

## 음소를 이용한 한국어의 인식

## Korean Speech Recognition using the Phoneme

\* 김 영 일 (Kim, Young Il)  
 \*\* 조 문 재 (Jho, Moon Jie)  
 \*\*\* 차 일 환 (Cha, Il Whan)

## ABSTRACT

As Korean language(Hangul) can be phonemically classified according to the characteristic and structure of its pronunciation, Korean syllables are divided into the phonemes such as consonants and vowels. The divided phonemes are analyzed by using the methods of liner prediction, partial autocorrelation and formant.

In this study, 256 data are used for the analysis and the orders of linear prediction and partial autocorrelation coefficient are 15.

This experiment is performed for 84 Korean syllables with two speakers and, as for vowels, the recognition scores of linear prediction coefficient, partial autocorrelation coefficient and formant are 95.2(%), 92.9(%) and 97.6(%) respectively. As for consonants, the recognition scores of linear prediction coefficient and partial autocorrelation coefficient are 88.7(%) and 92.9(%) respectively. As for syllables, those are 83.9(%) and 86.3(%) respectively.

In this study, it is shown that Korean syllables, divided into the phonemes, are analyzed and recognition with the minimum data and short processing time. Furthermore, it is shown that Korean syllables, words and sentences are analyzed and recognized in the same way and it is possible to synthesize Korean language by the regular combination of the phonemes.

## 要 約

본 연구는 한국어의 발음상의 특징과 구조에 의해서 한국어를 음소별로 분리할 수 있음에 착안하여, 자음과 모음으로 구성된 한국어 단음을 자음의 음소와 모음의 음소로 각각 분리하여 인식하는 새로운 방법에 관한 연구이다.

특정 화자 2명에 대하여 한국어 단음 84자를 모음의 음소와 자음의 음소로 각각 분리하여 인식한 실험 결과, 모음을 인식한 경우에는 선형 예측 계수를 이용하면 인식률이 95.2(%)이고, 편자기 상관 계수로 92.9(%), 폴란트로 97.6(%)의 인식률을 얻었고, 자음을 인식한 경우에는 선형 예측 계수로 88.7(%), 편자기 상관 계수로 92.9(%)의 인식률을 얻었다. 또, 자음의 음소와 모음의 음소를 결합시킨 단음을 인식한 경우에는 선형 예측 계수로 83.9(%), 편자기 상관 계수로 86.3(%)의 인식률을

\* 연세대학교 대학원 전자공학과

\*\* 한국 표준 연구소

\*\*\* 연세대학교 공과대학 전자공학과 교수

얼었다. 이 때, 각 음소들의 데이터의 수는 256개이고, 선형 예측 계수와 편자기 상관 계수의 예측 차수는 15차이다.

이와 같이 한국어를 자음의 음소와 모음의 음소로 분리하면 작은 데이터 양으로 처리 시간을 단축시켜 한국어의 모든 단음, 단어, 연속음, 문장 등을 분석하고 인식할 수 있고, 또한 각 음소들을 규칙적으로 결합시켜 모든 한국어의 합성이 가능함을 알 수 있다.

1. 서 론

한국어는 표음적인 음소 문자로 어느 음운 체계에도 알맞도록 제작되었으므로, 소리말로서 각 소리마다 입의 모양과 혀의 위치에 따라 발생된다. 그러므로, 모음인 경우는 입의 모양에 따라 정해지지만, 자음과 모음이 합쳐진 형태인 한국어의 단음은 혀, 잇몸, 입술에서 먼저 자음을 발생하고, 어느 정도 시간이 경과한 후 모음의 입 모양으로 변천한다.

지금까지 연구된 한국어에 대한 연구는 주로 숫자음의 분석과 인식에 관한 것으로, 이들의 연구는 단어 자체의 비교에 의해서 몇 개의 숫자음만을 인식하였기 때문에 대량의 기억 용량이 필요하고, 또한 처리 시간이 비교적 많이 소요되었고, 인식할 수 있는 단어가 제한되었다.

그러므로, 본 연구에서는 한국어의 발음 상의 특징과 구조에 의해서 한국어는 음소별로 분리가 가능함에 착안하여 자음과 모음으로 구성된 한국어 단음을 자음의 음소와 모음의 음소로 각각 분리하고, 선형 예측법과 편자기 상관법 그리고 폴만트주파수에 의해 분석하여 분석된 데이터를 이용하여 한국어 단음음 음소법에 의하여 인식할 수 있는 새로운 방법을 제시하고자 한다.

이러한 음소법으로 한국어를 자음의 음소와 모음의 음소로 분리하는 방법에 의해서 처리하면, 작은 데이터 양으로 처리 시간을 단축시켜 한국어의 모든 단음, 단어, 문장 등을 분석하고 인식할 수 있으며, 또한 음소들을 규칙적으로 결합시켜 모든 한국어를 합성할 수 있다.

