

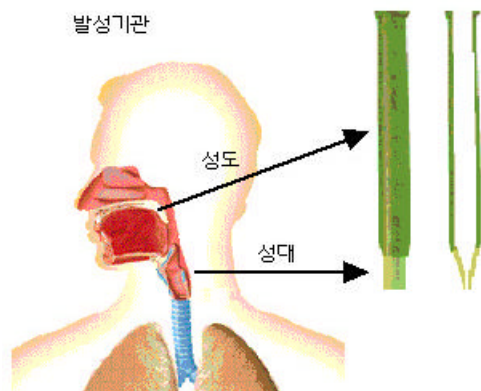
## 성문분석에 의한 화자확인이란 무엇인가?

양병곤 (동의대학교 영어영문학과 교수)

요즈음 테이프에 담긴 목소리의 주인공이 누구인지, 유괴범이 전화로 여러 번 돈을 요구할 때 녹음한 목소리와 체포된 유력한 용의자의 목소리가 같은지를 성문분석을 통해 조사하는 화자확인 문제가 사회적인 관심을 끌고 있다. 사람의 목소리는 어떻게 나오고 개인의 특징은 어떻게 나타나는가? 또, 사람의 목소리를 어떤 과정을 통해 비교하는가? 성문분석의 한계와 문제점은 어떤 것이 있는가? 많은 독자들이 이런 의문을 한번쯤 가져보았을 것이다. 독자들의 이런 궁금증을 해결하는데 조금이라도 도움을 주기 위해 이 글을 써본다.

우리는 일상생활에서 의사소통을 하기 위해 목소리를 주로 사용한다. 누구나 쉽게 발음하고 다른 사람의 목소리를 자연스럽게 알아듣는다. 요즈음 등장하는 음성을 알아듣는 컴퓨터나 음성인식 기술개발에 엄청난 비용을 들여 어느 정도 사람의 목소리를 알아듣고 상당히 자연스럽게 말하는 장치가 판매되고 있지만, 여전히 사람처럼 자연스럽게 듣고 말하는 장치는 개발되지 못했다. 그만큼 사람의 목소리는 복잡하고 정교한 반면 분석에 의한 합성 기술은 여전히 발음기관이나 청각기관의 기능을 제대로 파악하지 못하고 있다. 사람의 목소리는 크게 성대의 떨림과 성도의 공명으로 분리하여 생각해 볼 수 있다. 그림 1은 사람의 얼굴 옆 단면을 해부적으로 나타낸 그림과 버들피리의 모양을 비교하여 나타내고 있다.

그림 1. 발성기관과 버들피리의 관계



봄에 물기가 한껏 오른 버드나무 가지를 꺾은 뒤 껍질을 속 빼어서 끝부분의 거친 부분을 깎아내고 불면 소리가 난다. 사람의 목소리도 버들피리의 울림과 같은 원리로 나온다. 버들피리의 입술을 대고 부는 곳이 사람의 성대에 해당하고 긴 대롱이 바로 성도의 모양에 해당한다. 사람의 성대는 그림에서처럼 얇은 막으로 허파에서 올라오는 공기압력에 의해 열렸다가 빠른 속도로 지나가면서 기압이 낮아지며 서로 붙게 된다. 종이 장을 두개 서로 붙여서 그 사이에 바람을 불어보면 떨림이며 진동소리가 나는 것도 같은 원리이다. 보통 남자의 성대는 1초에 120번 붙었다 떨어지는 동작을 되풀이하는데 이를 기본주파수라고 한다. 기본주파수는 성대의 크기와 무게, 탄성도에 따라 달라지고 발음할 때 근육을 당겨서 팽팽하게 하면 더 빨리 진동한다. 이것은 고무줄을 팽팽하게 당겨서 통길 때와 약간 늦추어서

