Technical Disclosure Commons

March 2020

Spatially and Temporally Directed Noise Cancellation Using Federated Learning

Shantanu Goel

Piyush T. Itankar

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Spatially and Temporally Directed Noise Cancellation Using Federated Learning

ABSTRACT

Machine learning models can be trained to cancel noise of diverse types or spectral characteristics, e.g. traffic noise, background chatter, etc. Such models are trained by feeding training data that includes labeled noise waveforms, which is an expensive and time-consuming procedure. Further, the effectiveness of such machine learning models is limited in canceling types of noise absent from training data. Trained models occupy significant amounts of memory which limits their use in consumer devices. This disclosure describes the use of federated learning techniques to train noise canceling models locally at diverse device locations and times. With user permission, the trained models are tagged with timestamp and location, such that when a user device has time or location matching a particular noise cancellation model, the particular model is provided to the user device. Noise cancellation on the user device is then performed with a compact machine learning model that is suited to the time and location of the user device.

KEYWORDS

- Noise cancellation
- Noise waveform
- Machine learning
- Headphones
- Federated learning
- Context-based learning
- Unified ML model
- Geolocation

BACKGROUND

Traditional noise cancellation, e.g., performed during phone calls, music playback, during flights or other travel, etc., uses adaptive arrays of microphones to detect noise with specific directional or spectral patterns and nulls out the detected noise. Such noise cancellation schemes have several drawbacks such as: requirement of an array of microphones and related hardware, which is costly and may not fit in the tight spaces of certain applications such as mobile devices; infeasible for use in certain contexts such as earbuds; added power consumption to drive the additional hardware; etc. In some cases, e.g., for noise cancellation during playback or public-address, additional microphones (and their related costs and power consumption) are needed not for accepting audio input but to measure ambient noise.

One approach to noise cancellation is to train machine learning (ML) models to cancel noise of diverse types or spectral characteristics, e.g. traffic noise, background chatter, etc. This approach uses a low amount of power, can work without additional hardware, and is suitable for devices with tight space constraints. However, the effectiveness of such machine learning models is limited in canceling types of noise absent from training data. Also, such models are trained by feeding labeled noise waveforms, which is an expensive and time-consuming procedure. Further, the resulting model, which encompasses a wide variety of noise types, can impose significant memory requirements, which may not always be met in consumer devices.
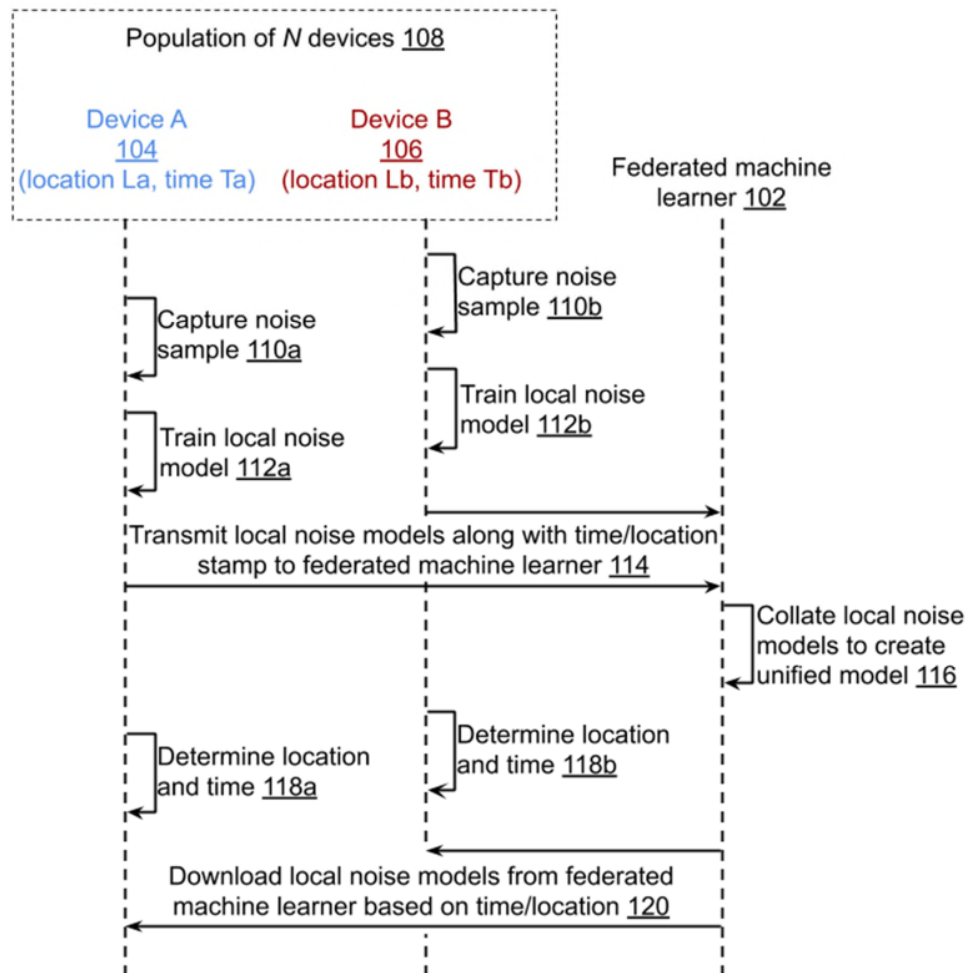
DESCRIPTION