2. 한국어의 음소

성대에서 입술까지인 성도를 그림 1과 같이 한 개의 음향관으로 볼 수 있으므로, 음성은 성대의 진동 또는 조음점의 근처에서 생긴 난기류

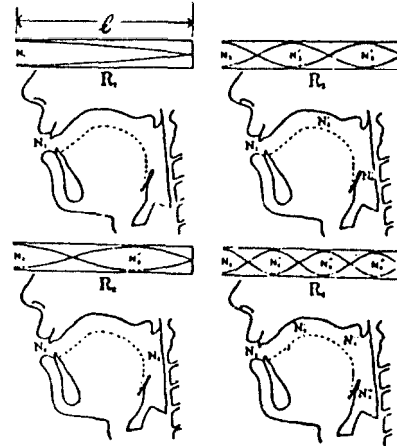


그림 1. 성도의 모델  
Fig. 1. Model of the vocal tract

에 의해서 음원이 성도를 통과할 때 성도의 모양에 따라 음향적 특성을 갖는다. 따라서, 각 소리마다 성도의 모양이 다르기 때문에 각각 다른 음향적 특성을 갖고 입술에서 외부로 방사된다.<sup>1), 2), 3)</sup>

한국어의 자음을 발음할 때, 각각의 자음에 대해서 조음 위치와 조음 상태 능이 다르기 때문에 자음들은 각각 서로 다른 음향적 특성을 갖는다. 또, 동일한 자음은 개인성과 시간성에 의해 조음 위치와 조음 상태에 약간의 차이가 있지만 조음 위치와 조음 상태가 거의 비슷하므로 그들의 음향적 특성은 유사하다.

한국어의 모음을 발음할 때, 각각의 모음에 대해서 입술의 모양이나 입안의 모양, 혀의 위치 능이 다르기 때문에 모음들은 각각 서로 다른 음향적 특성을 갖는다. 또, 동일한 모음은 개인성과 시간성에 의해 성도의 모양에 약간의 차이가 있지만, 모양이 거의 비슷하므로 그들의 음향적 특성은 유사하다.

한국어 단음은 표 1과 같이 자음과 모음들이 결합하여 이루어지며, 단음이 모여 단어를 구성

표 1. 한국어 단음의 구조

Table 1. Structure of Korean syllables

	단음의 구조	보 기
1	자음+모음	가
2	자음+모음+자음	각
3	자음+모음+모음	와
4	자음+모음+모음+자음	환
⋮		

하고, 단어가 모여서 문장이 구성된다.  
 단음이 발생될 때, 먼저 자음이 발생된 후 뒤

이어서 연속되는 모음이 발생되어 하나의 단음을 형성하므로, 단음을 자음의 음소와 모음의 음소로 각각 분리할 수 있다. 즉, /가/를 “ㄱ”과 “아”인 음소로 나눌 수 있다.”

음성을 마이크로폰으로 수신하여 전기적 신호로 변환시켜 오실로스코프 또는 플로터로 음성 파형을 관측할 수 있다. 그러므로, 한국어의 모음 /아/, /어/, /오/, /우/, /으/, /이/의 음성 파형은 그림 2와 같다. 그림 2에서와 같이 각 모음의 파형은 동일한 파형이 반복되지만, 각 모음의 파형에서 1주기만을 취하여 서로 비교하면 모음들의 파형은 서로 다르다.

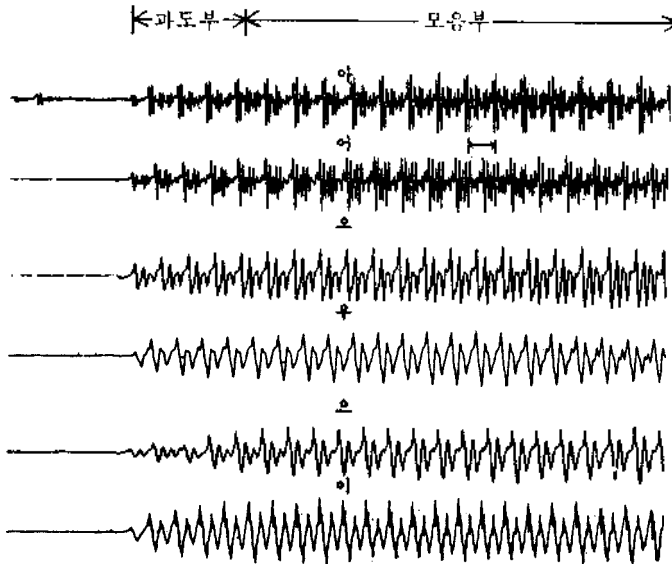


그림 2. 모음의 파형

Fig. 2. Waveform of Korean vowels

한국어 단음 중에서 /가/, /다/, /사/, /자/, /하/, /나/, /마/의 음성 파형은 그림 3과 같다. 이러한 음성 파형들은 먼저 자음 파형이 나타나고, 과도 부분을 거쳐 단모음 /아/의 주기적인 파형이 나타난다. 각 단음들의 파형 중에서 모음 파형은 동일함을 알 수 있다.

한국어 단음 중에서 /가/, /저/, /고/, /구/, /그/, /기/의 음성 파형은 그림 4와 같으며, 계속되는 모음의 종류에 따라 자음 파형은 조금 변화하고 있다.

단어나 문장을 발음할 때에도 각 모음과 자음 부분의 음성 파형은 짧아지지만, 각각의 모음과

자음 부분은 독립하여 발음할 때와 비슷하다.

이상에서와 같이 한국어 단음을 분석·합성·인식할 경우 음성 파형에서 자음의 음소와 모음의 음소로 각각 분리하여 분석·인식·합성을 함으로써 처리할 데이터 양과 처리 시간을 줄일 수 있고, 더 나아가서 단어, 문장 등 모든 한국어를 작은 데이터 양과 빠른 처리 시간으로 분석·인식·합성할 수 있다.