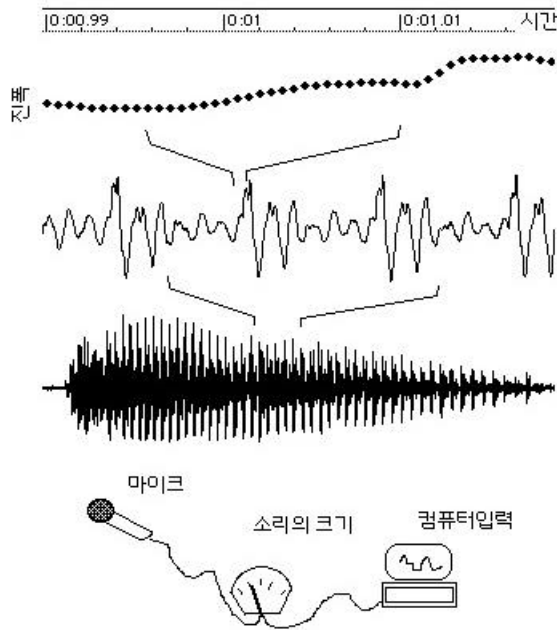
통길 때의 진동수가 달라지는 것과 같다. 남자의 성대에 비해 여자나 어린아이의 성대는 크거나 길이가 짧기 때문에 더욱 빠르게 진동한다. 여자의 기본주파수 평균은 230 Hz이고 아이들은 300 Hz이상이 된다. 서로 말다툼을 할 때는 피치를 올린다고 하는데 물리적으로는 성대가 팽팽히 당겨지고 빠르게 진동하는 현상을 말한다. 이 값의 변화범위도 매우 넓다. 소프라노가수인 경우는 1000 Hz나 된다. 그래서 사람마다 늘 하던 발음으로 하면 성대의 크기와 무게 등의 정보가 기본주파수에서 나타나게 된다. 따라서, 사람의 목소리를 들어보면 아이인지 어른인지, 또 남자인지 여자인지 쉽게 알 수 있는 것도 이런 기본주파수를 판단 기준의 하나로 삼기 때문이다. 그래서, 남자가운데서도 키가 작고 성대가 작은 사람은 전화로 이야기할 때 여자로 오해를 받는 경우도 종종 있다. 특히, 성대의 질병으로 혹이나 암조직이 생기면 목소리가 탁하고 거친 소음이 많이 섞이게 되는데, 규칙적인 음성신호와 불규칙적인 소음의 비율을 구하면 병적인 환자와 정상인을 쉽게 구별할 수 있다.

성도는 어떤 역할을 할까? 사람이 발음할 때 보면 턱이 오르내리기도 하고 혀를 매우 자주 움직이는 모습을 쉽게 볼 수 있다. 혀의 움직임이 눈에 띄게 많다. 이런 움직임은 입안의 공간을 여러 가지로 바꾸는 동작이다. 앞서 버들피리는 크기와 모양이 일정하지만, 사람은 이런 혀나 턱·입술과 같은 조음기관을 끊임없이 움직이면서 원하는 피리모양의 조합을 만든다. 우리가 모음을 발음할 때 입안의 모양을 살펴보면 ‘아’일 때는 입을 많이 벌리고 혀도 낮게 깔리며, ‘이’일 때는 입을 다물고 혀의 앞부분을 입천장 쪽으로 가까이 둔다. 이런 크기와 모양으로된 버들피리를 두개 겹쳐서 붙어도 같은 소리가 난다. 버들피리가 굵을수록 또 길이가 길수록 불기도 힘들고 매우 낮은 주파수로 진동을 한다. 정상적인 사람이라면 누구나 두 가지 다른 피리소리를 들려주고 어느 쪽인지 쉽게 선택할 수 있고, 그 굵기와 길이도 짐작할 수 있다. 머리가 좋은 독자라면 벌써 사람의 목소리에도 이렇게 성대와 성도의 신체적인 특징들이 실려있구나 짐작했을 것이다. 실제로 우리는 얼굴은 보지 않고 사람의 목소리만 들어도 누구인지 쉽게 알아맞힌다. 하지만 비슷한 길이와 굵기의 버들피리를 들려주면 쉽게 구분하기 힘들 것이다. 마찬가지로 두 사람의 발성기관이 비슷한 경우에는 그저 귀로서 들어서만 같은지 다른지 확인하기가 힘들다. 이런 경우에는 컴퓨터를 이용한 성문분석도구를 이용해야만 한다.

