

## 남성의 숫자음 발성에 나타난 화자변이\*

### Speaker Variation in Number Production by Males

양 병 곤\*\*  
Byunggon Yang

#### ABSTRACT

The author analyzed acoustic parameters of ten Korean numbers produced by ten male students using Praat. Variations of  $f_0$ , F1, F2 and F3 within and between speakers were examined by determining an average and standard deviation of the parameters of each number and by comparing the acoustic values with one another. Results showed that each subject produced the numbers within a certain range of variation across time. Thus, speaker identification can be more certain using dynamic information of the acoustic parameters within each vocalic segment. Also, percent difference of within-subjects' variation to that of between-subjects can be utilized to determine which sounds would be better stimuli for speaker identification. According to the criteria, the number '2' proved the best stimulus while the number '7' was the worst. Future studies will be necessary to explore robust methods of speaker identification under noisy conditions.

**Keywords:** Production, Korean Numbers, Speaker Variation, Speaker Identification

#### 1. 머리말

사람의 발음은 서로 다른 크기의 성대와 성도 모양 때문에 기본주파수( $f_0$ )나 포먼트와 같은 음향적 분석값이 달라지게 된다(Fant 1960, 1975). 이러한 화자 변이는 언어적 요소와 비언어적 요소로 나눌 수 있는데(Traunmueller 1988), 언어적 요소는 방언이나 사회언어학적 요소들을 말하고, 비언어적인 요소는 나이와 성별, 정서적 차이 등을 말한다. 따라서, 동일한 목표음이라도 정서의 변화에 따라서도 물리음향적으로 달라지게 된다. 이런 음성 신호의 변화성은 자동음성인식 장치 개발에서는 최소화시켜서 동일한 목표음을 동일하게 처리할 수 있어야 하고, 웹기반의 전자 상거래와 같은 화자확인 장치에서는 최대화시켜서 처리해야 한다. 특히, 화자 확인을 하기 위해서는 화자내의 변화가 적고, 화자간의 차이가 많은 음향적 변수를 구해야 한다. 일반적으로 발성한 음성신호의 음향적 변수들은 화자 안에서도 수시로 변하지만, 합성실험(양병곤 2000, 2001)에서 살펴보았듯이, 무조건 변하는 것이 아니라 자신의 귀

\* 본 연구는 한국과학재단 목적기초연구(R01-1999-00229) 지원으로 수행되었음.

\*\* 동의대학교 영어영문학과

로 들어서 동일한 음성으로 지각하는 범위에서 변한다. 구체적으로 말해서, 동일하게 지각하는 변화 범위의 중심점과 발음할 때의 포먼트 값의 중심점은 매우 높은 상관관계를 보였다. 그렇지만, 보다 많은 화자들과 서로 구분되는 특징을 찾으려면 단순히 한 지점의 측정값만으로는 중복될 확률이 높게 된다. 20 명의 한국인과 미국인의 모음의 포먼트 값을 조사한 Yang (1996)의 측정값에서도 남녀 각 집단 내에서는 서로 일치되는 값들이 나타났다. 비록 남녀화자의 차이는 기본주파수( $f_0$ )와 포먼트 등 여러 가지 면에서 매우 현저한 차이를 보이기 때문에 화자확인 장치에서는 쉽게 구별되나 동일한 피치값을 가진 남녀 그룹 내에서는 서로 중복되는 경우가 많다. 이 현상은 이들의 발성이 거의 비슷한 성대와 성도 모양을 가지고 있기 때문이다.

이 논문에서는 양병곤(2000)에서 제시한 숫자음의 연구 방법을 이용하여 방언차이와 같은 언어적 요소의 영향을 받지 않는 열 개의 숫자음을 열 명의 화자가 각각 열 번씩 비슷한 음조로 발음하게 하여 음향적 변수값들을 구해 화자내 차이와 화자간 차이를 살펴보고자 한다. 이러한 연구를 통해 화자내의 변화가 적고, 화자간의 변화는 많은 숫자음을 결정하는 방법을 제시하여 화자확인에 필요한 기초자료를 제시하고자 한다. 구체적으로 다음과 같은 연구문항을 중심으로 살펴보기로 한다.

1. 각 숫자음에 대한 음향적 변수의 화자내 변이는 어떠한가?
2. 각 숫자음에 대한 음향적 변수의 화자간 변이는 어떠한가?
3. 숫자음 가운데 화자확인에 이용할 수 있는 숫자음의 우선 순위를 어떻게 결정하는가?

이 연구문항에 답하기 위해 열 명의 남성화자가 열 번씩 발성한 숫자음 열 개에 대해  $f_0$  값이 구해지는 유성음 구간을 20 등분하여 기본주파수( $f_0$ )와, 제 1, 2, 3 포먼트( $F_1$ ,  $F_2$ ,  $F_3$ )로 된 80 개의 변수를 개별 화자내 또는 화자간 비교에서 나온 절대값 차이로 탐구해 보기로 한다.

