

음성언어 의학자를 위한 음성 수집 분석 통계 처리 개관

부산대학교 사범대학 영어교육과 양병곤

I. 머리말

일반적으로 우리는 행동과학적 연구를 통해 현상을 기술하고, 밝혀진 사실을 바탕으로 설명을 시도하며, 이런 과정에서 나타난 원리를 파악하여 앞으로의 일을 예측할 수 있게 된다(손충기 2003). 음성과학에 대한 연구도 이와 같이 적절한 자료 수집을 거친 다음, 자료 분석을 통해 얻은 측정값을 통계적 기법을 통해 살펴봄으로써 환자집단별로 공통된 문제점을 파악할 수 있는 기준을 마련할 수 있다. 또한 정상과 비정상적인 집단의 경계점이 밝혀지면 이것을 기준으로 내원환자의 발음이 정상치에서 얼마나 벗어나 있는지 파악할 수 있고, 수술이나 일정기간의 재활 후에 정상치에 얼마나 근접하였는지 밝혀볼 수 있다. 이런 과정을 장기간에 걸쳐 많은 수의 환자에게 되풀이 하여 관찰하게 되면 여러 가지 수술이나 치료방법에 대해 상호 비교가 가능하고 어떤 방법이 가장 우수한 치료효과를 가져올 것인지를 파악할 수 있을 것이고, 궁극적으로는 병원에 방문한 환자의 상태를 진단하여 어떤 치료를 얼마의 기간에 걸쳐 해야 하고, 예후는 어떻게 될지를 말해 줄 수 있을 것이다. 이와 같이 사람의 음성을 다루는 음성언어 의학자들에게는 환자의 음성에서 어떤 변화가 있었는지를 측정하여 통계처리를 통해 환자 집단에 대한 특징을 파악하고, 정상인의 측정값도 동시에 연구하여 기준값으로 정해두면, 앞서 언급한 기술, 설명, 예측과정에 대한 풍부한 지식과 정보를 축적할 수 있을 것으로 기대된다.

사람은 의사소통을 하기 위해 음성을 사용한다. 사람의 음성은 성대의 진동인 음원을 바탕으로 여과기의 일종인 성도의 모양을 다양하게 변화시키면서 원하는 목소리를 낸다. 발성의 원리와 음원과 여과기의 이론과 측정에 관한 개관에 대해서는 양병곤(1998a, 1998b)을 참고하기 바란다. 의학자로서는 음원에 해당하는 성대의 여러 가지 질병이나, 여과기에 해당하는 성도에 발생하는 질병에 따라 입술 밖으로 나오는 음성이 정상이 아닌 환자들을 다루게 된다. 이 글에서는 이와 같이 발성에 문제가 있는 환자의 수술 전후의 음성을 수집하고, 분석한 뒤, 측정값을 통계적으로 처리하는 세 가지 과정을 개괄적으로 살펴봄으로써 환자의 진단과 치료, 재활에 대한 과정을 체계화하는데 도움을 주고자 한다. 구체적으로는 음성 수집에 널리 사용할 수 있는 켈빈(kalvin)이라는 소프트웨어를 이용하여 음성을 수집하는 방법을 소개하고, 수집한 음성을 프라트(Praat)라는 소프트웨어로 불러와 성대와 성도의 정보를 어떻게 측정하며, 마지막으로 수집된 음성분석 측정값을 통계 소프트웨어인 SPSS로 집단별 비교에 흔히 사용되는 T-검정을 통해 어떻게 처리하고 해석하는지 또 주의해야 할 점은 무엇인지 살펴보려고 한다.

