좁은대역 스펙트럼의 차이값과 상관계수에 의한 화자확인 연구*

A Study on Speaker Identification by Difference Sum and Correlation Coefficients of Narrow-band Spectrum

양 병 곤** · 강 선 미***
Byunggon Yang · SunMee Kang

ABSTRACT

We examined some problems in speaker identification procedures: transformation of acoustic parameters into auditory scales, invalid measurement values, and comparability of spectral energy values across the frequency range. To resolve those problems, we analyzed the acoustic spectral energy of three Korean numbers produced by ten female students from narrow-band spectrograms at 19 proportional time points of each voiced segment. Then, cells of the first five spectral matrices were averaged to form a matrix model for each speaker. The correlation coefficients and sum of the absolute amplitude difference in each pair of the spectral models of the ten subjects were obtained. Also, some individual matrix models were compared to those of the same subject or the other subject with a similar spectral model. Results showed that in numbers '2' and '9' subjects could not be clearly distinguished from the others but in number '4' it shed some possibility of setting threshold values for speaker identification if we employed the coefficients and the sum of absolute difference. Further studies would be desirable on various combinations of the range of long-term average spectra and the degree of signal pre-emphasis.

Keywords: Speaker Identification, Individual Variation, Narrow-band Spectrogram

1. 머리말

인터넷을 통한 사이버 학습공간이나 상거래에서 접속하고 있는 사람을 음성으로 확인하기 위해서는 등록한 이용자의 음성의 음향적 특징을 수집하여 모델로 저장해두고 나중에 접속할 때 발음한 음성의 특징과 이미 저장된 모델과의 일치여부를 확인해야 한다. 지금까지 화자확인을 위한 파라메터 측정 연구에서는 다른 화자와 구별되는 일정한 음향적 특징이 화자의 음성 속에 분명히 들어있다고 가정하고 이를 분석해 왔다(김태식 2001, 2002; 최홍섭 2000; 안

* 본 연구는 한국과학재단 목적기초연구(R01-1999-000-00229-0) 지원으로 수행되었음.
** 동의대학교 영어영문학과
*** 서경대학교 컴퓨터과학과

성주 외 2000; 양병곤 2001; Ko and Kang 2001; Ko 2002). 하지만, 지금까지 사용해온 여러 가지 파라메터들은 측정값이나 분석 방식의 한계점 때문에 화자확인에 모델을 만드는데 어려움이 있었다. 몇 가지 문제점들을 들어보면 다음과 같다. 첫째, 음향적 측정값을 활용하는데 문제가 있다. 음성의 음향적 특징은 푸리에 변환을 통해 각 시간점에서의 스펙트럼에서 성도의 변화 특성인 포먼트값을 구하고 또 자기상관법에 의해 성대의 특성인 피치값을 구하는 것이 주된 처리과정이었다. 화자확인과정에서 이렇게 구한 값을 mel, bark와 같은 청각척도로 변환하여 처리한다면, 화자확인에 필요한 화자 자체 내에서의 비언어적인 요소들마저 없애는 것이 된다. 오히려, 화자확인에서는 이런 비언어적인 요소의 극대화를 통해 화자고유의 조음동작의 특징을 부각시켜야할 것이다. 두 번째, 현재의 대다수 음성분석 소프트웨어는 타당한 음향적 측정값을 안정적으로 제시하지 못한다는 점이다. 조음기관은 매우 유연하고 연속된 움직임을 보이나 음향적 측정값은 발성기관의 자연스럽고 느린 변화라는 기본 가정에 어긋나는 급작스런 오류값들이 많이 나타난다. 오류가 많은 자료값에 근거를 둔 화자확인 모델은 실용성이 낮을 것이다. 특히, 여성의 발음에 대한 피치와 포먼트는 오류가 많다. 또한 푸리에 분석을 통한 스펙트럼을 근거로 한 모델을 만들 때도 분석 설정에 따라 값이 많이 달라진다. 예를 들어보면, 그림 1은 여성화자가 발음한 모음 '아'를 넓은대역과 좁은대역으로 스펙트로그램을 구한 뒤 피치값이 구해지는 총 지속시간을 20 등분하여 각각의 시간점에 대한 스펙트럼의 일부를 보여주고 있다(좁은대역과 넓은대역의 특성에 관해서는 구희산 외 1998:373을 참고하기 바람.).