## 2. 연구 방법

### 2.1 피험자

동의대학교에 재학하는 건강하고 청각에 이상이 없는 남학생 10 명이 숫자음 녹음에 참여했다. 피험자의 나이와 키의 분포는 표 1과 같다. 나이의 평균은 24 세이고 키의 평균은 176 cm이다.

표 1. 피험자의 나이와 키

피험자번호	나이(세)	키(cm)	피험자번호	나이(세)	키(cm)
s1	24	167	s6	24	179
s2	23	170	s7	24	180
s3	21	172	s8	24	180
s4	25	174	s9	25	181
s5	24	177	s10	23	182

## 2.2 녹음 및 분석 과정

녹음은 조용한 연구실에서 AKG 마이크로 입력된 음성을 Shure Prologue 증폭기를 통해 G3노트북에 22,050 Hz의 표본 속도로 저장했다. 음성을 녹음하기 전에 피험자에게는 실험 용도를 설명하고 자연스럽게 발음하되, f0의 변화를 줄이기 위해 거의 비슷한 음조를 사용해주길 요청했다. 이들은 실제 인터넷에 연결된 컴퓨터를 향해 자신의 비밀 숫자음을 발음하듯이 또박또박 발음했다. 열 개의 숫자음은 열 번씩 되풀이 되게 만든 뒤 이를 엑셀의 난수생성 함수를 이용해서 섞었고, 동일한 숫자음이 연이어 나타나지 않도록 배열하여 한 종이에 크게 인쇄하여 화자가 보며 발성하도록 했다. 녹음된 음성은 각 숫자음별로 하나씩 잘라서 저장했으며, 음성 분석은 프라트의 스크립트를 작성하여 자동 처리했다. 음향적 파라미터의 구체적인 추출 방법은 양병곤(2000, 그림 1 참고)에 나타나 있다. 이 연구에서는 숫자음의 f0값이 0보다 큰 범위를 구한 뒤, 보다 정밀한 동적 변화를 측정하기 위해 다섯 개의 시간점 대신에 20 개의 시간점에서 f0, F1, F2, F3값을 구했다. 시작부분과 끝 부분은 변화량이 많기 때문에 각각 40 ms씩 제외했다. 이 과정은 모두 컴퓨터에 의한 자동작업으로 진행되었으며 각각의 데이터는 임의로 수정하지 않았음을 밝혀둔다.

수집된 음향적 변수로 총 80,000 개의 자료(10 명×10 번씩 발음×10 개 숫자음×20 개 시간점×4 개 음향변수)를 구했다. 이어서 이 값들은 엑셀을 이용하여 재배열하고, 화자내 변이를 구하기 위해서 첫 번째 화자가 발성한 숫자음의 음향적 변수를 모델로 정하고, 나머지 9 명을 각각 한 번씩 대조하여 각 시간점별로 음향적 변수의 절대 차이값(음수와 양수의 부호를 없앤 차이값)을 구했다. 이어서, 각 숫자음별로 이 절대 차이값을 모두 더하거나(숫자음별로 본 화자내 차이), 각 화자별로 더했다(화자별로 본 화자내 차이). 이어서, 화자간 차이를 살펴 보기 위해서는 s1이 발성한 숫자음의 음향적 변수를 기준으로 하고 나머지 9 명의 화자 가운데 한 사람씩 차례로 비교하여 20 개 지점의 4 개의 변수에 대한 절대적 차이값의 합을 구했다. 마지막으로 화자확인에 필요한 숫자음을 선택하는 우선 순위를 구하기 위해서는 각 숫자음별로, 화자간 절대 차이값의 합의 평균을 화자내 차이의 합의 평균으로 나눈 비율을 이용했다.

## 3. 분석 결과와 토론

### 3.1 화자내 변이

숫자음별로 또는 화자별로 전반적인 화자내 변이를 다루기 전에 자료의 특성을 세밀히 관

찰하기 위해서, s1이 발성한 숫자음 중 이중모음이 들어간 '0'과 단모음으로 된 '4'를 살펴보기로 한다. 그림 1은 s1이 열 번 발음한 숫자음 '0'의 음향적 변수의 궤적을 나타낸다.

