

Signalalyze에 의한 모음 포먼트 분석 방법 고찰

양 병 곤*

목 차

- | | |
|---------------------------|------------------|
| I. 서론 | V. 음성의 스펙트로그램 분석 |
| II. 성도면적과 포먼트의 관계 | VI. 결론 |
| III. Signalalyze의 포먼트추정방식 | 참고 문헌 |
| IV. 복합음의 추정값 비교 | Abstract |

I. 서론

사람의 음성은 성대에서 발생하는 음원을 성도의 공간인 구강 인강 비강 등을 변화시켜 원하는 모음을 생성한다. 모음의 음향적 특성은 이들 공간에 대한 공명특성인 포먼트로 나타나게 된다. 모음의 포먼트값을 측정하기 위한 여러가지 소프트웨어가 개발되어 있지만 매우 다양한 메뉴로 분류된 기능과 방법들의 유효성에 대한 연구가 별로 없다. 예를 들어, 선형예측계수(Linear Predictive Coding: LPC) 방법에 의한 측정에서는 남녀노소 화자에 따라 분석차수를 달리해야만 정확한 값을 구할 수 있음에도 이를 적절히 지정하지 못해 틀린 값을 그대로 수용하는 잘못을 범하는 경우가 많다. 따라서, 본 논문에서는 포먼트값과 성도의 면적변화에 대한 상관관계에 대해 간단히 살펴보고 이를 Macintosh용 Signalalyze 3.12를 이용해 정확히 측정하는 방식을 찾기위해 정현파합성기를 통해 만든 500Hz, 1500Hz, 2500Hz의 세 개의 정현파를 합성하여 만든 복합음을 여러가지 스펙트럼분석방식을 통해 점검해보고 이들 가운데 문제점과 유의할 점에 대해 고찰해보고자 한다. 아울러, 실제 두 화자가 발성한 모음 /이, 아, 우/의 분석 결과도 비교해 보기로 한다. 이 연구결과는 앞으로 모음 포먼트 연구의 기초적인 지식을 제공하여 다른 유사 음성분석프로그램에서도 쉽게 응용하여 기기활용의 극대화를 도모할 수 있으리라 여겨진다.

II. 성도면적과 포먼트의 관계

Fant(1960)의 음원-여과기 이론에 따르면 사람의 음성은 음원인 성대의 떨림을 여과기인 성도에 의해

* 동의대학교 인문대학 영어영문학과 부교수

걸러서 원하는 목소리를 낸다고 한다. 후두는 성대의 진동으로 모든 유성음의 음원이 된다. 공명체인 성도는 다른 공명체들처럼 입력된 에너지에 선택적으로 반응하는 특징이 있다. 그것은 마치 체가 일정한 크기의 알맹이만을 통과시키는 것처럼 일정한 주파수의 에너지를 잘 통과시키는 성질을 말하며 이런 의미에서 성도를 여파기라고 한다. 즉, 혀의 위치, 턱을 벌리는 정도, 입술 모양 등에 의해 성도 모양이 달라질 때마다 다른 공명체가 되어, 일부 주파수에서는 소리 에너지의 전달을 억제하고 다른 주파수에서는 최대의 에너지를 허용한다. 그래서, 기류의 방해없이 성도를 통과하는 모음의 경우, 허파로부터 성대를 통과하기까지는 동일한 음이지만, 다른 모양의 성도를 지나면서, 즉, 다른 모양의 여파기를 거치면서, 에너지의 분포가 변해 다른 모음으로 들리게 된다. 이 때 각 모음마다 성도 공명에 의해 선택된 에너지의 정점을 포먼트라고 하고 낮은 주파수부터 제1, 제2, 제3 포먼트 등으로 부른다.

