

통계적 방법과 인지실험을 통한
한국어 초성파열음의 음소단위 분석에 관한 연구

A Study on the Phoneme Based Analysis of Korean
Initial Plosives Using Statistical Method
and Perception Tests

조 철 우*, 이 우 선**, 이 규 호***,
김 종 안****, 임 광 일***** 이 태 원***

(Cheol-Woo Jo, Woo-Sun Lee, Gyu-Ho Ree, Jong-Ahn Kim,
Gwang-Il Lim, Tae-Won Rhee)

요 약

본 논문에서는 한국어의 규칙합성에 관한 연구중 파열음의 합성 파라미터를 추정하기 위하여 사용한 통계적 방법과 인지실험에 의한 방법에 관하여 기술하고 있다. 합성기로는 직렬 포먼트 합성기를 구성하여 사용하였고 통계적 분석에 사용된 음성시료는 9개의 초성 파열음과 8개의 모음으로 구성되는 72개의 독립 CV형태를 갖는 단음절을 단일 화자를 통하여 수집하였다. 음성의 분석은 시간 및 주파수 평면에서 파라미터의 변화를 중심으로 행하였으며, 인지실험을 통한 파라미터 추정방법을 통하여 독립적으로 포먼트 파라미터의 변화에 관하여 조사하였다.

ABSTRACT

This paper describes a statistical methods and perception test for extracting the parameters to be used for the synthesis-by-rule of Korean plosives. Formant synthesizer is chosen for the synthesis of the phonemes. Speech materials for the analysis consists of 72 CV monosyllables from the single male speaker. The analysis is done mainly focused on the variation of parameters in time and frequency domain, then perception tests are executed to estimate the effects of variations of the formant transitions.

I. 서 론

* 국립중앙대학 제어계측공학과
** 국립중앙대학 전자계산학과
*** 고려대학교 전자전산공학과
**** 한국전기통신공사 사업지원단
***** 대우공업전문대학 전자공학과

규칙에 의한 음성합성은 음성합성의 궁극적인 목표로 생각되어 많은 연구가 진행되고 있고 외국의 경우는 상품화까지 될 정도로 발전되어 있다.¹⁾ 그러

나 우리말인 한국어의 경우는 아직 음성의 특성에 관한 정보나 조음결합에 관한 많은 부분이 아직도 명확히 조사되지 못하고 있어 연구과제로 남아 있다. 한국어를 규칙에 의해 합성할 수 있게 하기 위해서는 이러한 합성을 위한 음성의 분석이 선행되어야 할 것이다. 특히 음소조각 연결방식이 아닌 파라미터에 의한 규칙합성을 행하기 위하여는 각 음소에서의 합성 파라미터의 변화 양상을 조사해야 할 필요가 있다.

한국어의 음소들 중에서도 파열음은 전체 음소의 종류중 가장 빈도수가 높은 것으로 조사된 바 있으며 전체 자음 19개중 9개로 절반정도를 차지하고 있고 한국어만의 특징적인 음소인 경음울 3개 포함하고 있어서 한국어에서 차지하는 비중이 크므로 본 논문에서는 파열음의 합성파라미터를 추정하기 위한 실험을 행하고 합성에 의해 그 결과를 확인하였다.⁽²⁾⁽³⁾ 분석에는 먼저 신호처리적인 방법에 의해 음성시료를 분석하여 측정하고자 하는 파열음의 지속시간을 측정한 뒤 인지실험에 의해서 각 파열음의 주파수 대역에서의 특성을 측정하는 방법을 사용한다. 이때 지속시간은 각 파열음의 시작점에서 변이 구간의 시작점까지로 정하였다.

본 논문이 적용한 인지 실험에 의한 방법(4)은 실제 음성으로부터 분석된 정보를 여러 사람의 청각적 인지도를 통하여 확인해 볼 수 있으므로, 규칙합성에 필요한 음성 파라미터의 변화에 대한 음소변화 정도를 측정하는 가장 유효한 방법중의 하나가 될 수 있다.

이상과 같은 방법을 통하여 수집한 음성시료에 대한 분석을 행하고 한국어 파열음을 포맷트 합성법에 의해 합성해 봄으로써 포맷트 합성법에 의해 한국어 파열음을 합성할 경우 필요한 파라미터의 영역을 추정하고 그 방법의 유효성을 확인하였다.