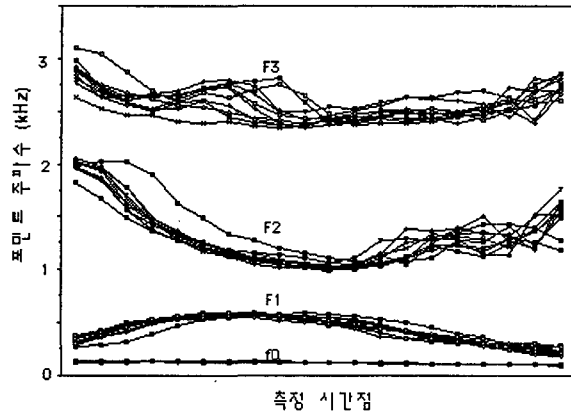


그림 1. s1이 열 번 발음한 숫자음 '0'의 f0, F1, F2, F3의 궤적

그림 1에서 살펴보면 성대의 진동을 나타내는 f0값의 변화는 매우 적고, F1도 중앙 부위에서는 매우 안정적인 모양을 보이고 있다. F2도 안정적이거나 하나가 약간 높게 예외적으로 발생되었고 14 번 지점 이후는 변화가 많다. F3은 처음 시작부가 2 개의 발화를 제외하고는 안정적이거나 6-11 번의 값에서 변화가 많다. 그림 2는 동일한 화자가 열 번 발성한 숫자음 '4'의 f0와 포먼트값의 변화를 보이고 있다.

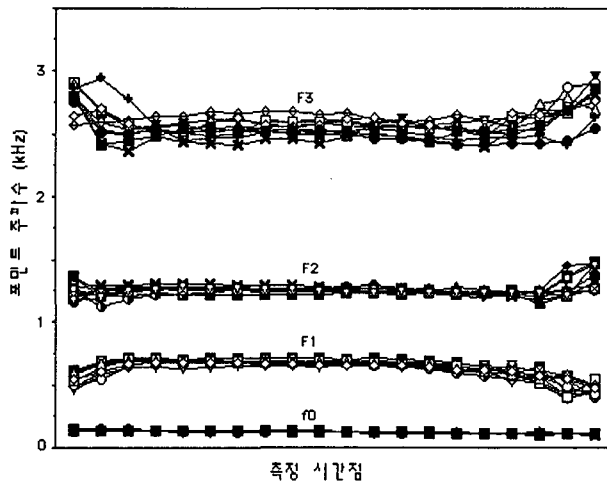


그림 2. s1이 열 번 발음한 숫자음 '4'의 f0, F1, F2, F3의 궤적

그림 2에서도 f0는 처음부터 끝까지 열 번 발음의 변화가 적고, F1과 F2는 시작과 끝 부분의 세 개 시간점에서 약간의 변화를 보이지만 중간에서는 거의 변화가 없다. F3은 F1과

F2에 비해 보다 넓은 폭으로 변하는 모양을 보인다. 시작과 끝 부분은 각각 40 ms의 부분을 제거했음에도 여전히 변화율이 높음을 알 수 있다. 이것은 인접한 초성자음 'ㄱ'에 의한 공동 조음의 영향과 끝부분의 음성신호의 진폭이 낮기 때문에 생긴 측정 에러 때문으로 여겨진다. 앞으로 음성신호 가운데 일정한 진폭값 이하는 분석에서 제외시키면 더 정확한 화자확인이 되는지 연구해보아야 할 것이다.

화자마다 모든 숫자음에 대해 그림을 그려보면 보다 더 명확한 경향을 파악할 수 있겠지만, 우리의 관심은 연구문항 1에서 제기했던 바와 같이 화자내에서 얼마만큼의 변화를 보이는 가다. 화자별 변화 정도는 각 숫자음별로 음향 변수의 표준편차값을 살펴보면 된다. 일반적으로 표준편차값이 클수록 발음할 때마다 변화가 많고, 적을수록 일정한 범위에서 안정적으로 발음한다고 말할 수 있다. 그림 3은 동일한 화자가 열 번 발성한 각 숫자음의 f0와 F1값의 표준편차값의 분포를 나타내준다.

