

# 特定 帶域 에너지를 利用한 韓國語 基本 數字 音聲의 自動 認識에 關한 研究

## (A Study on the Automatic Recognition of Korean Basic Spoken Digit Using Energy of Special Bandwidth)

韓 熹\*, 金 淳 協\*\*, 朴 圭 泰\*\*\*

(Hee Han, Soon Hyob Kim and Kue Tae Park)

### 要 約

基本 母音의 分析을 爲한 특징 파라미터로 特定 帶域의 에너지의 比를 利用하는 方法을 使用하여 이 파라미터와 零通過率(zero crossing rate;ZCR) 그리고 에너지 파라미터의 論理 組合으로 韓國語 基本 數字 音聲의 認識을 시도하였다.

本 實驗을 爲해서 音聲 信號는 차단 주파수 10KHz의 低域 濾波器로 濾波되었고 20KHz의 표본화율로 표본화 되어 IBM 370으로 시뮬레이션 되었다. 本 시뮬레이션에서는 리메츠 교환 알고리즘<sup>[13],[14]</sup>에 의해 61차, 120차, 25차, 25차 등 4개의 FIR 디지털 濾波기를 설계하여 使用하였다. 실험 결과 3인의 話者에 대해서 92%의 認識率을 얻었다.

### Abstract

Through the use of energy ratio of special bandwidths of basic vowels, recognition of Korean basic spoken digit is performed in logical combination with a zero-crossing rate and an energy parameter.

In the experiments for recognition of the digits, the speech signal of spoken digits is filtered by a lowpass filter of which the cutoff frequency is 10KHz, and then sampled at 20KHz of sampling rate, In the speech signal processing, we used four FIR digital filters, and the order of filter lengths is 61, 120, 25, 25 respectively. The filters are designed by using Remetz exchange algorithm.<sup>13, 14</sup>

As a result, the recognition rate of 92% for the three speakers is obtained.

### I. 序 論

音聲의 研究는 Jones<sup>[1]</sup>와 같은 實驗的인 音聲學者

들에 依해서 初期에 進行되어 왔다. 이 같은 研究는 主로 音素의 質과 調音이 혀의 위치나 입술의 진동과 어떤 關係를 갖는가 하는 것이었다.

母音 音聲이 음향 공진기에 依해서 發生될 수 있다는 사실은 Helmholtz와 Rayleigh<sup>[2]</sup>에 依해서 알려졌다. 그러나 이들의 스펙트럼이 聲道 面積 函數에 따라서 變한다는 사실은 Russel<sup>[3]</sup>에 와서야 알려졌다.

고전적인 研究에서는 Peterson과 Barney<sup>[4]</sup>가 여러 사람들의 音聲에서 母音에 대해서 第一, 第二 formant 周波數를 測定하였다. 그러나 이들 二個의 for-

\* 正會員, \*\*\*正會員, 延世大學校 工科大學 電子工學科 (Dept. of Electronics Engineering, Yonsei Univ.)

\*\*正會員, 光雲工科大 應用電子工學科 (Dept. of Application Elec. Eng. Kwanwoon Univ.)

接受日字: 1981年 8月 20日

mant 周波數만으로는 접치는 것이 너무 많아 認識을 爲해서는 第一, 第二 以上の formant 周波數에 對한 情報가 必要하다는 것이 밝혀졌다. 음향 신호가 인가되었을 때 人間이 이를 받아 들이는 方法에 關한 理論中에 合成에 依한 分析 方法이 MIT에서 Hale과 Stevens<sup>[5]</sup>에 依해서 제시되었다.

音聲認識을 爲한 하드웨어에는 1963년 Nippon Electric Co.에서 Nagata<sup>[6]</sup>에 依해 報告된 日本 數字音聲의 認識機와 RCA의 Ross<sup>[7]</sup>에 依한 單語 認識機 등이 있으며 最近에는 話者에 無關한 英語 數字音聲 認識機가 Bell 研究所에서 개발되었다.<sup>[8]</sup> 이것은 騒音이 차폐된 곳에서는 2.7%의 誤認識을, 또한 騒音이 있는 컴퓨터실에서는 5.6%의 誤認識을 하였다.

여기서 使用한 특성 파라미터는 10msec당 얻어진 零通過率, 에너지, 2차 선형 예측(two-pole linear predictive coding; LPC), 선형 예측 추정 오차 등이었다.

이외에도 1970년 이후에 60個 이상의 研究報告가 있었다. 그러나 이들의 一般의인 특징은, 前에 개발된 기법을 使用하여 性能을 개선하는 시도였을 뿐 이 문제에 關하여 基本的인 새로운 생각을 제시하지는 못하고 있다는 것이다.

韓國語 基本 數字音에 對해서는 주로 오영환에 依해서 이루어졌고,<sup>[9],[10],[11]</sup> 1980年初에 日本 東京工業大의 博士學位 論文으로 提出된 것이 가장 最近의 것이다. 여기서는 선형 예측 계수 方法에 依해 認識을 시도하여 3人的 話者에 對하여 96% 이상의 좋은 結果를 얻고 있다.

本 論文에서는 周波數 平面的 特定 帶域幅에서의 에너지 분포에 依한 基本 母音의 分析을 시도하는 方法을 使用하여, 零通過率과 대수 에너지의 測定 結果와 的 論理 組合에 依해 數字 音聲의 認識을 시도하였다. 이를 爲해서는 音聲 信號는 20KHz의 표本化율로 표本化 되어 처리되었으며 이때 使用한 디지털 濾波器는 리베즈 알고리즘에 依한 25차 2個, 61차, 120차, 각각 1個씩의 FIR 디지털 濾波器를 설계하여 使用하였다.

## II. 有·無聲音의 識別

音聲信號의 有·無聲音의 區分을 爲한 여러 가지 기법이 개발되어 왔다. 이들 대부분은 대형 컴퓨터의 도움으로 실현 가능했고 각기의 복잡성과 설계방식에 따라서 여러 가지 차이를 보이며 대부분 實時間內的 처리가 不可能하였다.

最近에 들어서 비록 10~40 msec의 지연 시간이 요

求되지만 實時間에 처리되는 시스템이 제시되었다. 이 같은 實時間 디지털 하드웨어의 하나는 L. R. Rabiner<sup>[12]</sup> 등에 依해서 제시된 피치 검출기이다.

