The Study on Korean Phoneme for Korean Speech Recogintion

Young-Soo Hwang
Division of Information Electronics Engineering,
Kwan Dong University
KangWon-Do, Yang Yang, KOREA
FAX:(0396)670-3409
Email: hysoo@mail.kwandong.ac.kr

ABSTRACT

In this paper, we studied on the phoneme classification for Korean speech recognition.

In the case of making large vocabulary speech recognition system, it is better to use phoneme than syllable or word as recognition unit. And, in order to study the difference of speech recognition according to the number of phoneme as recognition unit, we used the speech toolkit of OGI in U.S.A as recognition system.

The result showed that the performance of diphthong being unified was better than that of seperated diphthongs, and we required the better result when we used the biphone than when using mono-phone as recognition unit.

1. Introduction

Human-Machine Communication is a very active research field and is one of the main areas in which computer manufacturers, telecommunications companies. The best way to communicate between Human and Machine is speech. The research topics of Spoken Language communication with machine are speech recognition and synthesis as well as speaker recognition.

In order to obtain good performance of speech system, many pattern matching methods (DTW[1], VQ[2], Neural Network[3], HMM[4], Fuzzy theory and hybrid systems) have been developed. And these many methods have been used in many countries which have their own languages, and they obtained some good results. But , in order to obtain better performance, I think that each country should be ready database of its own language and statistics of its speech parameter. Speech recognition units are classified into word, syllable, phoneme and biphone. In generally, because of the problem of memory and processing time, phoneme or biphone are used in large vocabulary recognition system and continuous speech recognition system. And we studied Korean phonemes fitting for Korean speech recognition system.

2. Recognition System and Phonemes used in this Paper

2-1. Speech recognition system used in this paper

In this paper, we used OGI speech tool kit as recognition system. Fig 1. represents block diagram of this speech recognition. In Fig 1., the first step is to classify time frames as phonetic categories by using Neural Network and the second step is to match category scores to target words by using Viterbi search and vocabulary grammar.

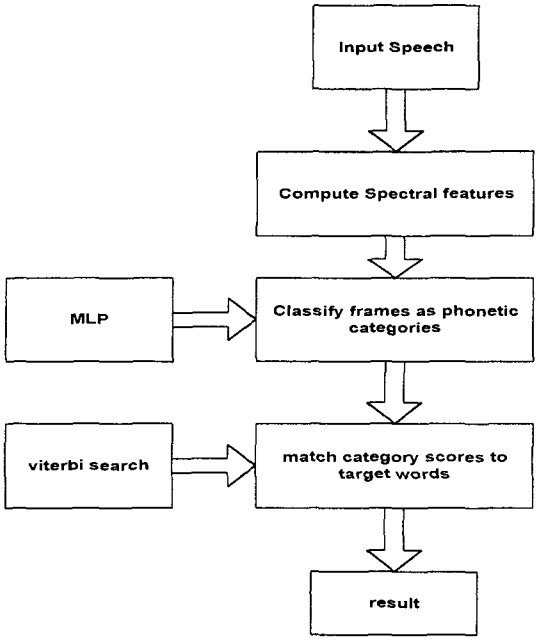
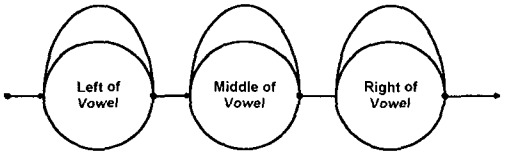
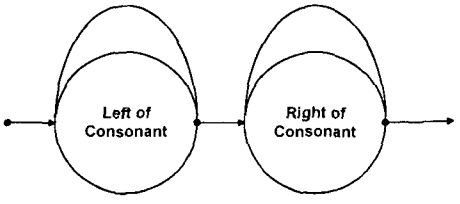


Fig 1. Block diagram of speech recognition used in this paper. (OGI speech tool kit)

In this paper, we used Neural Network and HMM as pattern matching methods. In case of speech recognition by using biphone, we used phoneme model having 3 states for vowel and 2 states for consonants. Fig 2. represents this phone model.



