THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.23, NO.4E 2004. 12 pp. 134~139

A Study on the Phonemic Analysis for Korean Speech Segmentation

Sou-Kil Lee*, Jeong-Young Song*
*Dept. of Computer Engineering, PaiChai University
(Received September 16 2004; rivised December 2 2004; accepted December 27 2004)

Abstract

It is generally known that accurate segmentation is very necessary for both an individual word and continuous utterances in speech recognition. It is also commonly known that techniques are now being developed to classify the voiced and the unvoiced, also classifying the plosives and the fricatives. The method for accurate recognition of the phonemes isn't yet scientifically established. Therefore, in this study we analyze the Korean language, using the classification of "Hunminjeongeum" and contemporary phonetics, with the frequency band, Mel band and Mel Cepstrum, we extract notable features of the phonemes from Korean speech and segment speech by the unit of the phonemes to normalize them. Finally, through the analysis and verification, we intend to set up Phonemic Segmentation System that will make us able to adapt it to both an individual word and continuous utterances.

Keywords: *Speech recognition, Phonemic segmentation, Phonemic analysis, Hunminjeongeum*

I. Introduction

This is a study on the setting up of the Phonemic Segmentation System for Korean speech recognition. The phonemes, which are the minimal meaning distinguishing units of a voice, play a leading role in effective speech recognition. Speech recognition is more effective and helpful by using phonemes than by using syllables or words. A small amount of countable works brings out the basis of the setup for an unlimited recognition of continuous speech.

It is generally known that accurate segmentation is very necessary for both an individual word and continuous utterances in speech recognition. It is reported that the study at this point is actively underway. It is also commonly known that techniques are now being developed to classify the voiced and the unvoiced, also classifying the plosives and the fricatives. The method for accurate recognition of the phonemes isn't yet scientifically established. The vowels are periodical, long in duration and powerful in energy. Contrary to the vowels, the consonants are relatively non-periodical, short in duration and less powerful in the spectral energy, which gives us the functions of importance to classify the voice[1].

Therefore, in this study we analyze the Korean language, using the classification of "Hunminjeongeum" and contemporary phonetics, with the frequency band, Mel band and Mel Cepstrum. We extract notable features of the phonemes from the Korean speech and segment speech by the unit of the phonemes to normalize them[2,3].

Finally, through the analysis and verification, we intend to set up the Phonemic Segmentation System that will make us able to adapt it to both an individual word and continuous utterances in speech recognition.

Corresponding author: Sou-Kil Lee (sklee2@mail.pcu.ac.kr)
Lab of Jeong-Young Song of Computer Engineering Dept., Paichai University, Doma-Dong Seo-Gu, Daejeon 302-735, Republic of Korea

Table 1. Monophthong

Position / Height	Front	Middle	Back
High vowel	이[i]	으 [ɯ]	우[u]
High-mid vowel	에[e]	어 [ɔ]	오[o]
Low-mid vowel	애[ɛ]		
Low vowel		아[a]	

Table 2. Diphthong

rounded	위[wi],웨 [we],왜 [wɛ],워[wɔ],와[wa]
unrounded	예[ye], 얘[yɛ], 여[yɔ], 야[ya], 유[yu], 요[yo], 의[iy]

II. Korean Speech

2.1. Setting up of the Vowels

In general, vowels are classified as Monophthongs and Diphthongs in Korean Speech.

2.1.1. Monophthongs

Classifying of monophthongs depends mainly on the height and position of the tongue. Monophthongs are classified as in table 1 and when they are pronounced they keep similar vocal tract shapes from the beginning to the end.

2.1.2. Diphthong

Diphthongs are classified into two groups by the shape of the lips; rounded, unrounded as in table 2.

They are made when each vowel meets the semi-vowel. Semi-vowel is characteristically pronounced when the tongue is being moved from the location of a vowel to another. It looks like a vowel but works like a consonant. It has a feature of the consonants that could only be made before or after a vowel.

2.2. Setting up of the Consonants

The consonants can be classified simply by the voicing, the position and the manner.