성문분석에 관한 TV 뉴스보도에서 종종 어떤 연구자가 컴퓨터 화면에 떠있는 여러 가지 문양의 그림을 차례로 이동시키며 살펴보는 모습을 기억하는 독자들이 많을 것이다. 성문분석이란 한마디로 소리의 특징을 사람의 손금을 보듯이 그림으로 나타내거나 정밀한 숫자값으로 표현하는 방식이다. 컴퓨터에 소리를 입력하는 과정은 그림 2와 같이 마이크를 통해 소리의 크기를 컴퓨터의 음성처리카드가 받아서 숫자로 기록하여 저장하고 녹음을 시작한 시간을 기준으로 얼마의 시간이 지났는지 함께 기록한다.

<그림2여기에>

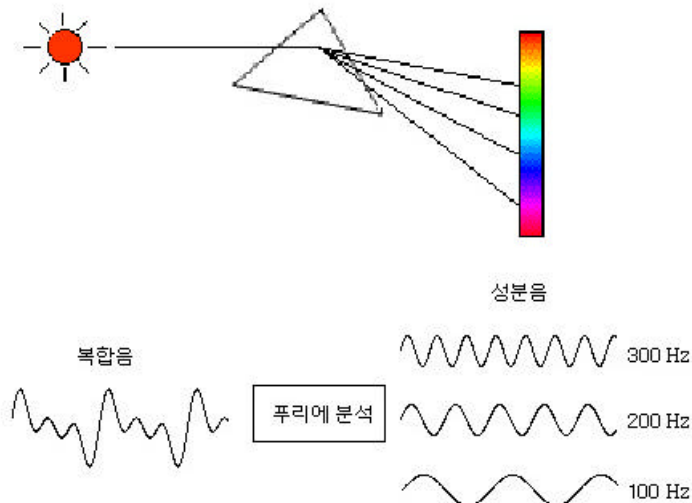
그림 2. 컴퓨터 음성입력과정(맨아래)과 파형의 점진적 확대



성대의 진동속도인 기본주파수에 덧붙여, 최근까지 많이 사용되어온 성분분석의 핵심은 푸리에변환에 의한 스펙트럼 분석이라고 할 수 있다. 푸리에변환이란 여러 개의 성분으로 이뤄진 복합파를 단절된 시간간격의 창으로 떼어내어 수학적 연산을 통해 각각의 성분음으로 분리하는 기법을 말한다. 모든 색깔의 빛을 합해져 있는 햇빛을 삼각유리막대로 통과시키면 그림 3과 같이 색깔의 띠로 분리되어 나타나는데 복합음처럼 나타나는 사람의 목소리도 이렇게 성분음으로 분리하여 나타낼 수 있다.

<그림 3여기에>

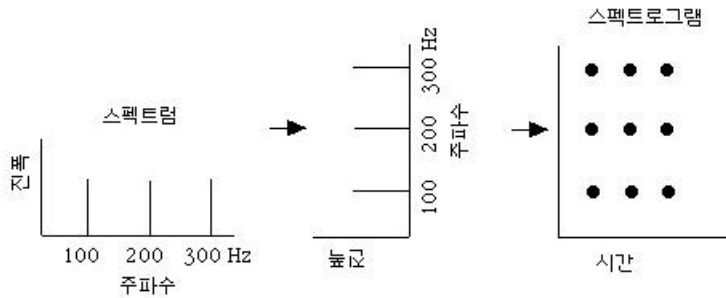
그림 3. 스펙트럼 분석과정을 나타낸 그림



푸리에 분석은 컴퓨터에 저장된 각 시간점에서의 진폭의 전후값을 다양한 간격으로 곱하거나 누적시키는 연산을 통해 규칙성을 찾아낸다. 마지막 단계에서 제일 기본이 되는

성분을 값(기본주파수가 된다)을 구하고 이 값의 정수배(1,2,3...배)에 해당하는 주파수의 진폭값을 구하여 그림 4과 같이 주파수값을 x축에 진폭값을 y축에 나타내면 스펙트럼이 된다.