(a) phoneme model of vowel



(b) phoneme model of consonant

Fig 2. Phoneme model used in this paper

2-2. Korean phonemes used in this paper

Korean alphabet(Han-Gul) consists of forty letters. Twenty-one of these represent vowels, such as,

ㅏ(a), ㅓ(ə), ㅗ(o), ㅜ(u), ㅡ(u), ㅣ(i), ㅐ(e), ㅔ(e), ㅑ(ya), ㅕ(yə), ㅛ(yo), ㅠ(yu), ㅒ(ye), ㅖ(ye), ㅘ(wa), ㅙ(we) ㅝ(wə), ㅞ(wə),ㅚ(wə), ㅟ(wi), ㅢ(ui).

Nineteen of these represent consonants, such as,

ㄱ(g,k), ㄴ(n), ㄷ(d,t), ㄹ(l,r), ㅁ(m), ㅂ(b,bc), ㅅ(s,t), ㅇ(N), ㅈ(ch,t), ㅊ(tS), ㅋ(kh), ㅌ(th), ㅍ(ph,bc), ㅎ(h,h_v,tc), ㄲ(kk,k), ㄸ(tt), ㅃ(pp), ㅆ(ss,tc), ㅉ(cc)

Twenty-four are basic, while the others are compounds of the basic letters.

The Korean language posesses a rich variety of vowels and consonants with ten simple vowels and three series of stops and affricates: plain, aspirated and glottalized. This gives difficulty to foreigners who are just starting to learn the language and also complicates the task of Romanization.

Among the vowels, most pepole under 50's pronounce vowels('ㅖ','ㅒ') as a vowel 'ㅔ' and we used these two vowels as one symbol. And we also used vowel 'ㅒ' as 'ㅖ', and 'ㅙ', 'ㅞ' as 'ㅚ'. And we did not use diphthong as 'semi-vowel+monophthong' but as one symbol.

As a result, we assigned 17 phonemes for Korean vowels, such as

monophthong : ㅏ ㅓ ㅗ ㅜ ㅡ ㅣ ㅔ
diphthong : ㅑ ㅕ ㅛ ㅠ ㅖ ㅘ ㅝ ㅚ ㅟ ㅢ

Among consonants, 'ㄹ' was used as different symbols according to position (intervocalic position or final position) which it was placed. 'ㄷ','ㅅ','ㅈ','ㅊ','ㅌ','ㅆ' and 'ㅎ' were used as one symbol when they were placed final position. 'ㅂ' and 'ㅍ' were used as one symbol when they were placed final position.

'ㄱ' and 'ㄲ' were used as one symbol when they were placed final position.

3. Experiments

3-1. Experimental Conditions

452 words were selected as database for recognition experiment. Nine persons(5 men and 3 women) participated in recording. The data were recorded two times per word and we got total 8136 word data. Four persons's (2 men and 2 women) first data were used as training data. Other 5 persons's(4men and 1 women) data and four persons's (participated for training data) second data were used as test data. Analog speech signal is sampled with 16KHz and converted to 16bit. The digitized speech is then end-pointed and analyzed with 13-th order Mel scale cepstral coefficients and added the first and second order time deritives to the base cepstral coefficients. This resulted in a total 39 values for every 10ms time slice.

HMM model was 5-state(3 observation states, an entry and exit state) left-to-right models for each of the phoneme models. And HMM model used in hybrid system was 3 states(1 observation state, an entry and exit state) and Neural Network had 1 hidden layer.

3-2. Simulation Results

In this paper, we made recognition experiments according to the number of phonemes and speech recognition unit. The recognition result according to the number of phonemes represents in Table 1. This experiment was used HMM and mono-phone. Table 1 shows that the result by using 41 phonemes is better than that by using 45 phonemes regardless of speakers participated in training or not. According to these results, we used 41 phonemes in hybrid system. Table 2 shows recognition result by using mono-phoneme and biphone as recognition unit.