II. 파열음의 조음특성 및 합성정보

성대의 진동에 의해 생성된 음원이 성도에서 성도를 공명시켜 생성되는 모음과는 달리 자음은 어떤 부분에서의 폐쇄나 협착에 의한 제한을 받고 나온다. 자음의 특성은 이러한 제한을 받는 지점의 위치

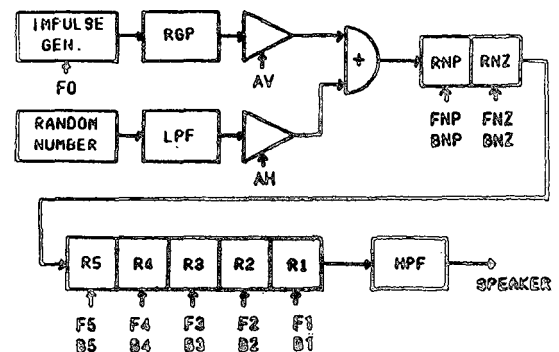
와 음원의 종류에 따라 결정된다. 자음은 조음위치에 따라서 또는 조음방식에 따라서 나누는 것이 일반적이다.

자음은 구강의 폐쇄로부터 발생되며 구강의 형태의 급격한 변화를 수반하기 때문에 모음에 비해 크기나 주파수성분등의 변화에 있어서 복잡한 조음 특성을 띠게된다. 파열음은 조음방식에 의한 자음의 분류에서 구강을 완전 폐쇄시켜 막힌 기류의 압력을 높였다가 폐쇄를 개방하면 압축된 공기가 방출되면서 나는 소리를 말한다. 한국어의 경우는 / ㄱ, ㄷ, ㅂ, ㅋ, ㅌ, ㅍ, ㅊ, ㅌ, ㅍ / 의 9개가 여기에 해당한다. 이중 한국어만의 독특한 조음방식인 경음은 / ㄱ, ㄷ, ㅌ / 가 있다.

파열음은 폐쇄의 시간과 성대의 진동시기를 기준으로 분류한다면 된소리, 예사소리, 거친소리로 분류될 수 있다. 경음은 폐쇄에서 성대진동까지의 시간이 가장 짧고 예사소리, 거친 소리의 순으로 길다.⁽²⁾⁽³⁾

III. 음성합성기의 구성.

분석된 파열음에 관한 분석자료를 확인해 보기 위하여 음소단위의 합성을 행할 수 있는 포맷트 합성기를 소프트웨어로 구현하였다.



- RGP: Glottal Pulse Resonator
- RNP: Nasal Pole Resonator
- RNZ: Nasal Zero Resonator
- R1-R5: Vocaltract Resonator 1-5
- F1-F5: Formant 1-5
- B1-B5: Bandwidth 1-5

그림 1 음성합성기의 블럭도

이 합성 프로그램을 이용하여 무성음과 그 후속 모음과의 결합을 분석된 결합규칙에 의해 합성할 수 있게 하여 분석된 자료에 대한 확인을 할 수 있게 하였다.

음성합성기는 그림1과 같은 구성을 갖도록 하였다. 본 연구에서는 직렬 공진기 형태로 포맷트 합성기를 구성하였다. 직렬 공진기의 형태는 각 공진부의 크기 파라미터를 일일이 지정해 주지 않아도 되고 인간의 성도 모델과 유사하다는 장점이 있다.

합성시는 유성음원 또는 무성음원에서 생성된 음원이 음원형성필터를 통하여 성도모델에 해당하는 필터 RNP, RNZ, R5, R4, R3, R2, R1들을 통과 함으로써 음성이 생성되게 된다.

IV. 음성시료 및 음성 수집환경

음성시료는 가능한 한 정확한 발성을 얻기 위하여 한국어의 발성훈련을 받은 바 있는 자를 대상으로 수집하였다. 음성시료의 선택은 음운환경을 CV형태

로 갖는 자음 / ㄱ, ㄷ, ㄴ, ㅋ, ㆁ, ㅌ, ㅍ, ㅊ, ㅍ / 와 모음 / ㅏ, ㅑ, ㅓ, ㅕ, ㅡ, ㅣ, ㅞ, ㅟ / 의 조합으로 택하였다.

