A PRIMARY EXPERIENCE: THE VECTOR QUANTIZATION TECHNIQUE IS AN EFFECTIVE TOOL FOR PHONETIC RESEARCH

ZHAO GUOTIAN

Speech Technology Lab.
Dept. of Radioengineering
Harbin Institute of Technology
Harbin, P.R. China

ABSTRACT

The effective phonetic symbol system representing the phonetic feature of speech exactly is important tool for speech processing technique. But the present spelling symbol system of Chinese is suitable for teaching only, but not for speech technology. In this paper, as a beginning, we have investigated the spelling symbol [i] in different environments by the method of vector quantization, which follows consonants [j, q, x, z, c, s, zh, ch, sh, t, b, p, l, m, n, r, y]. The results show that Chinese symbol [i] can be represented by International Phonetic Symbols [i, ɿ, ɻ] in detail.

INTRODUCTION

In phonetics, phonetic symbols are used to indicate phoneme, for instance, International phonetic symbols. The phonetic symbols system which is established for one language, in teaching language and studying pronunciation and correctly expressing language contents with speech, is very sufficient and clear for hearing with which people are satisfied. But, in the field of speech signal processing which is related to phonetics closely, the phonetic symbol system for teaching speech and studying pronunciation becomes too rough to represent the similarities and differences between symbols. Specially the Chinese spelling symbol system was established for Chinese teaching and it is simplitic. The requirements for speech signal processing can not be content with it. Therefore, the meticulous classification for phonetic symbols is required in the field of speech signal processing. For example, in Japanese, for phonetic symbol [n], only one is applied to teaching speech and studying pronunciation. But in speech signal processing, for instance, in digits recognition, it is divided into two [2].

There are 32 phonetic symbols in modern Chinese [1]. For some symbols, the problems similar to above exist, and is even more serious as compared with other languages, because the present symbols for Chinese are more sketchy than international symbol, probably. So it is necessary to research Chinese phonetic symbols deeply and meticulously to meet the needs of speech signal processing technique. In this paper, [i] of Chinese phonetic symbol is analysized and reseached. The method used is vector quantization.

The similarities and differences of [i] are obtained under the various possible environments. Finally, [i] is divided into three [ʅ], [ɿ] and [ɻ] in detail. They are suitable for speech processing, in present stage.

THE METHOD FOR ANALYSIS

As feature parameters, the linear prediction coefficient cepstrum are applied to our research. The experience demonstrated that they are more sufficient feature parameters for speech recognition.

The vector quantization technique has been developed in the field of speech compression for communication. But recently this technique has also been introduced to the field of speech recognition. The basic concept of vector quantization applied to speech compression is schematically in Fig. 1. Both training and input vectors are the same kind of speech feature vectors. In our research, the capacity of codebook is 64, and convergence threshold is 0.01 [3]. Each input vector is compared with the codewords in codebook and then is endowed with the code of the most similar one, so the code substitutes for the input. The further analysis for input vector will be simplified.

In speech analysis with above method, what is analysized statistically is the code quantized of speech festure vector, not the speech vector itself. According to the code, the similarities and differences for some symbol in different environments are found out. With this method, investigation for phonetic symbols gets more convenient and sufficient.

EXPERIMENT AND RESULTS ANALYSIS

In our experiment, the similarities and differences of symbol [i] has been investigated in different environments, namely in combination with different consonants. The consonants followed by [i] are [j, q, x, z, c, s, zh, ch, sh, d, t, b, p, l, m, n, r, y].

Speech samples for research are from seven people, including five men and two women, and everyone has 20 samples for each of 18 monosyllable. After being quantized, the code sequence of stationary part of [i] following [j, q, x, d, t, b, p, l, m, n, y, z, c, s, zh, ch, sh, r] are obtained, as shown, for examples, in Tab. 1 which belongs to two people respectively.

According to the tables, the codes of [i] following [j, q, x, d, t, b, p, l, n, m, y] are more similar to each other, and that following [z, c, s] and [zh, ch, sh, r] come to the same thing.

