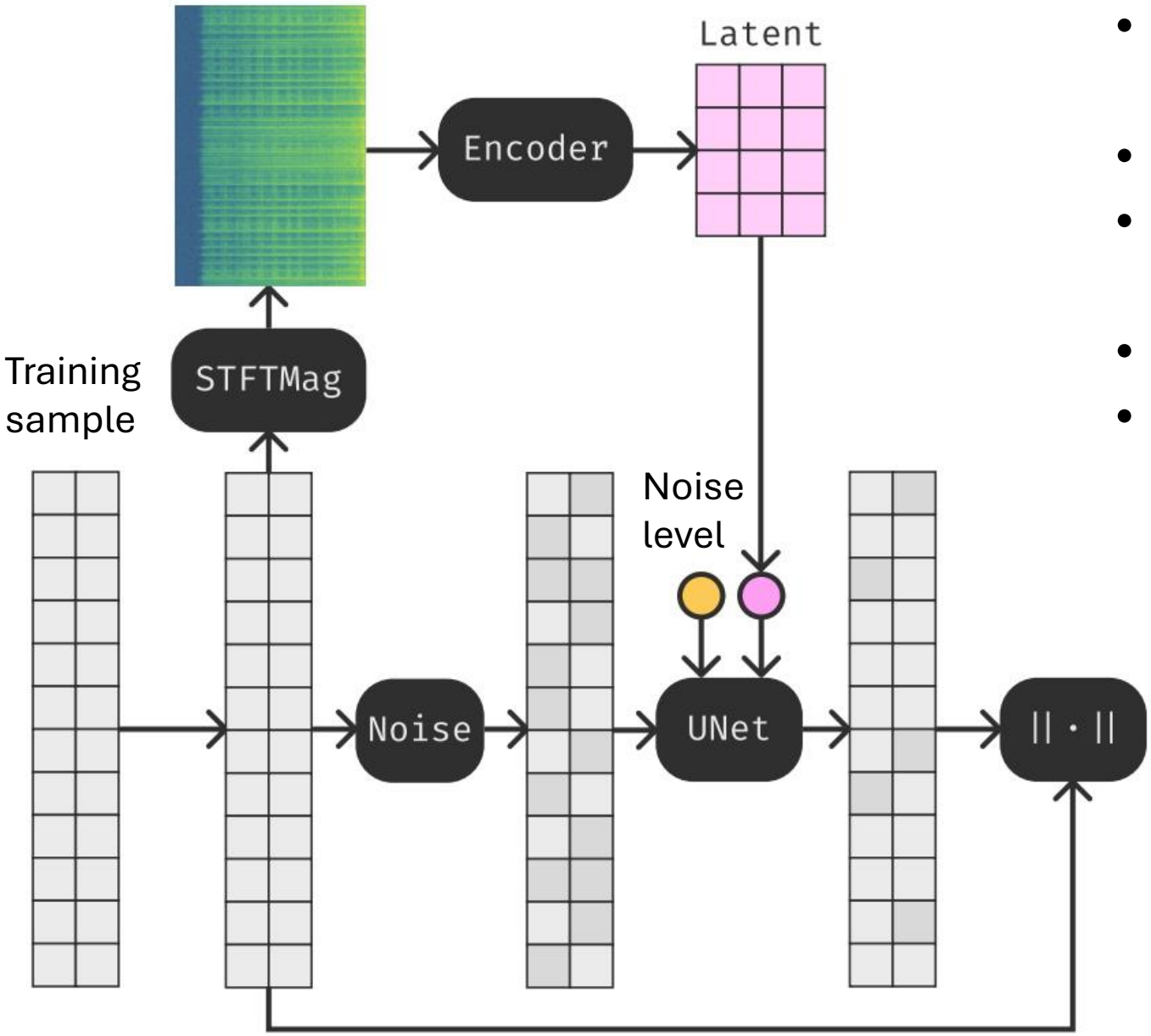


Concatenate channel impulse responses of different receivers and times together to form one complex-valued vector $\boldsymbol{y}_0$