II. 음성 자료 수집

먼저 음성 자료를 수집하기 위해서는 어떤 단어나 문장을 녹음할 것인지, 누구를 대상으로 해야 할지 결정해야 한다. 환자의 일상생활의 목소리를 모두 녹음하기 보다는 사전 연구를 통해 또는 평소에 진료하면서 관찰한 발음장애를 중심으로 목록을 만드는 것이 좋다. 녹음 대상은 가능하면 남자와 여자와 같이 비슷한 집단을 따로 묶어서 측정값을 구해야 평균 처리값이 의미가 있다. 대상인원은 많을수록 일반화하는데 도움이 되고, 너무 적으면 극단치가 평균값에 영향을 주어 올바른 결론을 내리기 어려울 수도 있다. 그러나 장기간에 걸쳐 음성 자료를 수집하고 여러 병원의 자료를 공유한다면 이런 문제들은 극복할 수 있을 것이다. 녹음하는 환경은 조용한 사무실이나 방음실을 이용하는 것이 좋다. 주위 잡음이 많이 들어갈수록 음성의 특징을 파악하기 어렵다. 또한 마이크, 녹음기나 컴퓨터와 같은 저장매체가 필요하다. 녹음용 마이크는 반드시 비싼 것을 구입할 필요는 없다. 좋은 마이크를 사용할수록 대체로 선명한 입력 파형을 보이긴 하지만 싼 마이크로 녹음한 음성이라고 해서 분석 결과가 청각적으로나 통계적으로 유의미한 큰 차이를 보이지 않는 경우가 많다. 오히려 녹음 환경을 조용한 곳으로 하고 주위에 선풍기나 냉장고와 같은 소음을 내는 물체가 없도록 하는 것이 더 중요하다. 마이크를 손으로 들고 녹음하게 되면 손과 마이크의 접촉 소리도 입력되므로 가능하면 옷핀으로 고정하거나 거치대를 구입하여 고정시켜 두는 것이 좋다. 마이크와 입술 사이는 약 10 cm 전후로 띄어서 거친 파열음의 소음이 강하게 입력되지 않도록 주의한다. 일반 녹음기는 녹음을 할 때 주위 잡음과 녹음기 자체의 소음이 함께 들어가는 경향이 있고, 녹음된 자료를 분석하려면 또 다시 컴퓨터에 입력하여 해당 부분을 따로 잘라낸 다음 파일마다 이름을 부여하는 지루한 작업이 필요하다. 필요한 부분을 찾을 때도 처음부터 들으며 찾아야 하는 불편이 있다. 이러한 단점을 극복하기 위해서는 처음부터 컴퓨터에 녹음하고 피험자마다 두문자를 부여하여 파일로 저장하는 도구인 켈빈을 사용해보기 바란다.

한글로 된 켈빈은 필자의 웹사이트 <http://fonetiks.info/alvin/>에서 kalvin.zip을 마우스로 선택한 뒤 이를 자신의 컴퓨터로 내려 받아 압축을 풀면 된다. Tcl이라는 스크립트 언어로 실험환경을 만들거나 통제할 수 있는 이 소프트웨어의 영문판은 참고문헌에 기재된 웹페이지에 들어가 받을 수 있다. 이 소프트웨어는 음성녹음 기능과 함께 주로 심리학에서 다루는 사람의 음성 지각에 대한 많은 실험메뉴를 갖추고 있다. 예를 들어, 두 가지 이상의 소리의 차이를 구별하는 구별실험, 어떤 단어나 발음인지 판단하여 이를 선택하거나 타자로 입력하는 확인실험이라는 청각실험, 음질의 판단, 양이 자극에 대한 이중지각 등이 있다. 각 실험메뉴에 대한 기본적인 설명과 설치하는 방법, *alvin* 폴더 구성과 기본적인 실험환경을 만드는 방법, 그리고 메뉴 등을 한글화하는 과정은 양병곤(2003a)을 읽어보고, 필자가 한글화한 켈빈에 나타나 있는 전체 메뉴와 폴더 구성 등에 대해서는 웹사이트에 있는 사용법 앵크로벳 파일을 참고하기 바란다. 이 글에서는 의학자들의 음성수집에 도움이 되는 두 가지 녹음저장 관련 메뉴에 대해 소개하고자한다.

단어나 문장을 들려주거나 보여주면서 녹음하게 하려면 켈빈의 초기화면에서 [단어 듣고 녹음

저장] 또는 [단어 보고 녹음저장] 가운데 하나를 택하여 실행해야한다. 피치의 변화를 조사하는 지터(jitter)값이나 음압의 크기변동을 나타내는 shimmer(shimmer)를 보다 안정적으로 포착하려면 거의 비슷한 높이의 목소리로 녹음을 시키는 것이 좋는데 이를 통제하려면 [단어 듣고 녹음저장] 방식의 모델음성을 활용하는 것이 좋다(양병곤 2003b). 먼저 [단어 보고 녹음 저장]은 컴퓨터 화면에 나타나는 우리말로 된 단어를 보고 피험자가 자신의 목소리를 하나씩 녹음할 때 사용한다. [단어 보고 녹음저장]의 실험 폴더 안에는 녹음과정을 통제하는 recording.tcl, 차례로 실행될 파일목록을 보여주는 text.stm이 있으며, recordings는 참가자가 녹음한 음성을 저장하는 폴더이다. recording.tcl을 메모장 프로그램으로 열어보면 앞부분이 다음과 같이 나타난다.

```
defineExperiment {
    -name "recording"
    -helpFile "Instructions.txt"
    -shuffle 1
    -controlFile "text.stm"
    -stimulusDir "samples"
    -recordingDir "recordings"
    -initialsRequired 1
    -presentation displayTextStimuli }
```

