論文 87-24-2-15

連續音 分類認識에서 G-peak를 이용한 鼻音의 分類

(The Extraction of Nasal Sound by Using G-peak in Continued Speech)

裵明振*, 鄭益周*, 安秀桔*

(Myung Jin Bae, Ik Joo Chung and Souguil ANN)

要 約

本 論文에서는 連續音에서 鼻音을 抽出해 내는 새로운 알고리즘을 제시하였다. 면적비교법에 의하여 pitch를 구하고 한 pitch 區間에서 G-peak의 面積과 Side-peak의 面積을 비교하므로서 鼻音을 추출하였다. 本 알고리즘은 처리속도가 향상되어 凡用 Microprocessor에 의해서도 실시간 처리가 가능하다.

Abstract

In this paper, we describe a new algorithm for extracting nasal sound in continuous speech. We obtain pitches by using Area Comparison Method and extract nasal sound by comparing the area of G-peak and the area of side peak in one pitch interval.

By using this method, the process can be speeded up. Therefore realtime processing is possible with a general microprocessor.

Ⅰ. 序 論

音聲認識이란 인간의 기본적인 의사전달 수단인 音聲을 Man/Machine Interface로 이용하고자 하는 것으로, 現代社會가 情報社會로 발달하고 디지탈 신호처리 기술과 반도체 기술의 급속한 성장에 따라 音聲認識에 대한 연구와 그 응용이 활발히 추진되고 있다.

音聲認識을 크게 분류하면 孤立單語 認識과 音素單位 認識으로 구분할 수 있다. 그리고 音聲認識의 궁극적인 목표는 連續音 認識이고 이것이 이루어질 때 그 파급효과는 대단히 크다. 韓國語 連續音 認識을 孤立單語 認識의 방향에서 접근하면 적은 單語 認識을 행할지라도 助辭 또는 語尾변화에 의하여 많은 data base가 필요하게 되고 認識 시간도 길어지게 된다. 이러한 문제점 때문에 韓國語의 連續音 認識은 音素單位 認識이 바람직하나 현재로서는 孤立單語 認識에 비하여 어렵고 認識率도 떨어지고 있다.

連續音을 音素單位로 認識할 때 preprocessing으로서 音素를 몇가지로 分類하는 分類認識 단계를 거쳐 最終認識을 하는 것이 일반적인 추세이다. 지금까지의 分類認識은 주로 有聲音(Voiced), 無聲音(Unvoiced), 默音(Silence)으로 分類하는 정도에서 그쳤다. 그러나 分類認識의 중요성이 강조되면서 分類認識 단계에서 좀 더 세밀한 分類가 요망되어지고 있다. 특히, 有聲音은 다시 여러 部類로 分類되며 鼻音은 그 특수한 성질 때문에 한 部類를 형성하고 있다. 따라서 有聲音 중에서 鼻音의 分類가 分類認識 단계에서 이루어지므로서 最終認識 단계를 줄일 수 있다.

Ⅱ. 分類認識

音素單位의 認識에서 最終認識을 하기 전에 認識되어질 音素를 몇가지 部類중에 하나로 分類하는 과정을 分類認識이라 한다. 分類認識 과정을 거치므로서 얻는 잇점은 다음과 같다.

1. 最終認識 과정에서 미리 分類된 성질대로 처리

*正會員, 서울大學校 電子工學科

(Dept. of Elec. Eng., Seoul National Univ.)

接受日字 : 1986年 7月 30日

함으로서 認識率이 높아진다.

2. 미리 分類가 되었으므로 最終認識 과정에서 認識의 범위가 좁아져 最終認識에 소요되는 시간이 줄어든다.