시료음성은 가능한한 잡음이 없는 상태에서 순수한 음성만을 수집하기 위하여 방음장치가 되어있는 대학 교육방송국의 스튜디오에서 녹음을 행하였고, 발성자는 발성훈련을 받은바 있는 교육방송국 아나운서중 20세의 남자로 택하였다.

녹음은 고려대학교 교육방송국 스튜디오에서 행하였으며 외부잡음은 거의 없는 상태에서 녹음을 하였다.

녹음은 미리 준비된 임의의 순서를 갖는 발성표를 보고 충분한 시간 간격을 두고 발성하게 하였는데 발성표는 임의의 순서로 배열한 음절을 5개씩 구분하여 큰 활자로 표기하였고 동일한 자음에 관련된 음절별로 구분하여 만들었다.

녹음에는 금속 테이프를 사용하였는데 스테레오 트랙중 한쪽만을 이용하였다.

녹음된 음성은 IBM-PC에 부착된 12 비트 A/D 변환기를 통하여 표본화 하여 PC내의 하드 디스크

```
** SPEECH -wd1.a- ANALYSIS SPECS **
SAMPLES = 256, SAMPLING FREQ. = 10000, LPC ORDER 14
WINDOW = HANNING, SHIFT = 100, DURATION PER TIC = 10 MS
DATA SIZE : 13690
```

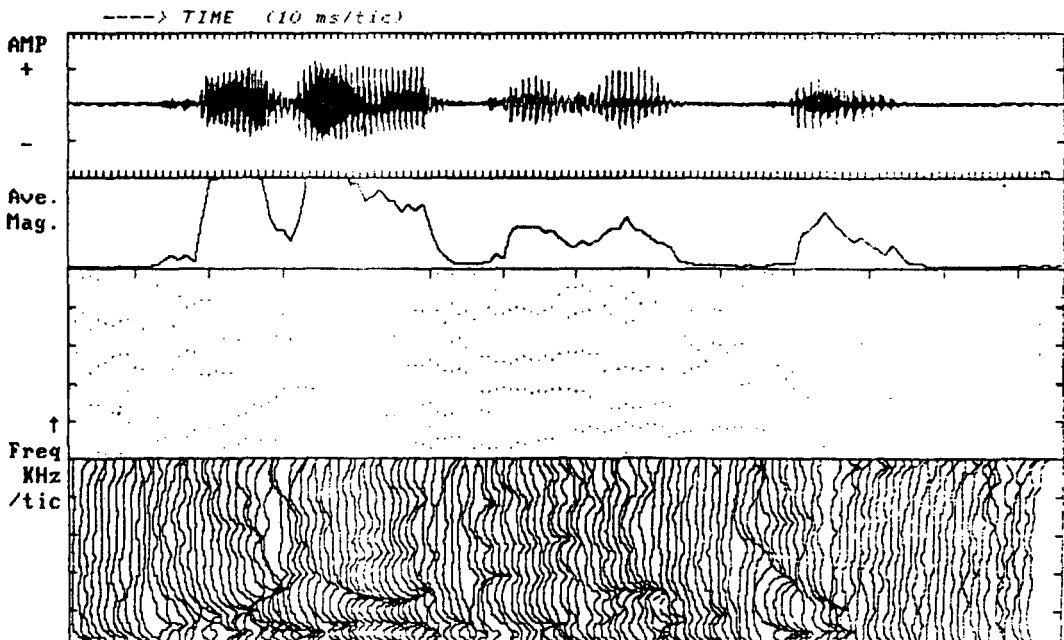


그림 2. 음성분석의 예("고려대학교"에 대한 음성분석)

상에 저장하였다.

V. 파열음의 음성분석

1. 통계적 방법에 의한 분석

파열음의 지속시간의 측정은 원래의 음성파형과 신호처리 방법에서의 분석된 파라미터 자료들 및 음성편집기에 의한 반복청취에 의해 행하였다. 신호처리방법에 의한 음성분석은 음성의 평균크기(Average Magnitude)와 연속 스펙트럼 분석으로 행하였다.⁽⁵⁾⁽⁶⁾⁽⁷⁾⁽⁸⁾ 그림2는 이러한 음성분석의 일 예를 보인 것이다. 자음의 지속시간은 음소가 위치해 있는 환경에 따라 여러가지로 측정될 수 있으나 본 연구에서는 초성 파열음의 지속시간의 측정에 한정하였다.

