

<< 이 논문은 1981년도 산학협동재단 학술연구비의 지원으로 연구되었음 >>

음성을 이용한 컴퓨터 입력장치 개발에 관한 연구

이 현 복 연구 책임자
안 수 길 공동 연구자

0. 머 리 말

음성언어의 자동인식은 전자산업시대에 들어선 오늘날 그 용도와 응용이 다양하기 때문에 세계 각국에서 연구에 열을 쏟고 있는 문제이다. 통신분야는 물론이거니와 많은 고객을 상대로 하는 은행이나 보험회사 따위의 업무는 "음성인식"을 이용하여 업무를 개인화할 뿐만 아니라 인건도를 높일 수 있으며 가정 및 산업용 전자기기는 특정 개개의 음성언어로 조정할 수 있게 되기 때문에 대단히 유용하다. 그러므로 앞으로 전자산업분야에서의 치열한 국제경쟁에서 살아 남기 위해서는 우리도 음성언어의 자동인식 문제의 연구에 박차를 가하지 않아서는 안될 것이다.

본 연구는 산학재단의 추천과제인바 우선 한국어의 음성인식에 필요한 기초연구를 목적으로 한다. 또한 이 과제는 언어, 특히 한국어에 관한 것이므로 음성언어를 전문으로 다루는 인문계의 언어학자와 음성언어의 음향학/전자공학적인 분석을 시도하는 전자공학자의 공동연구가 바람직하다. 양자의 연구태도와 방법은 상호보완적이기 때문에 자칫 범하기 쉬운 오류를 막고 효율적이고 종합적인 연구성과를 기대할 수 있을 것이다.

1. 목 적

현대의 통신 및 전자 산업분야에서 음성을 이용한 컴퓨터 입력이나 음성의 자동인식을 대단히 중요한 문제로 등장하였다. 그런데 언어마다 음성의 체계가 다르므로 한국에서 외에 말한 음성과 관련한 여러문제를 연구하고 해결하기 위해서는 무엇보다도 한국어의 음성체계를 분명히 밝혀 놓아야 한다. 현대 한국어의 기본적인 음성체계를 음성학 및 언어학적으로 분석, 규명함이 없이 한국어음성의 인식이나 컴퓨터 입력장치를 개발한다는 것은 불가능에 가깝기 때문이다. 따라서 이 논문은 우선 음성인식의 관건이 되는 한국어의 단손모음의 체계와 그 음가를 음성 언어학적으로 분석, 기술한 다음, 파라미터(Parameter) 분석 언어학적으로 기술한 한국어 단손모음의 특성을 정밀 분석하여 종합적이고 다각적인 연구를 시도함을 목적으로 한다.

2. 분 석 대 상

한국어라 할지라도 방언에 따라 음성의 체계가 다르므로 대상 언어의 범위를 한정하지 않을 수 없다. 여기서는 우리나라의 표준발음을 대상으로 하여 분석을 시도 하였으며, 파라미터분석은 연구자의 발음을 자료로 하여 진행되었다.

3. 표준 한국어의 단손 모음

현대한국어의 표준발음에서 쓰이는 단손모음 음소는 다음과 같이 9개로 나뉜다.

- (1) /이/, (2) /예/, (3) /애/, (4) /아/
 (5) /오/, (6) /우/, (7) /어/, (8) /으/, (9) /:/

이 중에서 /오/와 /우/는 원순 모음이며 그 밖의 다른 모음은 모두 평순모음이다. 또한 (9)의 /:/는 길이 음소로서 8개의 모음과 결합하여 장단의 음운론적인 대립을 보이고 있다. 가령 /이/와 /이:/는 "일"과 "일:", /아/와 /아:/는 "말"과 "말:"에서 장단에 따른 음운론적인 대립을 보인다. 그러므로 위에서 제시한 음소를 장단으로 대립시키면 도합 16개의 모음을 얻게 되며, 이는 실제로 표준 한국어에서 사용되는 모음들이다. 이밖에도 /위/와 /외/를 단순모음으로 추가하는 일이 있으나 필자는 이들이 /y/와 /ø/와 같은 단순모음보다는 이중모음 /wi/와 /we/로 발음되는 것이 오늘날의 일반적인 경향으로 보기 때문에 이중모음으로 처리한다.