Construct one training sample



Diffusion models developed based on audio were used



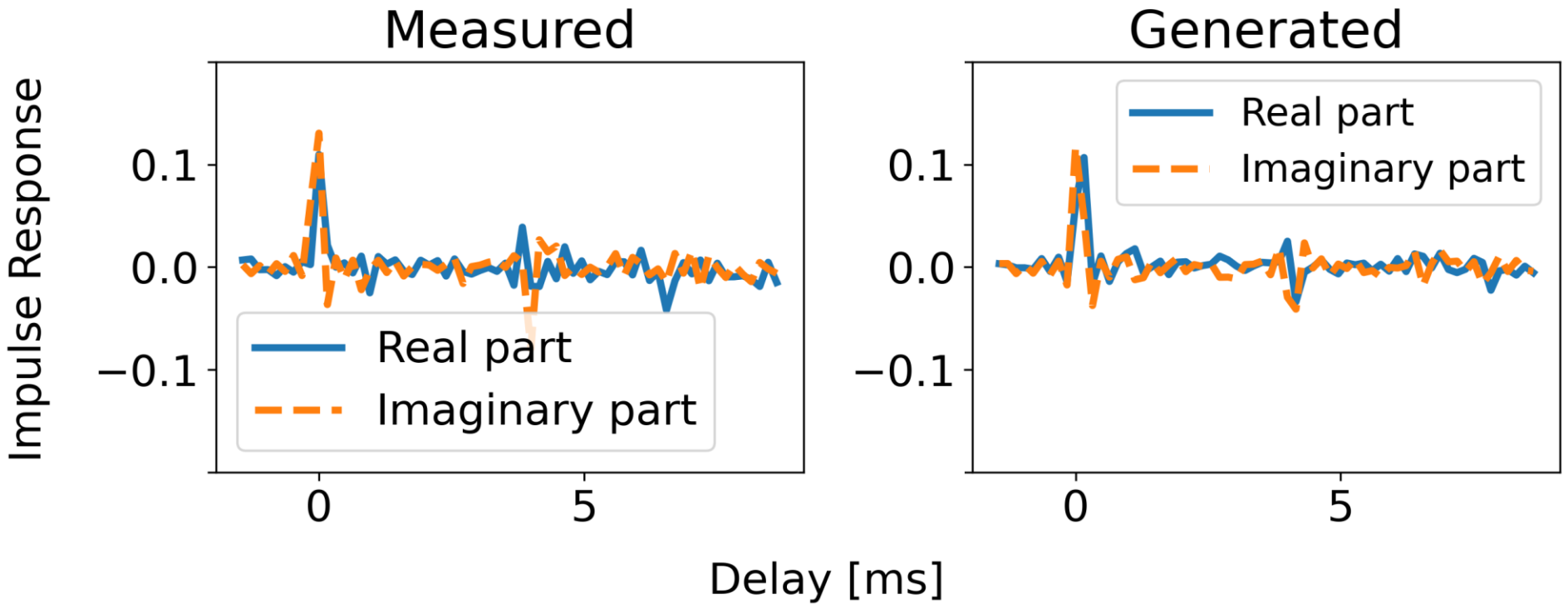
- 7 nested U-Net blocks with 32 channels in each block for diffusion models
- No pre-training networks were used
- 200 training samples were randomly selected for each training iteration
- Pytorch Adam is used
- 20 minutes of training time using NVIDIA RTX A6000 (18000 iterations)

Figure cited from Schneider, Flavio, et al. "Moûsai: Efficient text-to-music diffusion models." *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers). 2024.

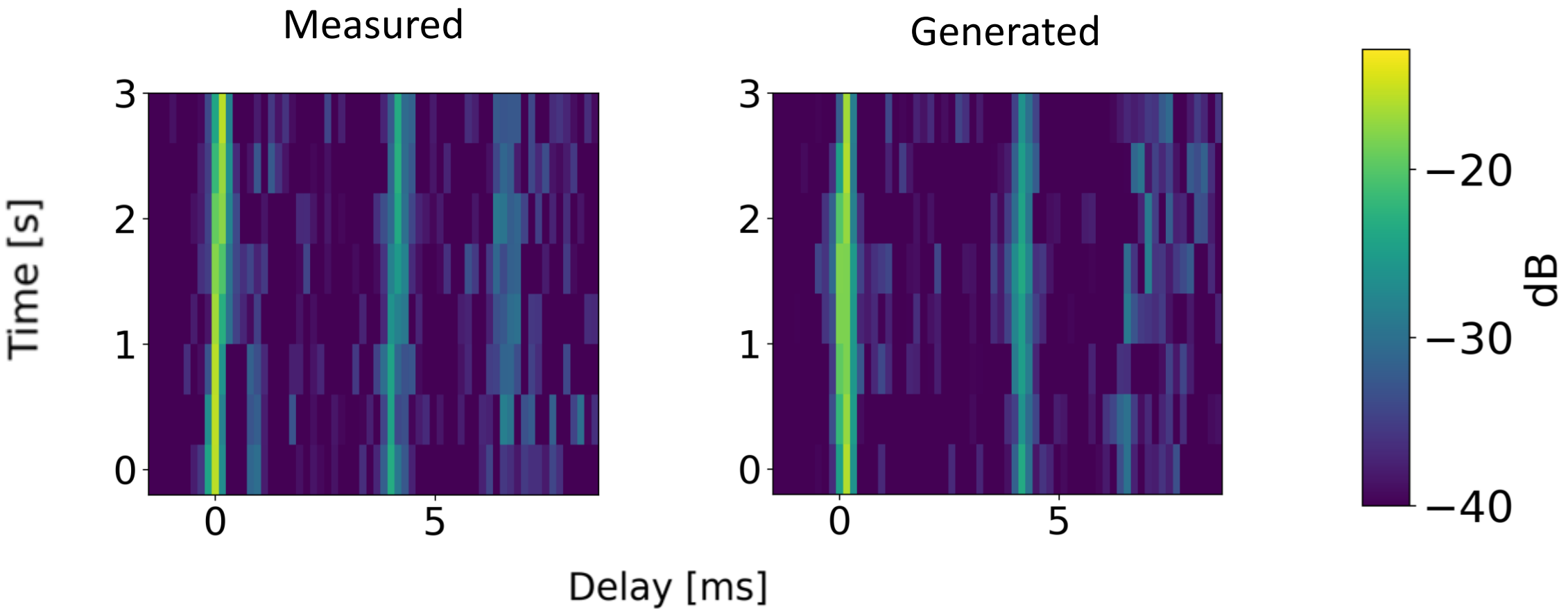
Trained diffusion model can be used for data generation