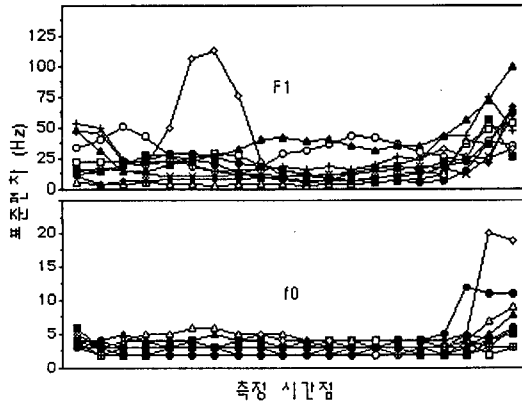


그림 3. s1이 열 번 발성한 각 숫자음의 f0와 F1값의 표준편차값의 분포

이 그림에서 보면 화자는 각 숫자음마다 피치값에서는 약 4 Hz 범위에서 안정되게 발음했다. 표준편차값은 통계적으로 정상분포곡선에서 68% 이상이 평균값  $\pm 4$  Hz 범위 내에 분포되어 있음을 나타내기 때문에(오택섭, 1996: 126), 지각실험(Yang, 2001)에 나타난 10.5 Hz의 허용범위에 근접한다. 이 허용범위는 합성용 모델 숫자음과 f0 변수를 올리거나 내리서 합성한 음을 연이어 들려주었을 때, 피험자들이 동일한 음성으로 판단한 범위다. 피치값의 변화율은 음절 말의 18~20 번 시간점에서 많은 변화를 보이고 있다. 이 부분에서 20 Hz의 편차를 보인 숫자음 '3'은 비음화된 영역의 피치값 추출에 변화가 많음을 나타내준다. F1의 편차값에서 가장 돋보이게 나타난 부분은 숫자음 '3'의 6~7 번 시간점이고 이는 비음성의 영향으로 추정된다. F1에서도 여전히 편차값이 숫자마다 달라지고 있음을 알 수 있고 시작과 끝 부분에 보다 높은 편차를 보이고 있다. 이 화자의 각 숫자음에 대한 f0 평균값을 그래프로 그려보면 그림 4와 같다.

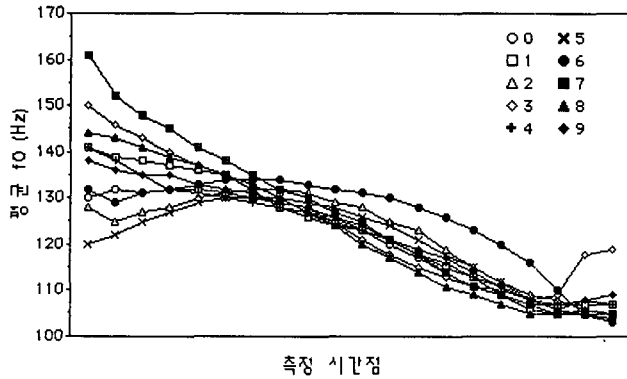


그림 4. s1이 열 번 발성한 각 숫자음의 f0 평균값.

이 그림에서 살펴보면 1 번에서 6 번 시간점까지의 피치값이 숫자음마다 다른 모양을 보이고 있으나 7~9 번 시간점에서는 매우 안정적으로 발음하고 있음을 알 수 있고 숫자음 '6'을 제외하고는 중반부의 안정도가 매우 높다. 그림 3의 표준편차 값이 11 번에서 18 번까지 매우 낮은 4 Hz 전후였음을 생각해보면 7~9 번 시간점의 피치값도 개인별 특성으로 지정할 수 있는 가능성을 보인다. 덧붙여, 초기의 피치값을 숫자별로 살펴보면 '7, 3, 8, 9'의 순서로 초성자음이 있는 숫자음들의 피치값이 높고, '6, 0'은 거의 비슷한 위치에, 이어서, '2, 5'의 어두 자음이 없는 숫자음으로 나타나 있다. 공동조음(coarticulation)의 영향으로 해석할 수 있기도 하지만, 인접 무성자음의 피치값 측정에 가끔씩 에러가 발생하기도 하기 때문에 이 값의 정확도를 차후에 검토할 필요가 있다. 그림 5는 s1의 F2와 F3의 표준편차값을 나타내고 있다.

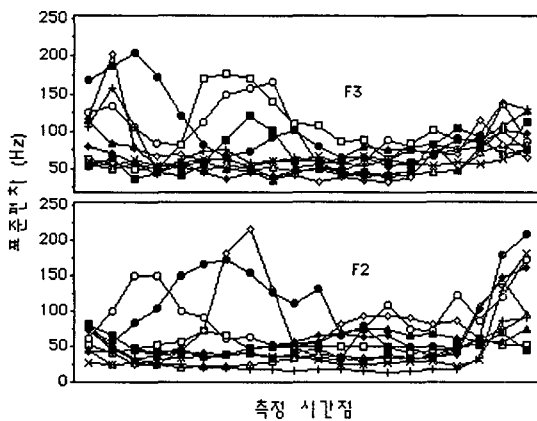


그림 5. s1이 열 번 발성한 각 숫자음의 F2와 F3의 표준편차값의 분포

그림 5에서도 전반부 시간점에서는 안정적이나 후반부에서는 불안한 숫자음이 많다. 가장 큰 변화를 보이는 값은 마른모꼴로 표시된 숫자음 '3'의 1/3 지점이고 '6'과 '0'도 이중모음의 특징상 초기부분에 많은 변화가 있었다. F3의 편차값은 F2와 거의 비슷한 패턴을 보였으며

전반적으로 편차값이 상승했다. 첫 두 개 시간점에서는 '0', '3', '6'에서 높은 편차를 보였고, 6~9 번까지 시간점에서는 '0'과 '1'의 변화율이 다소 높다. 하지만 후반부에서는 100 Hz 이하의 안정된 구간도 보인다.