여기서 shuffle을 1로 하면 원래의 text.stm에 나타나 있는 순서를 한번 섞어서 임의의 순서로 재생하게 한다. 이 값을 0으로 하면 원래의 순서대로 재생하고, 2이상으로 하면 해당 숫자만큼 섞어서 재생하게 된다. presentation displayTextStimuli라는 부분은 kalvin/tcl/basic_exp라는 파일의 본문에 있는 다음의 프로시저에 의해 실행된다.

```
proc displayTextStimuli { stimulusName afterPresentation }
    {global currentStimulus
    setvalue .stimulus [lindex $currentStimulus 0]
    $afterPresentation }
```

초당 음성표본속도를 조정하려면 recording.tcl 파일의 맨 끝에서 두 번째 줄에 있는 "startrecording 22050 .meter"의 22050이라는 숫자를 바꾸면 된다. 보통 음성 파일은 2의 배수의 표본속도로 지정하는데 이것의 반인 11025로 하면 컴퓨터에 저장하는 분량이 반으로 줄어든다. 이어서 text.stm을 열어보면 다음과 같이 나타난다.

```
"안녕하세요" 안녕하세요.wav
"반갑습니다" 반갑습니다.wav
"아버지" 아버지.wav
```

기본적으로 각 줄의 곱따옴표로 표시된 단어나 문장이 화면에 나타나게 되고, 두 번째 파일이름에 해당하는 음성이 samples폴더에 피험자의 두문자와 함께 저장된다. 만약 다른 단어를 추가로

녹음하여 저장하려면 text.stm을 열어 진하게 표시된 줄과 같이 추가한 뒤 저장하면 된다.

```
"안녕하세요" 안녕하세요.wav
"반갑습니다" 반갑습니다.wav
"아버지" 아버지.wav
"어머니" 어머니.wav
"누나" 누나.wav
"안녕하세요 반갑습니다" 인사.wav
```

이 때 확장자인 .wav가 반드시 있어야 윈도우즈용 웨이브 파일로 저장된다. 마지막의 예와 같이 인용부호 안에 긴 문장을 입력하여 피험자의 발음을 오랜 시간에 걸쳐 녹음할 수도 있다. 다만 화면에 출력되는 길이에 제한이 있으므로 적절한 위치를 지정하려면 recording.tcl의 내용 가운데 다음 줄의 좌표축을 나타내는 진하게 표시된 숫자들을 바꾸어야 한다.

```
label .stimulus -value "" -y 75 -w 100 -h 6 -align center -font .Arial24
```

[단어 듣고 녹음저장]의 실험폴더의 recording.tcl을 메모장 프로그램으로 열어 실험환경을 지정하는 defineExperiment 안의 내용을 보면 presentation playAudioFileWithText 이란 프로시저를 사용하는 것으로 되어 있다. 이 부분은 kalvin/tcl/basic_exp라는 파일의 본문을 열어보면 다음과 같은 스크립트 명령어를 볼 수 있다.

```
proc playAudioFileWithText { stimulusName afterPresentation }
{
    global currentStimulus
    setvalue .stimulus [lindex $currentStimulus 0]
    playfile $stimulusName $afterPresentation }
}
```

[단어 듣고 녹음저장]의 실험폴더와 다른 부분은 재생할 음성 목록을 recording.stm에 저장하고 이 파일에는 기본적으로 각 줄의 곱따옴표로 표시된 단어나 문장이 화면에 나타나게 되고, 두 번째 파일이름에 해당하는 음성을 samples 폴더에서 불러와 피험자에게 들려주게 된다. 만약 samples 폴더에 음성이 들어가 있지 않으면 실행이 중단되고 오류가 생긴다. 따라서, 현재 들어가는 음성이외의 다른 단어나 문장을 들려주려면, 프라트나 윈도우즈의 녹음기로 음성을 컴퓨터에 저장하여 윈도우즈웨이브 파일 형태로 samples폴더에 반드시 넣어줘야 한다.