이 方法의 長點은 종래의 方法들이 컴퓨터의 도움을 얻었음에 비해 단지 150個의 IC로서 하드웨어를 실현했다는 것이었다.

本 論文에서 使用한 方法은 가장 고전적인, 그러나 드물게 보고된 스펙트럼 에너지 분포의 기법 즉, 低域 및 高域 濾波된 音聲波形的 미소 시간 스펙트럼 에너지 분포의 동시 비교에 依해서 有·無聲音 決定하는 方法에 그 기초를 두고 있다. 이 같은 方法은 배후 잡음이나 다른 音聲 信號가 없는 환경에서는 좋은 結果를 주지만 實際의 환경에서는 처리되는 音聲 信號의 區間이 한 피치보다 적으면 매우 애매한 結果를 유도한다. 따라서 이를 해결하기 爲해서는 한 피치 區間보다 긴 音聲 信號 區間(10msec) 즉, 자기 상관을 音聲 信號에 연속적으로 적용할 수 있는 區間이어야 한다. 本 論文에서는 低周波 濾波器로써 차단 周波數  $f_c=1\text{KHz}$ 用과 高周波 濾波器로써 차단 周波數  $f_c=5\text{KHz}$ 用을 使用하였다. 1 KHz~5 KHz사이의 帶域은 피치 발생 동안의 高周波 성분과 第2, 第3, 第4 formant 周波數 같은 1 KHz이상에서 충분한 에너지를 갖는 有聲音 要素들이 存在하는 애매상태문에 使用하지 않았다. 使用한 濾波器는 리베즈 교환 알고리즘에 依해서 설계된 25차 FIR 디지털 濾波器로써  $\pi$ 로 normalize된 디지털 차단 周波數<sup>[14]</sup> (cut off frequency;  $f_c$ ), 정지 周波數 (stop frequency;  $f_s$ )가 각각 0.05 및 0.18인 低域 濾波器와 0.25 및 0.13인 高域 濾波器였다. 본 실험 과정에서 컴퓨터 시뮬레이션에서 설계하여 使用한 濾波器의 계수들을 부록으로 첨가하였다. 有·無聲音의 相對的인 스펙트럼 에너지에 必要한 weighting은 /아/, /이/, /오/, /이/ 등의 有聲音과 /스/, /츠/, /프/ 등의 無聲音 信號를 上記 濾波器에 인가하여 얻어진 출력을 비교하여 低域 濾波器의 出力이 高域 濾波器의 出力의 50배가 넘으면 有聲音으로 判定하였다. 표 1(a), (b)에 이들 濾波器의 有·無聲音에 對한 出力의 比를 나타내었다. 표 1 (a)의 RD는 信號를 10msec의 frame으로 나누어 200個씩의 데이터를 上記 高域 및 低域 濾波器에 인가하여 濾波한 結果의 比로써 아래와 같은 식으로 表示된다.

$$RD = \frac{\sum_{k=1}^N (y_{lk})^2}{\sum_{k=1}^N (y_{hk})^2}, \text{ 단 } k \text{는 표本の 번호이}$$

고  $y_l$  및  $y_h$ 는 각각 低域 및 高域 濾波器의 出力이다. 또 frame 번호인 NI가 8까지는 無聲音이고 9이하

는 有聲音 데이터에 關한 것이다. 표 1 (b)는 20msec의 frame마다 처리한 결과이고 NI가 4까지는 無聲音이다. 또한 本 論文에서는 上記의 有·無聲音의 決定結果와 組合하여 零通過率의 파라미터를 使用하였다. 一般의 10msec의 frame에서 有聲音은 10~4(회/frame), 無聲音은 50~60(회/frame) 이상의 零通過率을 보인다. 따라서 이들 零通過率은 有·無聲音을 區分하는 좋은 파라미터가 된다. 本 論文에서는 스펙트럼 에너지 기법의 結果와 零通過率 測定結果가 모두 同一한 決定을 하였을 때에 限하여 有·無聲音이 決定된 것으로 하고 둘의 結果가 일치하지 않을 경우는 애매성 또는 有聲音과 無有音의 경계 구간으로써 간주하여 다음 frame에 對해서 分析하도록 하였다.

표 1. 저역 및 고역 여파기의 출력의(a), (b)  
Table 1. Output ratio of low and high frequency band pass filters(a) and(b).

(a)		(b)	
NI	RD	NI	RD
1	0.54385D+01	1	0.85210D+01
2	0.11474D+02	2	0.48709D+01
3	0.53405D+01	3	0.14622D+02
4	0.42907D+01	4	0.15066D+02
5	0.15202D+02	5	0.15045D+03
6	0.14325D+02	6	0.34063D+03
7	0.15106D+02	7	0.90788D+02
8	0.44534D+01	8	0.95430D+03
9	0.28874D+03	9	0.37662D+03
10	0.18468D+03	10	0.26856D+03
11	0.35544D+03	11	0.11938D+03
12	0.32375D+03	12	0.16950D+03
13	0.27512D+03	13	0.27245D+03
14	0.49387D+02	14	0.14070D+03
15	0.10989D+04	15	0.13423D+03
16	0.90316D+03	16	0.15748D+03
17	0.42774D+03	17	0.49297D+03
18	0.33259D+03	18	0.39390D+03
19	0.81394D+03	19	0.42309D+03
20	0.15475D+03	20	0.28569D+03
21	0.15048D+03		
22	0.89042D+02		
23	0.17569D+03		
24	0.16108D+03		
25	0.34405D+03		
26	0.22658D+03		
27	0.19788D+03		
28	0.11546D+03		
29	0.22021D+03		
30	0.10100D+03		
31	0.26597D+03		
32	0.11658D+03		
33	0.88533D+03		
34	0.33056D+03		
35	0.11663D+04		
36	0.18548D+03		
37	0.12374D+04		
38	0.25881D+03		
39	0.83308D+03		
40	0.16990D+03		