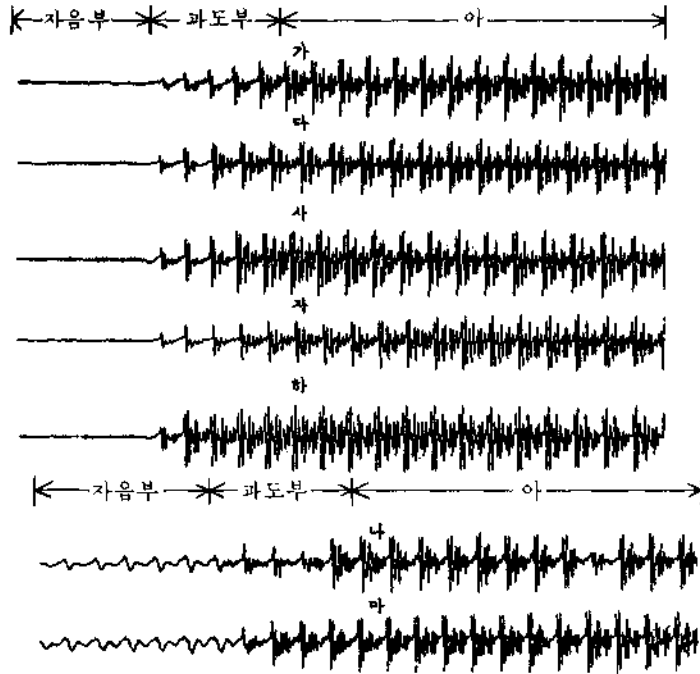


그림 3. 한글 단음의 파형

Fig. 3. Waveform of Korean syllables

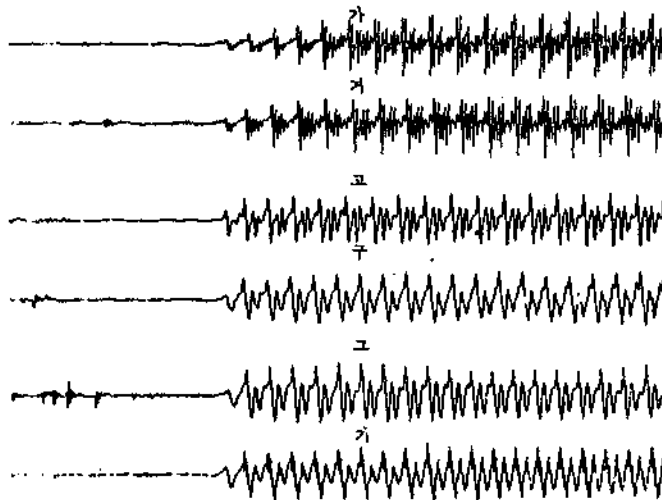


그림 4. 모음의 종류에 따른 자음의 변화

Fig. 4. Consonant variation by vowels

3. 실험 및 결과 고찰

본 연구에서 구성한 단음을 분석하고 인식하는 전체 시스템의 구성도는 그림 5와 같다.

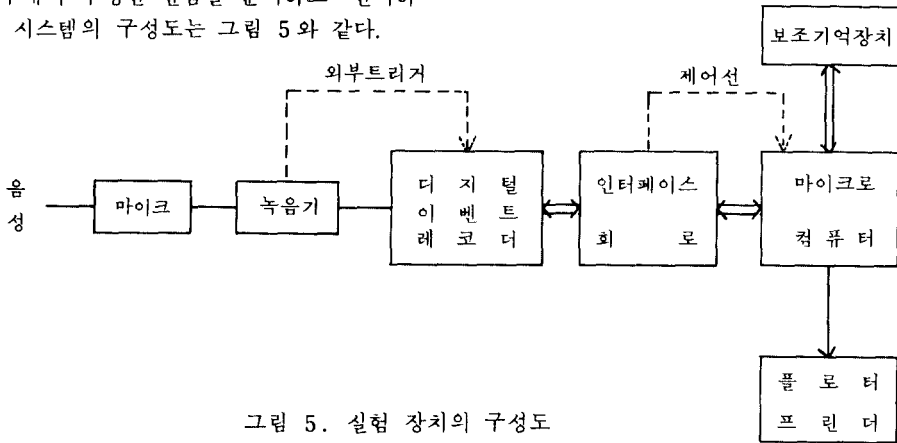


그림 5. 실험 장치의 구성도  
Fig. 5. Block diagram of experimental system

남자 아나운서가 보통 속도로 발음한 한국어 단음을 녹음하여 디지털 이벤트 레코더(digital event recorder)에 입력시켜 A/D 변환을 하였다.

A/D 변환시 샘플링 주파수는 10[KHz] 이고, 8 비트로 양자화하여 2의 보수의 데이터를 기억시켰다. A/D 변환의 시작 신호는 녹음기의 출력 신호의 레벨을 이용하여 외부 트리거(trigger)시켰다. A/D 변환된 데이터를 D/A 변환시켜 스피커 또는 오실로스코프를 사용하여 데이터를 확인하였고, 또 A/D 변환된 데이터를 인터페이스 회로를 사용하여 마이크로컴퓨터로 전송시켰다. 인터페이스 회로는 PPI(programmable peripheral interface) 칩 8255A-5 2 개를 사용하여 마이크로 컴퓨터와 데이터의 송·수신을 하였는데, 데이터의 전송 방법은 핸드 셰이크(handshake) 방식을 사용하였다.

