IMPLEMENTATION OF THE POW (PHONETICALLY OPTIMIZED WORDS) ALGORITHM FOR SPEECH DATABASE

Yeonja Lim and Youngjik Lee

Automatic Interpretation Section
Electronics and Telecommunications Research Institute(ETRI)
E-mail: yjlim@zenith.etri.re.kr

ABSTRACT

This paper proposes the concept of the POW (phonetically optimized words) set. To collect a speech database, all possible phonological phenomenon should be included. In addition, it is preferable to have the same phonological distribution as the general speech. For this purpose, we suggest a new algorithm for selecting a word set which has the properties that 1) it includes all phonological events, 2) it has the minimal number of words, and 3) the phonological similarity between the POW set and the high-frequency word set is maximized.

We extract the Korean POW set from 50,000 high-frequency words out of 3 million text corpus. The POW set is much more similar to the high-frequency word set than the PBW (phonetically balanced words) set with less number of words.

1. INTRODUCTION

Speech database is constructed in two different ways. One is to construct a speech database for a specific domain in order to maximize the recognition rate. In this case, the set of word in speech database is restricted within the domain.

Another way is to construct a speech database for the general purpose, e.g., large-scale word recognition and COC(context oriented clusters) in speech synthesis. To construct this kind of speech database, a phonetically representative set of words based on a large text database should be extracted. The PBW algorithm has been proposed for this purpose. It tries to have equal probability for all phonological events. In fact, the PBW set is far from this idea since only a few events have equal probability. To cope with this problem, we proposed the POW algorithm based on the information theory.

2. PBW (PHONETICALLY BALANCED WORDS) ALGORITHM

2.1. Definition of PBW Algorithm

The PBW set must include all the possible triphones. The triphones in the PBW set should have equal probability and it should have the minimum number of words. The definition of the PBW algorithm must satisfy the three ordered conditions as follows:

The PBW set must

1. include all the possible triphones that are composed of all kind of phonological phenomena,
2. be the minimum set of words, and
3. have equal probability for all triphones in the PBW set.

The algorithm for constructing PBW set has the following two concepts, i.e.,

1. maximization of entropy and
2. minimization of the number of words.

The two principles are trade-off relation. Therefore, there can be some problems for implementing.

2.2. The Problems of PBW Algorithm

In general, the PBW set have been gathered as a basic database for speech research. Such database, however, usually have some problem in the speech recognition, since the distribution of triphone in population can be different from that in PBW set. Having equal probability is the same error rate for the frequently-used word and rarely-used words. Thus, it lowers the recognition rate.

0-7803-2431-5/95 $4.00

In the next section, we propose a new algorithm to handle these problems.

3. POW (PHONETICALLY OPTIMIZED WORDS) ALGORITHM

3.1. Definition of POW Set

The POW set should have all the possible triphones. The distribution of triphones in the POW set should be similar to that in the population. The POW set should have the minimum number of words. The first two condition in section 2.1 can be applied to both the POW and PBW algorithms. But they have a different point of view in using the concept of entropy.

The POW algorithm should satisfy three ordered conditions as follows:

1. The POW set must includes all the possible triphones.
2. The POW set has the minimum number of words.
3. The distribution of triphones in the POW set should be similar to that in the population.

Here, the population means the set of high-frequency words (50,000 words) in large corpus (3 million words in this case). The following two concepts are included in the POW algorithm.

1. Maximization of the modified entropy
2. Minimization of the number of words

3.2. The Modified Entropy

The entropy used in the POW algorithm can be described as follows:

1. Analyze the population and find the frequencies of ith triphones A_i, $i = 1, 2, \ldots, N$, and the highest frequency, $\mathcal{M}_{max}$.
2. Let $B_i = \mathcal{M}_{max} - A_i$.
3. Let $a_i^{(k)}$ be the frequency of ith triphone in the temporary POW set at kth iteration.
4. Let
$$T^{(k)} = \sum_{i=1}^{N} a_i^{(k)} + B_i \qquad (1)$$
5. The modified entropy to be maximized is
$$H^{(k)} = \sum_{i=1}^{N} \frac{(a_i^{(k)}+B_i)}{Tk} \log \frac{(a_i^{(k)}+B_i)}{Tk} \qquad (2)$$

3.3. The POW Algorithm using ADD/DELETE Method

To maximized $H^{(k)}$, we use the following steps.

- Construction of the initial set
  1. Construct the initial set by gathering all words which have unique triphone (only one occurrence in the population).
  2. Exclude rarely-occurring words, e.g., words of foreign origin, etc. from the initial set.
  3. Construct the candidate set by ordering the rest of words. Here, the word with larger distinct triphones comes first.
- ADD step
  1. Choose a set of words with the largest number of distinct triphones. Pick a word which has the largest number of new triphones, and add to the POW set. Continue adding until no word in the set have any new triphones.
  2. Choose a set of words with the next largest number of distinct triphones. Add new words with the same method as in 1.
  3. Continue step 2 for all words in the candidate set.
- DELETE step
  1. Choose redundant words from the above POW set. Here, the redundant word means a word with no unique triphone. Thus, deleting such words does not affect the total number of distinct triphone in the POW set.
  2. From the set of redundant words, pick a word that in most redundant. For each word in the set, we measure $H^{(k)}$ after deleting that word. The most redundant word means a word that maximize $H^{(k)}$. Delete it from the POW set.
  3. Continue step 2 until we cannot increase $H^{(k)}$.