그림 4. 스펙트럼을 시간축에 농도로 나타내는 과정을 나타낸 그림



이 스펙트럼을 90도 시계반대 방향으로 돌린 뒤 이 정보 가운데 진폭값이 클수록 진하게 나타내고 작을수록 연하게 나타내어 첫 번째 스펙트럼 두 번째 세 번째 각각의 시간점에 표시한 것이 위의 스펙트로그램이다. 실제 음성 스펙트럼은 헤아릴 수 없이 많은 주파수와 진폭이 표시된다. 그림 5는 '아버지'라고 필자가 발음한 파형과 푸리에분석을 한 스펙트로그램이다.

그림 5. 아버지의 파형(위의 그림)과 스펙트로그램(아래 그림)

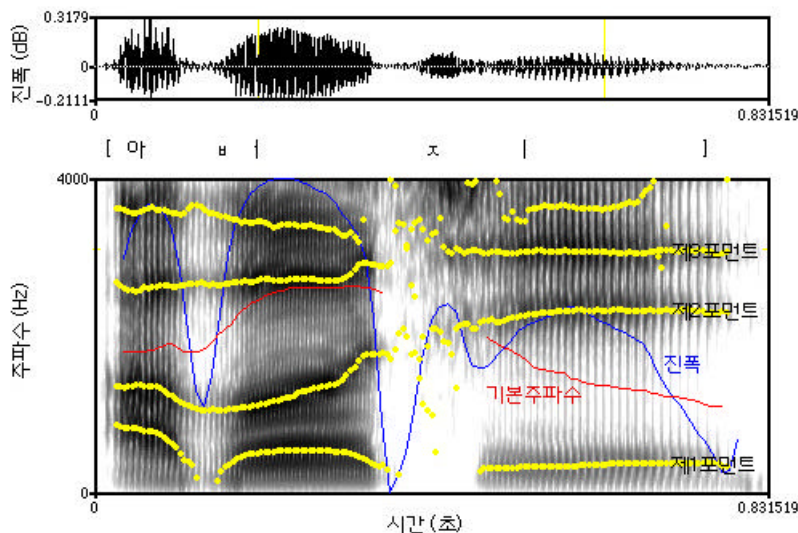


그림 5는 파형을 7밀리초 간격으로 넓은대역에 의한 푸리에변환을 시켜 분석한 스펙트로그램으로 파형에서 볼 수 없는 많은 화자의 개인적인 특성을 보여주고 있다. '아' 부분의 제일 아래에 위치하는 검은 띠 모양을 제1포먼트라고 하고 차례로 제2포먼트, 제3포먼트라고 표현한다. 이 그림에서는 컴퓨터가 스펙트럼의 꼭지점에 해당하는 각각의 포먼트를 노란 점으로 연속해서 인쇄했다. 여러분도 손거울을 들고 아버지하고 천천히 발음하면서 입 모양의 변화를 관찰해보기 바란다. 우선 턱의 움직임을 살펴보면 '아' 발음을 하기 위해 입을 많이 벌렸다가 '버'와 '지'로 가면서 점점 다물며 턱이 위로 올라가는데 스펙트로그램에서 제1