인터페이스 회로를 통하여 디지털 이벤트 레코더와 마이크로컴퓨터의 데이터를 전송할 때, 마이크로컴퓨터에서 어셈블리어를 사용하여 프로그램으로 모든 제어 명령을 수행하였다.

디지털 이벤트 레코더에서 마이크로컴퓨터로 전송된 음성 신호를 분석하여 특징 파라미터(선형 예측 계수, 편자기 상관 계수, 폴란트 주파수)를 추출하고, 이 추출된 특징 파라미터와 표준 패턴과 비교하여 음성 신호를 인식하는 과정은 마이크로컴퓨터로 수행을 하였다.

한국어 단음을 자음의 음소와 모음의 음소로 각각 분리하여 선형 예측 계수와 편자기 상관 계수로, 모음인 경우에는 폴란트도 포함하여 분석한다. 이 때, 분석한 자음과 모음의 데이터 수는 각각 256개이다. 선형 예측 계수의 예측 차수는 성도의 공진 특성인 폴란트 주파수의 추출과 밀접하게 관계가 있고, 예측 차수는 주로 샘플링 주파수에 의해서 정해지는데, 모음 /이/에 대하여 예측 차수를 1 차에서 21차까지 분석한 결과, 예측 차수가 5 차이하에서는 성도 정보의 탈락이 현저하였으며, 예측 차수 13차~16 차까지는 비슷한 특성을 나타냈지만, 예측 차수가 19차 이상인 경우는 다른 성도 정보의 추가가 현저히 나타났다. 그러므로, 자음의 무성음인 경우도 고려하여 선형 예측 계수와 편자기 상관 계수의 예측 차수를 15차로 한다.

선형 예측 계수와 편자기 상관 계수는 Durbin 법을 이용하여 계산하고, 폴란트는 피크- 피킹 법을 이용하여 추출한다. Durbin법을 이용하여 선형 예측 계수와 편자기 상관 계수를 구하는 순서도는 그림 6과 같다.

본 연구의 타당성을 확인 또는 검토하기 위해서, 모음의 한 주기만을 반복시키거나, 자음과 모음의 한 주기만을 취하여 이들을 각각 몇 주기 반복시키고, 자음과 모음 사이에 과도부를 연결하여 음소법에 의해서 자음의 음소와 모음의 음소를 규칙적으로 결합시켜 모음, 자음 + 모음

