Supplementary Information

The α and T values we used for the public datasets.

* The α and T values were kept consistent across all folds within each model and sampling combination during the evaluation for the ESC-50 dataset. We used a grid search from 0.1 to 0.5 to explore the optimal α, and for T, it ranged from 1 to 8. It was observed that α values between 0.6 and 0.8 resulted in similar distillation outcomes as the range between 0.1 and 0.5.

* The model training utilized a single NVIDIA GeForce RTX 3080 Ti GPU. In our study, we applied the same random seed throughout the entire study to ensure comparable and reproducible results across baseline and distilled models as well as different tested cases / folds:

```
torch.manual_seed(0)

numpy.random.seed(0)

random.seed(1)

train_loader = torch.utils.data.DataLoader(*other agrs,

shuffle = True, generator=torch.Generator().manual_seed(0))
```

* Early-stopping was conducted based on the target metrics – macro-averaged class accuracy for ESC-50 / TAU-2019-Mobile, and macro F1 score for user data.

ESC-50 (following the default 5-fold split of the dataset)

	T-S 2k-2k	T-S 16k-2k	T-S 1k-1k	T-S 16k-1k
CNN14	0.5,2	0.3,4	0.4,2	0.4,5
ResNet38	0.4,2	0.2,5	0.3,3	0.3,5
MobileNetV2	0.5,3	0.5,6	0.2,2	0.4,8

TAU-2019-Mobile (following the training-validation-test split)

	T-S 2k-2k	T-S 16k-2k	T-S 1k-1k	T-S 16k-1k
CNN14	0.5,8	0.5,3	0.4,2	0.5,5
ResNet38	0.5,1	0.5,2	0.5,5	0.4,5
MobileNetV2	0.5,5	0.3,6	0.5,6	0.5,6

Notation: T: the teacher generation, S: the student generation.