포먼트 주파수는 화자의 성대 전체길이에 반비례한다. 성대 길이는 나이와 성별에 따라 다르다. 여성은 보통 남성보다 짧은 성도를 갖고 있다. 어린애는 여성보다 훨씬 짧은 성도를 갖고 있다. Chiba와 Kajiyama(1941)는 그 비율을 남자를 1로 했을 때, 여성은 0.87이고, 아홉살 소년은 0.80 여덟살 소년은 0.70의 비율치를 가진다. 따라서, 모음음소를 비교적 동일한 성도 모양으로 발음하더라도 포먼트 주파수는 남성에서, 여성, 아이로 갈수록 높아진다. 또한 인강과 구강의 길이 비율도 화자간 변이의 요소가 된다. Chiba와 Kajiyama (1941)는 여덟살 소녀의 구강길이는 성인남자보다 30% 더 짧으며 소녀의 인강길이는 성인남자보다 56% 더 짧다고 했다. 이런 차이에서 Fant는 남녀 화자의 포먼트값은 비균일 척도의 관계임을 예측했다. 남녀사이와 같은 성별과 나이 또는 감정적인 변화에 따라 변하는 모음형성음을 정규화시켜 보려는 여러가지 수학적인 형성음의 정규화 알고리즘이 제안되어 왔다. 실제 생활에서 음성 신호는 화자내에서나 화자사이에서 상당히 다양하게 나타난다. 예를 들면, 화자는 결코 물리적으로 똑같은 발음을 두 번이상하지 않는다. 또한 서로 다른 두 화자도 조음적으로나 음향적으로나 똑같이 발음하지는 않는다. 이런 화자 변이는 주로, 방언적, 사회 언어학적 차이와 같은 언어적 요소와, 신체구조, 나이, 성별, 발화자의 정서 상태 등과 같은 비언어적 요소로 구분되었다. Lindblom (1990)은 화자가 청자를 의식해서 발성을 하기 때문에 환경이나 문맥에 의한 발음 변이도 고려해야 한다고 보았다. 덧붙여, 한 번 발성한 모음에서도 동일한 포먼트값으로 시작하여 끝나는 것이 아니라 모든 지점에서 계속 변하는 양상을 보인다. 따라서, 분석지점이 어딘가에 따라 측정값도 달라지게 된다. 이러한 변화는 연음이나 공동조음이 있을 경우에는 이중모음처럼 처음부터 끝까지 변한다 (양병곤 1996a) 실제 컴퓨터로 음성 분석을 실행해보면 애매한 부분이 나타난다. 이 때는 정확한 발음에 의하여 발성된 음성으로 이미 분석한 표준자료 등을 이용하여 판단하는 것이 좋다. 예를 들어, 영어나 한국어의 단모음에 관한 성별 나이별로 수집한 자료를 참고하여 예상되는 포먼트 주파수를 염두에 두고 수집을 하는 것이 좋으며, 만약 컴퓨터 자동 분석 프로그램을 개발할 때도 마찬가지로 그 범위에 대한 자료를 입력하여 예상 밖의 자료제시가 되지 않도록 유의해야 한다. 특히, 남녀 변수를 고려하지 않고 한꺼번에 처리하여 평균을 낸다면 그 결과치는 어느 그룹도 대표하지 못하는 별로 의미없는 자료가 될 수도 있다 (양병곤 1996b).

III. Signalize의 포먼트 측정방식

최근의 디지털식 신호 분석은 지속적인 변수를 분석하는 기존의 아날로그 방식과는 달리 이산 신호 (discrete samples)를 음성 파형으로부터 채집하여 푸리에 분석 방법을 사용하므로 보통 이산푸리에 변환 (Discrete Fourier Transform: DFT)이라 부른다. 디지털적 분석은 음성 파형의 진폭을 일정 시간 간격으로 숫자로 표시하기 때문에 아날로그 장치보다 더 효율적으로 신호를 합성, 변환 또는 보존할 수 있고 일정한 시간점에 대한 주파수 분석을 신속히 할 수 있다. DFT 과정에는 2진법이 사용되므로 표본화 단계는 2의 제곱수가 된다. 또한 단위 초당 취해진 표본의 수(sampling rate)는 그 발화음이 분석되게 될 가장 높은 주파수의 두 배를 택하는데 이는 신호처리방법에서 생기는 오류치를 막기 위한 범위이다. 따라서, 분석하고자 하는 발화음의 주파수를 5000Hz까지 포함하고자 할 경우 초당 10000개의 표본들을 취해야 한다. 그런데 이러한 DFT는 시간을 함수로 음향 특징을 나타내는 푸리에의 분석 원리를 따르므로, 스펙트럼과 같은 정보를 얻기 위해서는 여러 개의 표본들을 함께 수집하여 분석해야 한다. 특히, 이러한 작업 과정 중에 연속된 음성을 단절된 신호로 취급함으로써 생기는 인위적 요소를 피하기 위해 양 가장자리 표본들의 진폭을 서서히 감소시켜 나타내는 해밍 창이 많이 사용된다. 대부분의 음성 분석 기법은 성도내의 모습이 시간에 따라 서서히 일어나므로 이에 따른 신호변화도 서서히 일어난다고 가정하여 음성신호를 일정한 지속기간의 윈도우의 모양에 따른 가중치를 곱하는 방식을 취한다. DFT를 빨리 얻어낼 수 있는 연산 방식으로서 많이 사용되고 있는 것이 바로 고속 푸리에 변환 (Fast Fourier Transform: FFT)이다. 이것은 보통 64, 128, 256, 512, 1024개의 지정된 수의 표본들을 처리한다. 그밖에도 최근에 많이 사용되는 것으로는 LPC 분석법이 있다. 이것은 되도록 적은 오류로 음성 신호를 예측