지금까지 s1의 개인별 편차를 관찰하였는데 각 숫자음마다 다른 자모음 결합으로 되어 있기 때문에 측정 위치마다 차이를 보이고 있음을 확인할 수 있었다. 덧붙여, 비음으로 된 종성이나 이중모음에서 편차를 많이 보였으며, 음절의 시작부분과 끝부분은 인접하는 자음의 영향으로 측정값의 변화율이 매우 높음을 알 수 있었다.

다음으로는 첫 번째 연구문항에 대해 살펴보기로 하자. 각 화자별로 첫 번째 발음한 숫자음의 80 개 음향적 변수( $f_0$ ,  $F_1$ ,  $F_2$ ,  $F_3$  각각 20 개씩)를 모델로 하고 나머지 아홉 번 발성한 각 숫자음의 변수와의 절대 차이값(10 명 화자 $\times$ 10 개의 숫자음 $\times$ 9 번 비교=900)을 구한 뒤, 이들의 합계를 구하고, 각 숫자음별로 90 개의 합계값의 평균과 표준편차, 최대값, 최소값 등의 통계 분석값을 표 2로 나타내 보았다.

표 2. 숫자음 별로 살펴본 화자내의 음향적 변수의 통계 분석값

숫자음	평균	표준편차	최대값	최소값
4	2895	788	5022	1379
2	3304	1294	7914	1377
8	3556	969	5783	1675
1	3633	936	6021	1658
9	4039	1410	7639	2168
7	4202	1123	10343	2261
5	4468	1735	9999	1585
3	4765	1198	9648	2179
6	5065	1354	8923	2456
0	5673	1570	12460	2930

표 2를 보면 숫자음 '4'를 열 번 발음했을 때  $f_0$ ,  $F_1$ - $F_3$ 값의 변화가 가장 작음을 알 수 있다. 다시 말해서, 화자별로 이 숫자음을 매우 안정적으로 발음했다고 할 수 있다. 이어서, 숫자음 '2'가 따라오고, 숫자음 '0'에서 가장 많은 차이를 보였는데, 이것은 앞서  $F_2$ 와  $F_3$  측정값의 관찰에서 보았듯이 측정 에러 때문으로 여겨진다. 덧붙여, 이중모음이 들어간 숫자음이기 때문에 측정값이 단모음에 비해 시간에 따라 더 많이 변했을 것이다. 숫자음 '6'의 편차가 많고 숫자음 '3'이 그 다음으로 적은 편차가 이를 뒷받침해 준다. 따라서, 이중모음, 비음성의 순서로 편차가 많지며, 단모음일 때는 원순모음이 들어간 '5, 9'에서 편차가 다소 많은 편이다. 덧붙여, 10 명의 화자별로 이들 절대 차이값의 합을 구해 통계 분석값을 표 3으로 나타내었다.

표 3. 화자별로 살펴본 화자내의 음향적 변수의 통계 분석값

화자	평균	표준편차	최대값	최소값
s10	3,514	1,125	6,715	1,440
s7	3,618	1,106	6,648	1,658
s1	3,806	1,170	8,163	1,715
s2	4,062	1,626	8,923	1,377
s3	4,097	1,675	12,460	1,658
s9	4,193	1,428	9,608	2,135
s5	4,282	1,712	10,669	1,379
s8	4,386	1,497	9,999	1,956
s4	4,808	1,647	10,343	2,251
s6	4,834	1,301	8,488	2,321

표 3에서는 10 번 화자가 가장 적은 차이값의 합과 두 번째로 적은 표준편차를 보였다. 이는 10 번 화자의 발음이 매우 안정적으로 동일하게 발음하고 있음을 증명해 준다. 화자 확인에서는 가장 이상적인 화자로 취급될 것이다. s4와 s6이 많은 차이값의 합을 보였고, 이 중에서도 s4는 표준편차도 크고 최대값과 최소값의 범위의 차이도 크기 때문에 가장 나쁜 대상이라고 할 수 있다. 하지만, s10과 s4의 평균값의 차이는 1,290이고 이는 두 화자의 표준 편차값의 범위 내에 들어가기 때문에 큰 차이로 볼 수는 없다.

### 3.2 화자간 차이

화자마다 측정한 값들이 화자간의 차이를 가장 많이 보여주는 지점을 화자확인에 이용해야 하기 때문에 이번에는 각 화자가 발성한 열 번의 숫자음의 평균을 구하여 10 명을 한꺼번에 비교해 보고, 또한 두 번째 연구문항에 대한 답을 찾아보기로 한다. 먼저 숫자음 '0'의 평균값을 그래프로 나타내보면 그림 6과 같다. 이 그래프에서 살펴보면 피치값의 변화폭은 낮으나 F3으로 갈수록 변화율이 높아짐을 알 수 있다. 하지만, 서로 중첩되는 시간점이 많으며, 각각의 음향적 변수를 따로 이용하거나 한 지점의 값만을 구해서 화자 확인에 이용하려면, 앞서 화자 내 변이를 나타내는 표준편차값의 범위 내에 서로 겹쳐져 있으므로 어려움이 많을 것으로 예상된다.