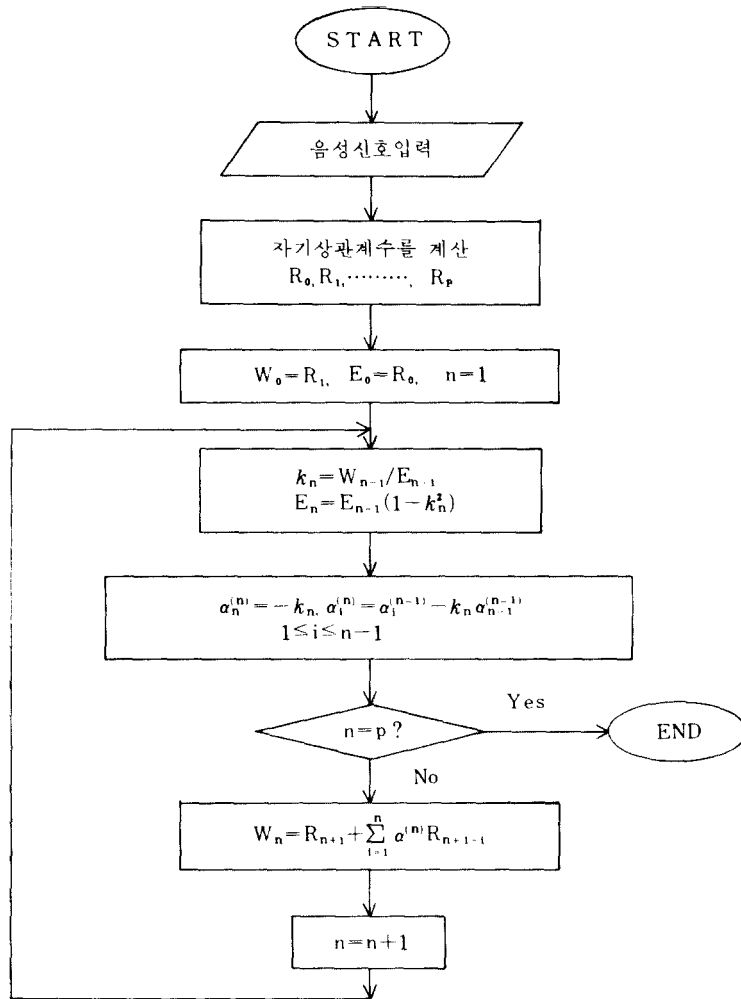


그림 6. 편자기 상관 계수와 선형 예측 계수를 구하는 순서도

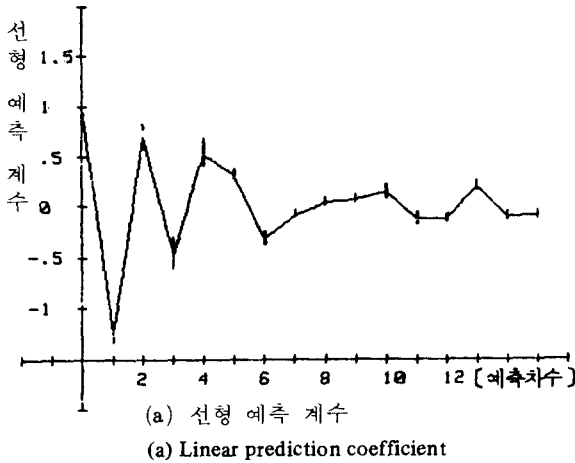
Fig. 6. The flow chart computing partial autocorrelation coefficient and linear prediction coefficient

형태의 단음, 모음+모음, 자음+모음+자음, 단어 등을 합성하였다.

음소변에 의해서 음성을 합성한 결과, 사람이 발음한 음성처럼 정확하지는 않았지만, 어느 정도 구별할 수 있었다.

그러므로, 한국어를 자음의 음소와 모음의 음소로 각각 분리하여 한국어 단음을 분석·인식 합성할 수 있음을 알 수 있다.

모음 /아/에서 과도부를 지나 정상 상태의 모음 파형을 순차적으로 분리한 후, 분리된 데이터를 선형 예측 계수, 편자기 상관 계수, 폴란트에 의해서 분석한 결과는 그림 7과 같다. 그림 7에서 보는 바와 같이 분석된 결과는 동일한 차수에서는 거의 같은 결과를 가지고 있으므로, 모음은 같은 성질이 반복됨을 알 수 있다. 그림 7의 (C)에서 고주파 부분의 파워스펙트럼



이 높은 것은 음성을 입력할 때 프리엠프시스 (pre-emphasis)를 행하였기 때문이다.

자음+모음의 형태로 된 단음, /무/의 파형 중에서 정상 상태의 모음 파형을 순차적으로 분리한 후, 분리된 데이터를 선형 예측 계수, 편자기 상관 계수, 폴만트에 의해서 분석한 결과는 그림 8과 같다. 그림 8에서 보는 바와 같이 동일한 차수 중에서는 거의 같은 결과를 가지고 있다. 그러므로, 단음 파형 중 모음 파형은 거의 같은 성질이 반복됨을 알 수 있다.

자음+모음의 형태로 된 단음 /기/, /니/, /디/, /리/, /미/, /비/, /시/, /이/, /지/, /치/, /키/, /티/, /피/, /히/에서 정상 상태의 모

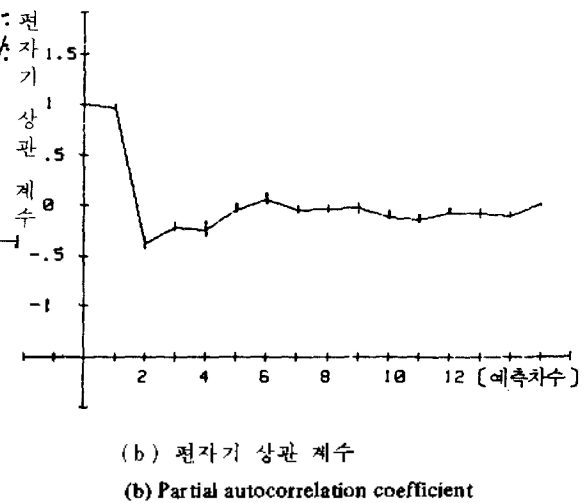
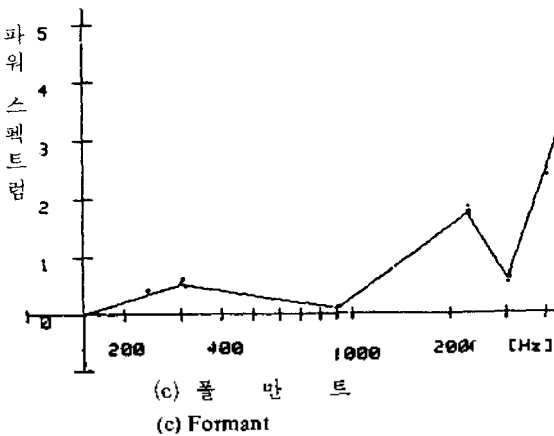
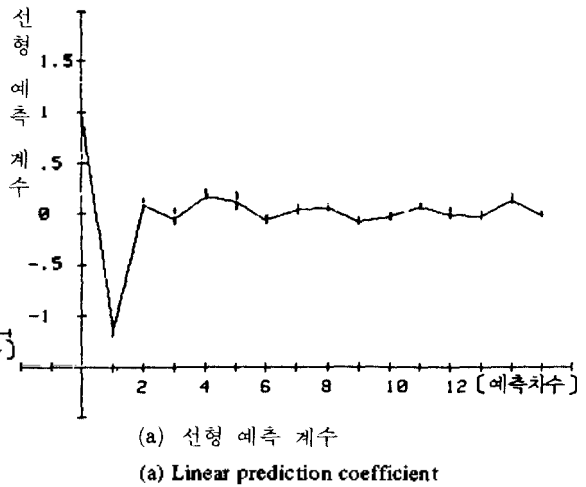
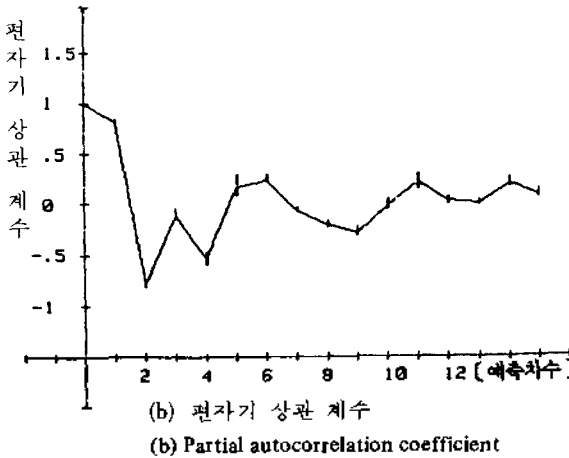