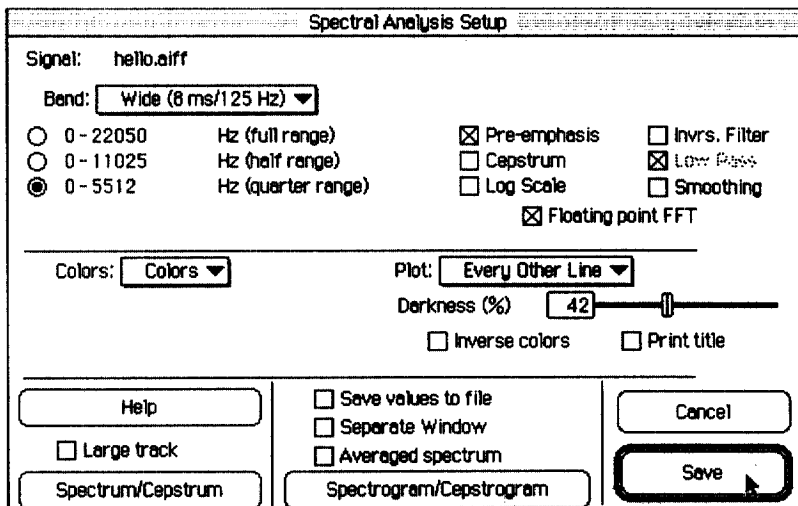


그림 1. 스펙트럼 분석 설정 대화 창

하고자 이전의 값들로부터 얻어낸 상관 계수로서 음성 신호를 나타내려는 분석법인데 이 상관 계수에 의해 만들어지는 스펙트럼은 역스펙트럼 분석 결과와 닮은 모양이 된다.

각시점에 해당하는 스펙트럼을 시간축위에 연속적으로 나타내면 스펙트로그램이 되는데, 이것은 시간, 주파수, 에너지의 세기라는 세 가지 변수를 한꺼번에 나타내 준다. 다시 말해서, 스펙트로그램의 가로축은 시간의 흐름을 나타내며, 세로축은 주파수의 범위를 나타내고, 에너지의 세기는 농도로 나타낸다.

Signalize에는 그림 1과 같은 스펙트럼 분석 설정 대화창이 나온다. 스펙트럼 분석과정은 먼저 분석 구간을 정하여 표본을 선택한 뒤 FFT 크기의 저장고에 넣고 모자랄 때는 빈자리에 0을 넣고 인접표본 값의 차이를 사용하여 사전강조(pre-emphasis)를 한 다음 해밍창의 비율에 따라 표본값을 곱하고 LPC 또는 FFT를 적용한다. FFT에는 2의 제곱수가 되어야하고 표본 속도는 조절할 수 없으므로 Signalize에서는 분석용 샘플수를 조정하여 적절한 스펙트럼 정밀도와 속도를 구하기 위해 넓은 대역 FFT는 256표본을 좁은 대역 FFT는 512표본을 사용하도록 고정되어 있다. 따라서 연구에 정밀한 값을 사용할 때는 가능하면 스펙트럼에서 구한 값을 이용하는 것이 좋다. 스펙트럼 분석에 그림 2에서 처럼 열 가지 종류가 있는데, 이들 가운데 피치 분석에 흔히 쓰이는 좁은 대역스펙트럼분석을 제외하고 포먼트를 구하는 데 사용되는 6가지를 살펴보면 넓은(Wide) 대역은 8ms의 신호를 125Hz 폭으로 분석한다. 남자의 포먼트를

Very Slim (100 ms/ 10 Hz)
Slim (50 ms / 20 Hz)
Very Narrow (33 ms/30 Hz)
Narrow (25 ms/40 Hz)
Very Wide (5 ms/200 Hz)
Extra Wide (3.3 ms/300 Hz)
LPC (15 ms)
Narrow Cone Kernel (512 pts)
Wide Cone Kernel (256 pts)

그림 2. 스펙트럼 분석 방식 종류

구하는데 적합하다. 아주 넓은(Very Wide) 대역은 5ms의 신호를 200Hz 간격으로 분석한다. 여성이나 아이들의 포먼트를 구하는데 쓰인다. 특별히 넓은(Extra Wide) 대역은 3.3ms의 신호를 300Hz 폭으로 분석한다. 대역에 따른 두 스펙트로그램의 처리 시간은 대역에 반비례한다. 따라서, 50Hz의 공명 주파대를 갖는 정밀 스펙트로그램은 각 배음 주파수들의 에너지 변화를 나타내는데 약 20ms가 걸리고, 300Hz의 공명 주파대를 가진 간이 스펙트로그램은 음성의 전반적인 에너지 유형인 포먼트를 나타내는데 약 3.3ms정도만이 소요된다. 피치가 높은 여성이나 아이들의 포먼트를 구하는데 사용한다. LPC법은 모든 종류의 목소리를 분석하는데 사용할 수 있으며 차수를 지정해 줘야 하고 결과 해석에 주의가 필요하다. 차수 지정은 자