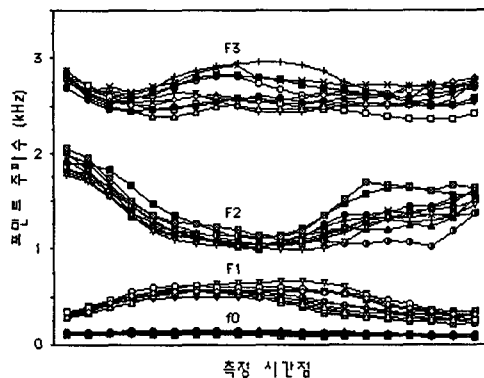


그림 6. 숫자음 '0'을 열 번 발음한 음향적 변수의 화자별 평균값



특히, 비음성을 나타내는 14 번째 시간점 이후에서 F2의 변화율이 높아서 개인별로 차이가 나타나고 있으나 이는 낮은 대역값에서 발생하는 측정상의 오류에 기인한 것으로 보아야 할 것이다. 왜냐하면, s1이 열 번 발성한 그림 2에서 보았듯이 이 부분의 편차도 높기 때문에 화자 확인의 도구로 사용하기에는 무리가 있을 것이다. 아울러, 이러한 변화 폭은 그림 1에서 살펴본 s1의 발성에 나타난 변화 범위와 비슷하다. 이것은 아마 개인 화자별로 동일하게 지각하는 포먼트 영역이기 때문으로 여겨진다. 다시 말해서, 이 범위의 값들은 누구나 동일한 숫자음으로 인식할 수 있는 범위가 된다. 다음으로는 단모음으로 된 숫자음 '4'를 발성한 열 명의 화자별 평균값을 그림 7로 나타내어 보았다.

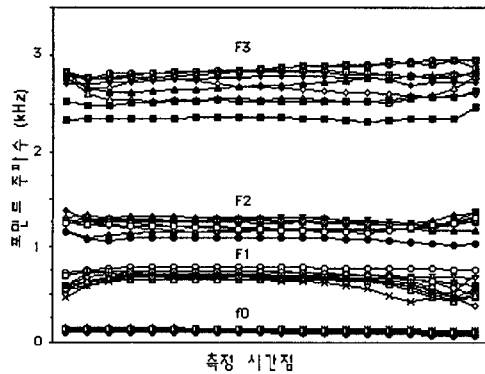


그림 7. 숫자음 '4'를 열 번 발음한 음향적 변수의 화자별 평균값

이 그림에서도 f0값은 사람마다 매우 비슷하며 F1에서 거의 비슷한 값을 보이고 F2에서 F3으로 갈수록 점점 더 화자간 차이가 나고 있다. 하지만, 이 변화율은 앞서의 개인별 변화율에 거의 근접하고 있다. 이 그래프는 평균값에 의해 개인별 발성의 특성이 둔화되었고, 4 개의 음향적 변수를 동시에 나타내야했기 때문에 개인별 차이를 쉽게 보기 어렵지만, 각각의 시간점에서 구한 4 개의 개별 변수값을 한꺼번에 고려해 보면 화자간의 차이를 극대화시킬 수 있을 것이다. 실제, 그림 6에서 보았던 거의 비슷한 f0값을 보인 s4, s6이 열 번씩 발성한 숫자음 '4'의 포먼트값을 나타내보면 그림 8과 같이 깨끗하게 분리된다.