III. 음성 분석

음성은 한 번 발성하는 순간 사라지게 되므로 이를 컴퓨터에 불러와서 분석함으로써 보다 객관적인 측정값을 구할 수 있고, 통계처리를 통해 임의로 선택한 피험자로 구성된 정상인이나 환자 집단별 특성을 파악하거나 상호비교를 시도할 수 있다. 시중에는 음성분석을 할 수 있는 다양한

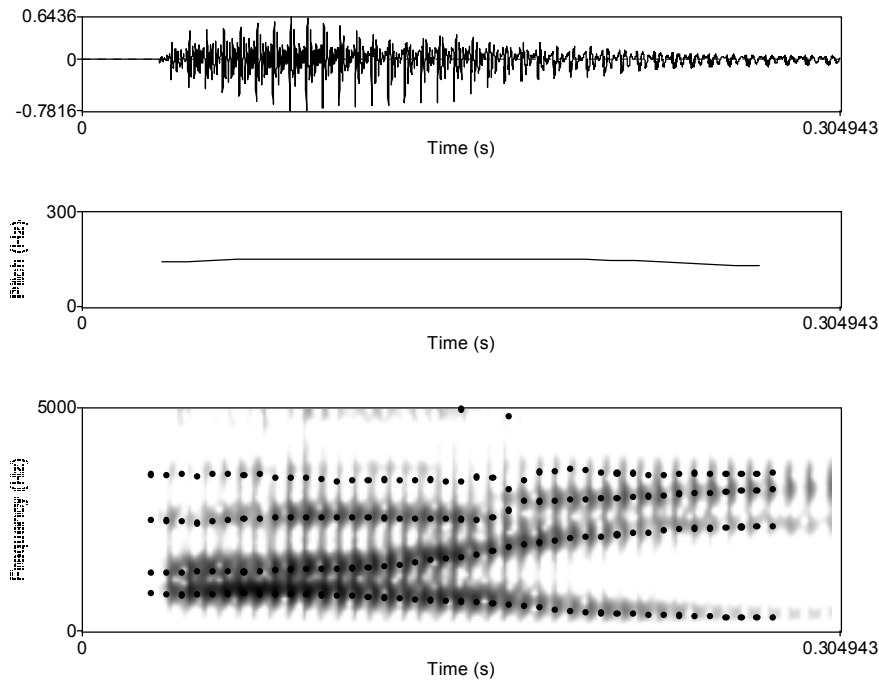
상용소프트웨어가 있다. 매우 비싼 가격에 비해 사용자의 편의를 고려하지 않아 익히는데 엄청난 시간이 걸리는 것도 있고, 너무 제한된 메뉴로 인해 필요한 분석을 제대로 하기 어려운 것도 있다. 여기서 소개할 프라트는 인터넷을 통해 무료로 다운로드할 수 있고, 적절한 분석과정을 이해하면 매우 정밀한 결과를 구할 수 있다는 장점이 있다. 그러나 분석할 때 설정변수를 제대로 지정하지 않거나 분석된 측정값의 타당성을 검토하지 않고 잘못된 수치값을 구하여 통계처리를 하게 되면 무의미한 결론을 내릴 수도 있으므로 조심해야 한다. 항상 정상인의 발음에 대한 기존 연구 자료를 참고하여 측정값이 옳은지 판단해야한다(양병곤 1998a, 1998b, 2000).

프라트는 http://www.fon.hum.uva.nl/praat/download_win.html에서 무료로 받을 수 있다. 설치하는 과정과 각 메뉴가 어떤 의미를 가지는지 그리고 음성분석에 필요한 이론과 실제 음성을 분석하는 방법에 대해서는 양병곤(2003c)의 전자파일을 참고하고, 되풀이된 작업을 자동으로 처리할 수 있는 스크립트를 만드는 과정은 양병곤(2004)을 참고하기 바란다. 이 글에서는 음성분석의 가장 기본적인 성대와 성도의 특징 분석에 대해서만 살펴보기로 한다.