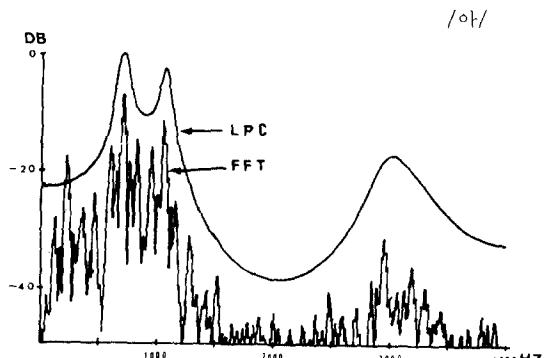
III. 母音의 識別

周波數 平面에서의 에너지 분포는 母音의 소리마다 거의 일정한 양태를 보여 준다. 즉, 직접적인 FFT 나

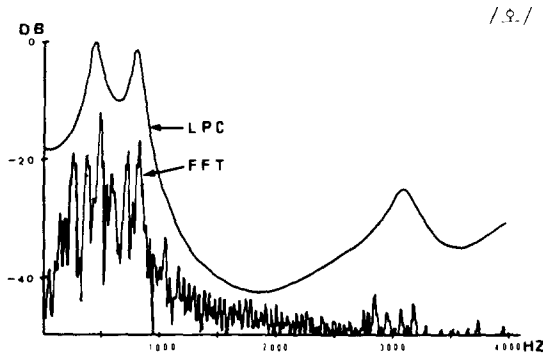
선형 예측에 依해서 추정된 各 母音들의 포먼트(formant)의 위치는 정확히 일정하지는 않으나 비교적 작은 周波數의 편이를 나타내는 특징을 보여 준다. 따라서 周波數 平面 分析 方法에서는 이들 포먼트(formant)의 직접 및 相對的 位置가 音素를 區分 및 認識하는데 대단히 중요한 파라미터가 되고 있다. 이들 formant는 해당 音素에서 에너지의 密集이 가장 強하게 나타나는 周波數이며 一般의 스펙트럼 위에서 이들 formant를 中心으로 한 어느 만큼의 帶域幅에 密集된 에너지의 比率이 다른 帶域幅에 존재하는 에너지의 比率보다 높게 나타나고 있다. 이들 帶域幅에 존재하는 에너지들의 比 하나로는 明確하게 모든 母音을 區分 認識할 수 없으나 母音을 /아/, /이/, /오/, 의 3가지 集團으로의 分類는 상당히 効果的인 尺度가 될 수 있다.

그림1에 基本 母音의 스펙트럼을 나타내었다. 여기서 살펴 볼 수 있는 것과 같이 /아/는 第1, 第2 formant가 1KHz 이하에 존재하며 第3 formant는 2.7KHz 근방에 있으므로 1KHz 근처 帶域에 존재하는 量보다 높게 나타나고 있다. /오/, /우/는 第1, 第2 formant가 700Hz 이하에 존재하고 第3 formant는 2.2~2.7 KHz 사이에 존재하므로 /아/의 경우보다는 적지만 1 KHz 帶域이 2 KHz 帶域보다 높은 에너지의 密集이 이루어지고 있음을 추정할 수 있다. /이/는 第1 formant가 300Hz 이하에 第2, 第3 formant가 2.2~3.2KHz 사이에 존재하므로 에너지의 密集은 1KHz 帶域보다는 2 KHz 帶域에 더 많은 量을 나타내고 있음을 추정할 수 있다.

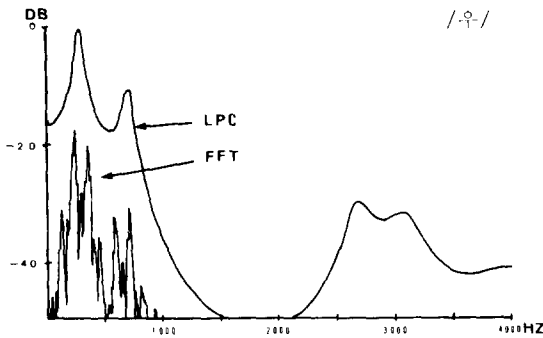
따라서 이들 3 集團으로의 分類는 1 KHz와 2 KHz의 어느 일정 帶域幅을 通過하는 에너지의 量으로서 區分 可能하게 된다. 비록 話者와 時間에 따라서 formant 周波數의 편이가 나타나지만 이들 편이의 結果



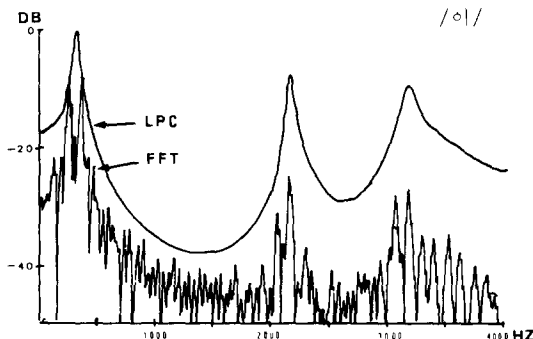
(a)



(b)



(c)



(d)

그림1. 한국어의 기본 모음의 FFT 방식에 의한 스펙트럼과 LPC 방식에 의한 스펙트럼의 비교  
Fig.1. Comparison of FFT spectra and LPC spectra for Korean basic vowels.

가 일정 帶域幅의 濾波된 結果와 重첩되지만 않는다면 이들의 結果를 區分하기 위한 파라미터로 定할 수 있음을 알 수 있다. 따라서 이들 여파기의 帶域幅을 어떻게 設定할 것인가가 문제가 된다. 本 論文에서는 4

표 2. 각 여파기의 주파수 특성

Table 2. Frequency characteristic of each band-pass filters.

