English Pronunciation Reasoning by NN Considering Frequency Distribution of Phonemes

Ikuo YOSHIHARA[1) Yusuke HIGASHI[2) Hanxi ZHU[3)
Kunihito YAMAMORI[4) Moritoshi YASUNAGA[5)

Abstract

English pronunciation is decided by not only alphabet but also the position in a word. Because to guess pronunciation of an unknown word is difficult, to develop a system is of great significance. Many methods have been proposed for English pronunciation reasoning (EPR). They have been reasoned only from spelling the words. We aim at improving reasoning accuracy by adding frequency of appearance of phonemes to spelling. The accuracy of EPR comes up from 85.43% to 86.39%

Key Words:
Neural network, EPR, frequency of appearance of phonemes

1 Introduction

In English, pronunciation is decided by not only alphabet but also the position in a word. The guess of English pronunciation of an unknown word is difficult. Since revision of a dictionary is not frequently performed even if pronunciation is investigated, it is not found easily. Therefore, development of the English pronunciation reasoning (EPR) system is important.

Many methods have been proposed for EPR. In 1987, for the first time, Sejnowski employed a neural network (NN) called NETtalk for EPR [1)]. DECtalk is commercial text-to-speech system, in which a string of phonemes is converted to sounds with digital speech synthesis [4)]. MBRtlak proposed by Stanfill,C. is a system by using "memory based reasoning" basing on a connection machine [5)]. LoDETT proposed by Yasunaga,M. who developed a reasoning hardware based on Genetic Algorithm (GA) [2)]. They have been reasoned only from spelling the words.

We aim at improving accuracy of EPR using NN. The improvement in accuracy is expected by adding frequency of appearance of phonemes to spelling. We validate the method of consideration of frequency of appearance of phonemes.

2 English Pronunciation Reasoning

English pronunciation is classified into 52 classes. String of letters are converted to corresponding elementary English speech sounds (phonemes). English is not phonogram, so pronunciation is decided by not only alphabet but also position in a word. Therefore, extracting rules of pronunciation is a difficult problem.

2.1 EPR database

The database used in this work is Webster English Pocket Dictionary, which is proposed by Sejnowski. This database involves 20008 English words and pronunciations. Pronunciation is described as shown in Table 1. Phonemes are represented by combination of 21 articulation features and five features. 21 articulation features are classified into articulation point, voiced silence, height of

Table 1 Phonemes and Example Words

Word	Phoneme
abalone	@bxloni
garnet	garnxt
loop	lu-p
over	ov-R

[1)]Professor,Dept. of Computer Science and System Engineering
[2)]Undergraduate student,Dept. of Computer Science and System Engineering
[3)]Doctoral student, Graduate School of Engineering
[4)]Associate Professor,Dept. of Computer Science and System Engineering
[5)]Professor,Dept. Institute of Information Sciences and Electronics, University of Tsukuba

vowel and etc. Five features represent accent and articulation boundary. Table 2 is an example of phonemes with their example words (Ex.word).

Table 2 Phonemes and example words

Phoneme	Ex.word	Phoneme	Ex.word
a	topic	D	that
b	blue	E	death
c	song	G	link
d	adobe	I	drip
e	able	J	just
f	food	K	sexual
g	game	L	angel
h	hack	M	imagism
i	easy	N	season
k	key	O	noise
l	lab	Q	quilt
m	make	R	vector
n	answer	S	sheep
o	old	T	math
p	sleep	U	book
r	run	W	bound
s	sky	X	matrix
t	article	Y	beauty
u	you	Z	jabot
v	voice	@	aback
w	wasp	!	pizza
x	welcome	♯	exit
y	yesterday	*	what
z	zoom	∧	above
A	five	+	devoir
C	chunk	-	(no sound)

3 Neural network for EPR

English pronunciation is reasoned using NN from spelling. In this paper, it is reasoned by two methods. First, it is reasoned from spelling. Next, it is reasoned by frequency of appearance of phonemes and spelling.

3.1 Multi-layer Neural Network

Multi-layer Neural Network consists of three kinds of layers; an input layer, hidden layers and an output layer. A neuron has weights between the neurons in adjacent layer.