2. 인지실험에 의한 분석

일반적인 신호처리방법에서 얻은 음성분석자료의 불확실한 점을 보완하고 인간의 청각에 의한 파라미터 변동의 영향을 측정하기 위하여 인지실험에 의한 파라미터 추정 방법을 사용하였는데, 이 방법은 임의의 파라미터로 합성한 음성을 임의로 선정된 피험자에게 들려준 다음 피험자로 하여금 그 음을 판별하게 하는 것으로 파라미터의 변경은 주로 측정된 파열음의 포먼트 변화를 기준으로 하였으며 자음부의 음원은 무성음원으로 하고 모음으로의 변화 과정에서는 혼합된 음원이 발생되도록 음원 및 포먼트의 전이구간을 초기와 끝값을 기준으로 비례적으로 증감하도록 결정하는 선형 보간법을 사용하였다. 변경 파라미터는 자음의 지속시간, 자음의 초기 포먼트 제적으로 한정하였다.⁽⁹⁾

실험환경은 방음장치가 되어 있지 않은 실험실에서 직경 5inch 정도의 고성능 스피커를 사용하여 약 2m 전방에 10명의 피험자들을 위치하게 한 후 합성기로부터 합성된 미지의 음을 들려주었는데 주위의 잡음은 거의 없었다. 이때 피험자들은 합성음을 이전에 들어보지 못한 사람들 중에서 택하였고 합성음의 음색을 미리 알려주기 위해 실험 시작전에 임의의 합성음을 수회 들려주었다. 또한 제시되는 자음과 모음의 범위는 미리 알려주고 판별 불능일 경우는 별도로 기록하도록 하였다. 실험은 다음과

같은 과정과 방법에 의해 행하였다.

- (1) 합성할 자음부의 초기 주파수는 제1포먼트를 고정하고 제2, 제3포먼트를 가변한다.
- (2) 제2, 제3포먼트의 범위는 실제 음성에서 분석된 주파수를 대부분 포함하는 범위로 정한다.
- (3) 제2 포먼트는 900-1800 Hz까지를 100Hz 씩 증가시킨다.
- (4) 제3 포먼트는 2000-3000Hz까지를 500Hz 씩 증가시킨다.
- (5) 자음부의 지속시간은 0,20,40,80 ms로 변화시킨다.
- (6) 후속 모음은 / ㅏ / 로 고정한다.
- (7) 합성음은 120개의 CV 음절형태로 한다.
- (8) 합성음의 제시는 임의의 순서로 한다.

VI. 분석결과

1. 통계적 분석결과

그림3은 측정된 파열음의 지속시간의 분포를 도시한 것이다. 결과에 의하면 된소리에 속하는 파열음인 / ㄱ, ㄷ, ㅃ / 는 예상했던 대로 10-20ms의 짧은 지속시간을 가짐을 알 수 있었다. 또 동일한 조음위치의 자음에 대해서도 된소리-예사소리-거친소리의 순으로 지속시간이 길어짐을 알 수 있다. 이러한 조음형태의 변화에 따른 지속시간의 양적인 변화는 조음점에 따라 차이가 있으나 된소리는 예사소리에 비해 약 30ms 정도 지속시간이 길며, 거친소리는 예사소리에 비해 약 50ms 정도 더 길게 지속됨을 알 수 있다. 또 / ㄷ / 계열이 다른 파열음에 비해 약간

표 1 파열음의 지속시간분포

파열음소	지속시간(ms)	파열음소	지속시간(ms)
ㄱ	30-42	ㅃ	20-70
ㅋ	80-110	ㅍ	70-110
ㆁ	8-20	ㅂㅂ	8-10
ㄷ	20-40	-----	-----
ㅌ	60-80	-----	-----
ㄸ	8-20	-----	-----

짧음을 알 수 있다. 경음의 경우는 지속시간이 짧아 주파수 분석에서는 충분한 해상도를 얻기가 힘들므로 시간과형과 포먼트의 전이구간의 시작점을 기준으로 관찰에 의해 측정하였다.