기상관계수의 차수를 말하며 예상 포먼트수마다 2개씩 할당하고 추가로 2-3을 더해준다. 간단히 정하는 방식은 표본속도 주파수의 반에 해당하는 유효범위내에서 1kHz당 남성일 경우에는 2개씩을 여성일 경우에는 1개씩을 배당한 뒤 2-3을 더한다. 예를 들어, 10kHz로 녹음한 남성의 목소리는 그것의 반인 5kHz가 유효범위가 되며 $5 \times 2 = 10$ 이 되고 여기에 2-3을 더한 12, 13차수가 적절한 범위가 된다. 만일 너무 많은 차수를 택하면 잘못된 포먼트가 나타나게 되고 너무 적으면 두세개의 포먼트가 뭉쳐져서 하나의 포먼트로 나타나게 되므로 화자의 성도길이에 따라 달리 설정해야 한다(Johnson 1997:88). 스펙트럼이나 스펙트로그램을 통해 포먼트의 대체적인 위치를 알아본 뒤 이 값을 사용하는 것이 좋다. Macintosh의 표본속도가 11128이나 22256Hz로 되어 있으므로 복합변환 기능을 이용하여 표본속도를 10000Hz로 변환한 뒤 14차수의 LPC 분석을 하면 보다 선명한 포먼트 궤적을 구할 수 있다. 모음 /아

에서 처럼 두개의 포먼트가 인접한 경우에는 LPC 분석이 편리하다.

범위(Range)는 스펙트럼의 어느 부분을 분석 결과로 보여주는지 지정할 수 있다. 이 범위는 FFT 분석에는 영향을 주지 않고 표시창에서만 적용된다. 보통 표본속도의 반인 유효범위를 보이거나 1/4 또는 1/8 범위를 지정할 수 있다. 음성신호는 주로 0-4 kHz 범위에 나타나 있기 때문에 넓은 대역의 분석일 경우에는 이 범위를 지정한다. 사전강조(Pre-emphasis)는 사람의 음성 자체가 고주파부분으로 갈수록 진폭값이 떨어지므로 저주파 부분은 감폭시키고 고주파부분은 증폭시켜서 고주파에 해당하는 포먼트를 선명히 나타내도록 한다. Signalalyze에서는 FFT 분석을 음성신호의 계수화된 진폭값을 그대로 사용하지 않고 한 표본값과 연이어 따라오는 표본값의 차이값을 FFT 분석에 사용하도록 함으로써 대체로 옥타브 당 6dB씩의 증폭하는 결과를 얻게된다.

이 방식을 이용한 경우에는 그림 3과 같은 모양의 스펙트럼을 볼 수 있다. 첫 번째는 사전강조가 없이 처리되었고 두 번째는 사전강조가 되고 세 번째는 이것을 부드럽게(Smoothing) 만든 모양을 보이고 있다. 대수척도(Log scale)를 사용하면 y축은 dB값의 진폭축으로 나타나 각 배음의 크기를 쉽게 짐작할 수 있다.

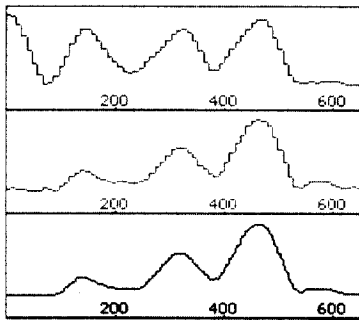


그림 3. 사전강조를 적용한 경우와 아닌 경우의 스펙트럼 비교

역여과기(Inverse Filter)는 FFT의 스펙트럼 분석이전에 자기상관 역여과기에 적용된다. 첫 번째 두 개의 포먼트를 조사할 때 매우 유용하다. 스펙트럼이 부드럽게 처리되고 첫 번째 두 포먼트 정점이 강조되어 나타난다. 평균스펙트럼을 선택하면 스펙트로그램을 만들기 위해 사용한 216표본이나 512표본의 FFT 값의 평균을 구해서 지정시간의 장기간에 걸친 스펙트럼의 평균을 보여준다. 분석결과를 저해상도와 고해상도로 나타낼 수 있는데 신호창에 바로 그려지는 것은 저해상도로 되어

있고 매우 빠르게 분석결과가 그려지며 포먼트 지점을 커맨드와 시프트 키를 누른 채 마우스로 클릭하면 해당 지점의 주파수값을 편집창에 보낼 수 있다. 별도의 창을 선택하면 고해상도의 스펙트로그램을 구할 수 있으나 처리속도는 느리고 주파수 값을 구할 수 없다.