In contemporary phonetics, the consonants can be classified into 2 groups; voiced consonants, unvoiced consonants according to the vibration of the vocal band. And by a phonemic system they could be classified as Bilabial, Front tongue, Back tongue sound according to the position of tongue. And they could be classified as Plosives, Affricates, Fricatives, Nasals, Liquids according to the manner of production.

Table 3. Consonant System

Position / Manner	Bilabial	Front tongue	Back tongue
Plosives	ㅂ[b],ㅃ[pp],ㅍ[p]	ㄷ[d],ㄸ[tt],ㅌ[t]	ㄱ[g],ㄲ[kk],ㅋ[k]
ffricates		ㅈ[z],ㅉ[zz],ㅊ[ʧ]	
ricatives		ㅅ[s]ㅆ[ss]	ㅎ[h]
Nasals	ㅁ[m]	ㄴ[n]	ㅇ[ŋ]
Liquids		ㄹ[l]	

Table 4. Consonant System (Hunminjeongeum)

Position / Manner	Back teeth	Tongue	Lips	Teeth	Throat	Semi tongue
Lax	ㄱ[g]	ㄷ[d]	ㅂ[b]	ㅅ[s],ㅈ[z]		
Aspirated	ㅋ[k]	ㅌ[t]	ㅍ[p]	ㅊ[ʧ]	ㅎ[h]	
Tense	ㄲ[kk]	ㄸ[tt]	ㅃ[pp]	ㅆ[ss],ㅉ[zz]		
Sonorant	ㅇ[ŋ]	ㄴ[n]	ㅁ[m]			ㄹ[l]

Table 4. was made and introduced by King Sejong in 1446, which is called Hunminjeongeum. This table could be classified with the Back teeth, Tongue, Lips, Teeth, Throat, Semi tongue sound according to the utterance position. And they could be classified as Lax, Aspirated, Tense, and Sonorant according to the manner in which they are made. Consonants will be classified according to Hunminjeongeum's way.

III. The Setup of Phonemic Segmentation System

3.1. Condition

In this test we used Pentium IV, pronouncing in speeches are speaker-dependent. We also the voices sampled at 16kHz, 16bit and in mono. Speeches were made at office by a researcher, who is male and 49 years old, who uses standard language with no dialectal features.

Figure 1. depicts the process of the phonemic segmentation system. From the input to the output, several technical steps are required to gain the results we are pursuing.

After doing an FFT (1) conversion of speech, we sought frequency spectrum. And then by using MF (2) that could transform the acoustic frequency scale into an auditory one, we converted and normalized them again by doing a DCT (3)

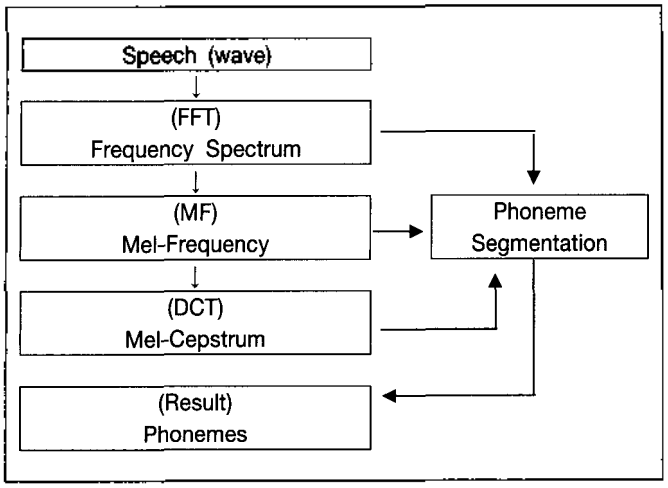

Figure 1. Phonemic Segmentation System

conversion resulting in getting Mel Cepstrum[4-9].

$$F = FFT(wave) \quad (1)$$

$$\overline{S_k} = MF(F) = 1000/\log 2 * \log(1+F/1000) \quad (2)$$

$$MC = DCT(\overline{S_k}) = \sum_{k=1}^{K} \log(\overline{S_k}) \cos(n(k-0.5)\frac{\pi}{K}) \quad (3)$$

3.2. Segmentation

We classified the voicing feature into several groups as shown in table 5. The voiced group is classified into Vowels, Sonorant, Voiced Consonant. The unvoiced group is classified into Lax, Aspirated, Tense.