- Use white Gaussian noise as input, we can generate realistic channel impulse responses

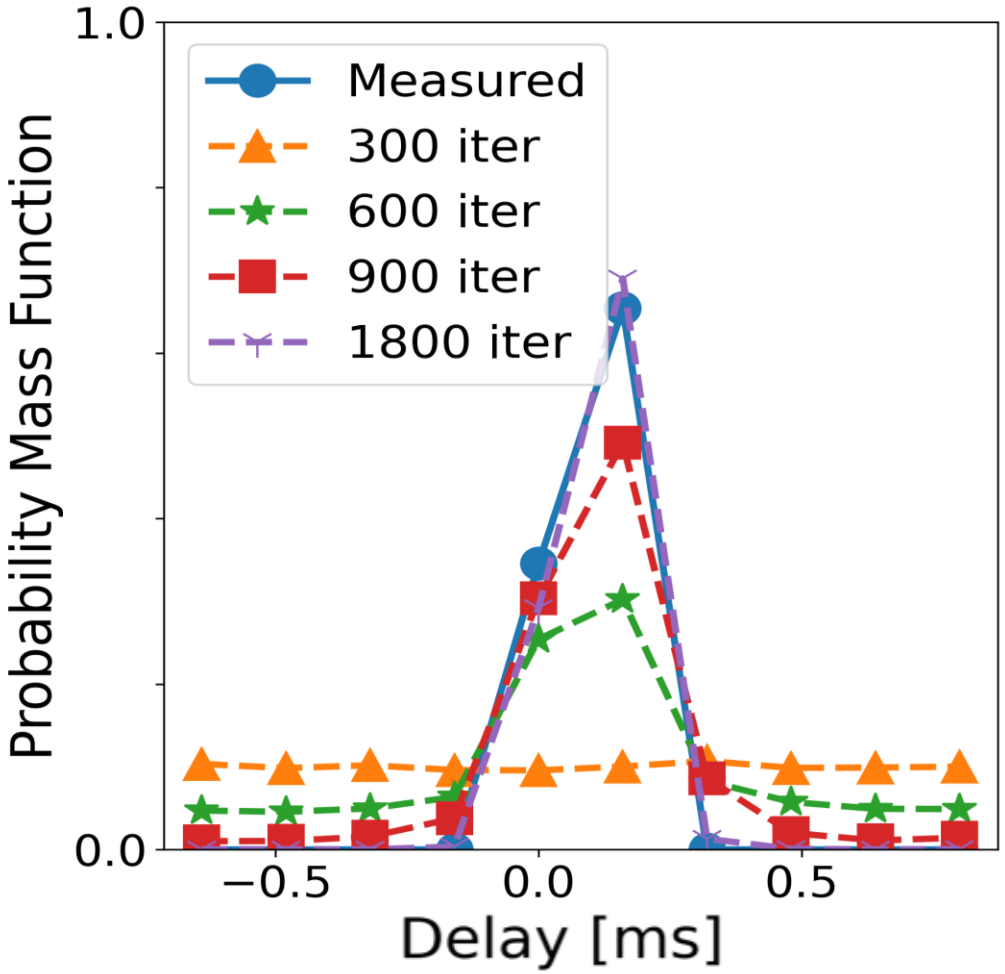
One realization of one channel impulse response:



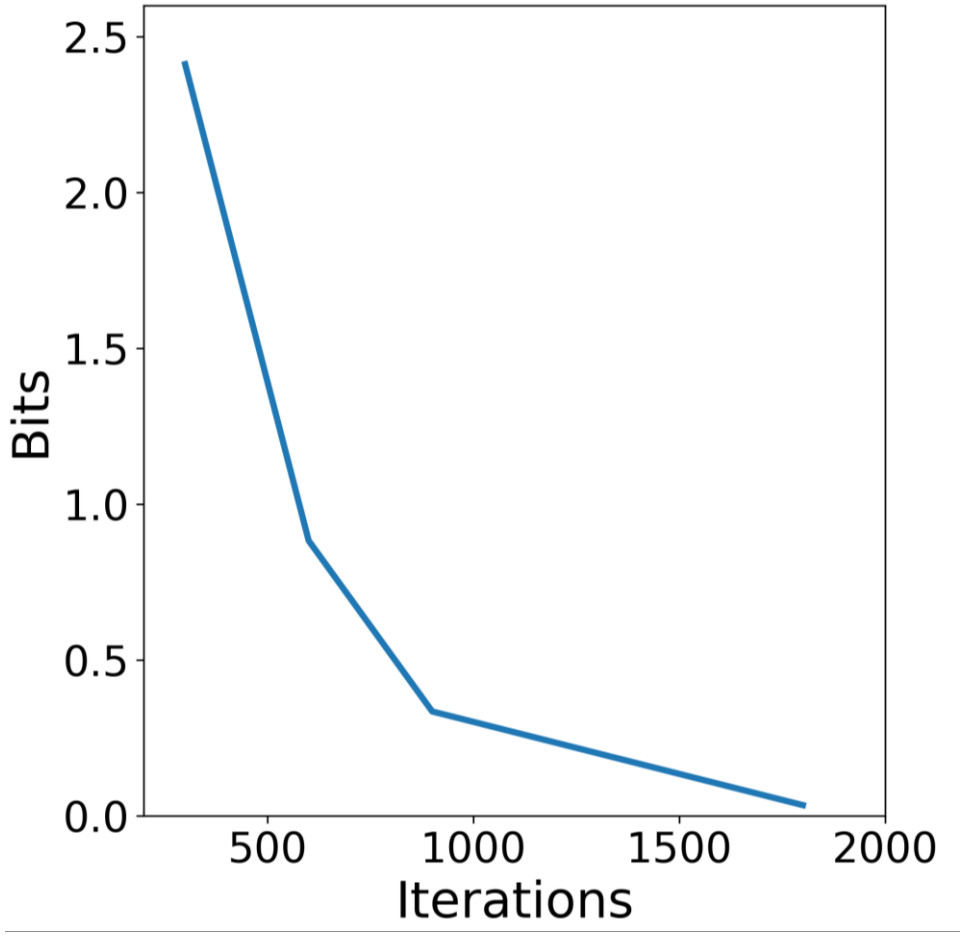
One realization of temporally correlated eight responses



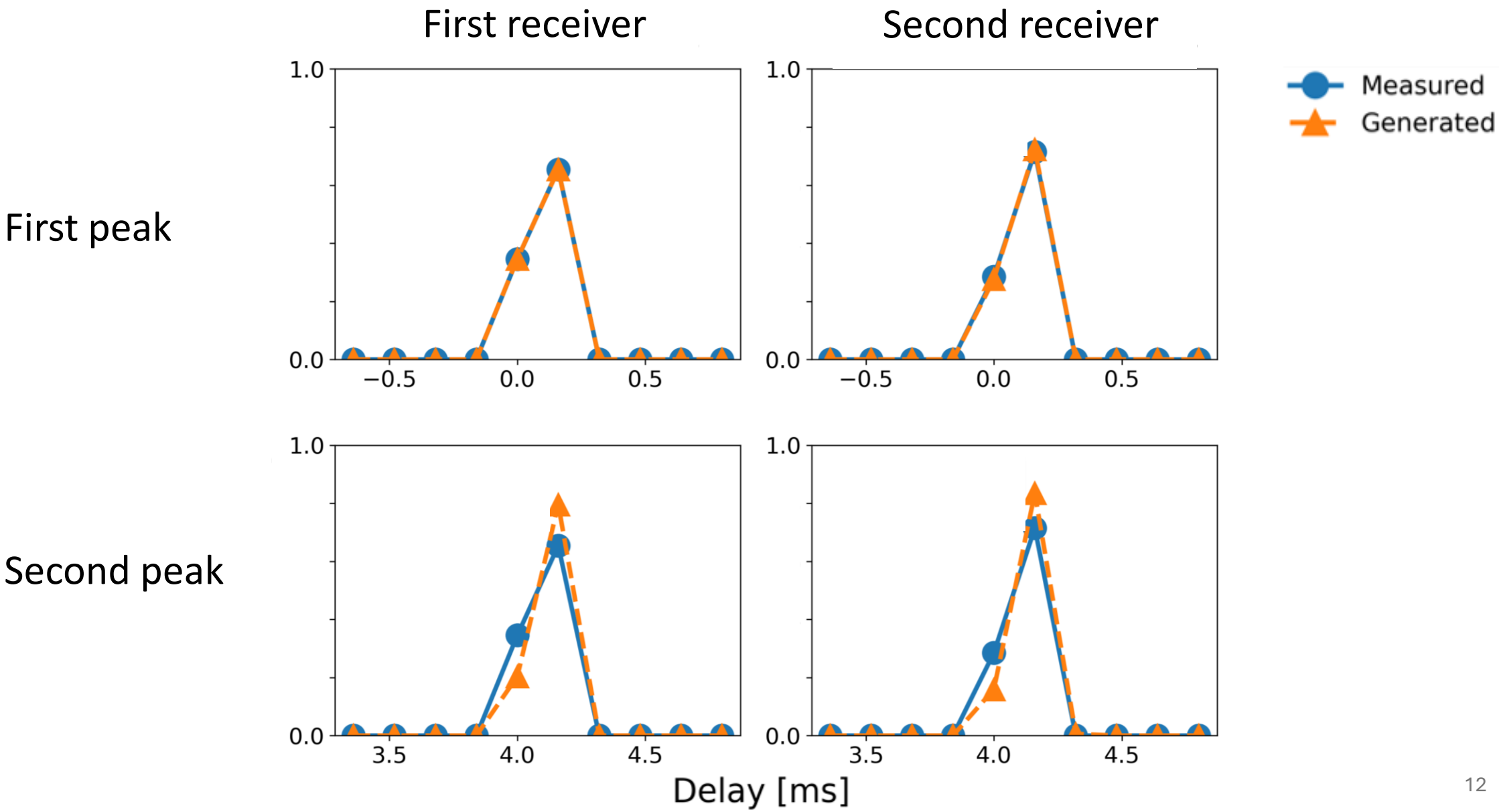
Probability Mass Function (histogram) of arrival time of the 1st peak