우선 韓國語의 音을 言語學的으로 分類하면 그림1과 같다.[12]

그러나 실제 音聲認識을 위한 分類認識에서는 言語學的 分類보다는 보통 그림 2 와 같이 分類한다.[11]

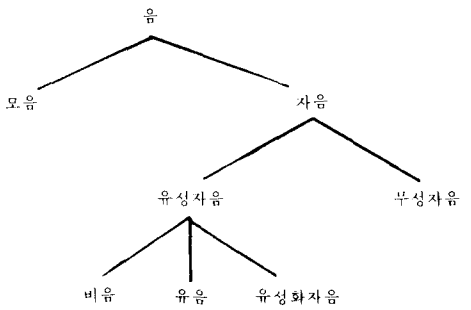

그림 1. 음의 언어학적 분류
Fig. 1. A classification of phoneme based on linguistics.

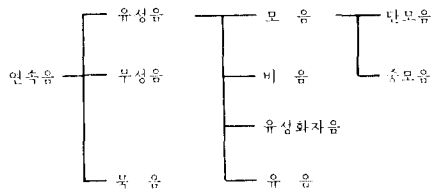

그림 2. 음성인식에 대한 음소의 분류
Fig. 2. A classification of phoneme based on a speech recognition.

1차 分類認識에 해당하는 有聲音, 無聲音, 默音의 區分은 Patterm recognition Approach, LPC distance measure를 이용한 방법 등이 소개되었고[1] 비교적 높은 認識率(95~97%)를 보여 주고 있다.[11] 이러한 알고리즘을 韓國語에도 적용하여 좋은 결과가 보고되고 있다.[11]

鼻音은 子音에 속하므로 子音을 다시 한번 살펴보면 子音은 鼻音(ㄴ, ㅁ, ㅇ), 流音(ㄹ), 有聲化子音, 無聲子音으로 이루어진다. 鼻音과 流音을 제외한 子音은 원칙적으로 모두 無聲音이다.[12] 그러나 音素가 놓이는 위치에 따라 有聲化되어 有聲化子音이 된다. 가령 "김", "돌", "불"에의 "ㄱ", "ㄷ", "ㅂ"은 無聲音이나 "밑음", "공부"에서의 "ㄷ", "ㅂ"은 母音과 母音 사이, 鼻音과 母音 사이에 오므로서 有聲化되어 有聲化子音이 된다.

分類認識에서 매우 중요한 요소는 수행 속도이다. 最終認識과는 달리 分類認識 단계에서는 빠른 속도로 分類하여 最終認識 단계에 넘겨 주어야 한다. 또한 分類認識의 範圍를 정하는 것도 중요하다. 왜냐하면 分類認識 단계에서 너무 세밀히 分類하려 하면 그것에 필요한 여러가지 parameter들이 필요하게 되며 결국은 最終認識에 匹敵할 정도로 처리과정이 복잡해지고 많은 시간을 要하게 되어 分類認識으로서의 의미가 없어지게 된다.

Ⅲ. 鼻音의 特性

우리의 發聲器官은 音을 發生시키는 音原(Sound source)과 거기서 發生된 音이 거치는 聲道(Vocal tract)로 이루어져 있다. 그림 3 은 發聲器官을 나타낸 그림이다.

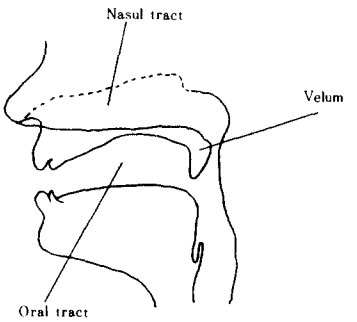

그림 3. 발성기관
Fig. 3. An articulatory organ.

有聲音과 無聲音의 구분은 音原에 따른 分類이다. 즉 聲帶(Vocal cord)의 떨림은 有聲音의 音原이며 White noise와 같은 Noise 成分은 無聲音의 音原이 된다.[13] 그리고 근육을 통하여 聲道의 모양을 바꾸므로서 여러가지의 音素를 發音할 수 있는 것이다.