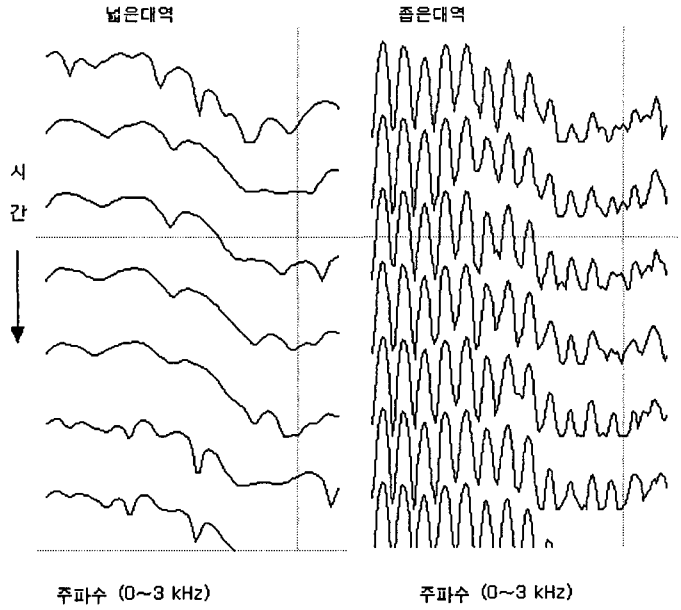

그림 1. 넓은대역(0.007 초 분석구간)과 좁은대역(0.029 초 분석구간) 스펙트로그램으로 분석한 숫자음 '4'의 연속 스펙트럼

각각의 스펙트럼에서 포먼트에 해당하는 지점은 대체로 일치하고 있으나 일부 시간점에서는

이웃하는 스펙트럼과 많은 차이를 나타내고 있다. 따라서, 이런 변화율이 높은 스펙트럼을 그대로 또는 후처리하여 모델로 설정한다면 화자개인에서도 변화율이 높고 화자간에는 서로 중복되는 부분이 많을 것이다. 반면, 동일한 발음을 25 ms의 좁은대역으로 분석한 스펙트럼의 변화를 살펴보면 다소 안정적이다. 고주파 영역으로 갈수록 주위 잡음의 크기가 커지면서 약간의 차이를 보이고 있으나 넓은대역에서 발견되는 급작스런 변화는 보이지 않는다. 따라서, 스펙트럼 정보를 화자확인모델로 사용하려면 좁은대역 분석방법을 사용하는 것이 바람직할 것이다. 셋째, 스펙트럼 정보를 이용하여 화자의 특징을 비교할 때 모든 주파수에서 서로 비교할 수 있는 크기가 되지 못하는 경향이 있다. 일반적으로 분석된 스펙트럼은 성문파형이 한 옥타브마다 12 dB씩 떨어지는 특성에 따라 저주파부분의 진폭값이 매우 높고 고주파로 갈수록 낮아진다(Kent and Read 1992:18). 하지만, 사람의 지각특성을 살펴보면 저주파보다는 고주파 영역에서 작은 진폭값이라도 귀는 민감하게 반응하기 때문에 입력된 음성을 지각의 특성에 맡게 고주파 영역의 진폭을 확대해 주는 것이 필요하다. 외이도의 길이가 보통 성인에게서 2.5 cm이기 때문에 한쪽이 막힌 관의 공명으로 1/4 파장이 가장 잘 울린다고 가정한다면 약 3,400 Hz 부근의 음의 진폭이 높게 되고 저주파음은 보다 높은 진폭값을 가져야 제대로 소리의 차이를 인식하게 될 것이다. 따라서, 개별 화자의 음성의 특징을 잘 살펴보기 위해서는 고주파대역강조(preemphasis)를 적용시켜 주는 것이 좋을 것이다. 미국보건성에서 조사한 동일한 강도수준을 나타내는 곡선분포(김기호 외 2000:47, 그림 3.17)를 살펴보면 저주파영역에서는 귀가 둔감하고 고주파영역에서는 민감한 특성을 반영하되, 실제 음성의 주파수 지역별 특성이 진폭값에 따라 달라지기 때문에 무조건 선형적으로 높여서도 안될 것이다. 또한, 첫 배음의 주파수는 고모음에서는 제1 포먼트에 가까운 값이 되므로 너무 증폭되지 않도록 할 필요가 있다. 따라서, 약 600 Hz 이후의 부분이 증폭되는 것이 좋을 것으로 생각된다. 그림 2는 모음 '아'의 발음의 1/3 지점의 스펙트럼을 고주파영역 증폭을 하지 않은 경우와 한 경우를 연속해서 나타내고 있다.

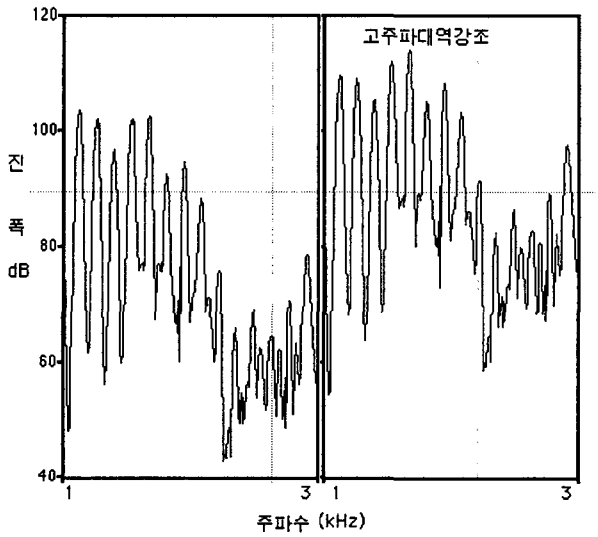

그림 2. 고주파대역강조를 한 경우(오른쪽)와 안한 경우(왼쪽)의 스펙트럼 비교

비록 고주파 부분을 증폭했지만 제2 포먼트의 배음 정점은 여전히 차이를 보이고 있다. 만약 스펙트럼 정보를 이용하여 모델을 설정할 때 비교할 두 음성의 저주파 영역의 진폭 차이값은 고주파 영역에 비해 매우 높은 비중을 차지하게 될 것이다. 한편, 모음 '이'일 경우에는 300 Hz 부근과 2,200 Hz 부근이 거의 같은 높이의 진폭을 보여서 상대적으로 대등한 진폭값의 차이를 보여준다. 따라서, 포먼트가 저주파 영역에 모여 있을 때와 고주파 영역에 서로 인접하여 배치되어 있을 때 달리 처리해야 하므로, 각 주파수영역별로 서로 비교될 수 있는 진폭값을 환산할 필요가 있다.