고해상도를 지정했을 때 16색, 256색, 또는 256 회색 등급으로 나눌 수 있다. 최신의 컴퓨터는 수천만 색상의 처리가 가능한데 아직 Signalalyze에서는 256등급까지 지원이 되므로 모니터 색상조절을 통해 256색으로 설정해야 분석결과를 볼 수 있다. 스펙트럼 간격은 빨리 윤곽을 보기 위해서는 하나씩 건너뛰어 표시하게 하거나 아주 빠르게 하려면 하나씩 건너뛰어 되풀이하게 할 수 있다. 농도는 보통 10%-20%로 설정이 되어 있으나 녹음강도가 약한 경우에는 높은 값으로 올린다. 너무 높은 값이 되면 포먼트 사이의 경계가 희미해지므로 주의한다. 반면 너무 낮은 농도를 주면 포먼트의 위치가 사라질 수도 있다. 보통 높은 진폭이 짙게 나타나고 낮은 진폭은 열게 나타나는데 이것을 반대색으로 설정하면 스펙트로그램의 농도를 바꿀 수 있다. 스펙트럼 분석에서는 보다 낮은 표본속도가 더 선명하게 나타나기 때문에 보다 낮은 표본속도로 변형하여 분석을 하는 것도 필요하다. 즉 동일한 갯수(1024)의 FFT를 행할 때 10000Hz로 녹음한 샘플에서는 +/-10Hz 간격으로 분석되나 20000Hz로 녹음한 샘플에서는 +/-

20Hz의 간격이 되어 더욱 성긴 스펙트로그램을 만들게 된다.

IV. 복합음의 측정값 비교

그림 4는 SoundEdit16의 정현파생성기 메뉴를 이용하여 합성한 음파를 분석한 것이다. 맨 위의 창은 흔히 생기는 네 개의 포먼트값에 해당하는 500Hz, 1500Hz, 2500Hz, 3500Hz의 네 개의 정현파음을 1초의 지속시간으로 각각 만든 뒤 이를 16비트, 22250Hz의 표본속도로 합성했다. 사전 합성 결과 음성분석에서 사전강조에 의해 고주파에 해당하는 음의 진폭이 증가됨을 감안하여 500Hz음의 진폭은 25%의 크기로 1500Hz는 20%, 2500Hz는 15% 3500Hz는 10%로 하여 과도한 합성음이 생기지 않도록 했다. Aiff파일형태로 저장된 합성음을 Signalyze에서 열어 여러가지 방식으로 분석해 보았다. 첫 번째 창문은 8ms 간격으로 125Hz씩 훑어 가는 방식을 취했고 사전강조를 선택하지 않았다. 정점의 포먼트 값을 측정해 본 결과 첫 번째 포먼트 값은 505Hz, 두 번째는 1514Hz, 세 번째는 2510Hz, 네 번째는 3507Hz로 거의 정확하게 측정이 되었다. 두 번째 창과 세 번째 창에서도 복합음의 분석값이 거의 일치하는 모양을 나타내고 있다. 대역이 넓어짐에 따라 스펙트럼의 윤곽도 넓어짐을 확인할 수 있다.

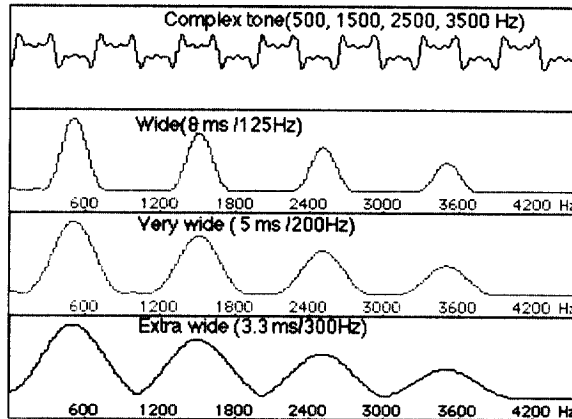


그림 4. 복합음에 대한 스펙트럼 분석 방식의 비교

덧붙여 Signalyze에 부가된 기능으로 512개의 표본을 중심으로 포먼트를 그려나가는 좁은 폭의 삼각뿔모양으로 스펙트럼을 분석하는 방식과 256개의 표본을 이용한 보다 넓은 폭의 스펙트럼을 이어서 스펙트로그램을 만들어보면 다음 그림 5와 같다. 이 그림에서 보면 좁은 삼각뿔형태가 3500Hz에서 약간 벗어난 것을 제외하고는 거의 정확히 포먼트 궤적을 그려줬고 넓은 폭의 그림에서는 두 개의 띠가 생성되어 중간지점을 취하면 거의 포먼트에 가까운 값을 구할 수 있다.