聲道는 Nasal tract와 Oral tract로 나누어 진다. 이러한 구분은 입천장 뒷 부분에 있는 軟口蓋(velum) 를 분기점으로 해서 나누어 진다. 鼻音이 아닌 音을 發音할 때는 이 軟口蓋가 Nasal tract의 入口를 막고 있다. 그러나 鼻音을 發音하게 되는 경우는 이 軟口蓋가 열려 대부분의 音聲 에너지가 Nasal tract를 거쳐 코로 나오게 된다. 鼻音과 鼻音이 아닌 有聲音 모두 音原은 聲帶의 떨림이므로 이를 區分짓는 것은 바로 그것이 어떠한 tract를 거치느냐에 의해서이다.

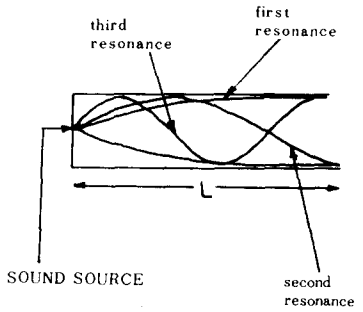

그림 4. 성도의 파이프 모델
Fig. 4. Pipe model of the vocal tract.

聲道는 매우 복잡한 구조를 가지나 그림4와 같이 한쪽에 音原이 있고 또 한쪽은 열린 管으로 Modeling 할 수 있다.

管의 길이를 ℓ이라 하면 이 管에 의하여 생기는 resonance의 波長은 4ℓ, $4\ell/3$, $4\ell/5$,…가 되고, 각각 그들의 周波數는 $C/4\ell$, $3C/4\ell$, $5C/4\ell$,…이 된다.[13] 여기서 C는 sound의 속도이다. 成人의 聲道의 길이를 대략 17cm, sound의 속도를 340m/sec라 하면 resonance 周波數들은 500, 1,500, 2,500…(Hz)가 된다. 이러한 聲道의 resonance 現象은 Speech wave의 에너지 스펙트럼에서 fromant들로 나타난다. 그림 5는 이러한 formant들을 나타낸다.

Nasal tract와 oral tract의 현저한 차이점은 Nasal tract가 oral tract 보다 길다는 점이다. 이는 그림 4에서 resonance 波長(ℓ)이 길어진다는 것을 의미하고 따라서 formant 周波數가 낮아지게 된다. Formant 周波數 F1은 F2보다 일반적으로 10dB 이상 높으므로

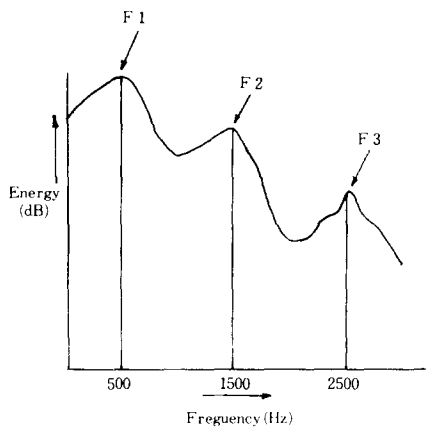

그림 5. 음성의 에너지 스펙트럼
Fig. 5. The energy spectrum of speech.

speech wave는 周波數 Domain 상에서 resonance 周波數 F1의 gain와 F1의 bandwidth의 영향이 지배적이 된다.[3] 각 音素들의 F1을 구해 보면 표1과 같다.

표 1. 음소들의 F1값
Table 1. F1 of phonemes.

Phoneme	F1(Hz)
아	720
에	560
이	360
오	600
우	380
ㄹ	380
ㅁ	190
ㄴ	190
ㅇ	190

위의 표에서 알 수 있듯이 鼻音의 F1은 모두 190Hz 정도로서 여타의 母音들 보다 작다는 것을 알 수 있다. 이러한 特徵은 Time domain 상에서 鼻音의 waveform에 곧바로 나타난다. 그림 6은 鼻音이 아닌 有聲音 (으)에서 鼻音 (ㅁ)으로 넘어가는 과정의 Speech waveform이다. "으"에 해당하는 部分은 "ㅁ"에 해당하는 部分보다도 3배 정도 높은 周波數로 減衰振動하는 것을 알 수 있다.