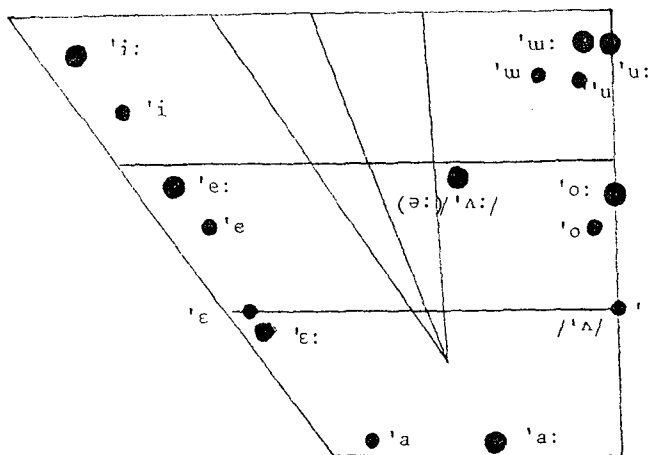
4. 단순모음의 조음음성학적 분석

위에 든 단순모음은 짧을 때와 길 때의 소리값이 다르다. 즉, 장단에 따라 모음의 음가가 다르며 이는 곧 생리적으로 혀의 위치의 차이(성도의 크기와 모양의 차이)와 음향적으로는 포르만트 구조의 차이로 풀이된다. 먼저 한국어모음의 음가를 장단별로 나누어 모음사각도위에 표시하면 그림1과 같다(이현복 1971: 이 사각도위의 모음표시는 원본보다 단순화한 것이다).

5. 단순모음의 음향음성학적 분석

위에서 조음음성학적으로 기술한 모음의 음가를 음향분석기(Spectrograph)를 이용하여 음향음성학적으로도 분석하였다. 여기서는 장단에 따른 모음의 음가를 음향분석기로 분석하여 제1포르만트의 위치를

그림 1



파악하고 이를 토대로 하여 음향도표(Acoustic chart)를 작성함을 목표로 한다. 이제 /이:/와 /이/, /예/와 /예/, /오/와 /오/, /으:/와 /으/의 포르만트 1과 2를 예로 들면 다음과 같다(이 현복, 지민재 1981:부분적으로 제시함).

| | | | | |
|-------|------|------|------|------|
| | /이:/ | /이/ | /예:/ | /예/ |
| F . 2 | 2420 | 2350 | 2100 | 2100 |
| F . 1 | 285 | 350 | 450 | 500 |
| | /오:/ | /오/ | /으:/ | /으/ |
| F . 2 | 800 | 900 | 1130 | 1200 |
| F . 1 | 450 | 500 | 400 | 450 |

위의 음향도표를 앞서 제시한 모음사각도와 비교하면 흥미있는 현상을 찾아 볼 수 있다. 모음사각도는 발음할 때의 혀의 최고점의 위치를 조음음성학적으로 표시한 것이고 음향도표는 음파의 음향학적인 구조를 나타낸다고 볼 수 있으나, 두 도표는 각 모음간의 상대적 위치나 장단에 따른 음가의 차이를 일관성 있게 보여 줄뿐 아니라 전체적인 윤곽에 있어서 거의 일치하고 있다. 다만 평순모음 /으:/와 /으/, /어:/와 /어/가 두 도표상에서 차지하는 위치에 비교적 큰 차이를 보이는데 이는 모음사각도 상에는 원순과 평순의 구별이 안되는 반면에 음향도표에는 원순과 평순의 음향적 특성이 반영되며, /어:/와 /어/의 음가차이가 짧은 세대(이 실험 자료의 발음자는 20대의 청년임)의 말에서는 잘 구별이 안된다는 현상으로 설명될 수 있을 것이다.

6. 성도의 파라미터적 변학을 추출에 의한 단모음의 분석

오래전부터 음성통신분야에서는 음성신호의 파형을 그대로 처리하기 보다는 음성 신호중의 중요한 특징만을 일정한 시간 간격으로(약 10 msec) 추출하여 송신하고 수신하는 쪽에서도 동일한 정보를 이용하여 다시 합성하는 파라미터 분석(Parametric representation) 방법이 많이 쓰인다. 특히 선형예측방법(Linear Predictive Coding Method)은 가장 효율적인 파라미터 분석법으로 알려져 있다. 따라서 이 논문에서는 선형예측방법에 의한 모음의 파라미터를 구하기 위해 파라미터분석기를 제작하여 사용하였다. 이 실험에 이용한 음성파라미터 분석기의 구성은 그림 3과 같다. 그리고 이 실험에 사용한 음성자료는 한 음절에서 네 음절까지의 낱말 또는 말호막 21개로 구성되어 있는데, 아래에 제시한바와 같이 이 안에는 8개의 긴모음과 8개의 짧은 모음이 나타나도

