6. APPENDIX

6.1. Computing Infrastructure

All experiments are conducted by CPU/GPU simulation. The simulation experiments are conducted on a computing server with eight GPUs. The server is equipped with AMD EPYC 7502 32-Core Processor and 1024G memory. The GPU is NVIDIA RTX A4000.

6.2. Hyperparameter Settings

We searched for the client learning rate in a range from 10^{-6} to 10^{0}, the server learning rate in a range from 10^{-4} to 10^{0}, and input batch size in a range from 5 to 30, and total training round in a range from 50 to 5000. After hyperparameter searching, we fixed the batch size for all datasets to 16 and the local epoch to 1 for all experiments.

6.2.1. Training with clean signals

When we train the model with clean signals, we select both FedAvg and FedOpt as the server aggregator functions. For FedOpt, we select the ADAM as the server optimizer. The other hyperparameters are shown in Table 6.

Table 6. Hyperparameters for training FL models with clean audio signals.

Dataset	Agg.	Sample rate	lr	Server lr	Round
Google Command	FedAvg	5%	0.1	-	5000
		10%	0.3	-	5000
		20%	0.2	-	5000
	FedOpt	5%	0.05	0.001	5000
		10%	0.01	0.001	5000
		20%	0.01	0.001	5000
IEMOCAP	FedAvg	100%	0.01	-	200
	FedOpt	100%	0.01	0.001	50
Crema-D	FedAvg	10%	0.1	-	200
		30%	0.1	-	200
		50%	0.1	-	200
	FedOpt	10%	0.1	0.001	200
		30%	0.1	0.001	200
		50%	0.1	0.001	200
Urban Sound	FedAvg	20%	0.075	-	300
		50%	0.075	-	300
	FedOpt	20%	0.1	0.001	300
		50%	0.1	0.001	300

6.2.2. Training with SNR noisy signals

When we train the model with SNR noisy signals, we select FedAvg as the server aggregator function. For Google command related experiments, the client learning rate is 0.1 and the communication round is 5000. For IEMOCAP related experiments, the client learning rate is 0.01 and the communication round is 200. For CREMA-D related experiments, the client learning rate is 0.1 and the communication round is 200. Finally, for Urban Sound, the client learning rate is 0.075 and the communication round is 300.

6.2.3. Training with label noisy signals

In this section, we select FedAvg as the server aggregator function. For Google command related experiments, the client learning rate is 0.1 and the communication round is 5000. For IEMOCAP related experiments, the client learning rate is 0.01 and the communication round is 200. For CREMA-D related experiments, the client learning rate is 0.1 and the communication round is 200. Finally, for Urban Sound, the client learning rate is 0.075, and the communication round is 300.