3.2 Reasoning by spelling

When using the NN for EPR, NN are learned by spelling (7 letters information), 7 letters sequences are produced from a word by shift operation, and a set with the phoneme to the central letter (or fourth) of the pattern. Table 3 is an example of pattern setting for the word "alphabet"(phoneme:"@lf-xbET"). The input pattern is

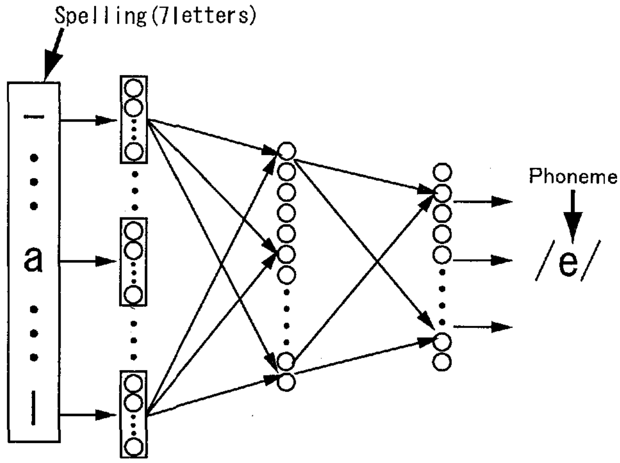

Fig. 1 NN for EPR by spelling

Table 3 Pattern setting for the word "alphabet"

Input pattern	Phoneme
- - - a l p h	@
- - a l p h a	l
- a l p h a b	f
a l p h a b e	-
l p h a b e t	x
p h a b e t -	b
h a b e t - -	E
a b e t - - -	T

represented by combination of 26 alphabets and a virtual letter. Each letter of the input pattern is represented with 27 units, the unit corresponding to letter is given 1 and other units are given 0. Number of units in input layer is 189(27×7).

Phonemes are represented by 52 kinds of symbols (Table 2). 52 units are set to output layer in the NN. Each unit corresponds to phonemes. The unit corresponding to phoneme output 1 and the other units outputs 0. The number of units in output layer is 52.

Figure 1 shows NN for EPR, which is trained by BP algorithm.

3.3 Reasoning by spelling and the frequency of appearance of phonemes

Reasoning from only spelling has the limit of accuracy. It is necessary to consider a new method for improvement in accuracy. In this paper, method of consideration of the frequency of appearance of phonemes is proposed as a method for the improvement in accuracy. It is adding the frequency of appearance of phonemes to spelling.

3.3.1 Frequency of appearance of phonemes

The frequency of appearance of phonemes is computed by training data. The input pattern with the same central

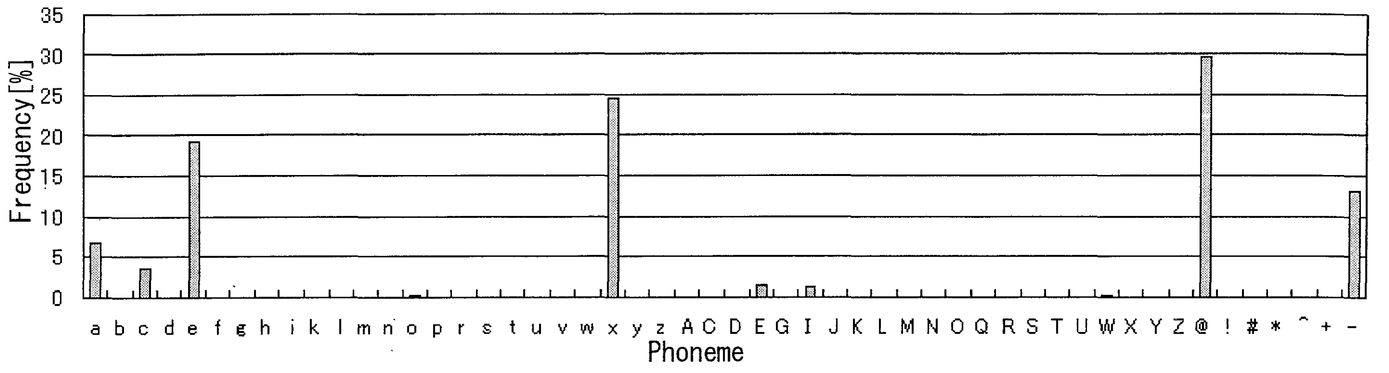

Fig. 2 Phonemes involved in letter 'a'

