NOISE REDUCTION OF SPEECH BY NEURAL NETWORKS AND VECTOR QUANTIZATION

Kazuo NAKATA and Akihiko SUGIURA

Faculty of Technology, Tokyo University of Agriculture and Technology
(2-24-16, Nakamachi, Koganei-shi, Tokyo, JAPAN)

ABSTRACT
In order to reduce the effects of noise mixed with speech, a method of noise reduction is proposed. The method is composed of a combination of neural networks and vector quantization. A multi-stage neural networks is proposed. Each stage has a small number of output categories, and its performance are evaluated. The scheme is effective for noise reduction in the range of SNR better than 0 dB, and developed to be practicable by the introduction of reject and re-try process.

1. INTRODUCTION
Noise from environments is unavoidable in actual speech recognition. The technique to reduce the effects of noise is called "noise reduction" or "speech enhancement". A neural networks, multi-layer Perceptron, shows high performances in pattern classification. A vector quantization provides effective and reasonable means to cluster speech signals, and quantize them into a sequence of the finite number code vectors.

In this paper, a new method of noise reduction of speech is proposed. The method combines two new technologies "neural networks" and "vector quantization".

2. BASIC PRINCIPLE OF NOISE REDUCTION
2.1. Technical Backgrounds
If the configuration of neural networks are properly designed, and they are well trained, they can perform a pattern classification function equivalent to the multi-templates pattern matching without any increase of memory for many templates and of processing time for finding the best matching among many of them. (1),(2).

The vector quantization provides with reasonable and effective means for continuous to discrete data conversion by proper clustering. (3).

2.2. Basic Principle of Noise Reduction
It is worthwhile to reduce the effect of noise at the level of auto-correlation function or coefficients because the basic data of linear predictive analysis are auto-correlation coefficients. (4).

The basic principle of our noise reduction scheme is illustrated in Fig.1.

An input is a sequence of the set of normalized auto-correlation coefficients derived from noisy speech at every analyzing frame.

The neural networks can classify them properly into one of the vectors in the code book, frame by frame.

The output (vector quantized sequence of the set of noiseless auto-correlation coefficients) are used as the data for further processing.

Fig.1. Basic Principle of Noise Reduction by Neural Networks and Vector Quantization

3. BASIC CHARACTERISTICS OF NEURAL NETWORKS
3.1. Classification of Five Japanese Vowels by Neural Networks
As the neural networks, the two-layer Perceptron type (input, output and one hidden layer between) is used all through the following experiments.

The following conditions must be set up in advance for carrying out experiments by neural networks.
1) Dimension of input vector : N.
2) Dimension of output categories : K.
3) Number of neural elements in hidden layer : H.
4) Range and way of connection of neural elements.
5) Initial values of the connecting weights.
6) Output function of each neural element.
7) Sensitivity gain of learning in BPL process.

All of these parameters are illustrated on the configuration of the neural networks in Fig.2.

The most popular Sigmoid function of the formula (1) with the parameter $\theta$ equals to 0.25 is used as the output function:

$$f(z)=1/(1+exp(-z/\theta))$$

where $z$ is the sum of all weighted inputs to the element, and no smoothing of connecting weights in BPL is used in our case. Sensitivity gains of learning are chosen to be small enough for steady and non-oscillatory convergence of learning.

Details of our experimental conditions are given below.
1) Input : normalized auto-correlation coefficients $r(nT)$ at every analyzing frame. $T$ is the sampling period.
2) Number of output categories: K is set to five. They are "a, i, u, e, and o" of Japanese vowels.
3) Initial values of connecting weights: in our case, the initial values are chosen as;
   \[ \pm 1/(\text{number of input neural elements}) \]  \quad \text{(2)}
   + or - is set alternatively.
4) Number of neural elements in the hidden layer H and sensitivity gain: they are both optimally determined by experiments.

A white random noise is generated numerically in a computer and is added to digitized speech data.
1) Performance of neural networks for the classification of noisy vowels by the learning of noiseless vowels only.
The results are shown in Fig.3.
2) Learning conditions of vowels are SNR of 10, 5, 0, -5, and -10 dB.
a) For all learned inputs: the results are shown in Fig.3.
b) For SNR in the midst of the learned values: the results are shown in Fig.3.