We could let each picture show its own F/kHz within 8kHz, second, MF, MC. This is for concurrent comparison.

With the MC graphics changes in MC will be easily noticed.

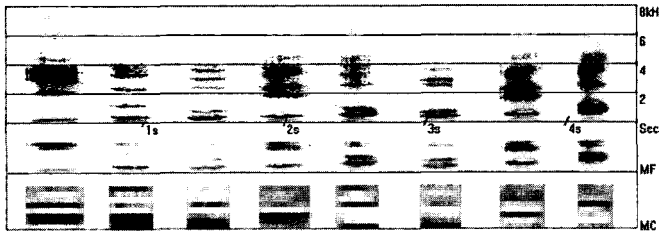

Figure 2. Monophthong : 이,으,우,에,어,오,애,아[i, ,u,e,ə,o, ,a]

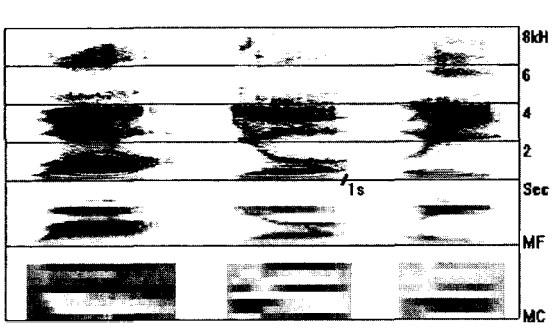

Figure 3. Diphthong: 와, 여, 의[wa,yə,iy]

Table 5. Speech System

Voiced	Vowels (monophthong, diphthong)
	Sonorant
	Voiced Consonant
Unvoiced	Lax
	Aspirated
	Tense

3.2.1. Classifying the Voiced Segments

We classified the voiced segments into 4 groups as follows: Monophthong (Figure 2), Diphthong (Figure 3), Sonorant (Figure 4,5), Voiced Consonant (Figure 6).

As in picture Figure 2, it is easily noticeable that the formant and spectral energy of vowels remain consistent. Each monophthong also shows us its shape in Mel Cepstrum through DCT conversion which remains consistent.

Figure 3 is 와[wa], 여[yə], 의[iy] which are a part of

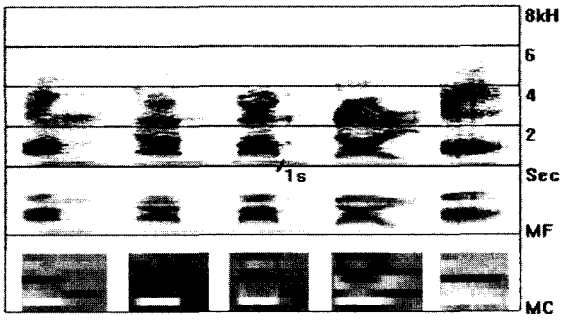

Figure 4. Sonorants: 앙,난,맘,랄,아[a ŋ ,nan,mam,lal,a]

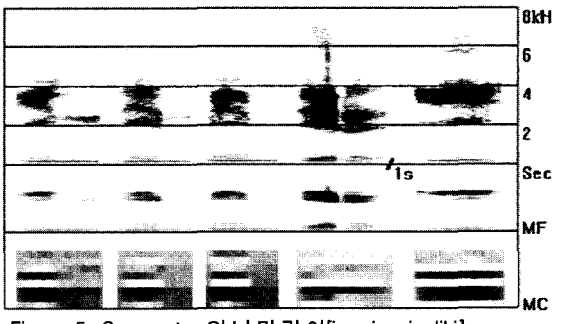

Figure 5. Sonorants: 잉,닌,밈,릴,이[i ŋ ,nin,mim,lil,i]