각여파기의 차수	$\pi$ 로 정규화(normalize)된 디지털 주파수			
	$f_{c1}$	$f_{s1}$	$f_{s2}$	$f_{c2}$
61	0.05	0.075	0.125	0.15
70	0.065	0.085	0.125	0.145
101	0.0125	0.025	0.075	0.0875
120	0.0125	0.025	0.061	0.0735

(단, 표본화 주파수=20KHz)

개의 帶域通過 濾波器를 使用하여 각 濾波器의 出力의 比를 求한 후에 최대의 편이를 나타내는 2개의 濾波器의 出力의 比를 認識에 必要한 기준치(threshold)로 잡았다. 이때 濾波器는 리에즈 교환 알고리즘에 의한 FIR 濾波器로 디지털 차단 주파수  $f_c$ 와 정지 주파수  $f_s$ 가 표2와 같은 61차, 70차, 101차, 120차의 디지털 濾波器를 설계하여 使用하였으며 이들의 여파기 계수를 부록으로 첨가하였다. 그리고 認識에 필요한 기준치(threshold)를 정하기 위해서는 아래와 같은 실험을 수행하였다. 즉, 基本 母音 /아/, /이/, /오/, /우/, /어/ 등에 對해서 上記의 디지털 濾波器를 通過시켜서 이들 濾波器의 出力의 比를 求하였다. 이때 처리된 信號는 각 音聲當 100msec 정도이고 프레임(frame)의 길이는 10msec로써 20KHz로 표본화되어 200개의 표본 데이터에 對해서 각기 처리되었다.

이들실험의 結果를 표3에 나타내었다. RA는 120차 濾波器에 對한 61차 濾波器의 出力의 비이고 RB는 120차에 對한 70차, RC는 101차에 對한 61차, RD는 101차에 對한 61차, RD는 101차에 對한 70차 濾波器의 출력의 비이다. 표3에서 알 수 있는 바와 같이 RA가 各 音素마다에 가장 현저한 차이를 나타내며 이들의 차이는 認識을 위한 기준치(threshold)로써 충분하다. 프레임(frame)의 數 NI가 10까지는 /아/ 소리에 關한 것이고 NI가 11에서 18까지는 /오/ 소리, NI가 19에서 27까지는 /이/ 소리, 28에서 36까지는 /우/ 소리 37 이하는 /어/ 소리에 關한 것이다. 따라서 本 論文에서는 出力의 比 RA가 1보다 적으면 /아/로 10보다 크면 /이/로 이들 두 값 사이에 있으면 /우/의 集團으로 認識하도록 하였다.

총래의 formant 周波數를 찾아서 인식에 使用하는 方法들은 선형 예측 기법등에 依해서 비록 많은 계산의 불을 줄였으나 아직도 많은 계산을 필요로 하며

대부분 實時間 처리가 不可能하고 미니 컴퓨터(mini-computer) 이상의 장비를 필요로 하게 된다. 이들 방법이 비록 정확하지만 數字 音聲의 認識같이 制限되고 特殊性을 가지고 있는 音素의 認識에는 非經濟的이다. 이에 比하여 本 論文에서 使用한 特定 帶域의 濾波에 依한 方法은 上記의 120차 및 60차 FIR 濾波器 만으로도 有用한 結果를 얻을 수 있으므로 간단히 아나로그 電子回路로써 구성이 可能하며 따라서 實時間 처리가 可能하고, 비용이 매우 절감된다는 長點이 있다.

표 3. 각 대역 여과기의 출력비

Table 3. The output ratio of each bandpass filter.

NI	RA	RB	RC	RD
1	0.525170E+00	0.445710E+02	0.433100E-01	0.533100E+01
2	0.529200E+00	0.498520E+02	0.394910E-01	0.508660E+01
3	0.327450E+00	0.616240E+02	0.318420E-01	0.597170E+01
4	0.339870E+00	0.650920E+02	0.361640E-01	0.692260E+01
5	0.552300E+00	0.676570E+02	0.452040E-01	0.553750E+01
6	0.708270E+00	0.106570E+03	0.312480E-01	0.472780E+01
7	0.502010E+00	0.661430E+02	0.468490E-01	0.508710E+01
8	0.357700E+00	2.668110E+02	0.304980E-01	0.562690E+01
9	0.392930E+00	0.719320E+02	0.261840E-01	0.476830E+01
10	0.549320E+00	0.685040E+02	0.308590E-01	0.484680E+01
11	0.164710E+01	0.246920E+03	0.206700E-01	0.269770E+01
12	0.110630E+01	0.183960E+03	0.228450E-01	0.228350E+01
13	0.190450E+01	0.134220E+03	0.224290E-01	0.272120E+01
14	0.190450E+01	0.121480E+03	0.232580E-01	0.249320E+01
15	0.148770E+01	0.163490E+03	0.317630E-01	0.349950E+01
16	0.167900E+01	0.206720E+03	0.245680E-01	0.289260E+01
17	0.167900E+01	0.206720E+03	0.401580E-01	0.488630E+01
18	0.193220E+01	0.193460E+03	0.308810E-01	0.389280E+01
19	0.239750E+03	0.781910E+05	0.414620E-01	0.506340E+01
20	0.170610E+03	3.175340E+05	0.313840E-01	0.313080E+01
21	0.421730E+03	3.319360E+05	0.505860E-01	0.493700E+01
22	0.256880E+03	3.287420E+05	0.343290E-01	0.384180E+01
23	0.231840E+03	0.201510E+05	0.429730E-01	0.373540E+01
24	0.143900E+02	0.161830E+04	0.533980E-01	0.400540E+01
25	0.215070E+03	0.218810E+05	0.352510E-01	0.345490E+01
26	0.143100E+02	0.146130E+04	0.449040E-01	0.508990E+01
27	0.208640E+03	0.208200E+05	0.493410E-01	0.506690E+01
28	0.224680E+01	0.240690E+03	0.315780E-01	0.338570E+01
29	0.421160E+01	0.379610E+03	0.362740E-01	0.326960E+01
30	0.277230E+01	0.347970E+03	0.293110E-01	0.254940E+01
31	0.784690E+01	0.702200E+03	0.319960E-01	0.318150E+01
32	0.419840E+01	0.365940E+03	0.226590E-01	0.199640E+01
33	0.653980E+01	0.776310E+03	0.268050E-01	0.308880E+01
34	0.443010E+01	0.449530E+03	0.315980E-01	0.320780E+01
35	0.329840E+01	0.441450E+03	0.190230E-01	0.247220E+01
36	0.449340E+01	0.430330E+03	0.355370E-01	0.348340E+01
37	0.608240E+00	0.818340E+02	0.262260E-01	0.352880E+01
38	0.593610E+00	0.794380E+02	0.311220E-01	0.444490E+01
39	0.632760E+00	0.748370E+02	0.352010E-01	0.416260E+01
40	0.440920E+00	0.711750E+02	0.244820E-01	0.376420E+01