그림 7. 모음 파형의 주기성 조사 결과  
Fig. 7. Results of periodic measurement  
of Korean vowel

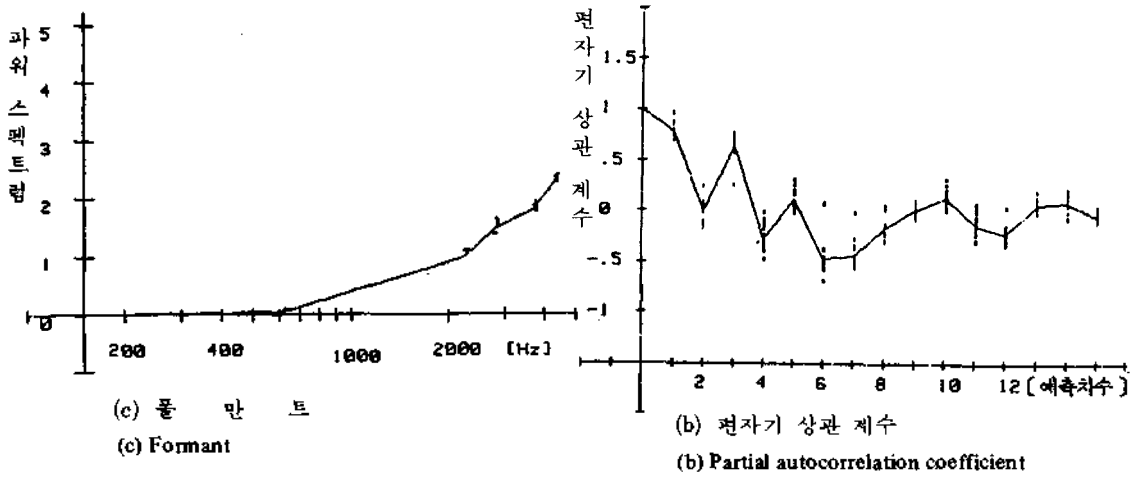


그림 8. 단음 파형 중 모음 파형의 주기성 조사 결과

Fig. 8. Results of periodic measurement of vowel in Korean syllables

음 파형을 각각 취한 후, 선형 예측 계수, 편자기 상관 계수, 폴만트로 분석한 결과는 그림 9와 같다. 그림 9에서와 같이 몇 개의 분석 결과만이 부분적으로 다르지만, 거의 비슷한 결과를 나타낸다.

따라서, 단음 파형 중에서 자음은 다르지만 모음이 같으면 모음 부분은 같은 성질을 가짐을 알 수 있다.

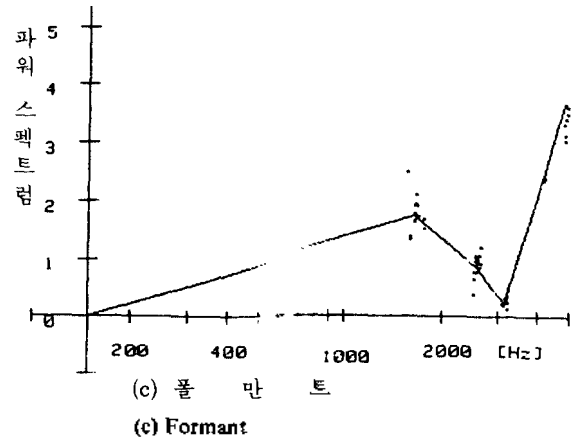
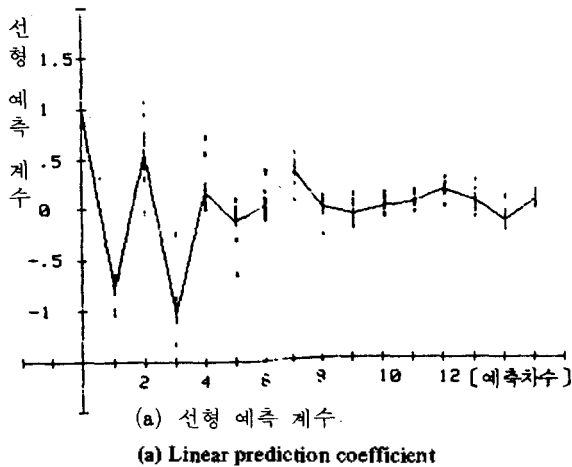