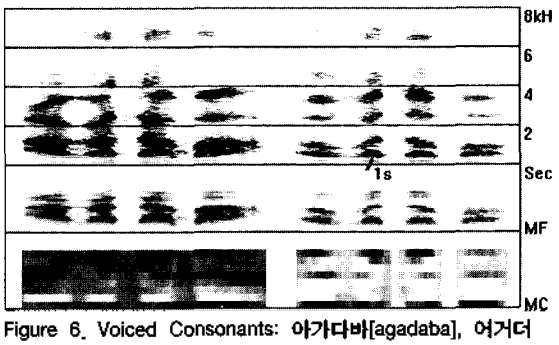

Figure 6. Voiced Consonants: 아가다바[agadaba], 어거더버[əgədəbə]

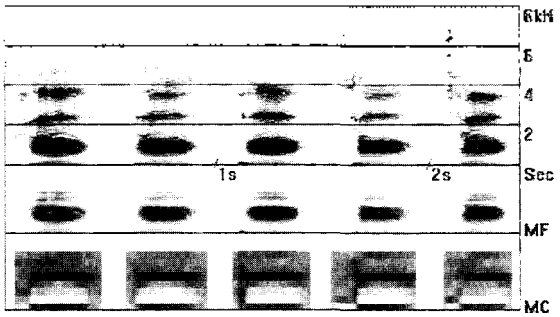

Figure 7. Lax 1: 가,다,바,사,자[ga,da,ba,sa,za]

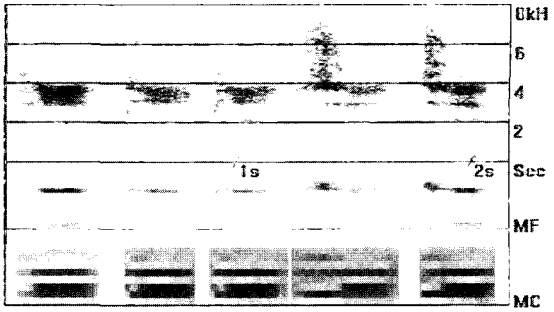

Figure 8. Lax 2: 기,디,비,시,지[gi,di,bi,si,zi]

Diphthongs. We can see the formant of the diphthongs in the picture is moving to the direction of a monophthong "ㅏ[a], ㅓ[ə], ㅣ[i]" at the point of the start. In diphthongs, the position of formants change.

Figure 4 shows Sonorants which meet to form monophthong "ㅏ[a]". The one on the far right is pure monophthong "ㅏ[a]". Figure 5 on the right shows Sonorants which meet to form monophthong "ㅣ[i]". The far right shows pure monophthong "ㅣ[i]". We used "ㅏ[a]" and "ㅣ[i]" to make comparisons easily. Each syllable appears a monophthongal shape as in Figure 2 in Mel Cepstrum which remains consistent.

Sonorants are classified by nasals and liquid. The nasals(ㅇ[ŋ], ㄴ[n], ㅁ[m]) make new formants and leave it remain, with a stronger spectral energy at the low frequency region. The formants of liquid(ㄹ[l]) move toward those of the following vowel.

Figure 6 shows Voiced Consonants between the vowels. The left part shows monophthong "ㅏ[a]" and ㄱ[g], ㄷ[d], ㅂ[b] meet monophthong "ㅏ[a]". The right part shows monophthong "ㅓ[ə]" and ㄱ[g], ㄷ[d], ㅂ[b] meet monophthong "ㅓ[ə]". This also shows us monophthong shape as in Figure 2. The voiced consonants between the vowels change the preceding and following vowel formants to separate directions.

3.2.2. Classifying of the Unvoiced Segments

We classified the unvoiced segments into the Lax, the Aspirated, and the Tense as following description. We used "ㅏ[a]" and "ㅣ[i]" to make comparisons easily. Especially, the reason why we used them is they have strong("ㅏ[a]") and weak("ㅣ[i]") energy.

Figure 7 shows the Lax(ㄱ[g], ㄷ[d], ㅂ[b], ㅅ[s], ㅈ[z]) meet monophthong "ㅏ[a]". Figure 8 shows the Lax (ㄱ[g], ㄷ[d], ㅂ[b], ㅅ[s], ㅈ[z]) meet monophthong "ㅣ[i]".