여기서 우리는 한 pitch 안에서 처음 오는 peak를 G-peak라고 정의하고[15] 나머지 peak들을 Side-peak라 하자. G-peak는 glottal 성분이 지배적인 peak라는 의미로서 일반적으로 side-peak들 보다 面積이 두배 가까이 크다.

그림 6에서 알 수 있는 鼻音의 特徵은 다음과 같다. 첫째 한 pitch 안에 G-peak만이 존재한다(예외 있음). 둘째 G-peak interval(D)가 鼻音이 아닌 有聲音 보다

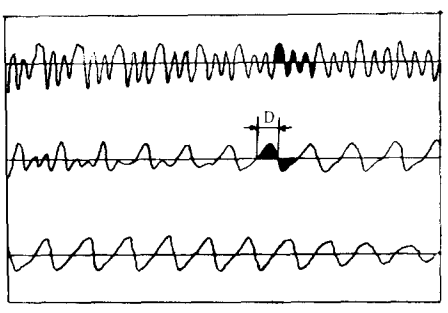

그림 6. 비음의 음성파형
Fig. 6. The speech waveform of a nasal sound.

길다. 이러한 性質들이 鼻音을 抽出해 내는 parameter로서 이용되어질 것이다.

마지막으로 鼻音의 音聲學的 特徵을 살펴보면 鼻音은 連續音에서 인접해 있는 다른 音素들을 鼻音化시켜 그 音素의 性質을 잃게 한다. 이것은 連續音 認識에 큰 장애 요소가 되고 있다. 특히 鼻音과 鼻音 사이에 오는 母音은 거의 완전히 鼻音化되어 고유의 性質을 잃게 된다. 또 鼻音과 "ㄹ" 또는 "ㅣ"가 인접하면 "ㄹ"이나 "ㅣ"가 鼻音의 性質을 띤다. 이는 "ㄹ"과 "ㅣ"의 F1이 여타의 音素들 중에서 鼻音에 제일 가깝기 때문이다.

Ⅳ. 알고리즘

Pitch가 긴 사람의 경우(150Hz 이하) 鼻音에서도 한 pitch 내에 G-peak 이외에 Side-peak가 존재하는 경우가 있다. 그러나 그림 7에서 보듯이 그러한 경우라도 매우 작은 Side-peak가 존재하게 된다. 이는 鼻音의 發生時 oral cavity에서 zero에 의해 pole이 상쇄되었기 때문이다.[1]

한 pitch 내에서 G-peak는 Side-peak 보다 amplitud가 크고 또 G-peak interval(D)가 크므로 이 두 변수를 하나로 합칠 수 있는 것은 面積이 된다. 따라서 鼻音은 한 pitch 내에 G-peak 만이 존재하거나 또는 G-peak의 面積대 Side-peaks의 面積들의 合의 比가 매우 크다는 것을 알 수 있다. 그림7-b는 鼻音과 여러 母音들의 G-peak의 面積, 같은 pitch 내에 존재하는 Side-peak들의 面積合 및 그들의 比를 정량적으로 나타낸 것이다. 여기서 비음 부분의 "*" 표시는 Side-peak가 존재하지 않았음을 나타낸다.

그림 8은 鼻音 抽出에 기본적인 flowchart를 나타낸다. 여기서 K는 통계적으로 구하여 20을 사용하였다.

Ⅴ. 實驗 및 結果

實驗을 위한 音聲 信號로는 "서울대 전자공학과 음

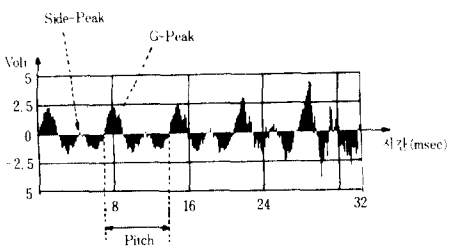

그림 7 (a). 비음의 G-peak와 Side-peak("네"音)
Fig. 7 (a). G-peak and Side-peak of nasal sound for "Ne" speech.