IV. 數字音聲 認識 알고리즘

本 論文에서 使用한 認識 알고리즘은 아래와 같다. 처음 4個의 區間 동안은 音聲 信號가 存在하지 않은 默音의 상태로 지속시켜서 배경 잡음의 信號 레벨을 測定하게 하여 音聲 信號가 始作되는 것을 發見하기 爲한 threshold를 決定하도록 하였다. 信號의 레벨이 threshold의 2 배를 넘었을 때 音聲 信號가 始作되는 것으로 보았다. 信號는 다시 低域 및 高域 濾波되어 有·無聲音 如否를 決定짓고 그 해당 구간에 對하여 零通過率을 계산하게 되며 零通過率이 60(회/frame)을 초과할 때 無聲音으로 35(회/frame)이하일 때 有聲音으로 이들 두 값 사이일 때는 애매한 상태로 감지하여 다음 구간에 對해서 다시 信號를 해석하도록 하였다. 일단 有·無聲音의 決定이 내려진 구간의

信號는 2 번째 앞의 區間的 零通過率과의 차이를 계산하게 되며 이렇게 해서 얻어진 零通過率의 변화율은 次后의 鼻音(nasality)을 區分하는 파라미터가 된다.

低域 및 高域 濾波의 比에 依한 決定과 零通過率에 依한 決定이 일치할 때만 해당 區間이 有·無聲音으로 決定된 것으로 보았고 일치하지 않는 경우는 불확실한 것으로 判定하여 그 다음 區間에 對해서 다시 해석을 하도록 하였다. 判定이 無聲音일 경우에는 無聲音이었음을 기록하고 다시 그 다음 區間에 對해서 判定을 계속하여 모음 즉, 有聲音이 나타날 때까지 계속 반복 하도록 하였다. 이들 無聲音 區間 이후에 有聲音이 나타나면 이 有聲音 區間은 다시 帶域 濾波器를 通過하여 출력의 比를 계산하고 이 값이 10을 넘으면 /i/로, 1 이하이면 /a/로, 1과 10 사이면 /o/ 계열로 判定하였다. 이렇게 하여 細分된 각 계열은 현재 처리하고 있는 區間앞에 無聲音 區間이 존재했는지 如否를 判定하여 認識의 決定을 내리도록 하였다. /i/ 계열은 처음에 無聲音 區間이 存在하였다면 “질”로 認識하고, 존재하지 않았다면 /i/가 존재했음을 인지하면서 다음 區間的 해석에 들어가며 이때 다음 區間들 중에서 /a/가 나타나면 “영”으로, /u/가 나타나면 “육”으로 認識하였으며, /i/가 나타나면 零通過率의 변화가 있으면 “일”로, 그렇지 않으면 “이”로 認識하도록 하였다.

또한, /o, 계열은 처음에 無聲音 區間이 존재하였다면 “구”로, 존재하지 않았다면 “오”로, 認識하게 하였다. 그러나 처음에 無聲音 區間이 존재한 후에 /a/ 계열이 나타나면 에너지의 변화율과 零通過率을 동시에 계산하여 에너지의 변화율이 threshold를 넘을 만큼 크면 鼻音이 존재하는 것으로 判定하여 “삼”으로 認識하고, 그렇지 않으면 零通過率을 基準으로 하여 信號가 基準 레벨 이하로 될때까지 threshold를 초과하지 않으면 단순 모음으로 끝나는 음으로 判定하여 “사”와 “팔”로 區分하도록 하였다. 이에 對한 흐름선도를 그림 2에 나타내었다. 또한 鼻音의 識別에 使用된 특징 파라미터인 零通過率의 threshold를 얻기 위해서는 數字 音聲 “영”에서 “구”까지를 入力 데이터로 하여 20msec의 frame당의 零通過率 및 에너지 변화율을 測定하여 그 基準으로 삼았다. 이때 零通過率

ZCR과 에너지 LE는 아래 식에 따른다.

$$ZCR = \sum_{k=1}^N [1 - \text{Sgn}(X_{k-1}) \cdot \text{Sgn}(X_k)] / 2$$

$$LE = 10 \text{ LOG} \sum_{k=1}^N (X_k)^2$$

여기서  $X_k$ 는 區間내의 표본의 크기이고,  $N=400$ ,  
 $Sgn(X) = \frac{X}{|X|}$  이다.

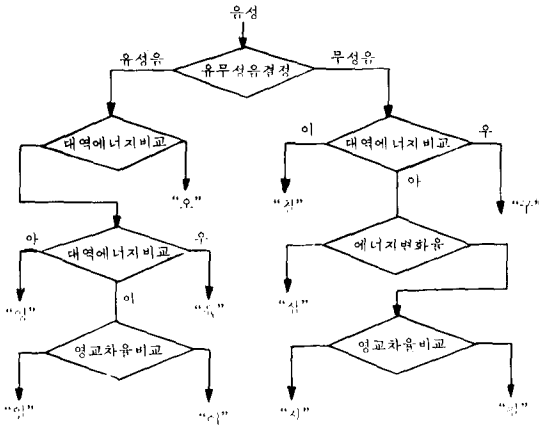
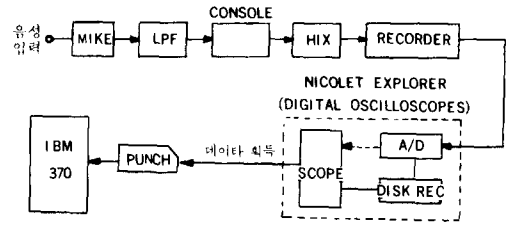


그림2. 인식 흐름선도  
 Fig.2. The flow chart of recognition.



Mike : Beyer dynamic m M67  
 Recorder : TEAC A-4300 SX  
 Tape : SCOTCH magnetic tape  
 LPF : 6pole butter worth  $f_c=10\text{KHz}$   
 A/D : 12 bits resolution  
 화 자 : 연세대학교 교육방송국 남자 아나운서  
 3명 (20, 22, 23세)  
 음 성 : 한국어 숫자 음성 10개

그림3. 실험 구성도  
 Fig.3. Block diagram of experiment.

