19

Recovering perfectly sounding speech by denoising

~High-quality speech enhancement using vast amounts of examples ~

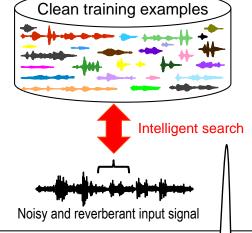
Abstract

A considerable number of studies have addressed speech enhancement (SE) algorithms to remove interference signals (e.g., noise, reverberation) from input speech signals. However, resultant signals often contain distortions such as residual interference and artifacts, which degrade overall audible quality.

Our proposed method establishes an SE mechanism that does not induce such distortions and utilizes vast amounts of clean training raw data (i.e., examples). It first seeks an example that best matches underlying clean speech sequence in test data and replaces them with examples. Resultant signals are generated essentially by collecting clean examples, and thus there is no place for interference to remain in output signals, which leads to high-quality enhancement.

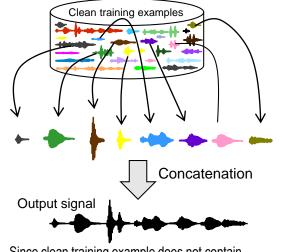
Concept of example-based speech enhancement

Step 1 Compare input signal and vast amount of clean training data (i.e., examples) to search for example in training data that best matches underlying clean speech sequence in input data. Clean training examples



Step 2 Generate output signal

by concatenating best matched longest examples. (**)



Since clean training example does not contain noise/reverberation, output signal generated in this way also does not contain noise/reverberation

What is intelligent search mechanism to look for examples in training data that best matches underlying clean speech sequence in input data?

Following complementary mechanisms have been proposed:

- Scheme 1 [1]: Search for best matched sequence while considering uncertainty (amount of noise in input signal, etc.) in matching process
- Scheme 2 [2]: Artificially generate a number of noisy examples paired with clean training examples. Then, compare noisy examples and input data by a fast search algorithm and find best-matched examples.

(XX) It is also possible to estimate an enhancement filter (e.g., Wiener filter) by utilizing resultant concatenated signal (to also make this entire scheme effective for speaker independent scenario.

Related works

- [1] K. Kinoshita et al., "Single channel dereverberation using example-based speech enhancement with uncertainty decoding technique", in Proc. of Interspeech, pp. 197-200, 2011.
- [2] A. Ogawa et al., "Fast segment search for corpus-based speech enhancement based on speech recognition technology", in Proc. of ICASSP, pp.1557-1561, May, 2014.

Contac

Keisuke Kinoshita Signal Processing Research Group, Media Information Laboratory E-mail: kinoshita.k(at)lab.ntt.co.jp