	G-Peak의 면 적	Side-Peak의 면적의 합	Side-Peak면적 / G-Peak면적 (%)
아	509	211	41.4
이	455	72	15.8
우	610	121	19.8
애	475	140	23.1
오	474	150	31.6
으	317	222	70.0
어	635	214	33.7
ㅁ	472	*	*
ㄴ	410	*	*
ㅇ	462	*	*
ㄹ	317	19	6.0

그림 7 (b). 한 pitch 구간 내에서 G-peak와 Side-peak의 평균 면적비
Fig. 7 (b). An average ratio of area between G-peak and Side-peak in a pitch interval.

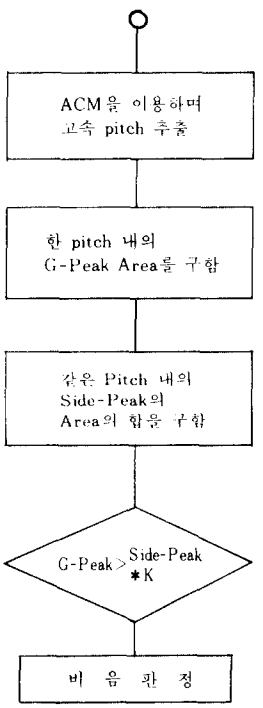

그림 8. 비음 추출에 관한 흐름도
Fig. 8. The flowchart of nasal sound extraction.

성신호 처리팀이다." "인수네 꼬마는 천재 소년을 좋아한다." "예수님께서는 천지 창조의 교훈을 말하셨다."의 각각 3초씩의 서로 다른 세 話者에게서 얻은 total 9초간의 sampling data를 사용하였다. Sampling 에는 12bit의 A/D 변환기를 사용하였고 12bit중 상위 8bit

만을 이용하였다. 이렇게 한 이유는 鼻音이 有聲音에 속해 에너지가 크고, 無聲音의 영향이 어느 정도 輕減되면서 처리 속도를 개선할 수 있기 때문이다. 모든 data의 처리는 16bit 퍼스날 컴퓨터인 IBM-PC/XT 를 사용하였다.

그림 9는 "서울대 전자공학과 음성신호 처리팀이다."와 "예수님께서 천지창조의 교훈을 말하셨다."의 Sampling data에 대한 結果로서 波形과 G-peak와 Side-peak의 면적 그리고 最終的으로 鼻音에 해당하는 部分을 찾아내었다. 結果에서 알 수 있듯이 모든 鼻音과 鼻音化된 音素, 그리고 경우에 따라 流音(ㄹ)을 찾아내고 있다. 다음 아래 文章들 중에서 굵은 글씨는 鼻音이 아닌데 鼻音으로 判定한 部分이다.

"서울대학교 전자공학과 음성신호 처리팀이다."

"인수네 꼬마는 천재 소년을 좋아한다."

"예수님께서 천지 창조의 교훈을 말하셨다."

위의 結果를 요약하면 제법 길게 發音되는 流音은

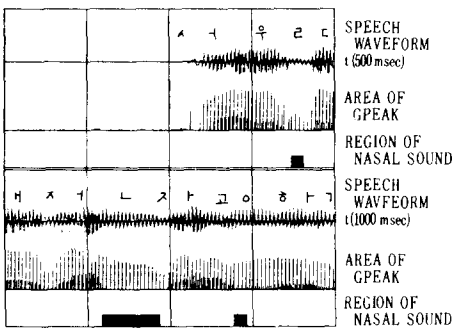

그림9 (a). 음성신호 "서울대 전자공학"에 대한 실험결과

Fig. 9 (a). The result for speech "seouldae jeonjagonghak".