이러한 문제점들을 보완하려면 어떤 파라메터에 초점을 두고 어떻게 측정해야 하는가? 발성의 측면에서 본다면 피치나 포먼트는 화자간의 구별을 하기에는 너무 단순한 모델이고 또한 자연스런 조음기관의 움직임을 정확히 포착할 수 없는 경우가 많기 때문에 이용하기가 어렵다. 그보다는 더 많은 화자개인의 정보가 담겨있는 스펙트럼 정보가 차선책이라고 할 수 있다. 스펙트럼 정보는 보다 안정된 좁은대역을 이용하되 고주파영역을 확대하여 저주파영역과 상응할 수 있는 값을 구해야할 것이다. 이런 문제점들을 보완하여 화자확인 모델을 설정하고 서로 비교하기 위하여 이 연구에서는 10 명의 여성이 임의의 순서로 10 번씩 발음한 숫자음 가운데 '2, 4, 9'를 좁은대역 스펙트로그램으로 분석한다. 이어서 피치값이 구해지는 유성음구간의 전체 지속시간을 20 등분하여 각 시간점에서의 스펙트럼을 구하고 다섯 번의 발음에서 구한 스펙트럼의 평균값을 구하여 모델로 설정한 뒤 서로 다른 화자와 비교한다. 이어서, 개별화자의 값들의 상관계수와 스펙트럼의 절대 차이값의 합을 구하여 비교해 봄으로써 화자확인에 필요한 음성분석과정의 연구 방향을 잡는데 도움을 주고자 한다.

2. 연구방법

2.1 화자 녹음과정

화자는 동의대학교에 재학하는 건강하고 청각에 이상이 없는 여학생 10 명을 임의로 선정했다. 화자의 나이는 평균 19 세이고 키의 평균은 160 cm이었다. 표 1은 화자의 나이와 키를 나타내주고 있다. 출신 지역은 부산 7 명, 대구 1 명, 경남 2 명이었다. 화자들에게는 인터넷을 통해 자신의 목소리를 저장하고 확인하는데 사용한다는 실험의 목적을 간단히 설명해 주었고, 양병곤(2001)에서 사용한 방법과 같이 한 페이지에 임의의 순서로 뒤섞어 인쇄된 '0'에서 '9'까지의 숫자음을 조용한 연구실에서 프라트를 사용해 G3노트북 컴퓨터에 AKG 음성녹음용 마이크로 22 kHz의 표본속도로 직접 입력하도록 했다. 이어서, 연속된 음성을 각각의 숫자음 마디로 분리하여 저장하였다.

표 1. 화자 정보

화자	나이(세)	키(cm)	화자	나이(세)	키(cm)
s11	20	172	s16	19	163
s12	21	157	s17	19	162
s13	21	151	s18	20	172
s14	21	165	s19	22	162
s15	19	162	s20	21	160

2.2 자료분석

이 연구에서는 여성화자에게서 피치값과 포먼트값의 에러가 너무 많이 나타나기 때문에 보다 안정적인 좁은대역의 스펙트럼 정보를 이용하였다. 특히, 모든 숫자음을 다 분석하기보다는 앞서의 연구(양병곤, 2001)에서 밝혀진 대로 화자간에 차이가 많이 나는 반면 개인적으로는 안정된 숫자음 '2'와 '4'의 발음을 택하고, 비록 차이비율은 낮지만 모음삼각도에서 모퉁이에 해당하는 숫자음 '9'를 추가하여 세 개의 숫자음에 대해 집중적으로 조사해보고자 한다. 분석과정은 발성한 음성의 피치값을 프라트를 이용하여 자기상관방법으로 구하고 이들 값이 0 이상인 유성음 영역의 처음과 끝의 40 ms를 제외한 총 지속시간을 구한 뒤 이를 20 등분하여 처음과 끝을 제외한 각 시간점에 대한 좁은대역의 스펙트럼을 3,300 Hz까지 구한다. 피치값이 초성에 짧게 잘못 측정된 경우가 300 개의 발음가운데 모두 7 개 있었는데 이것은 음성파일에서 제거한 뒤 나머지 부분에서 스펙트럼을 구했다. 좁은대역의 스펙트로그램은 0.029 초의 분석구간으로 3,300 Hz까지 0.001 초마다 20 Hz의 간격으로 가우시안 창을 씌워 처리했다. 이어서 배음 사이의 차이를 줄이기 위해 각 지점의 스펙트럼자료에 대해 피치값에 가까운 임의의 간격인 150 Hz마다 인접 스펙트럼의 장기간평균스펙트럼값 (Long-term average spectrum)을 22 개씩 구하여 하드디스크에 파일로 저장했다. 수집된 음향적 변수로 총 125,400 개의 자료(10 명×10 번씩 발음×3 개의 숫자음×19 개의 시간점×22 개의 스펙트럼 진폭자료)를 구했다. 10 번의 발음 가운데 처음 다섯 번은 스펙트럼 에너지의 평균을 구하여 모델로 설정하고, 이것을 화자간의 비교에 사용했다. 비교 방식으로는 개별화자마다 작성된 스펙트럼 행렬의 차이값을 구하고 +/-부호를 없앤 절대값을 모두 합하고 동시에 피어슨적률 상관계수를 구하였다. 상관계수값은 통계처리과정(김호영, 2000)을 프라트 스크립트로 작성하여 구하였으며 통계소프트웨어인 StatView+에서 처리된 결과와 차이가 없었다. 마지막으로 상관계수가 높은 화자끼리 모델에 사용하지 않은 다섯 번의 발음에 대한 스펙트럼자료와 서로 비교하여 보았다. 이 모든 분석과정은 프라트 스크립트를 작성하여 자동으로 처리하였으며 각 값들이 정확히 처리되었는지 엑셀로 재확인하였다.

3. 분석 결과와 토론

3.1 화자별 유성음 지속시간 비교

모든 화자의 음성의 피치가 구해진 유성음 구간의 길이는 표 2와 같다.