일반적으로 사람의 음성은 성대의 진동인 음원과 성도의 공명인 여과기로 분리된다. 음원은 프라트를 이용해서 f_0 라는 값으로 측정되며 기본주파수라고도 하며, 청각적인 측면을 강조하면 피치라 한다. 보통 남성의 평균 피치는 약 80 200 Hz, 여성은 150 300 Hz, 아이들은 약 200 500 Hz의 범위에서 변한다(Clark and Yallop, 1995:240). 대체로 키가 크고 우람한 체격의 사람은 피치가 낮고, 키가 작고 마른 체격의 사람은 피치가 높은 경향이 있다. 우리는 피아노의 건반에 따라 도미솔과 같이 음높이가 다르다는 것을 아는데 이 음의 높이가 음원에서 측정된다. 이에 비해 혀의 위치나 입 벌림 정도, 입술의 모양에 의한 성도의 변화는 여과기의 변화로 나타난다. 성도를 이루는 구강, 인강, 비강의 공간이 변해서, 공명강의 모양이 달라지고, 좁혀지는 위치가 어디인가에 따라 다양한 음성출력이 나온다. 여과기의 특성은 스펙트로그램에서 에너지가 가장 많이 모인 띠 모양을 보이는데 이를 포먼트라 한다. 일반적으로 모음 ‘이’와 같이 입 벌림이 좁으면 제1포먼트가 낮고, 모음 ‘아’와 같이 입을 많이 벌리는 발음은 제1포먼트가 높다. 제2포먼트는 혀가 앞으로 내밀어 발음하는 모음 ‘이’에서는 높지만, 혀가 뒤로 당겨지는 모음 ‘우’는 낮다. 또한 입술을 둥글게 하면 제3포먼트 값이 낮아진다. 그림 1은 필자가 “아이”라고 발성한 음성을 프라트로 분석한 다음 그림 창에 나타낸 것을 보여준다. 그림 1의 맨 위 부분은 음파의 모양을 보여주고, 중간 부분은 음원의 특징을 나타내는 피치곡선을 보이며 맨 아래 부분은 스펙트로그램 위에 에너지가 높은 부분을 컴퓨터가 자동으로 추적하여 5밀리초마다 점으로 표시한 포먼트의 궤적을 보여준다. 먼저 피치의 변화는 어느 정도 높은 값에서 약간 올라갔다가 서서히 내려오는 모양을 하고 있다. 이 그림에서는 거의 일직선처럼 나타나 보이지만 실제로 확대하여 살펴보면 음성의 피치는 시간에 따라 약간씩이지만 변화가 있다. 이러한 단위시간내의 변화량을 지터(jitter)라는 값으로 측정한다. 한편 단위시간내의 음의 크기변화를 심머(shimmer)라고 한다. 정상인의 변화율은 작는데 비해 성대의 진동이 불규칙적인 환자의 경우에는 이런 값들이 크다. 따라서 이 값을 이용하여 성대의 진동이

정상인지 여부를 판단하는데 활용할 수 있다.

그림 1. ‘아이’의 음성과형, 피치, 스펙트로그램의 포먼트 궤적



여과기의 특성인 포먼트의 변화는 모음 ‘아’에서는 제1포먼트 값이 높지만 모음 ‘이’에서는 낮아지고, 제2포먼트는 모음 ‘아’에서 ‘이’로 갈수록 높아지고 있다. 이러한 포먼트의 변화를 통해 우리는 시간의 변화에 따라 화자의 발성기관이 어떻게 움직였는지 관찰할 수 있고 동시에 각 지점의 측정값을 구하여 환자의 목소리의 어떤 부분이 얼마만큼 변화가 있었는지를 통계적으로 보여줄 수 있다. 덧붙여, 만약 환자의 성대의 근육이 비정상이어서 정상인에 비해 지속적인 성대 발성이 힘들 것이라 가정해보면, 한 번에 길게 발성할 수 있는 음성의 길이를 측정하여 정상인과 비교하거나, 일정 치료기간을 경과한 후에 얼마만큼의 발성지속을 할 수 있는지 총발성지속시간을 측정하여 통계처리를 통해 유의미한 결론을 끌어낼 수도 있다.

IV. 통계 처리

우리는 음성을 분석한 측정값을 통해 각 집단별로 어떤 특성이 있는지 일반적으로 기술할 수 있다. 여기서는 기존의 데이터베이스에 들어있는 정상인과 환자 집단의 목소리를 수집하여 측정된 피치와 지터, 심머값을 SPSS를 이용하여 통계분석하고 집단별 특성을 살펴본 뒤, 이들 집단을 통계적으로 비교한 결과를 해석하는 예를 보이고자 한다. 보다 더 전문적이고 자세한 통계학적 분석

일례와 주의할 점들은 노형진과 정한열(2002)을 참고하기 바란다. 표 1은 1994년 Kay Elemetrics사의 장애음 데이터베이스에 포함되어 있는 20명의 정상인과 환자 집단의 개인별 피치값, 지터값, 심머값을 보여준다. 이 자료는 정상과 환자로 분류한 각 집단에서 피치값이 다소 높은 여성들의 발음에 대한 데이터를 임의로 선택했고 소수점 2자리까지 나타냈다.