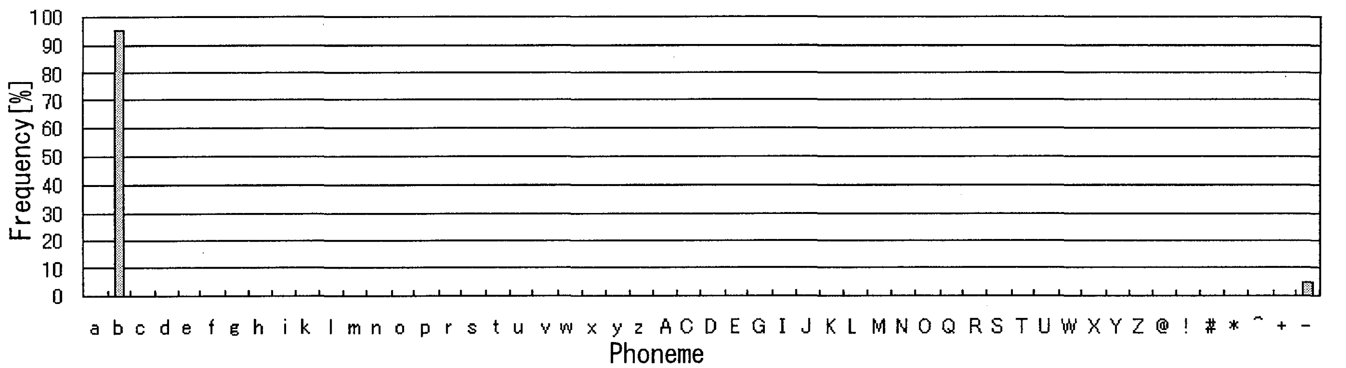

Fig. 3 Phonemes involved in letter 'b'

letter and corresponding phonemes are collected. Table 4 is example that collected input pattern with same central letter and corresponding phoneme. The frequency of appearance of phonemes is computed from the collected patterns.

Figure 2 and Figure 3 are the frequency of appearance of phonemes of 'a' and 'b'. Some features are found from two tables. For example, the features of 'a' (letter) has /e/, /x/, /@/ and /-/ (phoneme) with high frequency of appearance of phonemes. The 'b' (letter) has the features in which /b/ (phoneme) almost appears.

3.3.2 Input frequency of appearance of phonemes

The frequency of appearance of phonemes is normalized for input to NN. It is added to spelling. Figure 4 is an example which reasons a pronunciation by the method of consideration of frequency of appearance of phonemes. The number of units for a letter is 79 units that are 27 units of letter information and 52 units of frequency of appearance of phonemes. Therefore, the number of units of an input layer becomes 553 (79×7).

Table 4 Collected input patterns with the same central letter and corresponding phoneme

Input pattern	Phoneme
---able	e
-plant-	@
-board-	-
strain-	e
.	.
.	.
.	.
sonable	x

4 Reasoning accuracy experiments

The accuracy of EPR is searched by two methods. The first is the method of reasoning only from spelling. The second is the method of reasoning from the frequency of appearance of phonemes and spelling. Each accuracy of EPR obtained by two methods is compared.

The method of reasoning from spelling is called NN_n, and the method of consideration of frequency of appearance of phonemes is NN_p.

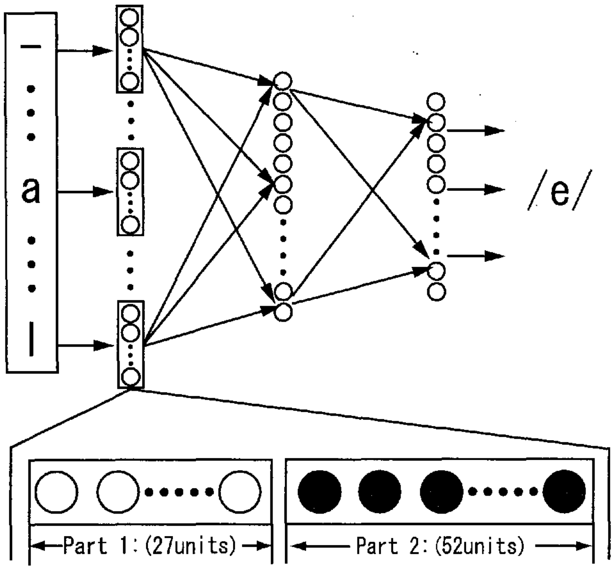

Fig. 4 Consideration of frequency of appearance of phonemes for NN

4.1 Experimental conditions

The data used in an experiment is set as follows. Each data is selected randomly from the database.