![Configuration of Two-Layer Perceptron Type Neural Networks](image)

By experiments and the analyses of those results of speaker specified vowels classification, the following conclusions are drawn in noiseless case.
1) In case of a small number of output categories, a two-layer Perceptron type neural networks is enough for the perfect classification.
2) Relatively small number of neural elements in the hidden layer is enough for a high rate of correct classification.
   They are probably between twice of the number of output categories and a half of the sum of the number of input and output.
3) For initial values of the connecting weights, the values given by the formula (2) are usable.
4) The sensitivity gain \( q \) of BPL is nearly optimum at the value inversely proportional to the number of forward connecting neural elements and the number of learning conditions.

4. DIRECT APPLICATIONS TO NOISE REDUCTION

4.1. Classification of A Single Speaker's Vowel With Additive Random Noise

A series of experiments are carried out to get some insights into the performance of neural networks in noisy cases.
After 10,000 times of learning of each condition, evaluation tests are carried out.

The rate of the correct classification is about 62%.

Into this neural networks, the speech of SNR of 1 dB steps from 10 dB to 1 dB, total 88*10=880 frames are fed, the rate of the correct classification is also about 62%.

5. MULTI-STAGE CLASSIFICATION METHOD AND ITS PERFORMANCES

5.1. Proposal of Multi-Stage Neural Networks (6)

From the results of five Japanese vowels classification, the following facts are understood as the basic characteristics of neural networks.

1) Classification performance is the better for the less number of output categories.
2) The range of adaptation of the learned performance is limited to relatively narrow.

Then, instead of a one-stage neural networks of 16 output categories, a serial connection of a two stage neural networks of 4 output categories each is expected to be effective for correct classification of frames of the word.

5.2. Performances of Two-Stage Neural Networks

Condition of experiments are the same as in the case of one-stage neural networks.

The first stage is 4-class gross classification.

The number of member frames of each gross class are: 8 for \#1, 25 for \#2, 15 for \#3, and 40 for \#4 in the noiseless correct classification.

The second stage is 4-class fine classification for each gross class of the first stage.

1) Experiments \#1—-a noiseless word.

To evaluate the basic performance of multi-stage neural networks, the classification experiments of the word " so endoo " is carried out in noiseless condition.

The performance is the perfect, as expected, at the first and the second both stages.

2) Experiments \#2---the first test of noisy word.

Conditions of SNR of 10, 5, and 0 dB in 5 dB step have been learned 10,000 times in each, and classification tests are carried out.

For the first stage, a percentage of the correct classification is 81 %, and for the second stage, an average percentage of the correct classification is 95 % under the assumption that the first stage classification is all correct.

3) Experiments \#3---the second test of the noisy word of SNR better than 0 dB.

All results of the classification tests are summarized in Fig.4.

The results of the one stage case are also shown in comparison.

These results show definite improvement of the correct classification rate in comparison with the one-stage neural networks.

In Fig.4., the results of usual pattern matching method by Euclidean distance are also shown in comparison. They evidently show the high performance of our proposed method.

4) Experiments \#4---introduction of rejection and re-try.

In practical applications, the available condition of learning are rather limited.

Then, the learning is carried out only at the one
one condition, for example SNR=5 dB. But in this case, rejection and re-try processes are introduced at the output of each stage. The flow of the processes are shown in Fig.5. The results are analyzed for SNR and shown in Fig.4. If the SNR are better than 2 dB, the average correct classification rate is 94%, and almost the same as the case of fine detailed learning shown in the same figure.

6. CONCLUSIONS AND FUTURE PROBLEMS

Basic characteristics of vector classification by the neural networks of two-layer Perceptron type are evaluated. Main findings are:

1) The performance deeply depends on the number of output categories. And when the categories are small, probably less than ten, very high performances are attained.
2) The performance is high for the trained/learned conditions, but an adaptive range of each learning is rather narrow. Then, the multi-stage neural networks of vector classification is proposed for noise reduction with the help of discrete pattern/vector clustering function of vector quantization. More than 90 % correct classification of frames in the word is possible by two-stage neural networks in case of SNR better than 0 dB.

A series of experiments are carried out in order to make the scheme to be practicable by the introduction of rejection and re-try rules, and results are very encouraging.

Our data of experiments are rather limited to the five Japanese vowels and one word of a single male speaker, then extensive experiments are further necessary to confirm and generalize conclusions obtained here.

But, at least in principle, a new method of noise reduction can be proposed and its performances are evaluated, and proved to be effective for the noise reduction in the range of SNR better than 0 dB, and to be practicable by the introduction of rejection and re-try processes.

7. ACKNOWLEDGEMENTS

This research works are supported by Grant-in-Aid for Scientific Research on Priority Area, Advanced Man-Machine Interface Through Spoken Languages, The Ministry of Education, Science and Culture, 1988. Many thanks to Mr. K. Yamada, graduate student, who engaged in this research in 1989.