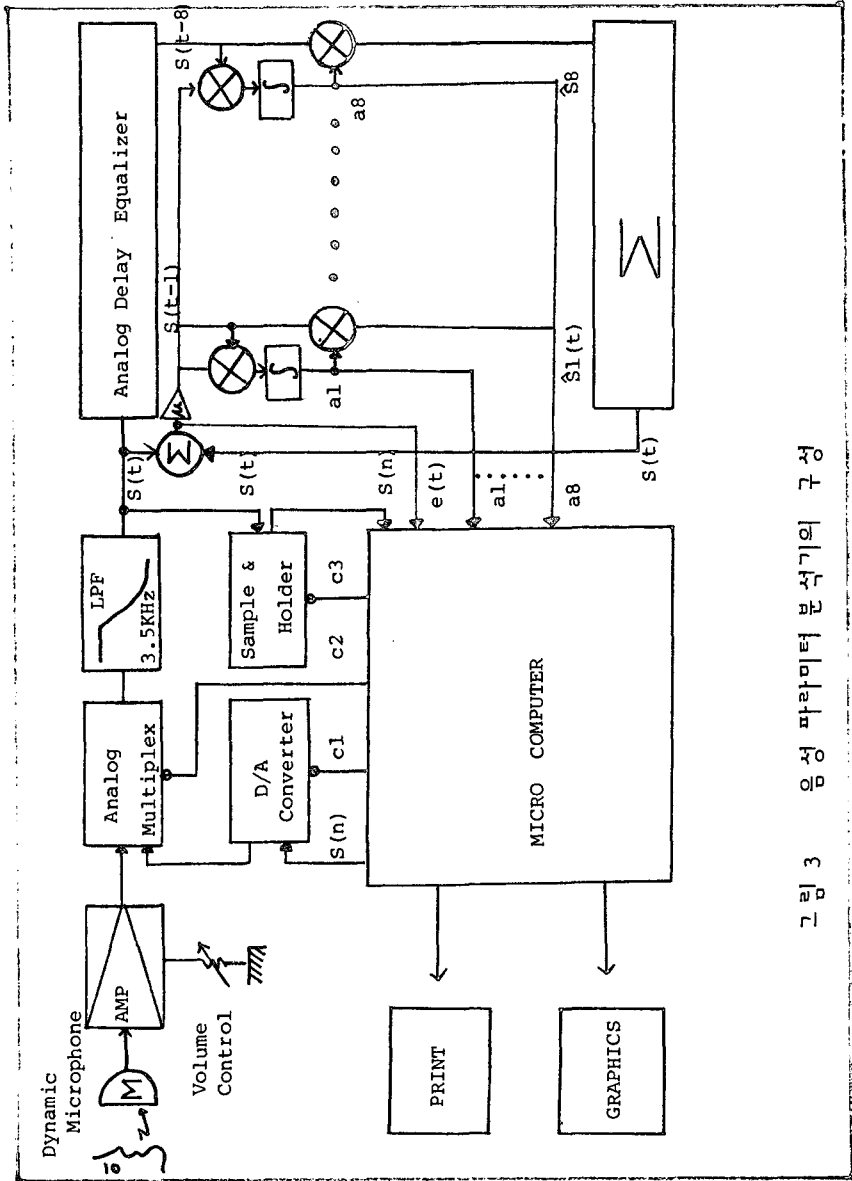


그림 3 음성 파라미터 분석기의 구성

록 마련되어 있다.

- | | |
|---------------|------------|
| 1. 비:료 | 12. 비로소 |
| 2. 베:다 | 13. 이 밥에 |
| 3. 배:다 | 14. (보:)배 |
| 4. 바:보 | 15. 바다가 |
| 5. 보:도 | 16. 보릭발 |
| 6. 부:자 | 17. 부모님 |
| 7. 별: | 18. '잘 씹어서 |
| 8. 벌(Punish)로 | 19. 없어도 |
| 9. 음:식 | 20. '입으너 |
| 10. 흠: | 21. 야! 이리봐 |
| 11. 그:림 | |

이 음성자료를 연구 책임자가 표준발음으로 몇 차례 읽어 녹음한 다음 녹음된 내용을 파라미터 분석기에 넣어 각 모음의 파라미터 값을 분석하였다.