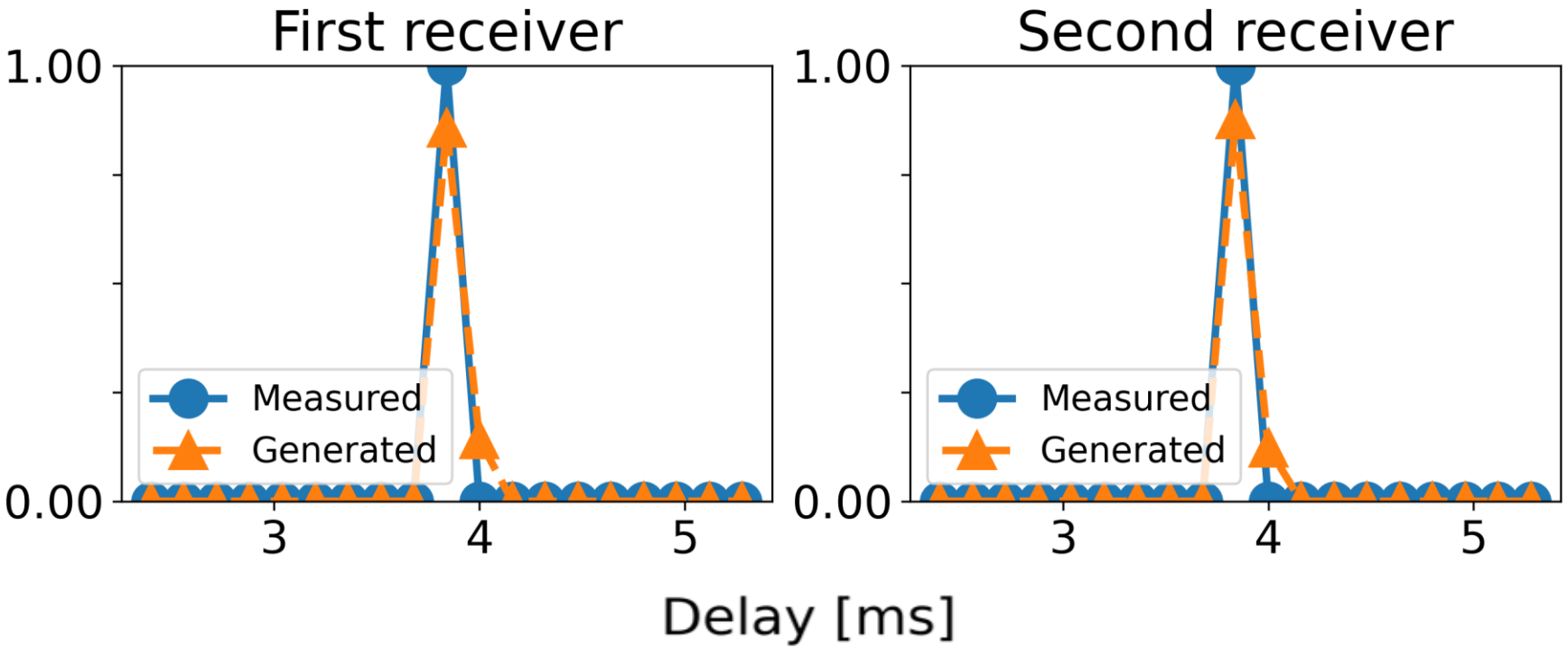
KL divergence between measured and generated responses