The "ㄱ[g], ㄷ[d], ㅂ[b]" shows a low and high frequencyfor a longer time. The "ㅅ[s], ㅈ[z]" shows strong high frequency for a longer time. The "ㅅ[s], ㅈ[z]" appear only at the initial part of a syllable.

Figure 9 shows the Aspirated(ㅋ[k], ㅌ[t], ㅍ[p], ㅊ[ʧ], ㅎ[h]) meet monophthong "ㅏ[a]". Figure 10 shows the Aspirated(ㅋ[k], ㅌ[t], ㅍ[p], ㅊ[ʧ], ㅎ[h]) meet monophthong "ㅣ[i]".

The "ㅋ[k], ㅌ[t], ㅍ[p]" shows stronger spectral energy. "ㅎ[h]" shows low and high frequency for a longer time. Especially, "ㅊ[ʧ]" shows a stronger spectral energy than the other two sets.

Figure 11 shows the Tense(ㄲ[kk], ㄸ[tt], ㅃ[pp], ㅆ[ss], ㅉ[zz]) meeting monophthong "ㅏ[a]". Figure 12 shows the Tense(ㄲ[kk], ㄸ[tt], ㅃ[pp], ㅆ[ss], ㅉ[zz]) meet monophthong "ㅣ[i]". The "ㄲ[kk], ㄸ[tt], ㅃ[pp]" shows low

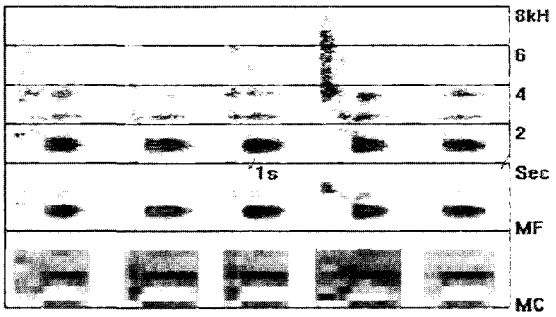

Figure 9. Aspirated 1:카,타,파,차,하[ka,ta,pa,ʧa,ha]

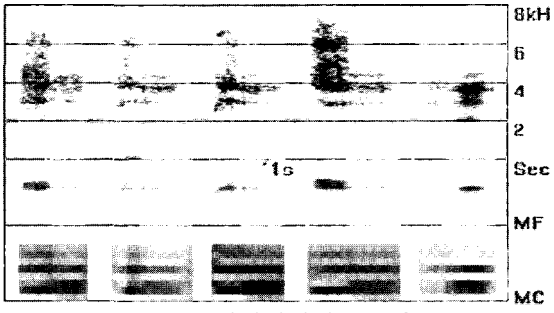

Figure 10. Aspirated 2:키,티,피,치,히[ki,ti,pi,ʧi,hi]

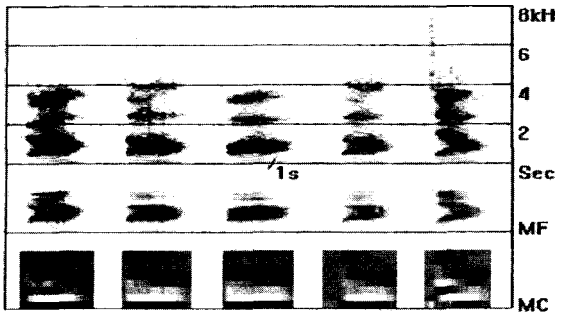

Figure 11. Tense:까,따,빠,싸,짜[kka,tta,ppa,ssa,zza]

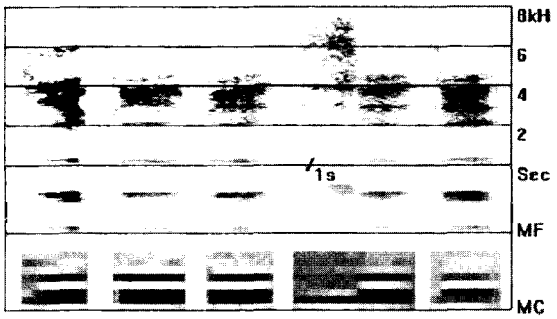

Figure 12. Tense: 끼,띠,삐,씨,찌[kki,tti,ppi,ssi,zzi]

frequency for a shorter time and changes the formants. The "ㅆ[ss]" shows high frequency for a longer time and changes the formant

The "ㅉ[zz]" shows high frequency for a shorter time.