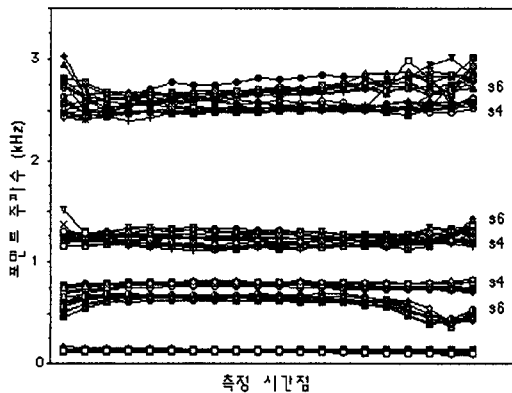


그림 8. s4, s6이 열 번씩 발성한 숫자음 '4'의 포먼트값 비교

이들 두 사람의  $f_0$ 는 매우 비슷하게 움직이고 있으나 F1에서는 s4와 s6이 서로 확연히 구분되는 영역에서 발음되고 있으며, F2는 중간부분에서 약간 구분이 되나 처음과 끝부분에서는 서로 엇갈리는 모양을 하고 있다. F3에서도 대체로 구분이 잘 되고 있으나, 처음부분에서는 서로 엇갈리는 모양을 하고 있다. 따라서, 화자 확인을 위해서는 단순히 한 지점의 포먼트만 비교해서는 안되고 동적으로 변하는 값들을 저장하였다가 이 모델을 중심으로 비교하고자 하는 화자의 음향적 변수와의 절대 차이의 합이 임의의 문턱값 이상이 되면 다른 화자로 판단해야 한다. 따라서, 이번에는 첫 번째 화자가 발성한 숫자음의 음향적 변수를 모델로 하고 나머지 9 명의 화자의 해당 시간점의 음향적 변수를 차례로 대비시켜서 그 절대 차이값(10명 화자×10 개의 숫자음×9 번 비교=900)을 구하고 이 값의 합을 각 숫자별로 구했다. 이어서 각 숫자별로 구해진 90 개의 값들의 평균과 표준편차, 최대값, 최소값 등 통계값을 구하여 표 4로 나타내 보았다.

표 4. 화자간 차이의 합의 통계적 특성

숫자음	절대 차이값의 합의 평균	표준편차	최대값	최소값
3	10,671	2,888	18,223	5,886
2	8,457	3,180	14,710	3,184
0	7,878	1,890	13,313	3,900
4	7,251	1,661	10,746	3,914
6	6,968	2,038	11,984	2,896
8	6,724	2,216	13,213	3,277
5	6,446	2,573	13,300	2,524
1	6,382	1,726	11,725	2,863
9	5,996	1,593	9,764	3,371
7	5,746	1,501	11,114	3,675

이 표에서 살펴보면 숫자음 3이 가장 큰 화자간 차이를 보여주고 있다. 아울러, 최대값과 최소값의 범위가 12,337 Hz이고 편차도 다소 높은 편이다. 이어서 숫자음 '0'도 화자간 차이가 세 번째로 크다. 표준편차값으로 보면 숫자음 '3'이 가장 크고, 다음으로 숫자음 '5'가 비슷한 값으로 나왔다. 숫자음 '7'이 가장 적은 평균과 표준편차값을 보여 화자간 차이가 가장 적다. 다음으로 숫자음 '9'가 거의 비슷한 값을 보이고 있으며, 최대값과 최소값의 범위도 6,393으로 가장 적다. 따라서, 화자간 차이가 가장 많이 나는 숫자음을 이용하여 화자확인을 한다면 비음이 섞인 숫자음 '3'과 '0'을 잘못 채택하는 오류를 범하기 쉽다. 앞서도 지적했지만, 화자확인에는 화자간의 차이도 많아야하지만, 그에 못지 않게 화자 내에서도 안정된 음성자료를 이용해야 한다.

### 3.3 숫자음별 차이비율

세 번째 연구문항은 화자 확인에 필요한 숫자음을 선택해야 한다면 어떤 기준을 이용할 수 있는가? 이다. 이 논문에서는 화자간 차이와 화자내의 차이 비율값(DR: Difference Rate)

을 제안한다. 이 비율은 다음과 같이 정의한다.

$$\text{차이비율(\%)} = \frac{[(\sum \text{화자간 절대 차이값}) - (\sum \text{화자내 절대 차이값})]{\sum \text{화자내 절대 차이값}} \times 100$$

차이비율은 화자간 차이값의 합이 클수록 높아지고, 화자내 절대값 차이값의 합이 클수록 낮아지게 된다. 따라서, 화자확인에서는 차이비율이 높은 숫자일수록 유리하다. 물론 개별 숫자음에 대한 표준편차값까지 고려해야하지만, 사전연구에서 편차값에 차이가 적었기 때문에 이 논문에서는 차이값의 평균만을 이용하기로 한다. 백분율로 나타낸 이유는 화자확인에서 상대적으로 현재값의 몇 %를 허용하는 범위로 정할지 결정하는데 도움이 된다.

표 5는 표 3과 표 4의 평균값을 이용해서 구한 차이비율이다.