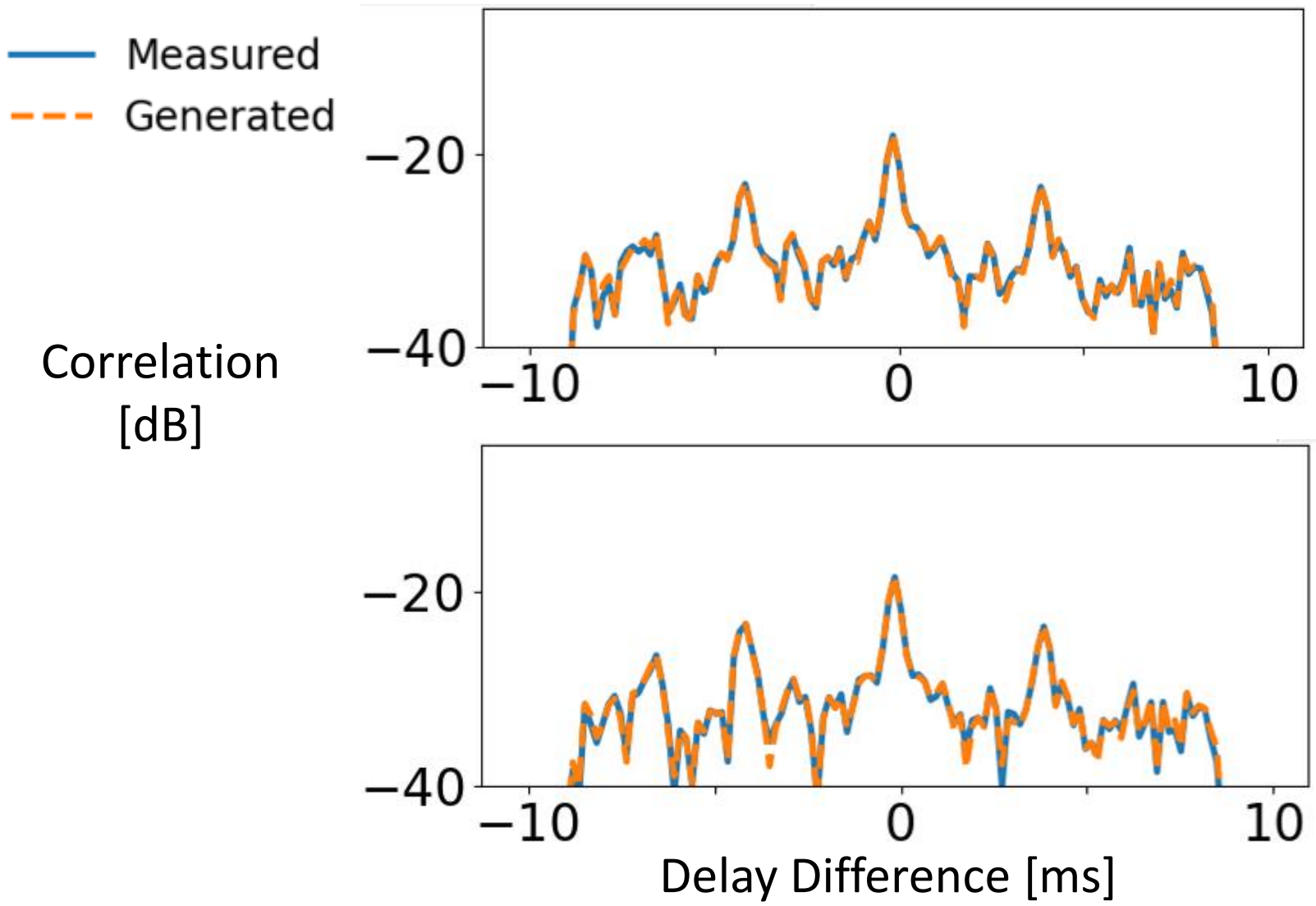
Probability Mass Function (histogram) of arrival time of the all peaks



Probability Mass Function (histogram) of arrival time difference of the 1st and 2nd peaks can show the diffusion model learns the correlation between different peaks