V. 認識實驗

1. 實驗方法

延世大學校 敎育 방송국의 방송 장치가 된 스튜디오에서 韓國語 基本 數字音 10個(“영”에서 “구”까지)와 母音 5個( /아/, /이/, /오/, /우/, /유/)를 녹음했다. 이때 마이크로폰을 通過한 信號는 10 KHz의 차단 주파수를 갖는 6차 버터-워스(butter-worth) 低域 濾波器를 通過시켜 고주파 성분을 제거시켰다. 이 信號는 다시 콘솔과 믹서를 지나서 마그네틱 테이프에 녹음되었고, 다시 이 녹음된 信號는 디지털 오실로스코프인 NICOLET EXPLORER에 인가되어 20KHz로 標本化된 후 카드에 펀치되어 IBM 370에 입력 데이터로 인가되었다. 이에 對한 構成圖를 그림3에 보였다. 分析 區間의 길이는 400데이터(20msec)씩 처리하였다.

2. 實驗結果

認識 實驗의 結果를 표 4에 보였다. 3人的 話者가 각기 3번씩 發聲한 90個의 數字 音聲에 對하여 7個의 誤認識이 나타났으며 이들은 주로 鼻音과 有聲子音에 對한 것이었다.

이에 對한 原因은 本 알고리즘에서 使用한 零通過率의 변화량에 依해서 鼻音을 區分하는 것이 完全하지 못한 특징 파라미터임을 나타내며 對數 에너지의 thres-

표 4. 인식 실험 결과

Table 4. The results of recognition experiment.

		음 성 신 호												
		0	1	2	3	4	5	6	7	8	9			
인 식 된 숫 자	0	8												
	1		8											
	2	1	1	9										
	3				6									
	4				1	9					1			
	5						9			1				
	6							9						
	7									9				
	8				2						7			
9													9	

hold레벨을 再調整함으로써 認識率을 높일 수 있음을 보여 주고 있다. 本 알고리즘에 上記의 보완책과 發聲者의 學習이 이루어진다면 더 높은 認識率을 얻을 수 있을 것이라고 기대된다.

VI. 結 論

本 論文에서는 韓國語 基本 數字 音聲을 對象으로 하여 特定 帶域의 에너지를 이용하는 方法으로 認識을

시도하였다. 종래의 周波數 平面에서 音素의 특징 파라미터를 추출하는 方法들은 많은 계산의 量을 必要로 하게 되며 따라서 많은 時間과 미니컴퓨터 이상의 하드웨어를 要求한다. 이들 方法이 비록 모든 母音에 적용할 수 있는 點들의 長點이 있으나 数字音의 認識같이 制限되고 特殊性을 가지고 있는 音素 認識에는 非 効果的이다. 이에 비하여 本 論文에서 使用한 特定 帶域에 존재하는 에너지의 比로써 특징 파라미터를 삼는 方法은 濾波器만으로 有用한 結果를 얻을 수 있었다. 또한 本 論文은 컴퓨터에 依해서 信號를 처리하는데 그 目的이 있는 것이 아니라 이 알고리즘의 여러 과정을 아날로그 회로로 實現하기 爲한 데 그 目的이 있었다. 따라서 이 같은 立場에서 살펴보면 本 論文에서 使用한 方法은 높은 표본화율 및 높은 차수의 디지털 濾波器를 使用하였으나 실제 아날로그 회로를 構成할 경우에 비교적 간단한 電子 回路로 製作이 可能할 뿐 아니라 實時間 처리가 可能하므로 컴퓨터가 認識機의 一部로써 使用되는 종래의 認識機에 比하여 내폭적인 構造의 單純化와 制作費의 절감을 가져 올 수 있어 認識機의 大衆化에 매우 밝은 전망을 주고 있다. 앞으로 實際의 환경 아래서의 認識 시도와 하드웨어 製作을 爲한 研究가 要望된다.

附 錄

FIR 디지털 필터의 계수  
(Coefficient of Digital Filter)

OREDR=25

- H( 1 ) = -4.47064E-04 =H( 25 )
- H( 2 ) = -2.04585E-03 =H( 24 )
- H( 3 ) = -5.61937E-03 =H( 23 )
- H( 4 ) = -.0112217 =H( 22 )
- H( 5 ) = -.0171319 =H( 21 )
- H( 6 ) = -.0193141 =H( 20 )
- H( 7 ) = -.0121643 =H( 19 )
- H( 8 ) = 9.25244E-03 =H( 18 )
- H( 9 ) = .0462146 =H( 17 )
- H( 10 ) = .0941573 =H( 16 )
- H( 11 ) = .142831 =H( 15 )
- H( 12 ) = .17932 =H( 14 )
- H( 13 ) = .192675 =H( 13 )

(a) 25차 高域 digital filter 계수

OREDR=25

- H( 1 ) = -4.55664E-03 =H( 25 )
- H( 2 ) = -6.12277E-03 =H( 24 )
- H( 3 ) = 1.27815E-03 =H( 23 )
- H( 4 ) = .0114243 =H( 22 )
- H( 5 ) = 4.99837E-03 =H( 21 )
- H( 6 ) = -.0204813 =H( 20 )
- H( 7 ) = -.0306257 =H( 19 )
- H( 8 ) = 8.70506E-03 =H( 18 )
- H( 9 ) = .0665326 =H( 17 )
- H( 10 ) = .0497575 =H( 16 )
- H( 11 ) = -.0953582 =H( 15 )
- H( 12 ) = -.292694 =H( 14 )
- H( 13 ) = .6145 =H( 13 )