표 5. 각 숫자음별로 살펴본 차이비율값

숫자음	차이비율(%)	숫자음	차이비율(%)
2	156	9	48
4	150	5	44
3	124	0	39
8	89	6	38
1	76	7	37

표 5에서 화자내 변이가 적고 화자간 차이가 큰 숫자를 선택한다면 '2, 4'의 순서가 된다. 숫자음 '2'는 화자간 차이가 화자내 차이보다 156%나 더 많다. 이때, 각 숫자음 별로 구체적인 표준편차값을 고려해본다면, '4'가 화자간, 화자내에서 매우 안정적이기 때문에 더 좋은 대상이 된다. 비음성이 포함된 숫자음 '3'도 매우 좋은 후보로 제시되었다, 이는 화자간 차이가 매우 많이 나타났기 때문이며, 화자내 차이는 '0, 6'과 함께 높기 때문에 주의해야 한다. 숫자음 '7'은 화자내 차이가 높은 편이고, 화자간 차이가 별로 없기 때문에 가장 나쁜 화자확인용 자극으로 판단된다. 이러한 숫자음의 비교는 시작 부분의 피치값의 측정변화 등을 고려해서 화자내 변화가 적고 화자간 차이가 많이 나는 시간점을 화자 개인별로 구축하여 모델로 설정한 뒤 이 모델을 매번 새로운 발음이 들어갈 때마다 재조정하는 방법을 이용하면 좋은 화자확인 장치를 개발할 수 있을 것으로 기대된다.

#### 4. 결론

이 논문에서는 10 명의 남학생들이 발성한 열 개의 숫자음의 음향적 변수를 구하여 화자내 변이와 화자간 변이를 통계적으로 분석하였다. 음향적 변수는 해당 시간점에 인접한 영역의 평균과 표준편차를 이용하여 걸렸으며, 프라트의 스크립트를 이용해 자동으로 측정하였다. 연구 결과로는 첫째 화자내 변이로서 비음성과 이중모음이 들어간 숫자음이 특정한 포먼트에

서 보다 많은 변화를 보였다. 변화량을 나타내는 표준편차값은 시작부분과 끝부분을 제외하고는 매우 안정적이었으며 동일한 숫자음으로 판단하는 지각 허용범위에 근접했다. 화자 내에서는 숫자음 '4'와 '2'가 적은 변화를 보였으며, '0'과 '6'은 편차가 많았다. 화자별로는 대체로 비슷한 변화 범위를 보였으며 숫자음별로 살펴본 통계값과 비슷한 분포를 보였다. 화자간 차이를 열 번 발음한 숫자음 '0'과 '4'의 평균값으로 비교했을 때는 f0에서 F3으로 갈수록 화자간의 차이를 많이 보여주었고, 일부 시간점에서는 일치하기도 했다. 따라서, 개별 변수값을 하나 하나 시간점 별로 비교하여 그 차이값의 합을 살펴본 결과 화자별 구분이 되기 때문에, 화자확인 은 동적으로 변하는 값들의 차이의 누적값을 이용할 수 있음을 밝혔다. 화자별로 비교한 결과에서는 숫자음 '3'이 가장 많은 차이를 보였고 '7'이 가장 적은 차이를 보였는데, 어떤 숫자음이 화자확인에 더 나은 자극음성인지 판단할 때 화자간 차이와 화자내의 차이를 동시에 고려한 차이비율을 제안했다. 이 차이비율로 비교해본 결과 숫자음 '2'가 가장 좋은 대상이고 숫자음 '7'이 가장 나쁜 대상으로 나타났다.

이 연구에서는 숫자음의 시작부분과 끝부분의 변화가 많은 부분을 걸러 줄 수 있는 장치를 제시하지 않았기 때문에 앞으로 개별화자마다 차이비율이 높은 시간점을 중심으로 모델을 설정했을 때는 얼마만큼의 개선이 가능한지 연구해 봐야 한다. 또한, 이러한 측정값이 다소 이상적인 발음환경에서 나왔기 때문에 앞으로 자연스러운 소음이 첨가된 환경에서의 처리과정도 연구해 볼 계획이다.

### 참 고 문 헌

- 양병곤. 2000. "Praat에 의한 숫자음의 음향적 분석법." *음성과학*, 7권 2호, 127-137.  
 오택섭. 1996. *사회과학 데이터 분석법*. 서울: 나남출판.  
 Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague. Netherlands: Mouton.  
 Fant, G. 1975. "Speech Production." *STL-QPSR*, 2-3, 1-19.  
 Traunmueller, H. 1988. "Paralinguistic variation and invariance in the characteristic frequencies of vowels." *Phonetica*, 45, 1-29.  
 Yang, B. 1996. "A comparative study of American English and Korean vowels produced by male and female speakers." *Journal of Phonetics*, 24, 245-261.  
 Yang, B. 2001. "Perceptual experiment on number production for speaker identification." *Speech Sciences*, Vol. 8 No. 1, 7-19.

접수일자: 2001. 7. 5.

게재결정: 2001. 8. 18.

#### ▲ 양병곤

부산광역시 진구 가야동 산 24 (우: 614-714)

동의대학교 영어영문학과

Tel: +82-51-890-1769, Fax: +82-51-890-1209

E-mail: bgyang@hyomin.donggeui.ac.kr