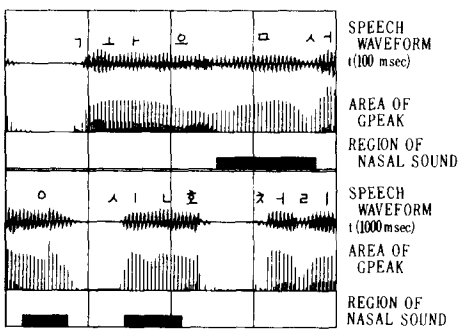

그림9 (b). 음성신호과 "음성신호처리"에 대한 실험결과

Fig. 9 (b). The result for speech "gwa eumseong sinho cheori."

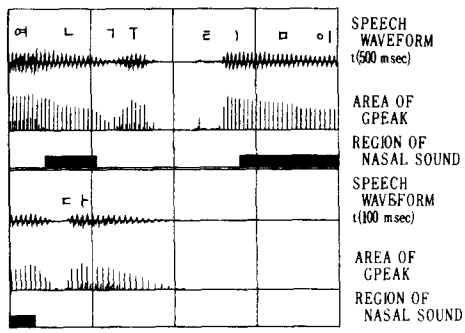

그림9 (c). 음성신호 "연구팀이다"에 대한 실험결과

Fig. 9 (c). The result for speech "yongutim ida."

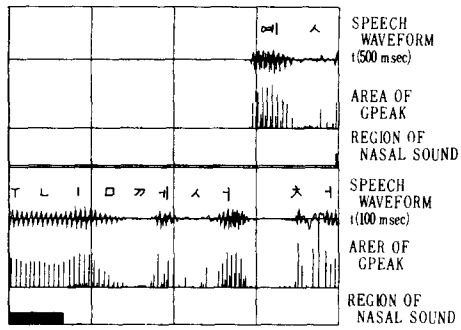

그림9 (d). 음성신호 "예수님께서 쳐"에 대한 실험결과

Fig. 9 (d). The result for speech "yeosunim kyeoseo cheo."

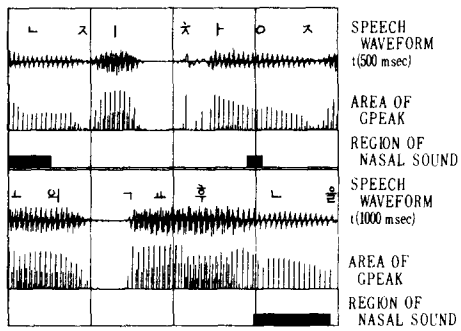

그림9 (e). 음성신호 "ㄴ지 창조의 교훈을"에 대한 실험결과

Fig. 9 (e). The result for speech "nji changjoyeoi gyohuneul."

鼻音과 유사한 性質을 띠어 鼻音으로 判定하는 경우가 있고 "는"에서의 "ㅡ"라든지 "년"에서의 "ㅕ"같은 音素는 鼻音사이에서 완전히 鼻音化되어 鼻音으로 判定하였다. 또 "다"에서의 "ㅏ"와 같이 amplitude가 점점 작아지면서 바로 silence가 장시간 연결된 경우 간혹

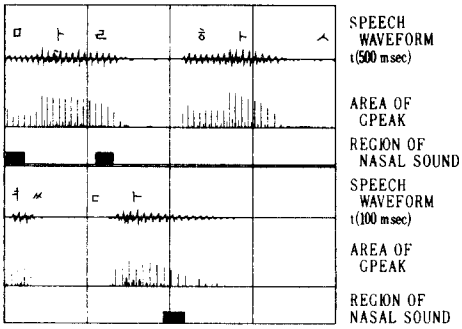

그림9(f). 음성신호 "말 하셨다"에 대한 실험결과
Fig. 9(f). The result of speech" malhasyotda".

鼻音으로 판정하는데 이는 energy의 느슨한 감쇄로 인해 鼻音과 유사해지기 때문이다.