포먼트는 높은 데서 낮은 데로 이동하고 있다. 즉, 제1포먼트는 대체로 턱의 오르내림을 반영한다. 혀의 움직임 관찰해보면 혀가 뒤로 간 ‘아’나 ‘버’ 보다는 앞으로 내밀어 올린 ‘지’에서 제2포먼트가 높다. 즉, 제2포먼트로 혀의 움직임을 추정할 수 있다. 한마디로 스펙트로그램을 통해 우리는 각 시간대별로 입안의 모양이 어떻게 변하고 있는지를 추정할 수 있다. 또한 ‘버’라고 발음할 때 입을 다물었기 때문에 아무 소리도 기록되지 않는 묵음구간과 성대가 열리고 규칙적인 진동을 하기 시작하는 지점들이 나타나는데 이런 정보들도 개인의 특성을 보여준다. 그 위에 그려진 빨간 선들은 성대의 진동을 나타내는 기본주파수값을 나타내준다. 필자의 발음에서는 ‘아’보다는 ‘버’에서 성대의 진동이 더욱 빠르다. 파랑색으로 나타낸 선들은 해당 시간점마다 소리의 강도를 나타내는 것으로 역시 기본주파수가 높은 ‘버’에서 가장 높다. 또한 ‘버’의 발음시간이 ‘아’보다 길고, ‘지’가 가장 짧다. 사람들은 기본주파수와 강도의 높이, 또는 발음을 길게 하여 자신이 주장하는 말을 돋들리게 하는데 이것을 개별적인 자모음의 특징인 분절음에 덧붙여져 있는 초분절음이라고 한다. 음성을 분석하는 과정은 이렇게 사람의 목소리를 컴퓨터로 기록하여 분절음 정보와 초분절음 정보의 특징을 음향적으로 살펴보고 측정값을 구하는 것이다. 그 측정값들을 이용하면 입안의 혀의 움직임과 이에 따른 입안의 공간 모양을 쉽게 예측할 수 있다. 음성학에서는 스펙트럼가운데 아주 강한 진폭을 지닌 포먼트 서너개로 음성의 특징을 요약하여 비교하고 논의하지만, 음성공학에서는 스펙트럼부분을 12개 이상의 부분으로 세분하여 화자확인에 사용하고 있다. 이런 스펙트로그램을 그려준 데이터는 모두 숫자로 저장되어 있는데 각 지점마다 농도의 차이가 매우 미세하게 되어있음을 알 수 있다. 화자확인과정에서는 이렇게 눈으로 확인하는 것에 덧붙여, 정밀한 숫자값으로 통계처리하여 개인마다의 발음 특징을 모두 분석하여 일치여부를 확인한다. 이렇게 목소리와 입안의 모양의 관계를 응용하면 외국어학습의 성취도를 측정하고, 학습자의 발성에 대한 분석결과를 보여줌으로써 언어교정에 활용할 수도 있다. 미국에 이민 온 외국아이들의 발음에서 측정한 값들이 체재기간이 길수록 미국아이들의 발음에서 측정한 값쪽으로 더 가까이 이동하였다는 보고나, 장애자와 같이 음성을 듣지 못하는 사람을 위해 마이크로 발음한 음성을 분석하여 이미 연구된 면적함수관계에 따라 현재의 입안의 모양이 어떤지를 보여 주며 발음 연습을 할 수 있는 학습소프트웨어가 외국에서 이미 판매되고 있다.

그런데 세상에는 많은 사람들이 거의 비슷한 키와 몸무게를 가지고 입안의 구조도 큰 차이가 없다. 쌍둥이의 말소리나 한 가족 구성원들의 말소리를 들어보면 거의 비슷하게 들리고 쉽게 누가 누군지 가리지 못하는 경우가 많다. 독자가운데는 멀리 시골에서 걸려온 친구 아버지나 형의 전화 목소리를 친구의 목소리로 잘못 듣고 농담을 건네다가 혼란 적이 종종 있을 것이다. 컴퓨터로 분석한 목소리의 자료에서는 이런 동일한 목소리를 구분해 줄 수 있을까? 결론은 만일 동일한 발음구조와 발음습관을 가지고 있다면 구분하기 어렵다고 말할 수 있다. 따라서, 비교할 두 사람의 목소리 가운데 단 한마디만 분석하여 동일한 화자인지 확인하려고 하면 일치되거나 동일하게 나타나기 때문에 적어도 20개 이상의 동일하게 발음한 단어를 비교한다든지, 통계적으로 확신할 수 있는 많은 데이터를 비교하는 방법을 사용해야 한다. 또 다른 문제점은 화자마다 상황에 따라 목소리를 다르게 한다는 점이다. 사람은 자신의 귀로 발음된 목소리를 동시에 들으면서 항상 상황에 맞게 조절한다. 우리는 시장바닥이나 파티장에 가서는 매우 큰 목소리로 또박또박하게 발음하여 의사를 전달하려고 한다. 하지만 이런 목소리로 조용한 도서관에서 외치게 되면 주위에서 공부하던 학생들이 모두 쳐다볼 것이다. 동일한 사람이 발음한 같은 단어의 목소리도 발성 환경이 달라지면 성문분석

에서도 다르게 나타날 수밖에 없다. 그림 6은 ‘아버지’란 단어를 조용히 발음한 경우와 다소 크게 발음한 목소리의 포먼트, 진폭, 기본주파수를 함께 그려놓았다. 점선으로 된 부분은 조용한 발음이고 실선으로 된 부분은 그림 5의 특징을 그대로 옮겨온 것이다. 빨간선은 피치값의 변화, 파랑선은 진폭값의 변화, 노랑선으로 표시된 포먼트 값은 작은 목소리 발음이고 분홍색은 다소 큰 목소리의 포먼트값의 변화이다.