IV. Analysis and Evaluation

4.1. Voiced

From the previous section, it can be observed that the formants and spectral energy of each monophthong remain consistent in MC. We can see that the formant of the diphthong is moving to the direction of a monophthong from the very beginning. The Voiced Consonants between the vowels change the preceding and following vowel formants to separate directions. The nasals make new formants and leave it remain, with a stronger spectral energy at the low frequency region. The liquid changes the formant of vowels.

Capturing the shape, the stableness and the changes of the formants well can be a leading key in segmentation of the phonemes.

4.2. Unvoiced

The Lax consonants show a low and high frequency for a longer time. The Aspirated consonants show stronger spectral energy. Especially, "ㅊ[ʧ]" shows stronger spectral energy. The Tense consonants show low and high frequency for a shorter time and changes the formants. Each unvoiced consonant followed by a vowel also shows us monophthong shape (Figure 2) in Mel Cepstrum which remains consistent.

Getting the spectral energy, the position and duration successfully can be also very important in segmentation of unvoiced.

4.3. Evaluation (Voice test)

We pronounced common words quickly and naturally for about two second. Figure 13 shows "아버지[abəzi], 어머니[əməni], 선생님[sənsæŋnim]" that is made of 3 words, and of 9 syllables, also of 19 phonemes.

Each vowel of "아버지(ㅏ[a], ㅓ[ə], ㅣ[i]), 어머니(ㅓ[ə], ㅓ[ə], ㅣ[i]), 선생님(ㅓ[ə], ㅐ[æ], ㅣ[i])" shows us its shape like Figure 2. But, we found it difficult to distinguish the "ㅓ[ə]" from the "선[sən]" because the shape of vowels was changed a lot under the influence of the consonants(ㅅ[s]), and we will regard it as an allophone.

All the consonants change the formants of vowels that is preceding and following. In "아버지[abəzi]", the latter part of "아[a]" and the first part of "어[ə]" are bent downwards and

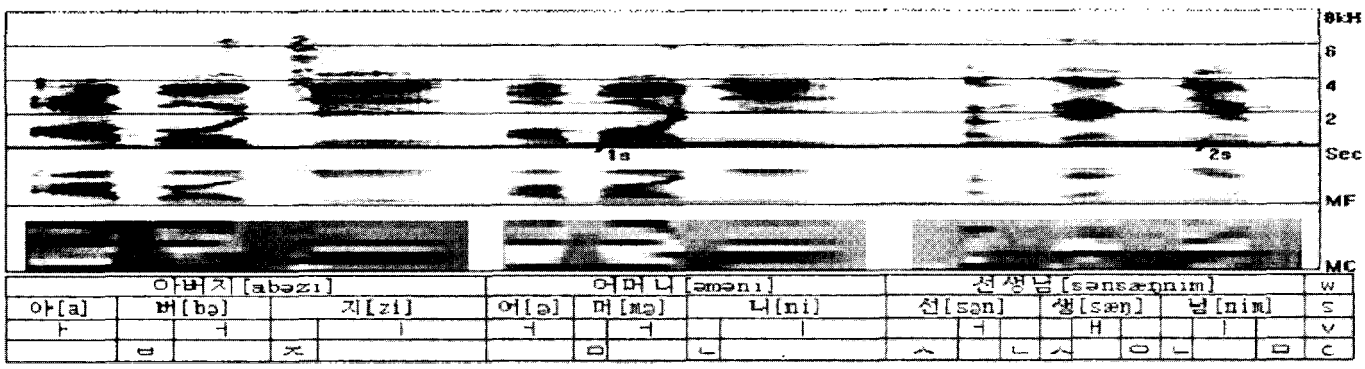

Figure 13. 아버지[abəzi], 어머니[əməni], 선생님[sənsæŋnim]

the latter part of "어[ə]" is bent upwards. This fact is also found in "어머니[əməni]".