그림 9. 단음 파형 중 같은 모음에서의 유사성 조사 결과

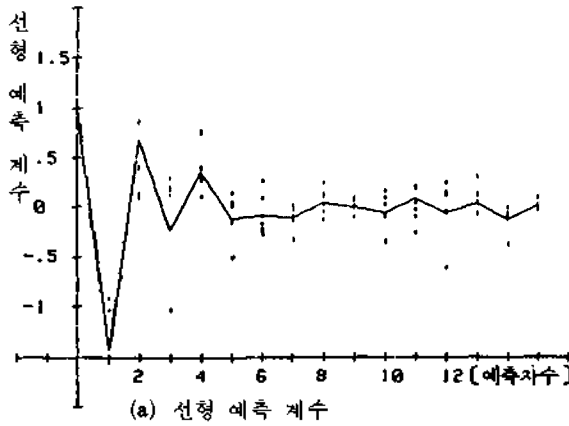
Fig. 9. Results of similar measurement of same vowel in Korean syllables

자음+모음의 형태로 된 단음 /가/, /거/, /고/, /구/, /그/, /기/에서 자음 부분의 파형만을 취하여 선형 예측 계수와 편자기 상관 계수로 분석한 결과는 그림 10과 같다. 계속되는 모음의 종류에 따라 /거/, /고/, /구/는 비슷하고, /가/, /그/, /기/는 같지 않음을 알 수 있다. 이것은 자음을 발음할 때의 조음 위치에서 계속되는 모음을 발생하기 위하여 혀의 이동이나 입모양의 변화에 의해서 영향을 받기 때문에 생긴다.

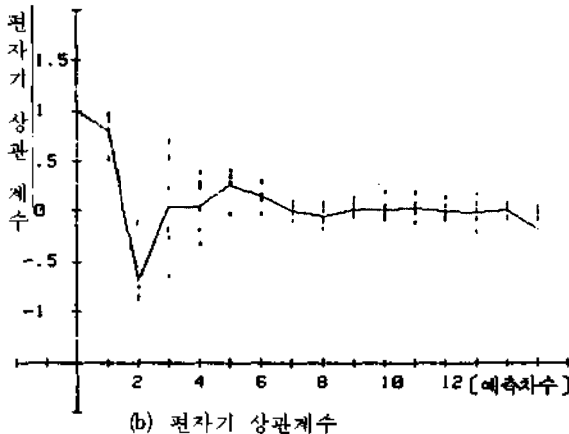
그러므로, 혀의 이동이나 입모양의 변화가 비







(a) Linear prediction coefficient



(b) Partial autocorrelation coefficient

그림 10. 단음 파형 중 같은 자음에서의 유사형 조사 결과

Fig. 10. Results of similar measurement of same consonant in Korean syllables

슷한 모음에 대해서는 자음 부분은 비슷함을 알 수 있다.

이상의 단음을 분석한 실험 결과에서 보는 바와 같이 단음을 자음의 음소와 모음의 음소로 분리하여 처리함에 있어서 동일한 자음의 음소, 동일한 모음의 음소는 그 특성이 비슷함을 알 수 있다.

그러므로, 이와 같은 성질을 이용하여 단음을 자음의 음소와 모음의 음소로 각각 분리하여 인식할 수 있다.

단음의 인식 과정은 그림 11과 같다.

단음의 파형에서 자음 파형보다 조금 높은 기준값을 설정하여 기준값과 비교하여 기준값(과도부 시작부분)에서 앞부분은 자음 파형 뒷부분은 모음 파형으로 분리했다.

특정 화자 2 명에 대해서 한국어 단음 84자의 인식 결과는 표 2 와 같다.

표 2. 2 명의 화자에 대한 인식 결과  
Table 2. Recognition results for two speakers

		화자1	화자2	계
모음	선형 예측 계수	96.4(%)	94(%)	95.2(%)
	편자기 상관 계수	92.9(%)	92.9(%)	92.9(%)
자음	선형 예측 계수	86.9(%)	90.5(%)	88.7(%)
	편자기 상관 계수	92.9(%)	92.9(%)	92.9(%)
단음	선형 예측 계수	83.3(%)	84.5(%)	83.9(%)
	편자기 상관 계수	85.7(%)	86.9(%)	86.3(%)

표 2 에서와 같이 모음의 경우에는 선형 예측 계수가 95.2(%), 편자기 상관 계수가 92.9(%)의 인식률을 얻었으며, 자음의 경우에는 선형 예측 계수가 88.7(%), 편자기 상관 계수가 92.9(%)의 인식률을 얻었고, 자음과 모음을 결합한 단음의 인식의 경우에는 선형 예측 계수가 83.9(%), 편자기 상관 계수가 86.3(%)의 인식률을 얻었다.

이상의 실험 결과에서 보면, 한국어를 자음의 음소와 모음의 음소로 각각 분리하여 인식을 행하면, 작은 데이터 양으로 처리 시간을 단축시켜 모든 한국어의 분석·인식·합성이 가능하다.

#### 4. 결 론

한국어의 발음상의 특징과 구조에 의해서, 한국어를 자음의 음소와 모음의 음소로 분리하여 처리하면 데이터 양과 처리 시간을 매우 감소시킬 수 있다. 그러므로, 한국어 단음을 자음의 음소와 모음의 음소로 각각 분리하여 분석하고 인식하는 실험을 행하였다.