표 1. 정상집단(10명)과 환자집단(10명)의 피치, 지터, 심머 데이터

정상집단			환자집단		
피치	지터	심머	피치	지터	심머
247	0.18	1.89	223	1.33	2.98
239	0.51	1.25	243	0.66	5.17
302	0.27	2.11	208	1.07	3.51
193	1.42	2.46	209	1.08	3.66
210	0.48	1.34	235	1.11	4.53
216	0.82	2.13	218	1.99	4.25
239	0.39	2.01	196	3.84	9.66
217	0.71	1.42	242	4.31	9.81
236	0.89	1.94	200	0.82	6.05
211	0.41	1.35	202	1.92	3.19

먼저 각 집단의 피치값의 평균과 표준편차를 SPSS 주메뉴인 [분석]아래의 [기술통계]라는 하위 메뉴를 실행해 구해보면 표 2와 같이 나타난다.

표 2. 정상과 환자 집단별 기술 통계값

구분		N	최소값	최대값	평균	표준편차
정상	피치	10	193	302	231	30.1
	지터	10	0.18	1.42	0.61	0.37
	심머	10	1.25	2.46	1.79	0.42
환자	피치	10	196	243	218	17.51
	지터	10	0.66	4.31	1.82	1.27
	심머	10	2.98	9.81	5.28	2.52

표 2에서 살펴보면 피치값의 평균은 약 13 Hz정도 차이가 나지만 앞 절에서 남자와 여자의 피치값이 약 100 Hz정도 차이가 난다는 사실에 비추어 볼 때 대체로 비슷한 그룹으로 생각된다. 하지만, 지터값과 심머값의 평균은 많은 차이가 나서 환자의 집단과 정상인의 차이가 매우 크다는 것을 짐작할 수 있다. 과연 이러한 두 집단의 차이가 통계적으로는 어떠한 차이를 보이는가를 조사하려면 SPSS의 독립표본 T검정을 사용하여 처리해 볼 수 있다.

표 3. 정상과 환자 집단의 음향적 측정값의 T-검정 결과표

	Levene의 등분산 검정			평균의 동일성에 대한 T-검정						
		F	유의확률	t	자유도	유의확률 (양쪽)	평균차	차이의 표준오차	차이의 95% 신뢰구간 하한	신뢰구간 상한
피치	등분산 가정됨	1.07	0.31	1.22	18.00	0.24	13.40	11.01	-9.74	36.54
	등분산이 가정되지 않음			1.22	14.46	0.24	13.40	11.01	-10.15	36.95
지터	등분산 가정됨	7.32	0.01	-2.88	18.00	0.01	-1.21	0.42	-2.08	-0.33
	등분산이 가정되지 않음			-2.88	10.48	0.02	-1.21	0.42	-2.13	-0.28
심머	등분산 가정됨	11.08	0.00	-4.32	18.00	0.00	-3.49	0.81	-5.19	-1.79
	등분산이 가정되지 않음			-4.32	9.49	0.00	-3.49	0.81	-5.31	-1.68