표 2. 숫자음별 발음의 지속시간 평균과 표준편차 (단위는 ms임.)

숫자음	지속시간평균	표준편차
'2'	452	95
'4'	425	80
'9'	434	83

표 2를 보면 각 숫자음 지속시간의 평균이 400 ms 이상의 긴 발음으로 했으며 편차값도 100 ms 이내로 나타났다. 자연스런 속도의 모음 발음의 지속시간이 약 300 ms(Yang, 1996)인 것을 고려해 본다면, 화자들이 숫자음을 의도적으로 길고 분명히 발음하여 확인과정에 통과하려했다고 말할 수 있다. 개인별 발음의 평균과 표준편차를 살펴보면 표 3과 같다.

표 3. 화자별 숫자음 발음의 지속시간 평균과 표준편차 (단위는 ms임.)

숫자음 / 화자	'2'		'4'		'9'	
	지속시간평균	표준편차	지속시간평균	표준편차	지속시간평균	표준편차
s11	504	51	456	43	435	45
s12	409	57	357	25	359	42
s13	335	38	341	26	354	29
s14	456	75	461	57	470	66
s15	530	109	487	23	476	45
s16	476	41	435	58	456	58
s17	457	53	451	43	473	29
s18	386	39	379	37	411	52
s19	376	47	336	19	337	32
s20	594	71	549	90	572	77

표 3에서 보면 화자마다 모음의 지속시간이 숫자음에 따라 차이가 있다. s13과 s19 화자는 다소 짧은 길이로 발음한 반면 s15와 s20은 약간 길게 발음하였음을 알 수 있다. 화자 내에서도 변화율이 높은 경우에는 109 ms나 되고 적은 경우에는 19 ms에 이르기까지 변화가 많다. 따라서, 분석구간을 절대값으로 처리하여 화자내의 모델로 설정한다면 본인의 모델도 시간에 따라 변화하는 특징을 찾기가 어려울 것이다. 이전 연구에서는 숫자음마다의 상관도가 매우 높았다 (양병곤, 2000). 이 연구에서 전체지속기간을 20 등분하여 각각의 시간점의 값을 구하는 것도 이렇게 길게 발음했을 경우와 짧게 발음했을 때 발음 동작 속도가 상대적으로 변하는 점을 반영하기 위해서다.

3.2 s11과 s12의 숫자음 '2'와 '4'의 모델

그림 3은 s11과 s12가 다섯 번 발음한 숫자음 '2'의 스펙트럼을 150 Hz 간격으로 장기간 평균스펙트럼을 구한 뒤 각각의 지점의 평균값을 구하여 나타냈다. 그림 4는 동일한 화자들이 발음한 숫자음 '4'의 스펙트럼 모델이다. 그림에서 음영이 같으면 동일한 범위의 진폭값을

나타낸다. 이 두 그림에서 살펴보면 s11의 숫자음 '2'에서는 모음 '이' 발성의 중앙부분에서 측정한 제1 포먼트 값이 약 370 Hz였는데, 이 부분의 진폭이 매우 높고, 제2 포먼트는 2,300 Hz, 제3 포먼트의 위치가 3,200 Hz로서 최대값에 가깝게 위치하고 있다. s12의 숫자음 '2'도 제1 포먼트는 390 Hz, 제2 포먼트 2,900 Hz, 제3 포먼트는 3,600 Hz가 넘어 그림에서는 제2 포먼트까지만 나타나 있다. 따라서, 두 사람의 스펙트럼은 매우 비슷하게 보이지만 제2 포먼트와 제3 포먼트에서 차이가 난다. 한편 그림 4에서 s11의 숫자음 '4'에서는 모음 '아'의 제1 포먼트값이 약 950 Hz이었는데, 이 부분의 진폭이 매우 높고 제2 포먼트는 1,480 Hz, 제3 포먼트의 위치는 스펙트로그램에서는 쉽게 측정되지 않을 정도로 매우 낮은 진폭값을 가지고 있었는데 약 2,800 Hz에 위치해 있었다. s11의 제1 포먼트는 1,000 Hz이고 제2 포먼트는 1,700 Hz이고, 제3 포먼트는 2,500 Hz 부근에 스펙트럼에너지정점이 나타났다.

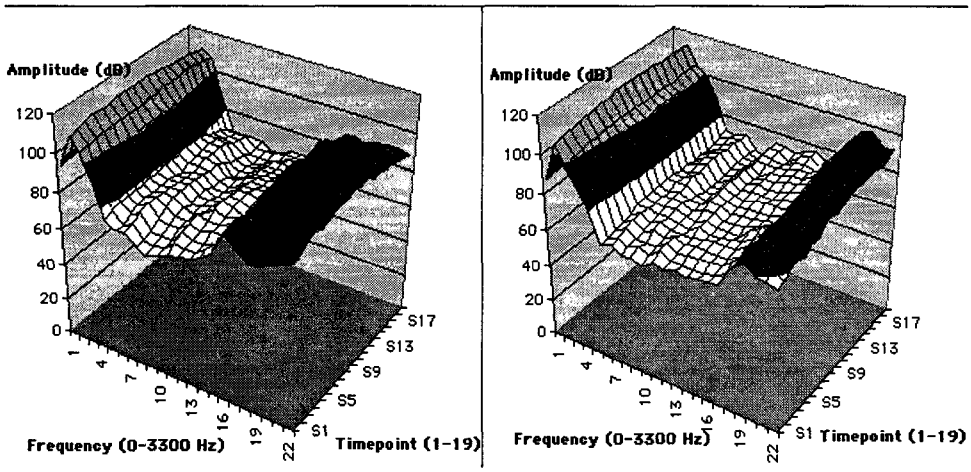

그림 3. s11(왼쪽)과 s12(오른쪽)의 숫자음 '2'의 모델. 20 개 측정시간점에서의 스펙트럼을 연속으로 겹쳐서 보여주고 있음. 주파수는 150 Hz 간격으로 평균스펙트럼을 구하였음.

