

CLAIMS

What is claimed is:

1. A computer-implemented method comprising:
 - obtaining, by a computer, a plurality of training audio signals including one or more lower-bandwidth audio signals having a first bandwidth and one or more corresponding higher-bandwidth audio signals having a second bandwidth, wherein the first bandwidth is comparatively lower than the second bandwidth;
 - training, by the computer, a bandwidth expander comprising a set of one or more neural network layers of a neural network, the bandwidth expander trained by applying the neural network on the plurality of training audio signals;
 - receiving, by the computer, an inbound audio signal having the first bandwidth; and
 - generating, by the computer, an estimated inbound audio signal having the second bandwidth by applying the bandwidth expander of the neural network on the inbound audio signal.
2. The method according to claim 1, wherein obtaining the plurality of training audio signals includes:
 - generating, by the computer, a lower-bandwidth audio signal having the first bandwidth by executing a codec program on a higher-bandwidth audio signal having the second bandwidth.
3. The method according to claim 1, wherein obtaining the plurality of training audio signals includes:
 - generating, by the computer, a simulated lower-bandwidth audio signal having a type of degradation by executing an augmentation operation for the type of degradation on a lower-bandwidth audio signal, the plurality of training audio signals further comprising the simulated lower-bandwidth audio signal, and
 - wherein the inbound audio signal has the type of degradation, whereby the estimated inbound audio signal generated by the computer is an enhanced inbound audio signal having comparatively less of the type of degradation.

4. The method according to claim 1, further comprising:
 - extracting, by the computer, one or more features from each of the training audio signals, wherein the computer applies the neural network on the one or more features of the training audio signals; and
 - extracting, by the computer, the one or more features from the inbound audio signal, wherein the computer applies the neural network on the one or more features of the inbound audio signal.
5. The method according to claim 1, wherein at least one higher-bandwidth audio signal of the plurality of training signals originated via a channel configured for the second bandwidth.
6. The method according to claim 1, wherein the computer generates the estimated inbound audio signal, in response to the computer determining that the inbound audio signal originated via a channel configured for the first bandwidth.
7. The method according to claim 1, wherein training further comprises
 - performing, by the computer, a loss function of the neural network according to a training estimated audio signal outputted by the neural network for a training audio signal, the loss function instructing the computer to update one or more hyperparameters of one or more layers of the bandwidth expander.
8. The method according to claim 1, further comprising
 - training, by the computer, a speaker recognizer comprising a second set of one or more neural network layers by applying the speaker recognizer on a plurality of second training audio signals comprising one or more clean audio signals and simulated audio signals;
 - extracting, by the computer, an enrollee voiceprint for an enrollee by applying the speaker recognizer on one or more enrollee audio signals of the enrollee;
 - extracting, by the computer, an inbound voiceprint for an inbound speaker by applying the neural network architecture to the estimated inbound audio signal; and
 - generating, by the computer, a likelihood score based upon the inbound voiceprint and the enrollee voiceprint, the likelihood score indicating a likelihood that the inbound speaker is the enrollee.
9. The method according to claim 8, further comprising executing, by the computer, one or more data augmentation operations on at least of a second training audio signal and an enrollee audio signal.

10. The method according to claim 9, wherein executing the one or more data augmentation operations includes applying the bandwidth expander on the at least of the second training audio signal and the enrollee audio signal.

11. The method according to claim 8, further comprising generating, by the computer, an estimated enrollee audio signal for the one or more enrollee audio signals by applying the bandwidth expander on an enrollee audio signal having the first bandwidth and originated via a channel configured for the first bandwidth.

12. A system comprising:

a non-transitory machine-readable memory configured to store machine-readable instructions for one or more neural networks; and

a computer comprising a processor configured to:

obtain a plurality of training audio signals including one or more lower-bandwidth audio signals having a first bandwidth and one or more corresponding higher-bandwidth audio signals having a second bandwidth, wherein the first bandwidth is comparatively lower than the second bandwidth;

train a bandwidth expander comprising a set of one or more neural network layers of a neural network, the bandwidth expander trained by applying the neural network on the plurality of training audio signals;

receive an inbound audio signal having the first bandwidth; and

generate an estimated inbound audio signal having the second bandwidth by applying the bandwidth expander of the neural network on the inbound audio signal.

13. The system according to claim 12, wherein to obtain the plurality of training audio signals the computer is further configured to:

generating, by the computer, a lower-bandwidth audio signal having the first bandwidth by executing a codec program on a higher-bandwidth audio signal having the second bandwidth.

14. The system according to claim 12, wherein to obtain the plurality of training audio signals the computer is further configured to:

generate a simulated lower-bandwidth audio signal having a type of degradation by executing an augmentation operation for the type of degradation on a lower-bandwidth audio

signal, the plurality of training audio signals further comprising the simulated lower-bandwidth audio signal, and

wherein the inbound audio signal has the type of degradation, whereby the estimated inbound audio signal generated by the computer is an enhanced inbound audio signal having comparatively less of the type of degradation.

15. The system according to claim 12, further comprising:

extract one or more features from each of the training audio signals, wherein the computer applies the neural network on the one or more features of the training audio signals; and

extract the one or more features from the inbound audio signal, wherein the computer applies the neural network on the one or more features of the inbound audio signal.

16. The system according to claim 12, wherein at least one higher-bandwidth audio signal of the plurality of training signals originated via a channel configured for the second bandwidth.

17. The system according to claim 12, wherein the computer generates the estimated inbound audio signal, in response to the computer determining that the inbound audio signal originated via a channel configured for the first bandwidth.

18. The system according to claim 12, wherein the computer is further configured to:

execute a loss function of the neural network according to a training estimated audio signal outputted by the neural network for a training audio signal, the loss function instructing the computer to update one or more hyperparameters of one or more layers of the bandwidth expander.

19. The system according to claim 12, wherein the computer is further configured to:

train a speaker recognizer comprising a second set of one or more neural network layers by applying the speaker recognizer on a plurality of second training audio signals comprising one or more clean audio signals and simulated audio signals;

extract an enrollee voiceprint for an enrollee by applying the speaker recognizer on one or more enrollee audio signals of the enrollee;

extract an inbound voiceprint for an inbound speaker by applying the neural network architecture to the estimated inbound audio signal; and

generate a likelihood score based upon the inbound voiceprint and the enrollee voiceprint, the likelihood score indicating a likelihood that the inbound speaker is the enrollee.

- 20.** The system according to claim 19, wherein the computer is further configured to:
generate an estimated enrollee audio signal for the one or more enrollee audio signals
by applying the bandwidth expander on an enrollee audio signal having the first bandwidth and
originated via a channel configured for the first bandwidth.