4. COMPARISON

We extract 50,000 high-frequency words from the phonetically transcribed version of the ETRI text corpus (3 million words). Then, we construct the POW and PBW sets from the 50,000 words. The number of distinct triphones in the POW set is exactly same as that in the PBW set. The total number of words in the POW set is 3,848, while that in the PBW set is 4,028. The total number of triphones in the POW and PBW sets are 25,795 and 56,784, respectively.

To see the similarity between the population and the word sets, we measure the divergence [3]. The divergence between the population and the POW set is 1.28, while that between the population and the PBW set is 5.62. The histogram in each case is drawn in figures 1, 2, and 3.

5. FURTHER WORK

The selected POW set will be used in the vocabulary-independent speech recognition system and T-t-S system using COC(context oriented cluster).

Applying this algorithm to pseudo-phoneme level will be beneficial to get more reliable POW set.

6. REFERENCES

[1] K. Shikano, "Phonetically balanced word list based on information entropy," *Proceedings of Acoustical society of Japan*, 1984.

[2] Hayamizu, Tanaka, Yokayama, and Ohta, "Generation of VCV/CVC balanced word sets for speech data base," *Bulletin of Electrotechnical Laboratory*, 1986.

[3] J. Lin, "Divergence measure based on the Shannon's entropy ," *IEEE Trans. Inform. Theory*, vol. 37, pp. 145-151, Jan. 1991.

Acknowledgement

This research is supported by Korea Telecom.

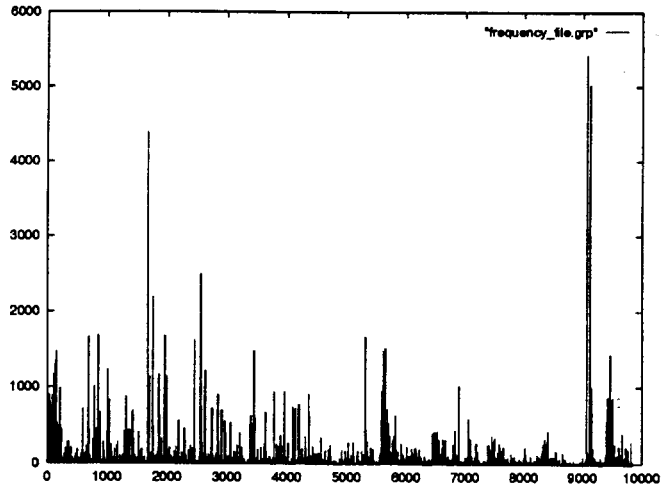

Figure 1: Frequency of triphones in the population

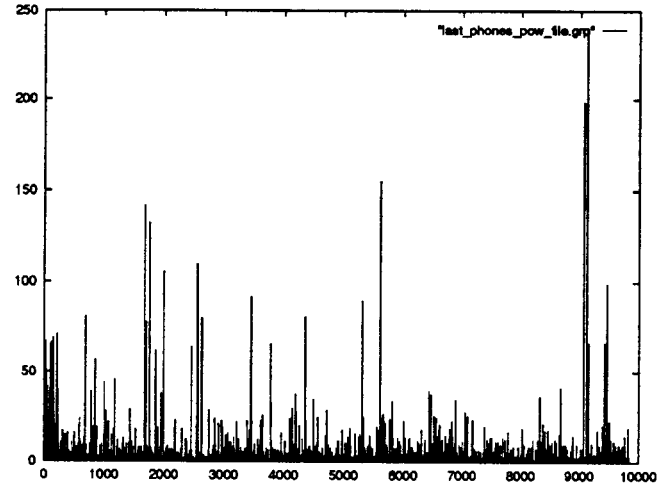

Figure 2: Frequency of triphones in the POW set

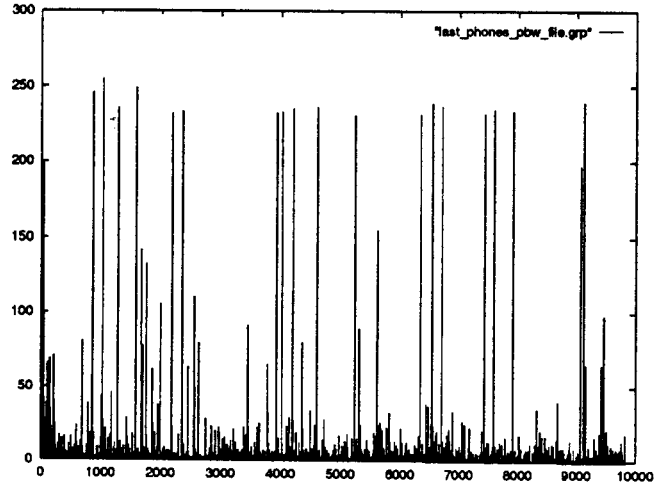

Figure 3: Frequency of triphones in the PBW set