Plosive Duration

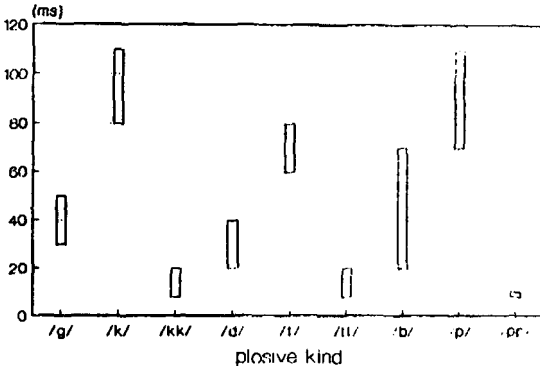


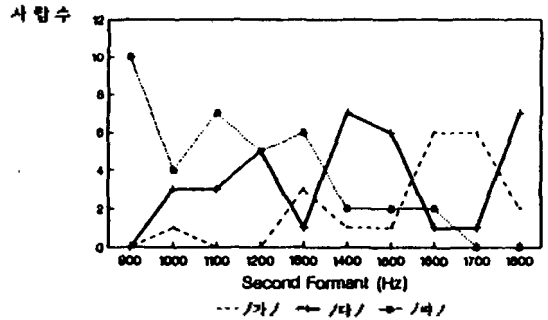
그림 3 파열음의 지속시간 분포

2. 인지실험에 의한 분석결과

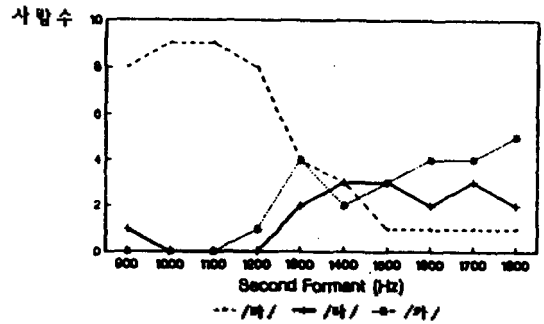
그림 4는 파열음의 파라미터 영역을 측정하기 위해 실시한 인지실험에 의한 분석결과이다. 결과에 의하면 파열음의 예사소리와 저친 소리간의 구분은 대개 자음부의 지속시간에 의해 가능하다는 것을 알 수 있었으며 F2의 영역이 /ㅂ/ 계열에서 가장 낮고, /ㄷ/, /ㄱ/의 순으로 높아짐을 알 수 있었다. 또 F3이 커질수록 /ㄷ/, /ㄱ/의 인지영역의 F2가 높아짐을 발견하였다. 가장 안정된 부분에서 얻은 값을 신호처리에 의한 분석결과와 비교했을 때 거의 일치하는 값을 얻을 수 있었다.

이 결과를 종합해 보면 /ㅂ/의 경우 제2 포먼트는 1200Hz 근처 제3 포먼트는 2500Hz 근처, 지속시간은 50ms정도의 값을 가지고 /ㄷ/의 경우 제2 포먼트는 1800Hz 근처, 제3 포먼트는 3000Hz 근처의 값에 지속시간은 30ms 정도의 영역에서 가장 뚜렷한 인지도를 나타내었으며, /ㄱ/의 경우는 제2 포먼트가 1800Hz 근처, 제3 포먼트는 2000Hz 근처, 지속시간은 30-50ms의 경우에 인지되었다. 그 밖에 동일한 파라미터에 대해 지속시간만을 변경하였을 경우에도 서로 다른 음소로 인식되는 경우가 있어 지속 시간의 조절이 자음소의 인식에 있어서 중요한 파라미터가 될 수 있음을 보여 주었다.

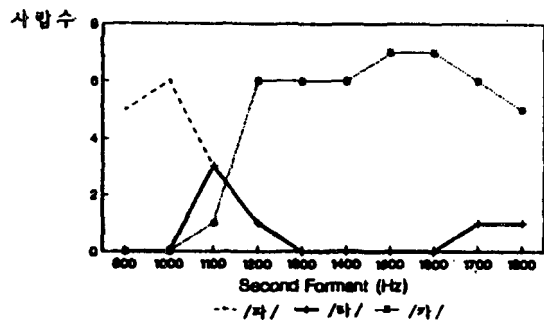
본 연구에서 행해진 인지실험에 의해서는 주로 파열음의 인지영역을 측정하기 위해 파라미터의 종류와 범위를 제한하였으나 본 연구에서는 고정된 합성기의 F4, F5 및 각 공진기의 대역폭에 대하여도 가변시켜 본다면 좀더 세밀한 파라미터의 추정이 가능할 것으로 생각되며 다른 음소에 대한 파라미터의 추정실험에도 유용할 것임이 이 실험의 결과를 통하여 확인되었다. 그러나 모든 자음에 관해 이러한 실험을 행하는 데는 많은 시간과 인원이 필요하므로 본 연구에서는 앞서 설명한 바와 같은 제한된 범위에 대하여 실험하였다.