So we can divide Chinese symbol [i] for teaching into three minimum: [i] following [j, q, x, d, b, t, p, j, m, n, y] is expressed as [i], and that following [z, c, s] and [zh, ch, sh, r] as [ɿ] and [ɻ] respectively.

CONCLUSION

According to our experimental investigation, the results show that representing by International Phonetic Symbol [i, ɿ, ɻ] is more proximal and representing by Chinese symbols used now for teaching is away from phonetic practice in principle, except depending upon environment. What we have done is as a part of investigation of phoneme representation for research and it is being continued.

Also, the experimentation shows the VQ technique is an effective tool for phonetic research, it can show similarity between phonemes evidently.

Tab. 1

Speaker: Mr. Zhao

Frame	11	12	13	14	15	16	17	18	19	20	21
Ji	51	52	52	63	63	62	63	52	56	56	64
Qi	59	60	60	60	60	60	60	60	57	57	57
Xi	50	50	51	51	51	51	50	63	64	64	64
Zi	21	17	17	17	19	19	20	20	20	20	20
Ci	20	20	20	20	20	20	20	20	20	20	20
Si	18	18	18	17	17	17	17	17	17	19	19
ZHi	1	1	4	4	1	1	1	1	1	2	3
CHi	27	13	4	1	3	3	3	3	1	1	1
SHi	12	11	11	11	11	11	11	11	11	11	11
Di	50	64	64	62	64	64	64	62	62	62	62
Ti	63	63	63	63	64	64	64	55	55	55	60
Bi	63	52	62	63	52	64	64	64	64	52	62
Pi	63	63	63	63	64	63	52	64	64	52	54
Li	58	62	58	63	62	62	64	62	64	58	64
Mi	56	56	53	53	53	53	53	53	53	53	53
Ni	55	56	56	54	56	54	54	54	54	54	51
Ri	3	3	3	3	3	3	3	3	3	3	3
Yi	62	62	62	62	64	56	56	56	56	54	54

Speaker: Mrs. Luo

Ji	56	56	56	56	56	56	56	56	56	56	56
Qi	41	44	42	42	42	42	56	56	54	54	56
Xi	42	42	47	47	47	47	54	47	54	45	54
Zi	3	3	4	4	4	4	4	4	4	4	4
Ci	14	10	3	3	3	3	3	3	3	3	3
Si	10	9	9	3	3	3	3	3	3	3	3
ZHi	23	23	23	23	23	23	23	23	23	23	32
CHi	41	44	17	22	27	22	27	22	22	22	22
SHi	43	15	15	15	15	15	22	21	27	27	27
Di	56	56	56	56	50	50	50	42	42	42	42
Ti	46	46	60	60	45	45	46	48	46		
Bi	55	55	55	55	55	57	57	50	50	42	42
Pi	45	45	45	45	45	45	58	58	58	58	58
Li	56	56	56	56	56	54	54	55	55	56	50
Mi	63	63	63	63	63	46	58	58	46	46	58
Ni	48	48	63	63	64	64	64	64	64	64	64
Ri	30	30	30	29	30	30	32	32	32	32	32
Yi	50	50	50	55	56	57	57	59	58	57	59

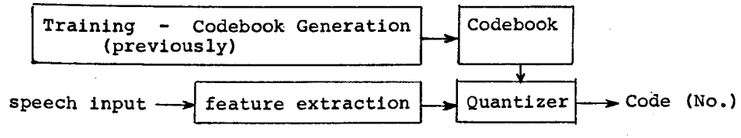

Fig. 1. Vector Quantization Process

REFERENCES

[1] Xu Shirong: On Putonghua Phonetics. 1980, Beijing.

[2] Masaki KOHDA et "Spoken Digit Mechanical Recognition System" Trans. of IEC, vol. 55-D No. 3 (Japanese).

[3] Anderes Buzo, Augustine H. Gray, Robert M. Gray and John D. Markel: Speech Coding Based Upon Vector Quantization. IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-28, No. 5, October 1980, pp. 562-574.