Each consonant "버(ㅂ[b]), 지(ㅈ[z]), 머(ㅁ[m], 너(ㄴ[n]), 선(ㅅ[s], ㄴ[n]), 생(ㅅ[s], 0[ŋ]), 님(ㄴ[n], ㅁ[m])" shows us its shape like previous Figure 4-8. The "ㅅ[s]" of "선[sən]" shows high frequency like Figure 7, 8. But, we found it difficult to distinguish the"ㅅ[s]" from the "생[sæŋ]" because the shape of spectrum was changed much under the influence of vowels, and we will also regard it as an allophone. Formants of "ㅇ[ŋ], ㄴ[n]" remain consistent like previous Figure 4,5.

In the voice test, though pronounced very quickly in this picture, we could see the shape of vowels change gradually under the influence of the consonants and segment the syllables by Frequency Spectrum, Mel Frequency, Mel Cepstrum and get phonemes as a result.

V. Conclusions

In this study we analyze the acoustic features in the Korean speeches on the basis of Hunminjungeum and contemporary phonetics.

And using Frequency band, Mel band, Mel Cepstrum through DCT conversion, to some extent we could segment speech by the unit of phonemes and normalize them and to make it recognized.

We could segment the voice by the unit of phonemes in Mel band and see the changes of formant in vowels. Each vowel shows us that Mel Cepstrum is very stable and we can distinguish the phonemes that have weaker energy.

Therefore, if the study on the transition of vowel formants resulted from the continuous sounds of Korean speeches, along with the study on the ending part of sounds is added. Thus, with the establishment of the Phonemes dictionary using Mel Cepstrum, it can be better applicable for a machine recognition of the speech more effectively and robustly than with the acoustic frequency scale information alone.

Reference

1. Lawrence Rabiner, "Fundamentals of Speech Recognition," Prentice Hall, 1993.
2. Juchae Bae, "Synopsis of Hangul phonology," Shingu Culture Publishers, 1997.
3. Changyun Yu, "Notes on Hunminjungeum," Hyoungsul Publishers, 1998.
4. Younghuan O, " Management of voice data," Heungryung Science Publishers, 1998.
5. Jinsoo Han, "Management of Voice Signal," Osung Media, 2000.
6. Peter B, Denes, Elliot N Pinson, "The Speech Chain," Hanshin Culture Books, 1999.
7. John Clark, Collin Yallop, "Phonetics & phonology" Hanshin Culture Books, 1998.
8. Byunggon Yang, "Theory & Practicals of Voice Analysis Using Praat," Mansu Publishers, 2003.
9. Sou-Kil Lee, Jeong-Young Song, "A Study on the Phonemic Analysis for Korean Speech," Proceeding of the Electronics, Information and Systems Conference Electronics, Information and Systems Society, I.E.E. of Japan, 2003.

[Profile]

•Sou-Kil Lee

Sou-Kil Lee received the M.S. degree in Computer Science from Hanbat University, Korea, in 2000. Since 2002 he has been studying in PaiChai University for speech recognition and is taking a course of getting a Ph.D. in that field. He is particularly interested for accurate speech recognition.

•Jeong-Young Song

Jeong-Young Song received the B.S. degree in Computer Science from Hannam University, Korea, in 1984, the M.S. degree in Electrical Engineering and information Engineering from Waseda University, Japan, in 1992 and the Ph.D. in Electrical Engineering and information Engineering from Waseda University, Japan, in 1995. From 1995 to 1997, he was a Professor, ChungUn University HongSeong, Korea. Since 1997 he has been a Professor in Dept. of Computer Engineering, PaiChai University DaeJeon, Korea. His current research interests are in Pattern Processing, Speech Processing and Human Interface Technology.