두 개의 그림에서 나타나 있듯이 다섯 번의 발음의 평균을 내었긴 하지만, 여전히 발성의 시작부분과 끝부분에 점진적인 스펙트럼의 변화를 보여주고 있음을 알 수 있다. 이것은 개인마다의 발음의 특성이라고 할 수 있다. 따라서, 단순히 한 지점의 스펙트럼 정보를 이용하여 화자를 비교하는 것은 개인적으로도 안정적인 값이 되지 않을 것이고 상대방과의 비교에서도 차이가 나지 않기 때문에 문제가 있을 수 있다.

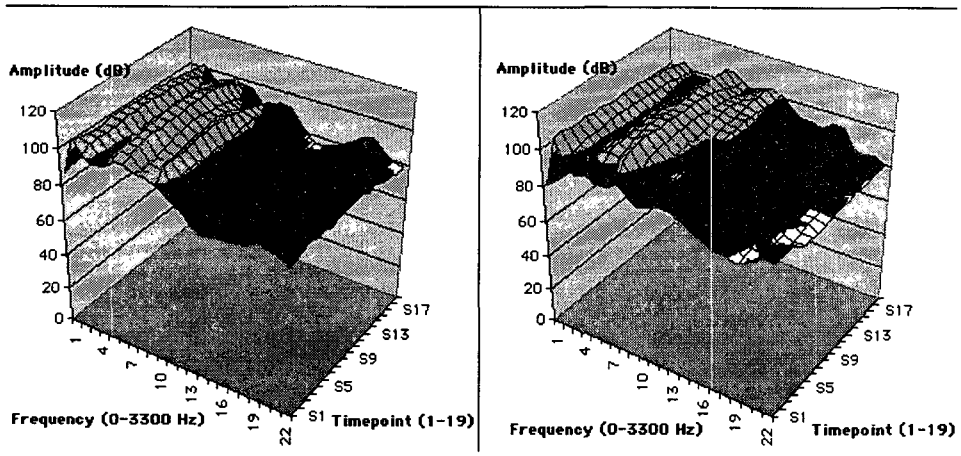

그림 4. s11(왼쪽)과 s12(오른쪽)의 숫자음 '4'의 모델. 20 개 측정시간점에서의 스펙트럼을 연속으로 겹쳐서 보여주고 있음. 주파수는 150 Hz 간격으로 평균스펙트럼을 구하였음.

그림 3과 4에서 본다면 측정시간점을 더 촘촘히 하면 세밀한 영상을 저장할 수 있고, 또한 주파수 영역을 150 Hz로 넓게 잡았는데 이를 더 좁은 간격으로 한다면 실제 스펙트럼 정보와 더 가까워질 수 있을 것이다. 하지만, 그만큼 정밀한 자료는 화자내의 변이를 더 많이 허용하기 때문에 안정된 모델을 설정하는데 어려움이 있을 것이다. 반면 낮은 해상도의 정보는 누구나 통과할 수 있는 모델로 될 가능성이 있다. 즉, 해상도와 화자마다의 특징 모델 설정의 효과 사이에는 일정한 역학관계가 있다. 앞으로 이 부분에 대해서는 장기간평균스펙트럼값의 범위의 변화와 주파수 간격의 변수를 조정하여 어떤 파라메터값의 조합이 화자확인의 모델 설정에 가장 적합한지 더 많은 비교 연구가 필요하다.

3.3 화자별 모델값의 절대 차이값 비교

표 4에서 9까지는 각 화자별로 설정된 모델값의 절대 차이값과 상관계수값을 나타내고 있다. 수집된 자료는 각 숫자별로 22 개 시간점에 19 개의 스펙트럼정보가 들어가기 때문에 모두 418 개의 행렬로 되어 있었는데, 사전 연구에서 각각의 값들의 절대 차이값의 합이 상대적으로 강하게 발음할 경우와 약하게 발음할 경우가 다르게 나타났기 때문에 상관계수값도 함께 구하였다. 상관계수가 높을수록 스펙트럼 모양이 같기 때문에 동일한 화자가 될 확률도 높을 것으로 기대된다.

먼저, 숫자음 '2'에 대한 상관계수의 평균은 0.91이었고, 표준편차는 0.03이었다. 서로 비교한 값들 가운데 최대 상관계수값은 s18과 s20의 발성모델에 대한 계수로 0.97이었고 최소 상관계수값은 s12와 s13 사이의 계수인 0.80이었다. 절대 차이값의 합은 평균 2,007 dB이었고, 표준편차는 439 dB로 화자마다 차이가 많다고 할 수 있다. 최대 절대 차이값의 합은 s12와 s16 사이의 3,007 dB이고 최소 절대 차이값의 합은 s14와 s17 사이의 1,271 dB였다. 이 두 화자의 상관계수도 0.95이기 때문에 아주 유사한 음성으로 처리될 것이다. 숫자음 '4'에 대한 상관계수의 평균은 0.72이었고, 표준편차는 0.1이었다. 최대값은 s12와 s16 사이의 발성모델에