Fig. 1: Spatially and temporally directed noise cancellation using federated learning

Fig. 1 illustrates spatially and temporally directed noise cancellation using federated learning, per techniques of this disclosure. With the prior permission of respective users, user devices (104, 106) within a population of devices (108) capture samples of noise waveforms (110a, 110b). In devices already equipped with noise cancelers, e.g., array-based or ML-based noise cancelers, noise waveforms can be captured as the output waveform of the noise canceler. In devices without existing noise cancelers, noise waveforms can be captured with user

permission as the ambient, e.g., environmental, sound present when the user is not in an active voice call.

The captured noise waveforms are used to train noise-canceling ML models (112a, 112b) locally on-device. Local training of noise-canceling models can be performed when the device is plugged in and idle, such that the power consumption and performance impact is minimal. With user permission, respective locally-trained ML models along with a tag indicating the time and location of noise-waveform capture (114) are transmitted to a federated machine learner (102). At the federated machine learner, which can be implemented on a central server, noise-cancellation models received from multiple devices across the population of devices (116) are collated to generate a unified model.

The noise-cancellation techniques disclosed herein leverage the observation that noise waveforms have particular spectral patterns at a given location and time. For example, at a street intersection at 9 AM on a weekday, the main background noise is that of traffic, e.g., automobile engines, honking, etc. On a Saturday evening in a downtown street, the main background noise is that of people partying to music. At a sports event in a stadium, the main background noise is that of crowds cheering.

A device in the field that deploys noise cancellation determines the present time and its location (118a, 118b). The device auto-downloads an up to date noise-cancellation model based on the time and device location (120). Noise cancellation is performed on the device with a compact machine learning model that is appropriate to the temporal and spatial parameters of the device.

With user permission, downloading of the noise-cancellation model can also be based on parameters such as the calendar of a user associated with the device, upcoming travel, etc. such

that noise cancellation models are prefetched and/or switched based on anticipated location of the device. For example, if it is known that the user is scheduled to be at the airport at a certain time, a noise cancellation model appropriate to the airport (and to the route thereto) can be prefetched. In this manner, noise cancellation models are downloaded ahead of time, rather than just in time.

The location tag of the noise cancellation model can be fine or coarse-grained depending on factors such as whether the device is indoors or outdoors; the technique used for geolocation, e.g., WiFi, GPS, etc.; the noise environment; user preferences, e.g., offline map caching; etc.

The techniques of this disclosure can provide noise cancellation that is localized and optimized to the time and immediate environment of the user. Compared to conventional techniques, noise cancellation can be performed with lower hardware complexity, reduced cost, and smaller size. Noise cancellation models are obtained that are lightweight enough to run ubiquitously, e.g., on any device such as smartphones, smart speakers, video-conferencing equipment, etc. The techniques apply to noise cancellation at the transmitting as well as the receiving end of voice calls, to playback on devices when a communicating end does not have noise cancellation support, etc.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user device location and time), and if the user is sent content or communications (e.g., trained noise cancellation models) from a server. In addition, certain data are treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity is treated so that no personally identifiable information can be determined for the user, or a user's geographic

location may be generalized where location information is obtained, so that a particular location of a user cannot be determined. Thus, the user has control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

   This disclosure describes the use of federated learning techniques to train noise canceling models locally at diverse device locations and times. With user permission, the trained models are tagged with timestamp and location, such that when a user device has time or location matching a particular noise cancellation model, the particular model is provided to the user device. Noise cancellation on the user device is then performed with a compact machine learning model that is suited to the time and location of the user device.

REFERENCES

[1] "RNNoise: Learning Noise Suppression" https://people.xiph.org/~jm/demo/rnnoise/ accessed on Jan. 10, 2020

[2] Valin, Jean-Marc. "A hybrid DSP/deep learning approach to real-time full-band speech enhancement." In 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), pp. 1-5. IEEE, 2018

[3] "Federated Learning" https://federated.withgoogle.com accessed on Jan. 10, 2020

[4] Kristjansson, Trausti and Matthew I. Lloyd. "Noise compensation using geotagged audio signals." European Patent Application 14162078.1, filed Mar. 22, 2011.

[5] Ang, Fan, Li Chen, Nan Zhao, Yunfei Chen, Weidong Wang, and F. Richard Yu. "Robust Federated Learning with Noisy Communication." *arXiv preprint arXiv:1911.00251* (2019).