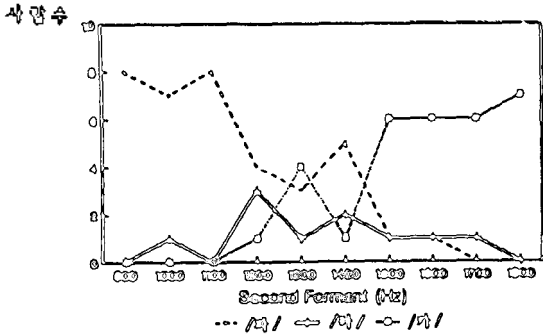
(a) 제 3포먼트=2000Hz, 지속시간=0. ms 인 경우



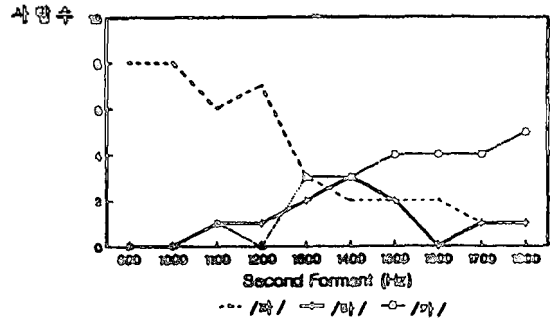
(b) 제3 포먼트=2000Hz, 지속시간=20 ms인 경우



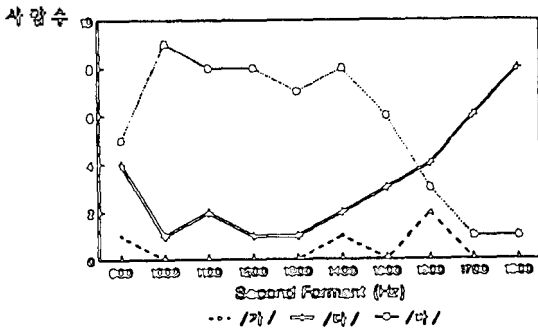
(c) 제3포먼트=2000Hz, 지속시간=40ms 인 경우



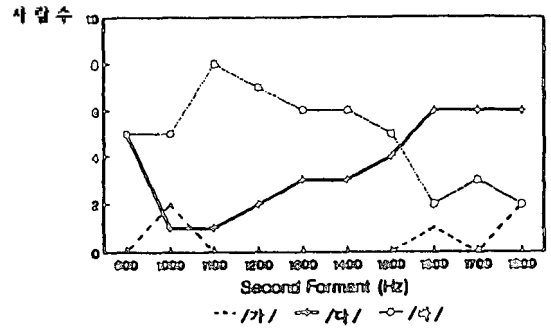
(d) 제3포만트=2000Hz, 지속시간=80ms 인 경우



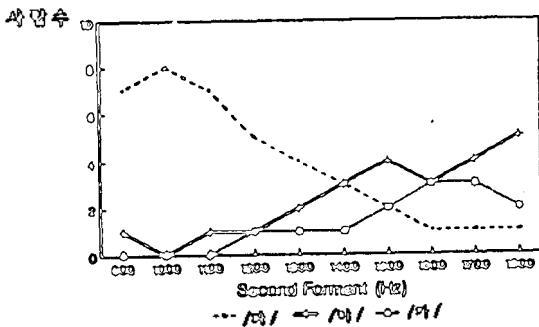
(h) 제3포만트=2500Hz, 지속시간=80ms 인 경우



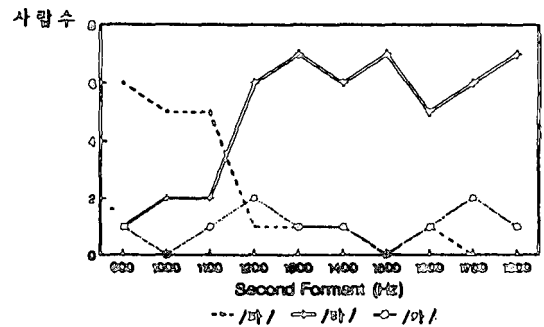
(e) 제3포만트=2500Hz, 지속시간=0ms 인 경우



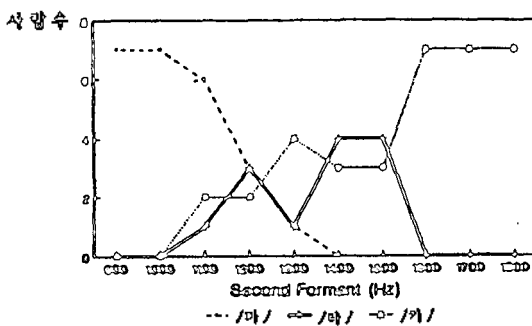
(i) 제3포만트=3000Hz, 지속시간=0ms 인 경우



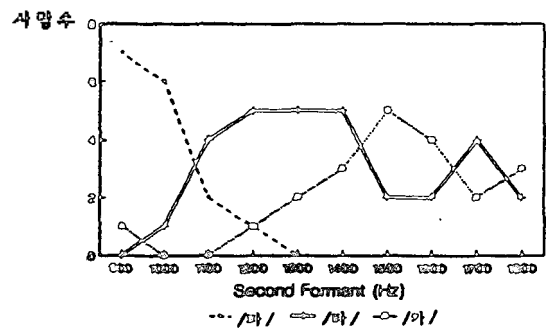
(f) 제3포만트=2500Hz, 지속시간=20ms 인 경우



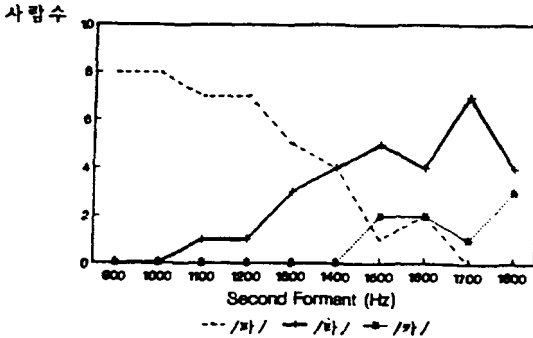