대한 계수로 0.91이었고 최소값은 s13과 s19 사이의 계수인 0.55이었다. 절대값의 차이의 합은 평균 2,198 dB이었고, 표준편차는 415 dB로 화자마다 차이가 많다고 할 수 있다. 최대 절대 차이값의 합은 s13과 s19 사이의 3,145 dB인데 이들의 상관계수도 최소값으로 나타났다. 최소 절대 차이값의 합은 s12와 s16 사이의 1,230 dB이었다. 이 두 화자의 상관계수도 최대값이기 때문에 유사한 음성으로 처리될 것이다. 숫자음 '4'가 보다 낮은 상관계수와 높은 차이값을 보인 것은 아마 제1, 2 포먼트 위치가 1,000 Hz 전후에 위치해 있어 상대적으로 화자마다 차이가 많이 났기 때문으로 추정된다. 즉, 숫자음 '2'나 '9'의 경우에는 저주파에 스펙트럼 에너지가 높기 때문에 고주파에서의 화자별 차이가 크게 반영되지 못하는 것으로 여겨진다. 숫자음 '9'에 대한 상관계수의 평균은 0.94이었고, 표준편차는 0.02이었다. 최대값은 s12와 s17 사이의 발성모델에 대한 계수로 0.97이었고 이들의 절대 차이값의 합도 최소값인 1,112 dB로 아주 동일한 발음으로 처리될 확률이 높다. 최소값은 s12와 s16 사이의 계수인 0.897이었다. 절대 차이값의 합은 평균 2,021 dB이었고, 표준편차는 536 dB로 화자마다 차이가 많다고 할 수 있다. 최대 절대 차이값의 합은 s16과 s18 사이의 3,446 dB인데 이들의 상관계수는 0.93으로 높은 관계를 나타내준다.

지금까지 살펴본 각 숫자음에 대한 상관계수와 절대 차이값의 합을 보면 숫자음 '2'와 '9'에서는 상관계수가 높고 동일한 음성으로 처리될 쌍도 많이 보였다. 반면 숫자음 '4'에서는 상관계수도 낮고 절대 차이값의 합도 높아서 화자확인에 이용될 수 있는 가능성을 보여준다. 앞서의 연구(양병곤, 2001)에서도 숫자음 '4'는 화자 내에서 가장 편차가 적었고 화자 사이에서도 서로 구별될 만큼 좋은 대상이었다. 따라서, 단순한 음 하나만으로는 서로 비교되는 값이 되기에는 부족하다면, 세 개의 값의 차이를 동시에 이용하는 방안을 더 연구해 보아야 할 것이다.

표 4. 각 화자가 발음한 숫자음 '2'의 모델에 대한 화자간 상관계수

화자	s12	s13	s14	s15	s16	s17	s18	s19	s20
s11	0.84	0.80	0.91	0.93	0.92	0.87	0.92	0.86	0.89
s12		0.89	0.90	0.87	0.89	0.92	0.90	0.84	0.93
s13			0.93	0.87	0.87	0.90	0.89	0.86	0.90
s14				0.94	0.93	0.95	0.95	0.91	0.94
s15					0.95	0.93	0.94	0.94	0.94
s16						0.92	0.95	0.88	0.92
s17							0.94	0.92	0.97
s18								0.90	0.93
s19									0.95

표 5. 각 화자가 발음한 숫자음 '2'의 모델에 대한 절대 차이값의 합

화자	s12	s13	s14	s15	s16	s17	s18	s19	s20
s11	2,750	2,702	1,805	1,448	1,757	2,080	1,817	2,408	2,447
s12		1,970	2,275	2,530	3,079	2,034	2,375	2,690	1,551
s13			1,923	2,092	2,553	1,970	2,207	2,390	1,881
s14				1,381	1,796	1,271	1,326	1,909	1,790
s15					1,523	1,594	1,549	1,750	1,979
s16						1,917	1,642	2,591	2,774
s17							1,645	1,870	1,555
s18								2,091	2,040
s19									1,567

표 6. 각 화자가 발음한 숫자음 '4'의 모델에 대한 화자간 상관계수

화자	s12	s13	s14	s15	s16	s17	s18	s19	s20
s11	0.78	0.62	0.73	0.69	0.83	0.70	0.77	0.88	0.79
s12		0.89	0.73	0.61	0.91	0.88	0.66	0.70	0.69
s13			0.66	0.59	0.79	0.87	0.60	0.55	0.68
s14				0.67	0.66	0.70	0.66	0.73	0.75
s15					0.61	0.57	0.58	0.66	0.74
s16						0.80	0.67	0.72	0.63
s17							0.67	0.65	0.70
s18								0.89	0.87
s19									0.85

표 7. 각 화자가 발음한 숫자음 '4'의 모델에 대한 절대 차이값의 합

화자	s12	s13	s14	s15	s16	s17	s18	s19	s20
s11	2,140	2,917	2,237	2,417	1,758	2,301	2,049	1,558	1,924
s12		1,931	2,123	2,635	1,230	1,611	2,389	2,502	2,190
s13			2,416	2,714	2,289	1,484	2,888	3,145	2,747
s14				1,920	2,179	2,044	2,132	2,191	2,091
s15					2,369	2,502	2,610	2,419	2,195
s16						1,869	2,215	2,302	2,398
s17							2,428	2,483	2,353
s18								1,493	1,461
s19									1,668

표 8. 각 화자가 발음한 숫자음 '9'의 모델에 대한 화자간 상관계수

화자	s12	s13	s14	s15	s16	s17	s18	s19	s20
s11	0.91	0.95	0.93	0.95	0.94	0.93	0.95	0.96	0.96
s12		0.91	0.93	0.90	0.90	0.97	0.90	0.90	0.94
s13			0.95	0.92	0.92	0.91	0.92	0.96	0.94
s14				0.93	0.92	0.94	0.94	0.96	0.95
s15					0.96	0.93	0.95	0.96	0.95
s16						0.91	0.92	0.93	0.94
s17							0.93	0.91	0.95
s18								0.96	0.96
s19									0.96