이와 같은 鼻音 判定은 連續音 認識과정에서 일어나는 音素間의 영향에서 起因하는 것이다. 그렇지만 鼻音을 鼻音이 아니라고 판정하지는 않았다. 分類認識 단계에서는 鼻音化된 音素와 鼻音이 區別되지 않으므로 이를 다시 分類하는 것은 훨씬 더 많은 parameter들이 요구되어지고 또한 매우 복삽한 처리 과정을 要하므로 이것은 最終認識 단계에서 처리하는 것이 바람직하다.

Ⅵ. 結 論

本 論文에서는 有聲音의 한 pitch 區間 중에 처음의 peak가 有聲音의 glottal 成分과 F1의 영향을 지배적으로 받는다는 것을 이용하여 G-peak를 정의하고 이를 이용하여 pitch를 추출하였다. 이렇게 얻은 pitch를 바탕으로 有聲音과 鼻音을 分類해 내는 parameter로서 G-peak의 面積 대 Side-peak들의 面積의 合의 比를 이용하여 最終的으로 鼻音을 分類해 내었다.

모든 것을 time domain에서 처리하고 특히 pitch를 抽出하는 과정에서 Sampled data의 summation에 해당하는 G-peak만을 이용하므로 비교적 빠른 속도로 처리할 수 있다. 鼻音을 認識하는 과정에서 鼻音의 고유적인 特性을 paramater로 이용하였으므로 鼻音은 물론 鼻音化된 영역도 分類認識이 가능하였다. 특히 pitch를 기본 單位로 처리하므로 話者에 independent한 分類認識이 가능하였다.

參 考 文 獻

[1] L.R. Rabiner and R.W. Schafer, "Digital processing of speech signals," *Prentice Hall, Inc.*, 1978.

[2] J.D. Markel and A.M. Gray, "Linear prediction of speech", *Springer-Verlag, Berlin* Heidelberg, New York, 1980.

[3] Myungjin Bae and Souguil Ann, "The high speed pitch extaction of speech signals using the area comparison method," *KIEE*, vol. 22, no. 2, pp. 101-105, Feb. 1985.

[4] A.E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *J. Acoust. Soc. Am*, vol. 49, pp. 583-590, 1971.

[5] H.K. Dunn and S.D. white, "Statistical measurements on conversational speech," *J. Acoust. Soc. Am*, vol. 11, pp. 278-288, January 1940.

[6] Myungjin BAE, "A study on the fundamental frequeny extracting of speech signals using specond Order rundown method." Seoul National University, MA Paper, Jan. 1983.

[7] Myungjin BAE and Souguil ANN, "The voiced-unvoiced-silence classification by Emphasized spectrum of speech signals," *JASK*, vol. 4, no. 1, pp. 9-15, June 1985.

[8] Myungjin BAE and Souguil ANN, "Low pass filtering on the high speed pitch extraction", *KIEE*, to be published, 1986.

[9] Myungjin BAE and Souguil ANN, "Inverse Rate type Filtering for the pitch Extraction," *JASK*, vol. 5, no. 3, sept. 1986.

[10] Myungjin BAE and Souguil ANN, "Data compression by elimination of redundancy in human speech signals." *Seoul National University Engineering Report*. vol. 17, no. 1, pp; 129-133, April. 1985.

[11] Chulhi LEE, "A study on the recognition of korean vowel in continuous speech," *Seoul National university*, MA paper, Jan. 1986.

[12] Sungchan BANG, "A study on the classification of Korean voiced into vowel, nasal, and voiced consonant," *Seoul National university*, MA paper, Jan. 1986.

[13] Ian H. Witten, "Principles of computer speech", *Academic Press*, 1982.

[14] 김영송, 우리말 소리의 연구, 과학사, 1981.

[15] M.J. BAE, J.Y. RHEEM, I.J. CHUNG, S.G. ANN, "A study on the Energy parameter by G-peak in speech signals," *KIEE* To be published, 1987. *