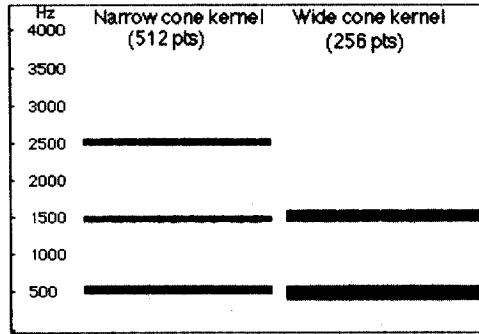


그림 5. 좁은 삼각뿔 모양과 넓은 삼각뿔 모양으로 스펙트럼을 분석한 결과 비교

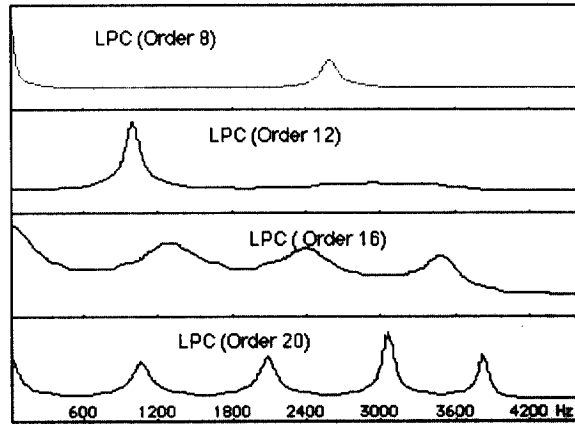


그림 6. 서로 다른 차수를 지정한 LPC에 의한 합성음 분석 비교

이번에는 LPC 분석을 통해 같은 합성음을 비교해 보았다. LPC의 차수를 임의로 최저차수인 8차부터 12, 16, 20차수값을 넣어 스펙트럼을 뽑았다. 그림 6은 이 결과를 나타낸다. 첫 번째 창에서는 8차로 지정했는데 2601Hz에 한 개의 포먼트만 나타났다. 두 번째 창문에는 12차로 지정했는데 여전히 996Hz에 해당하는 한개의 포먼트만 나타났다. 세 번째 창에서는 16차수의 경우로 네개의 포먼트가 각각 1294, 2407, 3481Hz로 앞서 보인 포먼트와 비슷한 값에 접근했으나 여전히 첫 번째 배음은 제대로 포착되지 못했다. 마지막 그림은 20차로 지정한 LPC로 제1포먼트는 1061, 2096, 3067, 3817Hz로 거의 1000Hz간격으로 허수의 포먼트 값이 나타났다. 이러한 분석에서도 알 수 있듯이 LPC를 사용하면 차수가 높을수록 많은 포먼트 개수를 제시해 주지만 이것이 반드시 실제 포먼트와 일치하지는 않는다는 점에 주의해야 한다. 보다 개선된 방법으로는 앞서의 넓은 대역스펙트럼을 통해 포먼트의 위치를 확인한 뒤 LPC로 분석하여 그 지점에 가까운 값을 읽는다. 가능하면 각 주파수에 해당하는 진폭값을 평균한 스펙트럼을 구하여 동시에 비교하면서 포먼트의 위치를 확인하는 것이 바람직하다.

V. 음성의 스펙트로그램 분석

다음으로는 40세 남성과 11세의 여자아이가 모음삼각도의 꼭지점에 해당하는 /이 아 우/ 세 개의 모음을 PowerPC로 입력하여 분석해 보았다. 그림 7은 남성화자가 발성한 모음 [a]를 넓은 대역 128Hz으로 8ms마다 분석했을 때 농도를 달리해 보았다. 농도의 차이에 따라 포먼트를 확인하기가 20% 정도일 때 적당함을 볼 수 있다. 30%일 때는 제1, 2포먼트가 너무 짙게 나타났기 때문에 두 개의 포먼트를 확인하기가 힘들다. 반면 10%에서는 제4포먼트 이상을 확인하는데 문제가 있다. 모음 [u]에서는 입술을 둥글게 하여 전체적으로 포먼트의 주파수가 낮아지는 경향이 있기 때문에 저주파가 강조되므로 농도를 40%이상 올려서야 제3포먼트를 확인할 수 있었다. 그림 8은 세 개의 모음의 스펙트럼을 보여주고 있다.