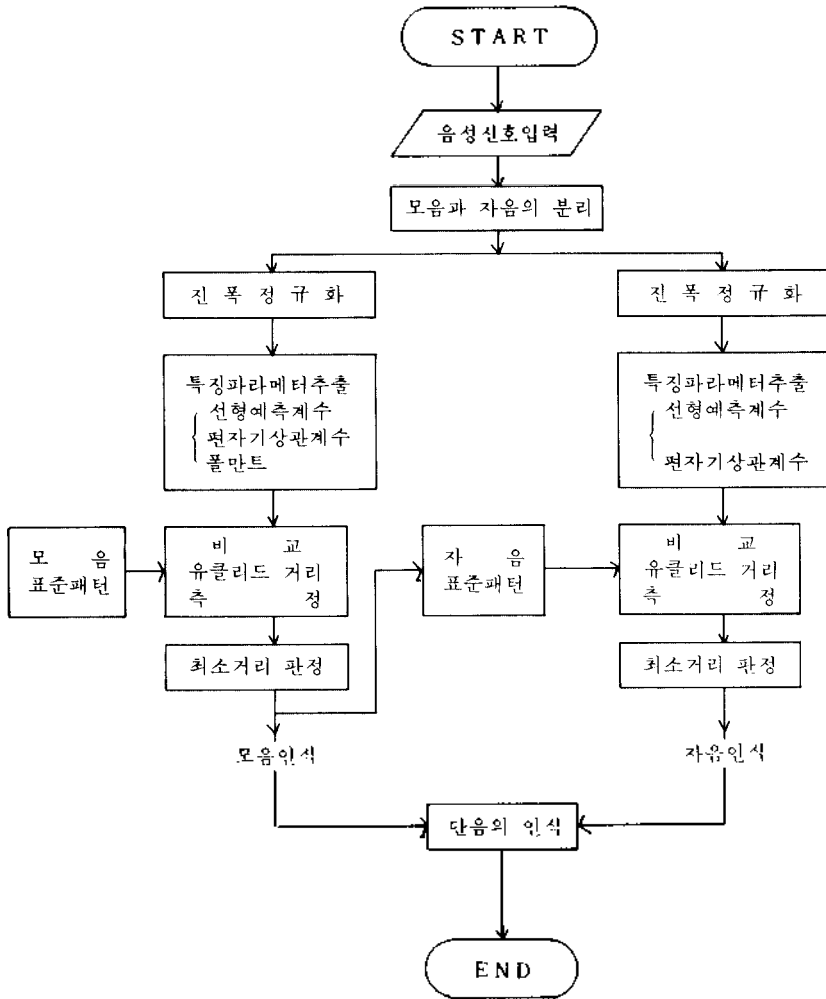


그림 11. 단음의 인식 과정

Fig. 11. The flow chart recognizing Korean syllables

먼저 한국어가 음소별로 분리가 가능함을 음소들을 규칙적으로 결합시켜 한국어불 합성하는 실험을 통하여 확인하였고, 자음과 모음이 합쳐진 한국어 단음에서 동일한 자음의 음소들과 동일한 모음의 음소들은 그 특성이 거의 유사하다는 것을 분석 실험을 통하여 확인하였다.

특정 화자 2명에 대해서 한국어 단음 84자를 모음의 음소와 자음의 음소로 각각 분리하여 인식한 실험 결과, 모음을 인식한 경우에는 선형 예측 계수가 95.2(%), 편자기 상관 계수가 92.9(%), 폴만트가 97.6(%의 인식률을 얻었고, 자

음을 인식한 경우에는 선형 예측 계수가 88.7(%), 편자기 상관 계수가 92.9(%의 인식률을 얻었다. 또, 자음의 음소와 모음의 음소를 결합시킨 단음을 인식한 경우에는 선형 예측 계수가 83.9(%), 편자기 상관 계수가 86.3(%의 인식률을 얻었다.

지금까지 연구된 한국어의 인식에 대한 연구는 단어 자체의 비교에 의한 몇 개의 숫자음만을 인식하였지만, 본 연구에서는 한국어 단음을 음소인 자음의 음소와 모음의 음소로 분리하여 작은 데이터 양으로 처리 시간을 단축시켜 한국

어의 모든 단음, 단어, 문장 등을 분석하고 인식할 수 있고, 또한 음소들은 귀적으로 결합시켜 모든 한국어를 합성할 수 있음을 알 수 있다.

#### 참 고 문 헌

1. L.R. Rabiner & R.W. Schafer, Digital Processing of Speech Signals. Prentice-Hall Inc., 1978.
2. J.D. Markel & A.H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, 1976.
3. J.L. Flanagan, Speech Analysis Synthesis and Perception, Springer-Verlag, 1972.
4. R.W. Schafer & J.D. Markel, Speech Analysis, IEEE Press, 1979.
5. N.R. Dixon & T.B. Martin, Automatic Speech & Speaker Recognition, IEEE Press, 1979.
6. 차일환, 음향공학개론, 한신문화사, 1976
7. 김영일, "한국어 단음의 분석과 인식에 관한 연구", 박사학위논문, 연세대학교 대학원, 1984. 12.