그림 6. 조용히 발음한 경우와 다소 크게 발음한 음성의 특징 비교

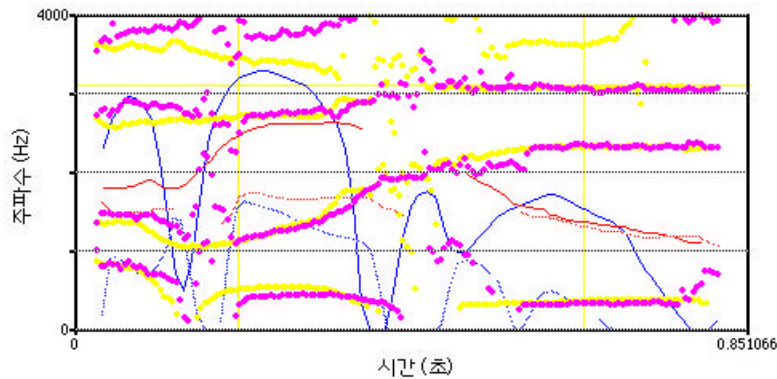
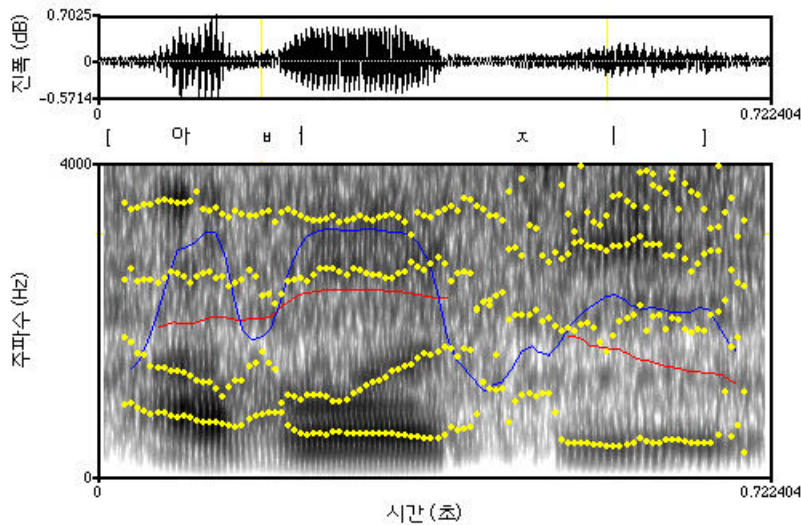


그림 6을 자세히 살펴보면 빨간색으로 표시된 성대의 떨림수의 변화나 파랑색으로 나타난 진폭값은 조용한 목소리는 상대적으로 매우 낮으나, 노랑색과 분홍색으로 나타낸 포먼트값은 거의 변함이 없다. 필자의 입안의 구조가 변하지 않았기 때문에 동일한 사람의 목소리로 판정할 수 있는 것이다. 특히, 앞에서 말한 것처럼 개인마다 자신의 목소리를 듣고 조정하기 때문에 무한정 발음이 달라지지는 않고 일정한 범위에서 변한다. 여기에 덧붙여 모르는 사람에게 말할 때와 친구사이에 말할 때에 따라 발음 속도를 달리할 수 있다. 속도가 달라지면 각 소리마다별로 값의 변화가 일정한 비율로 증가하거나 줄어든다. 녹음기의 특징도 개인의 특성을 뚜렷이 보이는데 어려움이 있다. 값싼 소형의 마이크는 주위의 소리를 최대한 수집하도록 설계되어 있기 때문에 목소리는 물론 주위의 잡음도 함께 받아들여진다. 그림 7은 주위에 라디오의 지지직거리는 소음이 나오는 상태에서 필자가 ‘아버지’라고 발음한 음성을 분석한 것이다.