7. 모음의 파라미터 분석결과

이미 앞에서 제시한 한국어의 모음 16개의 파라미터를 분석한 결과 그 수치는 다음과 같다.

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 |
|----|-------|-------|-------|------|-------|------|-------|-------|
| 이: | -0.15 | -0.51 | -1.13 | 0.34 | 0.31 | 0.38 | -0.17 | 0.07 |
| 이 | -1.05 | 0.16 | -0.73 | 0.70 | -0.17 | 0.09 | 0.17 | -0.09 |
| 에: | -0.58 | -0.46 | -0.27 | 0.45 | -0.17 | 0.13 | 0.09 | 0.03 |

| | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|
| 애 | -1.76 | 1.15 | -0.29 | 0.13 | -0.54 | 0.59 | -0.19 | 0.08 |
| 애: | -1.42 | 0.49 | 0.53 | -0.68 | 0.04 | 0.30 | -0.04 | -0.01 |
| 에 | -1.63 | 0.49 | 0.23 | 0.18 | -0.31 | -0.24 | 0.71 | -0.34 |
| 에: | -1.36 | 0.24 | 0.28 | 0.43 | -0.38 | -0.27 | 0.25 | 0.01 |
| 아 | -1.67 | 0.28 | 0.82 | 0.06 | -0.68 | -0.00 | 0.35 | -0.10 |
| 아: | -1.33 | -0.28 | 0.39 | 0.73 | -0.09 | -0.49 | 0.005 | 0.14 |
| 오 | -1.71 | 0.43 | 0.37 | 0.16 | -0.06 | -0.37 | 0.25 | -0.04 |
| 오: | -1.32 | -0.44 | 0.47 | 0.96 | -0.38 | -0.50 | 0.12 | 0.11 |
| 우 | -1.56 | 0.27 | 0.08 | 0.56 | -0.24 | -0.10 | -0.03 | 0.04 |
| 우: | -1.67 | 0.65 | 0.05 | 0.44 | -0.53 | 0.15 | -0.19 | 0.22 |
| 어 | -1.86 | 1.09 | 0.007 | -0.32 | -0.10 | 0.22 | 0.18 | -0.13 |
| 어: | -1.48 | 0.04 | 0.46 | 0.35 | -0.52 | 0.25 | -0.13 | 0.08 |
| 으 | -1.17 | -0.04 | -0.21 | 0.33 | -0.11 | 0.71 | -0.51 | 0.05 |

과 모음은 장단에 따라 파라미터의 값에 차이가 드러남을 알 수 있다. 그리고 이같이 추출한 성도 (Vocal tract)의 파라미터는 주파수 영역에서의 포락선 (Envelope)을 나타내므로, 이를 다시 FFT (Fast Fourier Transform)하여 주파수 영역에서의 진폭특성을 조사할 필요가 있다. 이렇게 하면 포르만트 특성과 밴드의 폭 (Band width)으로서 음성을 인식하기 위한 하나의 변수로 이용할 수 있다.

8. 맺는 말

1) 음성을 이용한 컴퓨터 입력 장치 개발이나 음성의 자동 인식 문제를 해결하기 위해서는 우선 한국어의 음성 / 음운의 체계와 음가를 정밀

하게 분석할 필요가 있으나, 지금까지 이에 대한 음성학/언어학적인 기초연구가 없이 실험위주의 부분적인 연구가 있었을 뿐이다.

2) 조음음성학 및 음향음성학적으로 분석한 결과 현대 한국어의 표준 발음에는 8개의 단순모음과 하나의 길어 음소가 설정되어야 한다. 즉, 장단별로 8개씩 모두 16개의 모음이 구별되고 있으며 이들의 장단별 음가는 모음사각도와 음향도표에서 분명히 차이를 드러낸다.

3) 선형예측 방법에 의한 파라미터의 변화율을 특별히 제작한 파라미터 분석기로 분석 검토한 결과도 위의 음성/언어학적인 분석 결과를 뒷받침함을 알 수 있다.

4) 앞으로 자음 및 고저와 강약등 악센트와 리듬에 관한 정밀 분석을 시도하여 종합적인 체계를 마련하는 일이 시급하다.

<이 현복 서울대 인문대 언어학과 교수>

<안 수길 서울대 공대 전자공학과 교수>