Table 1. recognition results (HMM,mono-phoneme)

(a) training speaker

# of phoneme	man 1	man 2	woman1	woman2
45	92.7	75.4	93.8	65.3
41	94.0	80.1	94.0	72.6

(b) Non-training speaker

# of phoneme	man3	man4	man5	man6	woman3
45	78.1	57.1	52.7	65.9	47.8
41	83.4	66.6	60.4	74.1	52.0

Table 2 Recognition result according to recognition unit
(U1: mono-phone, U2: biphone,
M: man, W: Woman)

	M1	M2	M3	M4	M5	M6	W1	W2	W3
U1	89.6	73.9	73.9	60.8	50.4	70.6	77.7	75.9	66.6
U2	97.6	97.3	91.6	79.4	76.3	81.6	88.3	89.8	72.3

In Table.2, we found that the result variation according to speakers was great(U1: 60.8%-89.6%, U2: 72.3%-97.6%). But the result by using biphone is better than that by using mono-phone regardless of speakers. These variation was shown speaker A's 6.7%-23.3%, speaker B's 10%-23.3%, speaker C's 3.3%-10% and speaker D's 4.8%-5.8%.

4. Conclusion

In this paper, in order to study the difference of recognition result according to the number of Korean phonemes, we performed speech recognition by using OGI speech tool kit(HMM ,Neural Network and hybrid system)

We made a experiment by using 45 phonemes (which were made by classifing the diphthongs seperately) and 41 phonemes(which were made by unifyng similar pronounced vowels). When we used HMM as recognition tool, the result by using 45 phonemes was 47.8%-78.1% and the result by using 41 phonemes was 52-83.4%. And, in the hybrid system, the result of using mono-phone was 60.8%-89.6% and the result of using biphone was 72.3%-97.6%. At the above result, we knew that performance of using 41 phonemes was better than the result of using 45 phonemes and the result of using biphone was better than that of using mono-phone.

In this paper, we showed the speech recognition result according to the number of Korean vowel phonemes, but we will study the another transformation of Korean phonemes which can be used in Korean speech synthesis as well as recognition systems.

Reference

[1] H.Sakoe, "Two-Level DP matching-dynamic programming based pattern matching algorithm for connected word recognition," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-27, pp.588-595, Dec. 1979.

[2] Y.Linde, A.Buzo and R.M.Gray, " An algorithm for vector quantizer design," IEEE Trans.on Com,Vol.COM-28,Jan.,pp.84-95, 1980.

[3] Y.H.Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley Pub.Com, 1989.

[4] L.R.Rabiner and B.H.Juang," An Introduction to Hidden Markov Models," IEEE ASSP Mag., Jan.1986.

[5] M McTear. "Software to Support Research and Development of Spoken Dialogue Systems" Eurospeech, Budapest, Romania, Sep 1999.

[6] J.Schalkwyk, J.H.de Villiers, S.van Vuuren, and P.Vermeulen, "Cslush: An extendible research environment," Eurospeech 97, Sep., 1997.

[7] Y.Yan, M.Fanty, and R.Cole, "Speech recognition using neural networks with forward-backward probability generated targets," Proceedings of the International Conference on Acoustic, Speech and Signal Processing, vol. IV, pp. 3241-3244, 1997.

[8] Sutton, S., Novick, D. G., Cole, R., and Fanty, M., "Building 10,000 spoken-dialogue systems." Proceedings of the International Conference on Spoken Language Processing, Philadelphia, PA, Oct., 1996.

[9] W.Y.Hwang, R.P.Lippmann and B.Gold," A Neural Net Approach to Speech Recognition," in Proc.Int.Conf. on Acoust., Speech, Signal Processing, pp.97-102,Apr.1988.

---

*본 연구는 2000년 관동대학교의 교내 연구비 지원에 의한 것임.