표 9. 각 화자가 발음한 숫자음 '9'의 모델에 대한 절대 차이값의 합

화자	s12	s13	s14	s15	s16	s17	s18	s19	s20
s11	2,406	1,866	1,683	1,915	2,985	2,196	1,493	1,376	1,522
s12		1,858	2,303	1,961	2,095	1,112	2,617	2,654	1,636
s13			1,727	1,839	2,293	1,816	2,267	1,768	1,466
s14				2,164	3,286	2,209	1,586	1,315	1,641
s15					1,728	1,773	2,318	2,129	1,529
s16						2,185	3,446	3,343	2,241
s17							2,351	2,452	1,534
s18								1,409	1,759
s19									1,708

3.4 화자의 모델과 자신의 발음 스펙트럼과의 비교 및 타자의 발음 스펙트럼과의 비교

이번에는 화자 자신이 다섯 번 발음한 숫자음의 평균스펙트럼 모델에 대해 나머지 다섯 번의 스펙트럼 값들과 어떤 관계가 있는지 알아보기로 한다. 개별 화자 내에서의 비교와 다른 화자와의 비교는 수많은 조합이 가능하므로 여기서는 앞서 상관계수가 높은 경우만 살펴보기로 한다. 먼저 숫자음 '2'에 대해서는 상관계수가 가장 높은 s18과 s20, 절대 차이값의 합이 가장 작은 s14와 s17을 비교해 보고, 숫자음 '4'에 대해서는 상관계수가 가장 높고 절대 차이값의 합이 가장 작은 s12와 s16을 비교해 보기로 한다. 이들을 선택한 이유는 이들 값에서 서로 구별된다면 다른 값에서는 쉽게 화자확인이 가능하기 때문이다.

먼저 s18이 다섯 번 발음한 숫자음 '2'에 대한 스펙트럼의 평균으로 작성된 모델과 나머지 다섯 번의 숫자음의 상관계수와 절대 차이값의 합을 구해보면 표 10, 11과 같다. 이 표에서 살펴보면 숫자음 '2'에서는 s18의 모델과 자신의 숫자음이 매우 높은 상관계수를 보여주고 있으나 여덟 번째 발음한 숫자음 '2'는 0.76으로 낮게 나타난다. 즉 자신의 발음이 항상 안정적이지 않기 때문에 이것은 통과되기 힘들 것이다. 반면 여섯 번째 발음한 '2'는 상관계수는 높지만 절대 차이값의 합도 동시에 높아서 통과시킬지 문제가 된다. 즉 s20과의 비교에서 열 번

째 발음의 상관계수가 0.92이고 절대 차이값의 합이 3,625 dB이기 때문에 상관계수와 절대 차이값의 합을 동시에 만족시킬 때 통과시키는 방안도 제시될 수 있을 것이다. 숫자음 '4'일 때 s18의 모델값과 동일한 화자의 상관계수값을 보면 모두 0.87 이상을 나타내고 있기 때문에 s20과의 비교에서는 모두 상관계수가 0.84 이하로 나타나 있다. 따라서, 숫자음 '2'에서는 두 화자가 구별되지 않지만, 숫자음 '4'에서는 구별이 된다. 표 11의 첫 번째 자료는 s12가 발성한 숫자음 '4'의 모델과 s16의 발음을 비교하고 있는데, 마지막 발음은 상관도가 0.87에 이르고 절대 차이값의 합도 1552 dB나 되어 동일화자로 처리될 가능성이 있다. 만약 상관계수를 0.88로 정하고 절대 차이값도 1550 dB으로 문턱값을 정해둔다면 두 번째 자료에 나타난 s12가 발성한 숫자음 '4'의 모델과 본인의 비교에서 9, 10번째 발음은 자신도 통과되기 어려울 것이다. 따라서, s16이 통과되지 못하도록 하려면 문턱값을 상관계수가 0.9 이상이고 절대 차이값의 합도 1500 dB 이하로 지정해야 할 것이다. 이것은 세 번째와 네 번째에 나타나있는 s18과 s20의 비교에서도 볼 수 있다. 즉, s18이 자신은 통과되고 s20은 통과되지 못하게 하려면 상관계수는 적어도 0.9 이상이 되어야 하고 절대 차이값의 합도 1390 dB이하가 되도록 설정하면 될 것이다. 머릿말에서 지적했듯이, 절대 차이값의 합은 서로 비교될 수 있는 스펙트럼 정보에서 좋은 잣대가 될 수 있으나, 저주파에서 차이가 많이 난다면 쉽게 이 값이 높아지게 되어있으므로 상관계수를 동시에 이용하는 것이 바람직할 것으로 여겨진다. 덧붙여, 안성주 외(2000)에서는 순차결정법에 의해 처음 발음에서 문턱값을 조정한 뒤 이를 다음 단어 발음에 적응시키는 방안을 제시했다. 즉 단순히 한 음절보다는 두 개 이상의 음절 또는 단어로 더 많은 스펙트럼 정보를 이용하면 더 정확히 판단할 수 있을 것이다.

표 10. 숫자음 '2'에 대한 상호 비교. s18m:s18(6)은 s18의 모델과 s18이 여섯 번째 발음한 숫자음 '2'의 스펙트럼자료를 비교한 결과를 나타낸다.