- Training data : 4000 words
- Testing data : 14008 words
- Checking data : 2000 words

The checking data are not used to calculate the weight changes during the training procedure, but used to determine the stop point of training. In order to prevent over-training, the training of the NN is stopped when the accuracy of the checking data begins to degrade.

The training parameter is set as follows.

- Iterations : 200 epochs
- Learning rate : 0.1
- Momentum : 0.8

Composition of networks is carried out as shown in a Table 5.

Table 5 Composition of networks

	NN_n	NN_p
Input layer	189	553
Hidden layer	60	60
Output layer	52	52

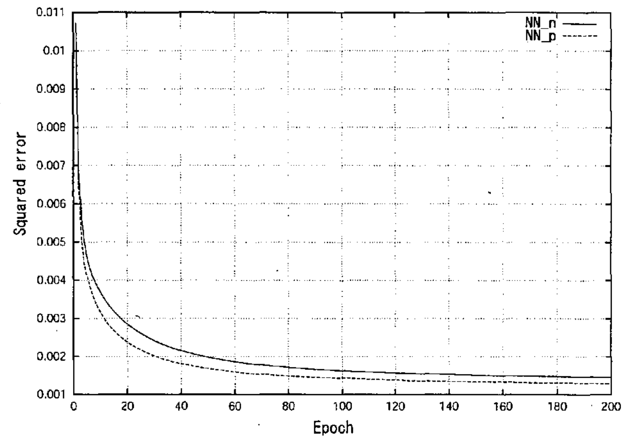

Fig. 5 Comparison of squared error

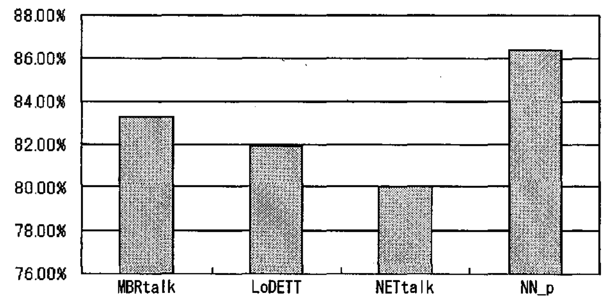

Fig. 6 Comparison with other methods

4.2 Reasoning results

The experiments were performed 10 times to each method under the conditions shown 4.1 (Experimental conditions). The average of the accuracy of EPR is shown in Table 6. As shown in Table 6, accuracy of EPR is improved 0.96%

Table 6 Reasoning accuracy

	NN_n	NN_p
Reasoning accuracy	85.43%	86.39%

4.3 Discussion

Figure 5 shows transition of squared error under study of NN_n and NN_p. Reduction of the squared error of NN_p is earlier than NN_n. The average accuracy of EPR is improved from 85.43% with NN_n to 86.39% with NN_p. The frequency of appearance of phonemes is considered to have carried out the effective action to the unknown pattern. Moreover, the method proposed in this paper is able to acquire accuracy higher than MBRtalk, LoDETT, and NETtalk. The accuracy of EPR obtained by each method is shown in Figure 6.

5 Conclusions

We developed a method of considering the frequency of appearance of phonemes for EPR. The frequency of appearance of phonemes had been computed from the training pattern. The accuracy of EPR in two methods was compared in the experiment.

The accuracy of EPR is improved from 85.43% by the spelling to 86.39% by adding the frequency of appearance of phonemes to spelling. The method developed this time was able to acquire accuracy higher than other methods (MBRtalk,LoDETT and NETtalk).

The future works are to reduce the calculation amount for training, etc.

References

[1] Sejnowski ,T ,J and Rosenberg ,C ,R : "Parallel Networks that Learn to Pronounce English Text" ,Complex Systems, Vol.1, pp.145-168 (1987)

[2] Yasunaga,M., Takahashi,M. and Yoshihara,I:"Reconfigurable Reasoning Hardware by using Evolutional Algorithm, Information Processing Society of Japan (IPSJ), Vol.40, No.7, pp.3031-3042(1999)

[3] Norio Baba, Fumio Kozima, Seichi Ozawa, : "Foundation and application of neural net" , Kyoritu syuppan. Co (1994) (in Japanese)

[4] Digital Equipment Corporation : "DECtalk DTC01 Owner's Manual" ,Digital Equipment Corporation, Maynard, Mass, Document number EK-DTC01-OM-002, (1986)

[5] Stanfill, C. and Waltz,D. : "Toward Memory-based Reasoning" ,Comm. ACM, Vol.29 ,No.12 , pp. 1213-1228 (1986)