Neural Network を使った波形入出力による雑音抑圧*

◉ 田村震一 アレックス ワイベル (ATR自動翻訳電話研究所)

1. INTRODUCTION

We propose a new noise reduction method using neural networks. Noise reduction can be viewed as a mapping from the set of noisy signals to the set of noise-free signals. Neural networks are attractive as mapping definition for the following reasons.

(1) An arbitrary decision surface can be formed in a multi-layered neural network[2]. So any complex mapping from the set of noisy speech signals to the set of noise-free speech signals can in principle be realized.

(2) Simple learning algorithms exist to construct a suitable mapping function using training samples [3].

(3) Neural networks have attractive generalizing properties [3].

2. NEURAL NETWORKS FOR MAPPING

A four-layer feed-forward network was chosen as an architecture for it can realize in principle any mapping function[2]. Each layer has 60 units and is fully interconnected with its next higher layer (Fig. 1).

The input and output of the network is given by the waveform itself, and the Back-Propagation algorithm was adopted as the network's learning algorithm. A unit element of the network first computes the weighted sum of all its inputs (including a bias input) and then deforms this sum by passing it through a nonlinear function, in our case the sigmoid function [3].

3. EXPERIMENT

3.1 Data

The speech data uttered by several male and female speakers (professional announcers) was digitized (16 bits) at a 20kHz rate and then down-sampled to 12 kHz.

As non-stationary noise, computer room noise was digitized to 16bit at a 12kHz-sampling rate.

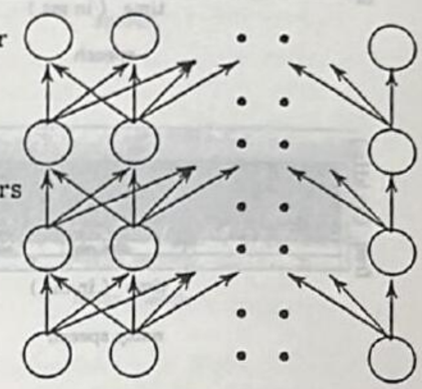

Fig 1 Network for noise reduction

Noisy speech data was generated artificially by adding the computer room noise to the speech data. The resulting S/N ratio is about -20db.

3.2 Learning

Using the waveforms of the 216 phoneme-balanced words uttered by one male speaker as target output and their noise added versions as the training input, the network scans each training utterance from beginning to end at a rate of 60 data points per input frame. This procedure is repeated until the network's squared error rate converges to a sufficiently small value.

The back-propagation learning procedure repeatedly adjusts the network's internal link weights in an attempt to find an optimal mapping between noisy and noise-free signals.

3.3 Results

The tests reported in the following were performed on a network that was trained on about 200 scans through the traininig utterances. Fig. 2 shows the result of training after about 200 scans. The input to the network is the noisy Japanese word "ikioi" from the training data. As can be seen, the noise has been reduced significantly, while the speech spectrum is preserved. Fig. 3 shows the network's ability to find a generalized noise reduction mapping

* Noise Reduction Through Waveform Input And Output Using Neural Networks, by Shin'ichi TAMURA, Alex WAIBEL (ATR Interpreting Telephony Research Laboratories)

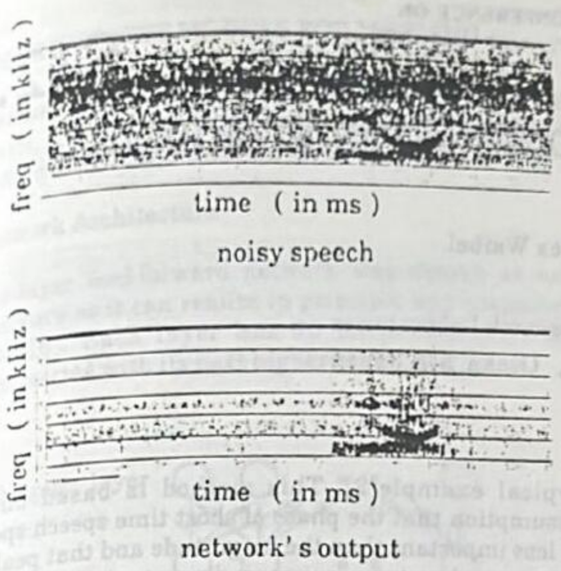

Fig. 2 Spectrograms of the training data

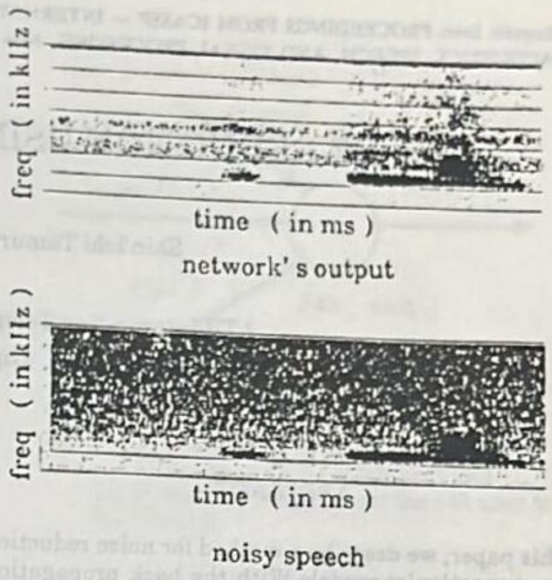

Fig. 3 Spectrograms of non training data and different noise

based on the observations in the training data. We show as input to the network the Japanese word "ikioi" which was uttered by a female speaker and was corrupted by computer-generated white noise. Despite the fact, that the network was trained on a male speaker and a different kind of noise, it produces a substantially cleaner output signal, without adversely affecting the speech signal.

Method	Score
Power spectrum subtraction	43.4%
Neural Network	56.6%

Fig. 4 Result of auditory preference test

Comparison with the conventional power spectrum subtraction method [1] was made. The frame length of the spectrum subtraction is 64 points long and the shift is also 64 points long. Noise suppressed speech was presented to 18 listeners in pairs and subjects were asked to mark the prefered speech sample. Fig. 4 indicates that our noise reduction method is comparable to or better than the conventional power spectrum subtraction method.

Our model produced a cleaner signal, it does, however, not appear to yield greater intelligibility. We believe that more focused learning of acoustic-phonetically important parts of the speech signal might lead to further improvements in intelligibility.

4. CONCLUSION

In a series of computer experiments we have shown that neural networks can learn the mapping between the set of noisy signals and the set of noise-free signals correctly. We have shown that the network produces noise-suppressed signals even for signals that differed from the training data in both the original speech input as well as the type of environmental noise.

ACKNOWLEDGEMENT

The authors are deeply grateful to Dr. Kurematsu, Dr. Shikano, Dr. Kuwabara, and all the members of Speech Processing Department for their constant help and encouragement.

REFERENCES

[1].S. F. Boll "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. on Acoustics, Speech and Signal Proc., vol. ASSP-29, pp.113-120, April 1979.

[2] R.P.Lippmann, "An Introduction to Computing with Neural Nets," IEEE ASSP magazine, pp. 4-22, April 1987.

[3] D.E.Rumelhart, J.L.McClelland and the PDP Research Group, *Parallel Distributed Processing* Vol.1, Chap.8, MIT Press, 1986.