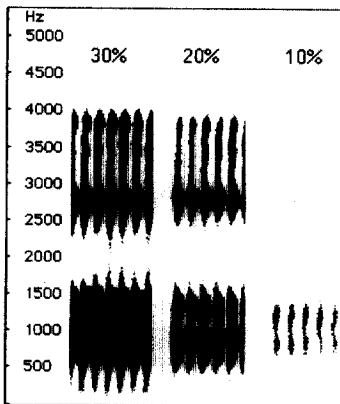


그림 7. 스펙트럼의 농도를 달리한 경우의 스펙트로그램 모양 비교

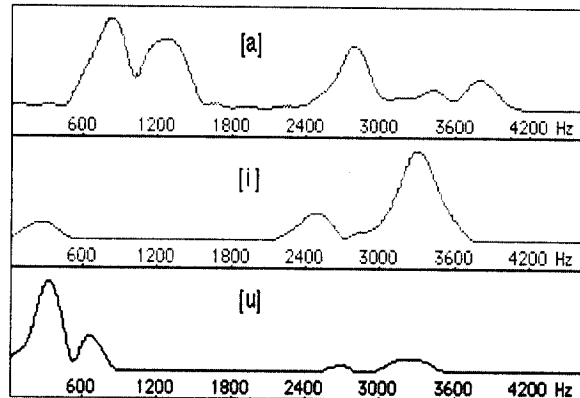


그림 8. 40세 남성의 모음 [아, 이, 우]에 대한 스펙트럼 비교

각각의 스펙트럼을 128Hz 간격으로 하여 나타내면 그림 8과 같고 이들 각각의 주파수는 모음 [a]의 포먼트는 각각, F1: 828Hz;F2: 1255Hz;F3: 2795Hz;F4:3416Hz이고 모음 [i]의 포먼트는 F1: 285Hz;F2: 2485Hz;F3: 3313Hz이고 [u]의 포먼트값은 각각F1: 336Hz;F2: 673Hz;F3: 2692Hz;F4: 3261Hz가 된다. 피치값의 변화는 129Hz에서 97Hz로 하강했다. 이 값은 국어 모음의 주파수와 일치한다 (Yang 1990).

모음 [u]의 경우에는 대역폭을 넓힐수록(예를 들어 200Hz이상) 제1, 2포먼트가 하나로 통합되어버려서 하나의 포먼트 값으로 처리되어 버리기 때문에 주의해야 한다.

그림 9는 11세 여아의 목소리를 넓은 대역으로 분석한 스펙트로그램이다. 먼저 피치를 분석해 본 결과 평균 297Hz에 해당하는 높은 수치를 보였다.

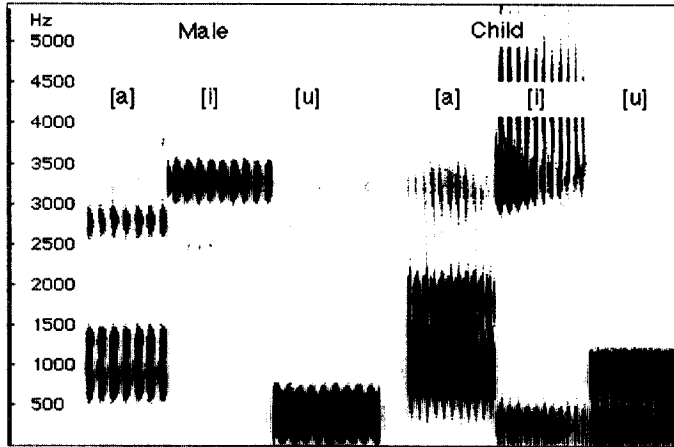


그림 9. 40세 남성과 11세 여아가 발성한 [아, 이, 우]에 대한 스펙트로그램

이것은 남성화자의 129Hz보다 훨씬 높은 값이다. 앞서 남성화자의 분석 방식을 이용해서 스펙트로그램을 구했을 때 많은 정점을 보여서 포먼트를 찾기가 어려웠다. 따라서, 분석구간은 200Hz 간격으로 좁혀가도록 지정했다. 스펙트럼을 분석해본 포먼트 값은 다음과 같다. 모음 [a]에서는 F1: 1100Hz, F2: 1682Hz F3: 3196Hz이다. [i] 에서는 각각 311Hz, 3377Hz, 3856Hz이고, [u]는 F1이 285Hz, F2가 932 Hz를 보였는데 입술이 나와서 성도길이가 길어지게 된 결과 저음이 강조되어 두 포먼트의 진폭값이 너무 강하고 제3포먼트의 모양이 거의 드러나지 않는다. 그러나 희미하지만 농도를 질게 하면 나타나므로, 이

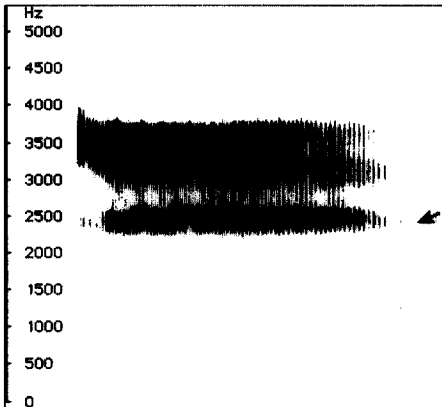


그림 10. 음성신호의 저주파 부분을 여과한 뒤 나타낸 모음 [우]의 제 3포먼트의 위치

변에는 제3포먼트의 위치를 구하기 위해 모음 [u]의 음성과형을 여유있게 1500Hz이하의 주파수를 걸러내어 고주파통과여과기로 처리함으로써 상대적으로 그 이상의 주파수가 돋보이도록 하였다. 그 결과 스펙트로그램은 그림 10과 같다.