그림 7. 잡음이 섞인 음성의 파형과 스펙트로그램



여기서 분석된 값들은 앞서 깨끗하게 녹음한 그림과 윤곽은 비슷하지만 잡음 때문에 포먼트 값이 오류가 많이 나타난다. 하지만, 빨간선으로 나타난 기본주파수의 변화나 파랑선으로 나타난 진폭의 변화는 거의 비슷하다. 포먼트 값의 변화에서도 그림 6과 큰 차이가 없다. 최근의 한 국제학회에서는 이런 잡음의 특성을 파악하고 희미하지만 뚜렷하게 나타난 스펙트럼의 특징만을 뽑아 재합성한 뒤 깨끗한 음성으로 들려주는 과정을 소개한 발표가 있었다. 먼지와 진흙에 뒤덮인 고대 유물을 깨끗이 세척하고 복원하면 수천년 전의 찬란한 모습이 생생히 드러나듯이, 잡음이 섞인 변형된 음성도 많은 컴퓨터 처리과정을 거치면 깨끗하고 명료한 소리로 어느 정도 복원할 수 있다. 여러번 테이프 복사를 한다든지 주위 잡음이 너무 심하면 특징을 제대로 분석할 수 없다.

마지막으로 음성 변조에 대해 생각해보자. 가끔씩 TV에 유명한 배우의 목소리를 흉내내는 성대모사에 특별한 재능을 보이는 사람을 본다. 이들은 몸동작과 표정까지도 같이 하려고 애쓴다. 하지만 어딘가 모르지만, 다르게 들리는 것을 느낄 수 있다. 만약 음성의 일부를 잘라내고 비슷한 목소리의 발음을 테이프로 편집했다면 컴퓨터로 자세히 확대하여 음파와 스펙트로그램을 관찰하면 쉽게 그 지점을 찾을 수 있을 것이다. 하지만, 컴퓨터에 저장된 음성자료는 모두 숫자로 되어있는데, 이 숫자파일은 마음대로 그림 편집하듯이 옮겨가거나 다른 곳에 붙일 수 있다. 아주 엄청난 시간이 걸리겠지만, 만일 숫자로 된 자료를 수학적 연산에 의해 시간적으로 자연스럽게 매끈하게 변화하도록 연결시킨다면 어디를 지우고 새로운 발음을 집어넣었는지 쉽게 알아내지 못할 것이다. 또한, 이미 작고한 유명한 역만장자의 목소리의 음성 특징을 모두 분석하여 재합성을 통해 역대의 연금을 고스란히 당신에게 드리기로 했다는 유연장을 작성할 수도 있을 것이다. 공들여 합성한 음성은 역만장자의 음성과 전혀 차이가 없을 것이다.

지금까지 우리는 사람의 음성이 어떻게 발생되고 음성분석은 어떤 과정을 거쳐 이뤄지는지, 또 문제점은 어떤 것이 있는지를 살펴보았다. 목소리에는 여러 가지 변수들이 존재하기 때문에 음성분석은 단순하지 않고 매우 복잡한 복병이 숨어있음을 알 수 있다. 따라서, 쉽게 동일한 목소리로 판정하기가 어렵다는 결론을 내리는 경우가 많다. 또한, 푸리에분석 방식에 의한 스펙트럼 분석은 실험자가 지정한 단절된 구간씩의 신호를 차례로 분석해나가

기 때문에 자연스런 조음의 변화를 나타내주지 못하는 단점이 있다. 이런 분석의 한계를 극복하기 위해 음성파형 자체의 통계적인 특징을 관찰하거나 동적인 변화를 단서로 화자확인을 시도하는 연구도 있다. 우리나라에서도 나이별로 성별로 방언별로 다양한 환경에서 수집된 음성을 정밀하게 분석하여 데이터베이스화하여 개인별로 어떤 특징들이 구별의 단서가 되는지 연구한다면, 성문분석에 의한 화자 확인도 전혀 불가능하지는 않을 것이다. 앞으로 많은 독자들이 사람의 목소리를 연구하는 음성학에 관심을 가지고 세종대왕의 한글창조에 버금가는 세계적인 연구성과를 밝힐 날이 오기를 기대해본다.

끝.