표 3에서 보면 피치값은 두 집단의 평균차가 약 13.4 Hz가 나지만 유의확률은 0.24이기 때문에 비슷한 집단으로 볼 수 있다. 덧붙여 차이의 95%신뢰구간에서도 상한(36.54)과 하한(-9.74) 사이에 0이 포함되므로 비슷한 집단으로 판단한다. 지터값에서는 Levene의 등분산 검정에서 유의미한 차이($p < 0.01$)를 보였기 때문에 등분산이 가정되지 않음을 나타내는 t값을 읽으면 -2.88이고 유의확률은 0.02이고 상한과 하한값의 범위에 0이 들어가지 않으므로 통계적으로는 유의수준인 0.05이하의 유의미한 차이를 보인다고 서술할 수 있다. 또한 심머인 경우에도 Levene의 등분산 검정에서 유의미한 차이가 나왔기 때문에 등분산이 가정되지 않는 t값인 -4.32를 채택하고 유의확률인 0.00에서 보듯이 통계적으로 유의미한 차이가 난다고 결론을 지을 수 있다. 결론적으로 정상집단과 환자집단의 지터와 심머에서는 유의미한 차이가 났다. 만약 수년간에 걸쳐 이러한 질병에 걸린 환자들을 대상으로 측정한 자료를 축적해 나간다면, 지터와 심머를 기준으로 정상집단에 속하는지 환자집단에 속하는지 어느 정도 정확하게 판단할 수 있게 될 것이다. 또한 장기간에 걸쳐 환자집단에 어떤 치료를 함으로써 지터와 심머값이 얼마나 정상집단에 가깝게 접근하는지를 살펴보면, 수술 전후의 환자집단의 지터와 심머값들을 측정하여 집단 내에서 비교해 보는 방법을 사용할 수 있다. 물론 환자집단에서도 질병의 진행정도에 따라 세분할 수 있을 것이다. 만약 단순히 정상과 환자라는 두 집단의 비교가 아닌 병증별로 구분된 세 집단이상의 평균을 비교하려면 ANOVA라는 분산분석 통계를 이용해야한다. 이 때 측정된 종속변수의 수에 따라 측정값이 하나일 경우에는 일변량 분산분석(analysis of variance:ANOVA)을 사용하고, 두 개 이상일 경우에는 다변량 분산분석(multivariate analysis of variance:MANOVA)을 사용하여 관찰할 수 있다. ANOVA와 MANOVA에서 독립변수의 수에 따라 일원 분산분석(one-way ANOVA), 이원 분산분석(two-way ANOVA), 다원 분산분석(multi-way ANOVA)으로 나뉘지고 독립변수간의 상호작용 등을 파악할 수 있다.

이러한 통계처리 과정에서 유의해야 할 점은 먼저 측정값으로 구한 지터와 심머값이 타당성이 있는가를 살펴보아야 한다. 악성 종양환자의 목소리일 경우 기존의 음성분석소프트웨어에서는 피치 측정값에 에러를 많이 보이기 때문에 잘못된 피치값에 근거한 지터값에 문제가 있을 수 있다. 또한 과연 이러한 집단별 측정값의 차이가 어떤 의미를 가지는가를 살펴보아야 한다. 일반적으로 선형적인 변화를 나타내는 물리음향적인 측정값과 비선형적인 변화를 보이는 주관적이고 심리적인 판단은 항상 일치하지는 않는다는 많은 증거들이 있다(양병곤 1997). 예를 들어, 위의 집단에서 지터값과 심머값의 차이가 실제 환자와 정상인에게는 얼마나 일상생활에 불편함을 주는지 주관적인 판단에 대한 연구가 있어야 한다. 즉 이러한 측정값의 수치적인 비교에 추가하여 환자의 만족도가 수술 전후에 얼마나 나아졌는지를 설문지를 이용한 정성적인 평가를 하거나, 전문가의 음질 평가척도를 이용하여 조사한 뒤 통계 처리하여 볼 수도 있다. 흔히, 수치의 비교에만 몰두하다보면 환자가 느낄 수 있는 수술 경과를 예측하기 어려울 것이다. 마지막으로 음성분석을 통해 쉽게 측정할 수 있는 피치나 포먼트값을 집단별로 비교할 때는 원래 피험자들의 발생기관인 성도의 신체적인 차이가 있기 때문에 나온 유의미한 차이인지, 아니면 질병의 유무에 따른 차이인지 잘 판단해야 한다. 특히, 남녀와 같이 성별이 다른 경우에 피치와 포먼트값은 남성보다 일정한 비율만큼 높은 경향이 있기 때문에, 남녀집단을 따로 하지 않고 하나로 묶어서 평균값을 내거나 통계비교를 하게 되면 정상분포나 집단의 동질성을 위반한 경우에 해당하므로 주의해야 한다(Yang 1996).