이 스펙트로그램에서 볼 수 있듯이 제3포먼트에 해당하는 부분이 화살표가 가르키고 있는 것처럼 질게 그려지고 있으며 이 값은 스펙트럼으로 위치를 확인하면 약 2536Hz가 된다. 제4포먼트는 3183Hz에 있다. 물론 고주파부분의 강도가 매우 낮기 때문에 제 3포먼트의 값은 청각에 큰 영향을 미치지 않을 수도 있으나 측정하는 방법은 고주파통과 여과기를 통해 가능함을 확인할 수 있다.

VI. 결 론

지금까지 **Signalyze**를 이용하여 사람의 성도의 특징인 모음의 포먼트를 측정하고 그 결과를 해석하는 방법에 대해 살펴보았다. 포먼트는 여러 가지 상황에 따라 달라질 수 있으나 본 연구에서는 주로 프로그램의 정확도를 검토하기 위해 합성음과 /이, 아, 우/ 세 개의 안정된 모음을 피치 차이가 많이 나는 남성화자와 여자어린이의 발성음을 분석하여 보았다. 포먼트를 구하는데 사용되는 각각의 분석방식에는 장단점이 있으며 이것을 잘 운용함으로써 가장 적절한 값을 구할 수 있을 것이다. 그러나 지금까지 단순한 복합과의 분석에서 살펴보았듯이 분석구간의 설정이나 LPC 분석 차수의 지정에 따라 문제점을 나타내고 있음을 보았다. LPC 차수는 너무 낮게 지정하면 포먼트를 놓치고, 높게 지정하면 실제 포먼트 외에 추가로 더 나타나기도 한다. 농도의 조정에 따라 포먼트의 구별이 어려워질 수도 있으며, 원하는 주파수 영역의 포먼트를 확인하기 위해서는 여과기의 기능을 사용하여 해당 범위의 부분을 걸러내는 방식을 이용하는 것도 가능성을 보였다. 이 연구는 모음의 포먼트 분석에 필요한 기본 지식을 제공하여 음성분석프로그램을 사용하는데 도움이 될 것이다.

참고 문헌

- 양병곤. 1994. 모음의 음향적 특성. 제 1회 음성학학술대회 자료집. 대한음성학회. pp. 113-124.
- 양병곤. 1995. 합성한 한국어 단모음의 지각실험 연구. 언어. 20권 3호, pp.127-146.
- 양병곤. 1996a. 합성한 한국어 이중모음의 지각실험 연구. 언어 21권 3호, pp. 829-843.
- 양병곤. 1996b. 음성학 실험의 실제. *Donguei International Journal*. No. 1, pp. 111-121.
- 양병곤. 1997. 영어모음 발음의 이론과 실제. 영어교육연구. 제 7집, pp. 127-137.
- Chiba, T and M.Kajiyama. 1941. *The Vowel—its nature and structure*. Tokyo: Kaseykan.
- Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Johnson, K. 1997. *Acoustic and Auditory Phonetics*. Cambridge, MA:Blackwell.
- Keller, E. 1992. *Signal Analysis for Speech and Sound: User's Manual*. Charlestown, MA:Infosignal Inc.
- Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H-H theory. In W.J. Hardcastle and A. Marchal(eds.). *Speech Production and Speech Modeling*. Dordrecht: Kluwer Publishers.
- Yang, B. 1990. *Development of Vowel Normalization Procedures: English and Korean*. Seoul: Hanshin Pub. Inc.

<Abstract>

A Study of Vowel Formant Analysis Methods by Signalize

Yang, Byung-gon

This paper reviews each item of the spectral analysis window of a speech analysis software Signalize on a Macintosh computer to determine correct formant values of vowels. Formant represents vocal tract resonance and is counted from the first maximal energy peak on the spectrum of vowel sound. A complex tone is synthesized to check whether various analysis methods of the software correctly determine the three formant values. Also, two subjects' production of corner vowels /a, i, u/ is analyzed using the methods. Results show that most wide band spectral analyses correctly produce formant values of each corner vowel but LPC analysis shows several false formants without a correct order setup. For vowels with weaker formant peaks such as the vowel /u/, high pass filtering is suggested to boost its weak formants by suppressing higher amplitude of its first two formants. This study will help researchers to correctly determine formant values of vowels using any sophisticated speech analysis software.