(b) 25차 低域 digital filter 계수

OREDR=61

- H( 1 ) = -.0102209 =H( 61 )
- H( 2 ) = -7.16377E-03 =H( 60 )
- H( 3 ) = -3.57245E-03 =H( 59 )
- H( 4 ) = 4.3474E-03 =H( 58 )
- H( 5 ) = .0135526 =H( 57 )
- H( 6 ) = .0192785 =H( 56 )
- H( 7 ) = .0176219 =H( 55 )
- H( 8 ) = 8.10509E-03 =H( 54 )
- H( 9 ) = -5.45592E-03 =H( 53 )
- H( 10 ) = -.0166741 =H( 52 )
- H( 11 ) = -.0201819 =H( 51 )
- H( 12 ) = -.0149678 =H( 50 )
- H( 13 ) = -5.22082E-03 =H( 49 )
- H( 14 ) = 2.19259E-03 =H( 48 )
- H( 15 ) = 2.37413E-03 =H( 47 )
- H( 16 ) = -.79305E-03 =H( 46 )
- H( 17 ) = -9.4502E-03 =H( 45 )
- H( 18 ) = -5.93544E-03 =H( 44 )
- H( 19 ) = .0108233 =H( 43 )
- H( 20 ) = .0357133 =H( 42 )
- H( 21 ) = .0551621 =H( 41 )
- H( 22 ) = .0537861 =H( 40 )
- H( 23 ) = .0238706 =H( 39 )
- H( 24 ) = -.0278426 =H( 38 )
- H( 25 ) = -.0807023 =H( 37 )
- H( 26 ) = -.108884 =H( 36 )
- H( 27 ) = -.0944632 =H( 35 )
- H( 28 ) = -.0379774 =H( 34 )
- H( 29 ) = .0397603 =H( 33 )
- H( 30 ) = .106196 =H( 32 )
- H( 31 ) = .132206 =H( 31 )

(c) 61차 digital filter 계수

\*\*\* OREDR=101 \*\*\*

- H( 1 ) = 8.98134E-03 =H( 101 )
- H( 2 ) = -5.84325E-03 =H( 100 )
- H( 3 ) = -6.59903E-03 =H( 99 )
- H( 4 ) = 6.10007E-03 =H( 98 )
- H( 5 ) = -5.12641E-03 =H( 97 )
- H( 6 ) = -2.98032E-03 =H( 96 )
- H( 7 ) = 2.6536E-05 =H( 95 )
- H( 8 ) = 3.22292E-03 =H( 94 )
- H( 9 ) = 6.17E-03 =H( 93 )
- H( 10 ) = 8.38613E-03 =H( 92 )
- H( 11 ) = 9.55163E-03 =H( 91 )
- H( 12 ) = 9.58742E-03 =H( 90 )
- H( 13 ) = 8.69884E-03 =H( 89 )
- H( 14 ) = 7.30815E-03 =H( 88 )
- H( 15 ) = 5.99972E-03 =H( 87 )
- H( 16 ) = 5.33126E-03 =H( 86 )
- H( 17 ) = 5.69534E-03 =H( 85 )
- H( 18 ) = 7.16235E-03 =H( 84 )
- H( 19 ) = 8.61006E-03 =H( 83 )
- H( 20 ) = 1.011908E-03 =H( 82 )
- H( 21 ) = .0137148 =H( 81 )
- H( 22 ) = .0140214 =H( 80 )
- H( 23 ) = .0121811 =H( 79 )
- H( 24 ) = 7.07867E-03 =H( 78 )
- H( 25 ) = 1.75893E-03 =H( 77 )
- H( 26 ) = 5.47837E-03 =H( 76 )
- H( 27 ) = .0109162 =H( 75 )
- H( 28 ) = .0106589 =H( 74 )
- H( 29 ) = .0017666 =H( 73 )
- H( 30 ) = .0211113 =H( 72 )
- H( 31 ) = -.017117 =H( 71 )
- H( 32 ) = .0095721 =H( 70 )
- H( 33 ) = -.29633E-03 =H( 69 )
- H( 34 ) = 3.47552E-03 =H( 68 )
- H( 35 ) = 6.67533E-03 =H( 67 )
- H( 36 ) = 4.98602E-03 =H( 66 )
- H( 37 ) = -2.33651E-03 =H( 65 )

- H(38) = -.0148157 =H(64)
- H(39) = -.0307009 =H(63)

- H( 40 ) = -.0471683 =H( 62 )
- H( 41 ) = .0697833 =H( 61 )
- H( 42 ) = -.0681215 =H( 60 )
- H( 43 ) = -.064457 =H( 59 )
- H( 44 ) = -.0544129 =H( 58 )
- H( 45 ) = -.032275 =H( 57 )
- H( 46 ) = 7.07867E-03 =H( 56 )
- H( 47 ) = .0322500 =H( 55 )
- H( 48 ) = .0307009 =H( 54 )
- H( 49 ) = .0958731 =H( 53 )
- H( 50 ) = .114301 =H( 52 )
- H( 51 ) = .121057 =H( 51 )

(e) 101차 digital filter 계수

\*\*\* OREDR=70 \*\*\*

- H( 1 ) = 8.21956E-03 =H( 70 )
- H( 2 ) = 9.69542E-03 =H( 69 )
- H( 3 ) = .0102643 =H( 68 )
- H( 4 ) = 6.57199E-03 =H( 67 )
- H( 5 ) = -5.16176E-04 =H( 66 )
- H( 6 ) = -7.7731E-03 =H( 65 )
- H( 7 ) = -.010209 =H( 64 )
- H( 8 ) = -7.32286E-03 =H( 63 )
- H( 9 ) = -2.09084E-03 =H( 62 )
- H( 10 ) = .0136445 =H( 61 )
- H( 11 ) = .015581 =H( 60 )
- H( 12 ) = .0218258 =H( 59 )
- H( 13 ) = .014676 =H( 58 )
- H( 14 ) = 4.48319E-03 =H( 57 )
- H( 15 ) = -2.87928E-03 =H( 56 )
- H( 16 ) = -3.89224E-03 =H( 55 )
- H( 17 ) = 1.67729E-04 =H( 54 )
- H( 18 ) = 3.84576E-03 =H( 53 )
- H( 19 ) = 1.44033E-03 =H( 52 )
- H( 20 ) = -8.40624E-03 =H( 51 )
- H( 21 ) = -.0204672 =H( 50 )
- H( 22 ) = -.0229784 =H( 49 )
- H( 23 ) = -.014643 =H( 48 )
- H( 24 ) = .0110268 =H( 47 )
- H( 25 ) = .0412969 =H( 46 )
- H( 26 ) = .0591352 =H( 45 )
- H( 27 ) = .0500067 =H( 44 )
- H( 28 ) = .0112063 =H( 43 )
- H( 29 ) = .041145 =H( 42 )
- H( 30 ) = -.0914942 =H( 41 )
- H( 31 ) = -.106741 =H( 40 )
- H( 32 ) = -.078848 =H( 39 )
- H( 33 ) = -.0171015 =H( 38 )
- H( 34 ) = .0519947 =H( 37 )
- H( 35 ) = .0970171 =H( 36 )