(j) 제3포만트=3000Hz, 지속시간=20ms 인 경우



(g) 제3포만트=2500Hz, 지속시간=40ms 인 경우



(k) 제3포만트=3000Hz, 지속시간=40ms 인 경우



(1) 제3포먼트=3000Hz, 지속시간=80ms 인 경우

그림 4. 인지실험에 의한 분석결과

V. 결 론

본 연구에서는 한국어의 규칙합성을 위한 기초적인 연구로서 파열음의 음소단위 음성분석을 행하였으며 분석에는 신호처리적 방법을 통하여 분석한 뒤, 통계적인 방법 및 인지 실험에 의한 방법을 사용하였다.

그 결과 한국어에 포함되어 있는 9가지 파열음에 대하여 포먼트 합성시 사용할 수 있는 특성 파라미터를 얻을 수 있었는데 지속시간의 길이면에서 보통 소리는 20-80ms, 격음은 60-110ms, 경음은 8-20ms 정도였고 인지실험을 통하여 제2 포먼트와 제3 포먼트의 위치가 파열음소간의 구분에 가장 중요한 요인임을 알 수 있었다.

사용된 방법중 인지실험에 의한 방법은 미지의 합성 파라미터에 따라 합성된 소리를 피험자들에게 들려주고 음을 구별하게 하는 방법으로 신호처리적인 방법에 의해 분석된 음성 파라미터와 인지실험에서 구해진 파라미터를 비교하여 구해진 파라미터의 유효성을 확인하는 방법으로 적합함을 실험의 결과에 의해 알 수 있었다.