비교대상	s18m:s18(6)	s18m:s18(7)	s18m:s18(8)	s18m:s18(9)	s18m:s18(10)
상관계수	0.95	0.91	0.76	0.96	0.94
절대차이값의합	3,931	2,661	2,961	1,527	1,510
비교대상	s18m:s20(6)	s18m:s20(7)	s18m:s20(8)	s18m:s20(9)	s18m:s20(10)
상관계수	0.90	0.92	0.90	0.82	0.92
절대차이값의합	2,575	1,813	2,018	3,438	3,625
비교대상	s14m:s17(6)	s14m:s17(7)	s14m:s17(8)	s14m:s17(9)	s14m:s17(10)
상관계수	0.87	0.87	0.92	0.85	0.95
절대차이값의합	2,098	3,337	1,754	2,399	1,299

표 11. 숫자음 '4'에 대한 상호 비교.

비교대상	s12m:s16(6)	s12m:s16(7)	s12m:s16(8)	s12m:s16(9)	s12m:s16(10)
상관계수	0.83	0.85	0.81	0.84	0.87
절대차이값의합	1,771	1,775	1,837	1,917	1,552
비교대상	s12m:s12(6)	s12m:s12(7)	s12m:s12(8)	s12m:s12(9)	s12m:s12(10)
상관계수	0.90	0.92	0.90	0.85	0.81
절대차이값의합	1,383	1,228	1,459	1,791	2,106
비교대상	s18m:s20(6)	s18m:s20(7)	s18m:s20(8)	s18m:s20(9)	s18m:s20(10)
상관계수	0.76	0.79	0.83	0.84	0.84
절대차이값의합	2,066	1,950	1,735	1,779	1,678
비교대상	s18m:s18(6)	s18m:s18(7)	s18m:s18(8)	s18m:s18(9)	s18m:s18(10)
상관계수	0.91	0.87	0.90	0.92	0.91
절대차이값의합	1,322	1,703	1,203	1,386	1,355

4. 맺음말

이 논문에서는 화자확인처리과정에서 음향적 척도의 변환과정과, 측정값의 유효성, 모든 주파수에서의 스펙트럼자료의 비교가능성과 같은 문제점들을 살펴보았고 이를 보완하기 위한 시도로 10 명의 여성화자의 발성을 좁은대역으로 분석하여 시간마다 변하는 동적인 스펙트럼 자료를 구하여 모델로 설정하고 각 화자의 모델끼리 스펙트럼값의 절대 차이값의 합과 상관계수를 구하여 비교해 보았다. 이어서, 상관계수가 높은 화자끼리 개별 발음과의 비교도 해보았다. 그 결과 숫자음 '2'와 '9'에서는 화자끼리 서로 동일하게 확인될 수 있는 여지가 많았지만, 숫자음 '4'에서는 상관계수와 절대 차이값의 합을 화자확인에 사용할 문턱값으로 정할 만한 가능성도 보였다. 특히, 스펙트럼 절대 차이값의 합과 함께 상관계수값을 동시에 이용함으로써 보다 안정된 구별 문턱값으로 이용할 수 있음을 살펴보았다. 앞으로의 연구과제로는 이들 값들이 제한된 수의 화자에게서 나타난 결과이기 때문에 비슷한 발성구조를 가진 더 많은 화자에 적용하여 검증할 필요가 있다. 또한 두 개 이상의 음절을 가지는 발음의 상관계수와 절대 차이값의 합을 찾아보고 또한 분석스펙트럼의 장기간스펙트럼평균값의 범위 조절과 고주파영역확대에 대한 조합을 다양하게 결합해 보아야 할 것이다.

참 고 문 헌

구희산, 고도흥, 양병곤, 김기호, 안상철. 1998. *음성학과 음운론*. 서울: 한신문화사.
김기호, 양병곤, 고도흥, 구희산. 2000. *음성과학*. 서울: 한국문화사.

김태식. 2001. "벡터 평균값을 갖는 스트레인지 어트랙터 기반 화자인식." *음성과학*, 8권 3호, 133-142.
김호영, 민수홍, 설인자, 유성렬. 2000. *알기 쉬운 통계*. 서울: 학지사.
김태식. 2002. "음성 특징 추출을 위한 스트레인지 어트랙터의 분석 방법." *음성과학*, 9권 2호, 147-155.
안성주, 강선미, 고한석. 2000. "가변문턱치와 순차결정법을 통한 문맥요구형 화자확인." *음성과학*, 7권 4호, 41-47.
양병곤. 2000. "Praat에 의한 숫자음의 음향적 분석법." *음성과학*, 7권 2호, 127-137.
양병곤. 2001. "남성의 숫자음 발성에 나타난 화자변이." *음성과학*, 8권 3호, 93-104.
최홍섭. 2002. "화자식별을 위한 파라미터의 잡음환경에서의 성능비교." *음성과학*, 7권 3호, 185-195.
최홍섭. 2000. "화자인증 시스템에서 배경화자 선정 방법에 관한 연구." *음성과학*, 9권 2호, 135-146.
Kent, R. D. & C. Read. 1992. *The Acoustic Analysis of Speech*. San Diego, CA: Singular Pub. Group.
Ko, D. 2001. "Voice Similarities between Sisters." *Speech Sciences*. Vol. 8. No. 3, 43-50.
Ko, D. & S. Kang. 2002. "Voice Similarities between Brothers." *Speech Sciences*. Vol. 9. No. 2, 1-11.
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: 2002. 7. 10.
게재결정: 2002. 8. 27.

▲ 양병곤
부산광역시 부산진구 가야동 산 24 (우: 614-714)
동의대학교 영어영문학과
Tel: +82-51-890-1227
E-mail: bgyang@dongeui.ac.kr
Website: http://www.dongeui.ac.kr/~bgyang

▲ 강선미
서울특별시 성북구 정릉동 16-1 (우: 136-704)
서경대학교 컴퓨터과학과
Tel: +82-2940-7291
E-mail: smkang@skuniv.ac.kr
Website: http://ihci.skuniv.ac.kr/