V. 맺음말

지금까지 화자의 음성을 수집하고 분석한 뒤 통계처리를 통해 집단별로 어떤 차이가 있는지를 연구하는 과정들을 개괄적으로 살펴보았다. 필자는 환자를 직접 다루는 의사가 아니기 때문에 구체적인 음성 자료 수집과 분석 내용에 필요한 단어나 문장, 또 음성분석에서 어떤 타당한 측정값을 활용해야 하는지에 대해서는 상세히 제시할 수가 없다. 이 부분에 대해서는 앞으로 음성언어의 학자들의 많은 연구와 발표가 있기를 기대한다. 이러한 분야의 발전을 이루기 위해서는 학제간의 연구가 매우 필요하다고 여겨진다. 의학 분야의 전문가들과 언어발달과정, 음성분석, 심리학, 언어처리 공학 등에 정통한 분들이 자신의 전문지식을 공유하고 동시에 다른 분야에도 많은 관심을 가지고 있어야만 한다. 예를 들어, 성대나 구강의 결절이나 연구개를 제대로 닫지 못하는 환자에게서는 어떤 발음이 문제가 되는지 등은 단순히 한 분야에 종사하는 연구자보다는 여러 분야의 전문가들의 토의를 거쳐 종합적인 음성수집 자료 목록을 만들 수 있게 될 것이다. 특히, 수년간의 음성수집결과를 연구자끼리 네트워크를 통해 공유하고 항상 열린 마음으로 이 분야에 종사할 때 국내외에서 인정받을 수 있는 성과를 거둘 수 있을 것으로 생각된다.

참고문헌

- 노형진, 정한열: *한글 SPSS 10.0 기초에서 응용까지*. 서울, 형설출판사, 2002
- 손충기: *교육·심리·사회 연구방법론*, 서울, 동문사, 2003
- 양병곤: 인간의 청각 척도에 관한 고찰. *음성과학* 제2권: 125-134, 1997
<http://fonetiks.info/bgyang/db/9712.htm>
- 양병곤: Signalyze에 의한 피치 분석방법 고찰. *동의 논집* 28:69-80, 1998a
<http://fonetiks.info/bgyang/db/98pitch.pdf>
- 양병곤: Signalyze에 의한 포먼트 분석방법 고찰. *동의 논집* 28:81-92, 1998b
<http://fonetiks.info/bgyang/db/98formant.pdf>
- 양병곤: Praat에 의한 억양 분석 방법. *동의논집* 32:29-39, 2000
<http://fonetiks.info/bgyang/db/00praf0.pdf>
- 양병곤: 청각 실험 도구 Alvin 사용법. 한국언어치료학회 제 12회 학술발표논문집, 2003a
<http://fonetiks.info/bgyang/db/alvin.PDF>
- 양병곤: 악성종양환자와 정상인이 발성한 모음의 좁은대역 스펙트럼값의 상관계수와 절대차이합 비교. *음성과학*, 10권 4호:189-200, 2003b
<http://fonetiks.info/bgyang/db/200312mal.PDF>
- 양병곤: 프라트를 이용한 음성 분석의 이론과 실제. 부산, 만수출판사, 2003c
<http://fonetiks.info/praat/praatdown.htm>
- 양병곤: *Praat Script*를 이용한 음성분석과 변형. 부산, 만수출판사, 2004
<http://fonetiks.info/praat/praatscript.htm>
- Clark, J. & C. Yallop: *An Introduction to Phonetics and Phonology*. Cambridge, MA, Blackwell Pub. Co., 1995
- Yang, B.: A comparative study of English and Korean monophthongs produced by male and female speakers. *Journal of Phonetics*, Vol. 24:245-261, 1996
<http://fonetiks.info/bgyang/db/96jp.pdf>

양병곤

부산대학교 사범대학 영어교육과

<http://fonetiks.info/bgyang>

e-mail: bgyang@pusan.ac.kr

전화: 051-510-2619, 010-9618-7636

Abstract

An overview of data collection, speech and statistical analysis for speech language doctors

Byunggon Yang

Professor

Dept. of English Education, Pusan National University

The purpose of this paper was to describe briefly the three major steps of speech research for speech language doctors: speech data collection, speech analysis and statistical analysis. For speech data collection, the author introduced two menus of *kalvin*, a software modified for Korean users who might find it useful to automatically collect and classify speech inputs of subjects by their initials. For speech analysis, the author mentioned some of the merits and demerits of the software, *Praat* illustrating a waveform and pitch and formant analysis of a speech sound. Finally, the normal and pathological data from a speech data base were statistically compared using the T-test of SPSS. Results showed that the normal group was significantly different from the disease group in parameters of jitter and shimmer. A few suggestions were made to ask the researchers to check the validity of measurement values and to consider what the statistical result would mean to the patients and normal people. Further studies on each step would be desirable to cure and rehabilitate patients with any speech-related diseases.