이와 같은 실험과정에서 앞으로 해결해야 할 문제점으로는 인지실험 과정에서 협조적인 피험자의 선택과 변경할 파라미터의 범위 및 증분치의 적절한 선택방법등을 들 수 있다. 향후의 피험자의 선택에 있어서는 사전에 실험의 목적을 충분히 이해시킬 필요가 있으며 파라미터의 증분을 보다 세밀히 하여 파라미터의 변동에 따른 음성인지영역을 더 자세히 추정할 수 있을 것이다.

본 연구의 결과로 얻어진 연구방법 및 자료는 차후 한국어 문자음성 합성 시스템에 관한 연구중 음소단위 합성에 관한 부분에 유용한 기여를 할 수 있을 것이다.

본 연구는 한국과학재단의 1987-1989 기초연구비 지원에 의한 연구의 일부로 수행되었습니다.

참 고 문 헌

1. D.H.Klatt, "Review of Text-to-Speech conversion for English", JASA Vol.82, No.3, sept 1987.
2. 김진우, "언어", 탐출판사, 1985.
3. 김영송, "우리말 소리의 연구", 과학사 1981.
4. G.J.Borden, K.S.Harris, Speech Science Primer, Williams & Wilkins, 1980.
5. I.H.Witten, "Principles of Computer Speech", Academic Press, 1982.
6. L.R.Rabiner, R.W.Schafer, "Digital Processing of Speech signals", prentice Hall 1978.
7. J.D.Markel, A.H.Gray, "Linear Prediction of Speech", Springer-Verlag, 1976.
8. S.Lawrence, Marple Jr., "Digital Spectral Analysis with Applications", Prentice Hall, 1987.
9. C.W.Jo, T.W.Rhee, "A Study on the Analysis of Korean Unvoiced Consonants for Synthesis-by-Rule", Proceedings of JTC-CSCC '88. pp.393-396, Nov. 19

△조 천 우



1961년 9월 12일생
1983년 2월 고려대학교 전자
공학과 졸업(공학사)
1985년 2월 고려대학교 교학
원 전자공학과 졸업(공
학석사) 대 ()
1989년 2월 고려대학교 대학
원 전자공학과 (공학박
사)

1989년 3월~현재 국립중앙대학 제어제측공학과
※관심분야: 음성의 규칙함성, 신호처리, 맨-머신
인터페이스, 의용신호처리

△이 우 선



1957년 4월 5일생
1981년 2월 고려대학교 전자
공학과 졸업(공학사)
1983년 2월 고려대학교 대학
원 전자공학과 졸업(공
학석사)
1988년 8월 고려대학교 대학
원 전자공학과(공학 박
사)

1987년 3월~현재: 국립중앙대학 전산학과
※관심분야: 음성처리

△이 규 호

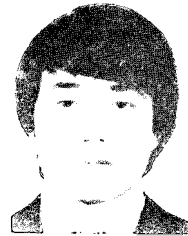


1960년 11월 21일생
1986년 2월 고려대학교전자
공학과 졸업(공학사)
1989년 2월 고려대학교 대학
원 전자공학과 졸업(공
학석사)
1989년 8월~현재: 고려대학
교 대학원 전자공학과
박사과정 재학중

△김 종 안

1961년 6월 21일생
1984년 8월 고려대학교 전기
공학과 졸업(공학사)
1988년 2월 고려대학교 대학
원 전자공학과 졸업(공
학석사)
1988년 2월~현재: 한국전기
통신공사 사업지원단

△임 광 일



1954년 8월 27일생
1981년 2월 고려대학교 전자
공학과 졸업(공학사)
1985년 8월 고려대학교 대학
원 전자공학과 졸업(공
학석사)
1985년 9월~현재: 고려대학
교 대학원 전자공학과
박사과정

1988년 3월~현재: 대우공업전문대학 전자공학과

△이 태 원



1931년 7월 27일생
1958년 서울대학교 통신공학
과(공학사)
1960년 서울대학교 대학원 전
자공학과(공학석사)
1975년 8월 서울대학교 대학
원 전자공학과 (공학박
사)

1963년~1971년 광운대학교수

1971년~1976년 중앙대학교 교수, 전자계산소장
1981년~1982년 미국Cornell대 방문교수
1977년~현재 고려대학교 전자전산공학과 교수
1985년~현재 고려대학교 전자계산소장
1987년~1988년 대한전자공학회장

※관심분야: 음성신호처리, 영상신호처리