Generated and measured channel responses have similar spatiotemporal correlation – some examples

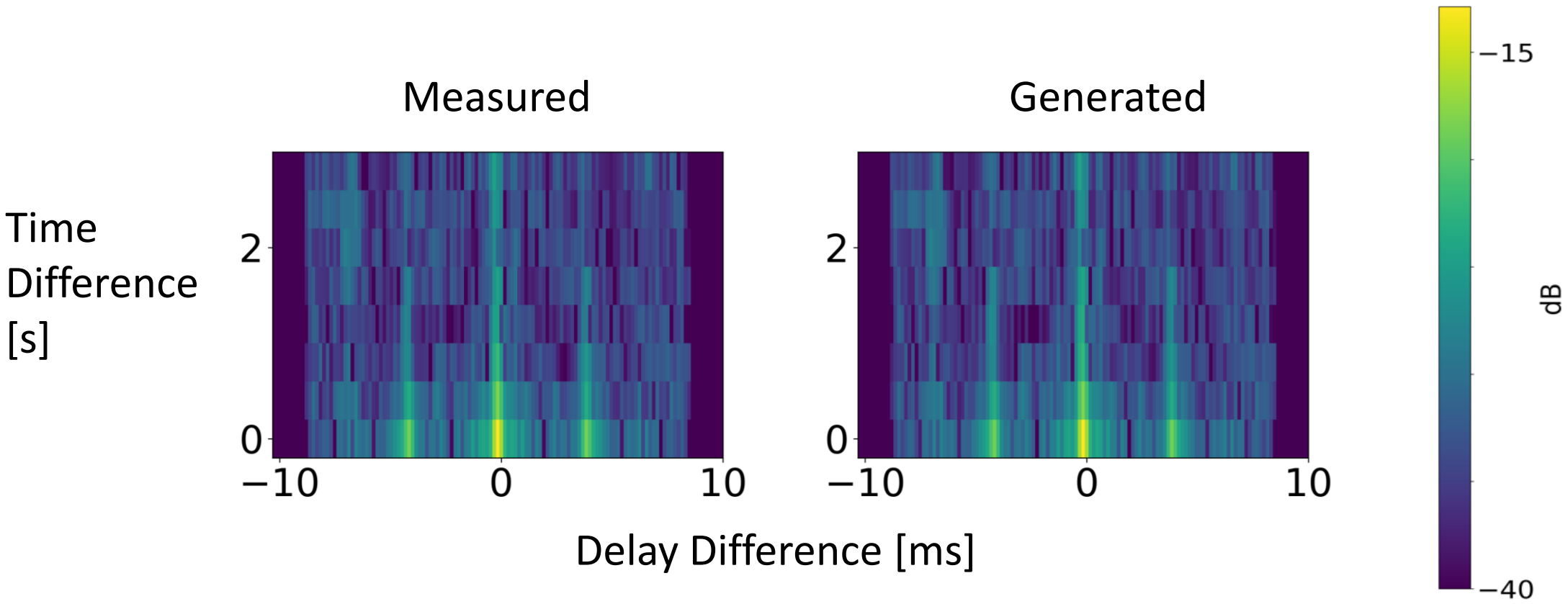


Magnitude of correlation at time difference 0.4 s:

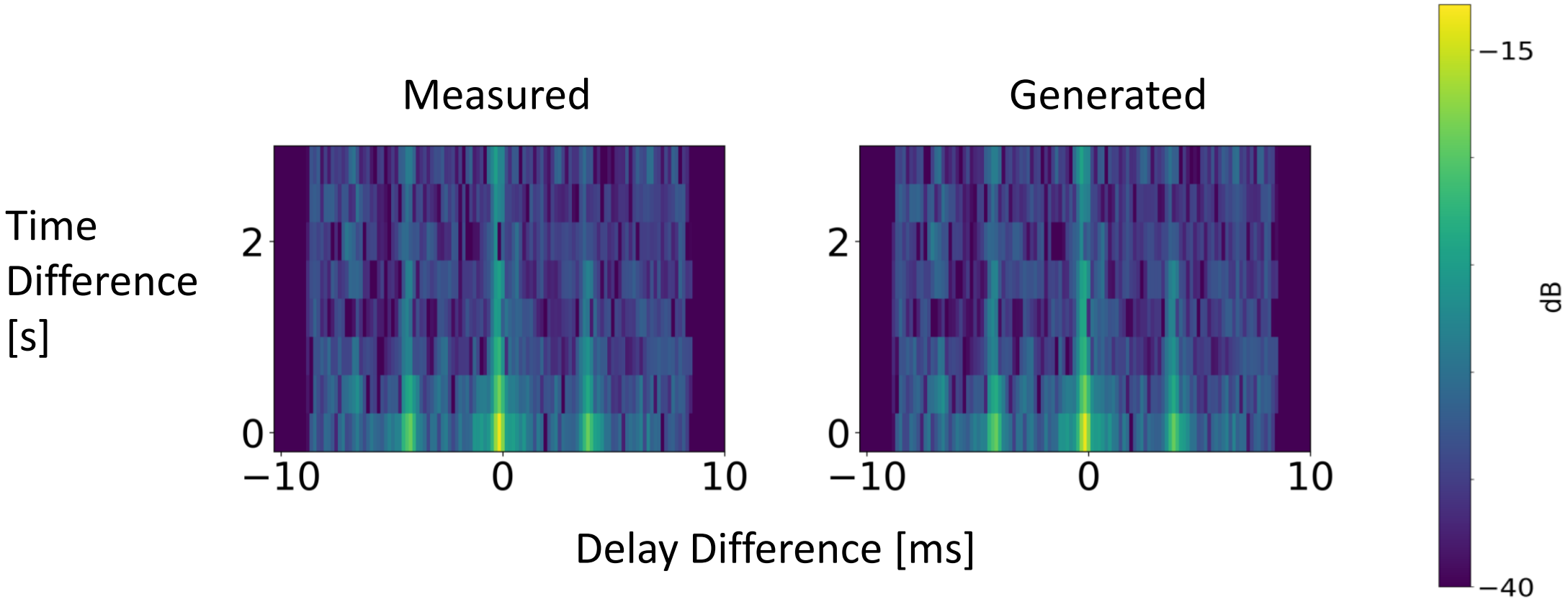
Of the same receiver location

Cross different receiver locations

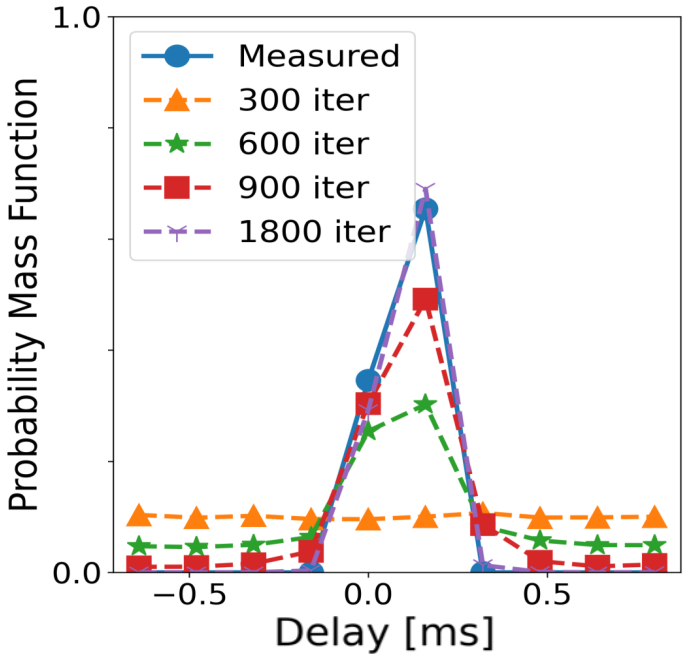
Generated and measured channel responses have similar temporal correlation – for the same receiver

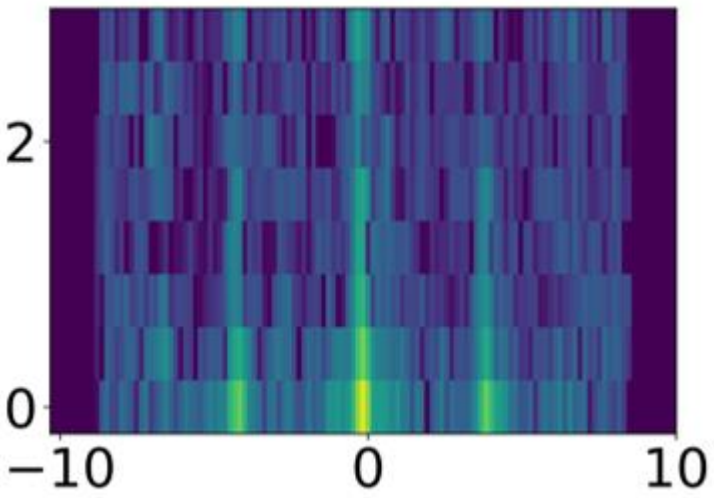


Generated and measured channel responses have similar temporal correlation – cross two receivers



Conclusion





- A deep generative method, diffusion model, is used to generate realistic channel impulse responses.
- The probability mass function (histogram) of the generated impulse responses matches with the measured responses:
  - For the arrival times of peaks
  - For the difference of arrival times between 1st and 2nd peaks
- The generated impulse responses have similar spatiotemporal correlation patterns compared with the measured responses