\*\*\* OREDR=120 \*\*\*

- H( 1 ) = -5.63149E-03 =H( 120 )
- H( 2 ) = -2.84509E-03 =H( 119 )
- H( 3 ) = -3.14359E-03 =H( 118 )
- H( 4 ) = 3.0404E-03 =H( 117 )
- H( 5 ) = -2.78301E-03 =H( 116 )
- H( 6 ) = -2.10814E-03 =H( 115 )
- H( 7 ) = -1.23616E-04 =H( 114 )
- H( 8 ) = -2.62576E-03 =H( 113 )
- H( 9 ) = 6.60057E-04 =H( 112 )
- H( 10 ) = 1.33297E-03 =H( 111 )
- H( 11 ) = 1.66106E-03 =H( 110 )
- H( 12 ) = 1.54913E-03 =H( 109 )
- H( 13 ) = 1.01842E-03 =H( 108 )
- H( 14 ) = 1.52522E-04 =H( 107 )
- H( 15 ) = -8.37165E-04 =H( 106 )
- H( 16 ) = -1.72282E-03 =H( 105 )
- H( 17 ) = -2.1948E-03 =H( 104 )
- H( 18 ) = -2.08218E-03 =H( 103 )
- H( 19 ) = -1.13349E-03 =H( 102 )
- H( 20 ) = 6.79687E-04 =H( 101 )
- H( 21 ) = 3.27737E-03 =H( 100 )
- H( 22 ) = 6.42602E-03 =H( 99 )
- H( 23 ) = 0.75807E-03 =H( 98 )
- H( 24 ) = .0178333 =H( 97 )
- H( 25 ) = .015177 =H( 96 )
- H( 26 ) = .0163903 =H( 95 )
- H( 27 ) = .016197 =H( 94 )
- H( 28 ) = .0143763 =H( 93 )
- H( 29 ) = .011853 =H( 92 )
- H( 30 ) = 6.91495E-03 =H( 91 )
- H( 31 ) = 2.09144E-03 =H( 90 )
- H( 32 ) = -2.64553E-03 =H( 89 )
- H( 33 ) = -6.63694E-03 =H( 88 )
- H( 34 ) = -9.31308E-03 =H( 87 )
- H( 35 ) = -.0102971 =H( 86 )
- H( 36 ) = -9.48859E-03 =H( 85 )
- H( 37 ) = -7.07603E-03 =H( 84 )
- H( 38 ) = -3.6418E-03 =H( 83 )
- H( 39 ) = 1.0476E-04 =H( 82 )
- H( 40 ) = 3.22715E-03 =H( 81 )
- H( 41 ) = 4.80593E-03 =H( 80 )
- H( 42 ) = 4.06492E-03 =H( 79 )
- H( 43 ) = 5.25877E-04 =H( 78 )
- H( 44 ) = 5.09344E-03 =H( 77 )
- H( 45 ) = .0147662 =H( 76 )
- H( 46 ) = .0252607 =H( 75 )
- H( 47 ) = -.0361153 =H( 74 )
- H( 48 ) = -.0458314 =H( 73 )
- H( 49 ) = -.023522 =H( 72 )
- H( 50 ) = -.0557742 =H( 71 )
- H( 51 ) = -.0535804 =H( 70 )
- H( 52 ) = -.0457099 =H( 69 )
- H( 53 ) = .010209 =H( 68 )
- H( 54 ) = .0147645 =H( 67 )
- H( 55 ) = 6.86201E-03 =H( 66 )
- H( 56 ) = .029394 =H( 65 )
- H( 57 ) = .0510533 =H( 64 )
- H( 58 ) = .0476545 =H( 63 )
- H( 59 ) = .0832684 =H( 62 )
- H( 60 ) = .0904589 =H( 61 )

(f) 120차 digital filter의 계수

## 參 考 文 獻

- [ 1 ] D. Jones, *An Outline of English Phonetics*. 9th ed. 1960.
- [ 2 ] J. W. Stutt (Load Rayleigh), *The Theory of Sound*. vol. 2, New York, 1945 (First published in 1878).
- [ 3 ] G. O. Russell, *The Vowel*. Ohio state Univ. press, 1928.
- [ 4 ] G. E. Peterson & H. L. Barney, "Central methods used in study of the vowels," *J. Acoust. Soc. Amer.*, vol. 24(2), pp. 175-184, 1952.
- [ 5 ] M. Halle & K. N. Stevens, "Speech recognition; A model & a program for research," *IRE Trans. Inform. IT*-8(2), pp 155-159, 1962.
- [ 6 ] K. Nagata, Y. Kato, "Spoken sigit decognizer for Japanese language," *NEC Res. Develop.* no. 6, 1963.
- [ 7 ] P. W. Rass, "A limited- vocabulary adaptive speech recognition system," *J. Audio Eng. Soc.*, vol. 15(4), pp. 414-419. 1967.
- [ 8 ] L. R. Rabiner, "A speaker-independent digit-recognition system," *The Bell System Technical Journal*. vol. 54, No. 1, pp. 81-102, Jan. 1975.
- [ 9 ] 吳, 安居院, "숫자음성 자동인식에 關한 一實驗," 대한전자공학회지, 제15권, 제6호, 12월 1978년.
- [ 10 ] 安居院, 吳, "한국어 숫자음의 인식 시스템에 대해서" 日本電子通信學會 全國大會 論文集, pp. 5-28, 1979.
- [ 11 ] 오영환, "한국어의 음성인식에 관한 기초적 연구" (博士學位 論文 東京工業大學 大學院), 102面, 1980.
- [ 12 ] R, W. Schafer, "Real time digital pitch detector," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. Assp-24, pp. 2-8, Feb. 1976.
- [ 13 ] L. R. Rabiner "FIR digital filter design techniques using weighted Chebyshev approximation," *Proceeding of the IEEE*, vol. 63, no. 4, pp. 595-610, April 1975.
- [ 14 ] A. Peled & B. Liu, *Digital Signal Processing* (John Wiley & Sons, Inc., 1976), pp. 123.