한국어 음성 DB의 구축 과정 및 활용

양 병 곤
동의대학교

A construction and application of Korean phonetic DB

Byunggon Yang
Dongeui University

1. 서론

사람이 의사소통을 하기 위해 내는 음성은 매우 복잡한 과정을 거쳐 생성되며 또한 이 음성은 청자를 항상 의식하고 발화된다. 따라서 이들 청자와 화자의 사이에서 일어나는 음성전달과정을 연구하는 분야는 조음기관인 혀와 입술등의 움직임과 근육운동등을 관찰하고 기술하는 조음음성학, 화자의 입술을 떠나 공기중에서 일어나는 음향학적인 특성을 연구하는 음향음성학, 그리고 이들 음성을 고막을 지나 달팽이관에서 처리되는 과정을 연구하는 청각음성학등이 있다 (양병곤, 1992). 이들 음성학의 구분에서도 알 수 있듯이, 음성을 단순한 발성의 측면에서만 연구한다거나 음향학적인 신호에 매달린다거나 기타 특정한 분야에만 치우칠 경우에는 올바른 결론을 얻기가 힘들다는 것을 알 수 있다. 다시 말해서, 음성학의 모든 분야에 대한 통합적인 자료와 이해가 선행되어야만 이를 응용한 한국어 음성처리 장치를 만들 수 있다.

이러한 음성학의 모든 분야에 대한 이해는 여러 각도에서의 연구가 필수적인데 오늘날 서로 다른 분야에서 많은 연구가 진행되었으나 이들을 통합할 수 있는 기회가 거의 없었고 빈약하고 단편적인 기초자료를 활용하여 종합적인 음성처리에 응용하려는 데서 많은 어려움을 겪고 있다. 또한 서로 다른 실험대상과 화자선정 방식의 차이 때문에 애써 수집한 자료를 통합하여도 별의미가 없거나 아예 통합할 수도 없는 경우가 많았다. 따라서, 본 논문은 누구나 연구할 수 있는 표준적인 한국어 데이터베이스의 구축과정을 전체적으로 설명하고, 이를 보다 집중적으로 분석연구하여 최종적으로는 이들 자료를 통합하여 우리의 음성에 대한 보다 종합적인 이해를 도모하고 다양하게 활용할 수 있는 방안을 제시하고자 한다.

2. DB구축 과정

2.1 발성 목록

이러한 중요한 DB를 구축하기위해서는 먼저 DB에 필요한 발성 목록을 선정해야 한다. 이 발성목록에는 제한된 수의 분절음인 자음과 모음의 개별적인 소리를 중심으로 단음절단위로 이뤄진 자료가 필요하고 또한 억양과 같은 초분절음의 특성을 수집하기 위해서는 어구 이상의 문장의 단위의 크기로 확대한다. 자음과 모음의 결합에 의한 음의 변형과정에 대한 자료수집을 위해서는 한국어의 모든 음운적 변화현상에 대한 종합적인 검토가 필요하다. 예를 들어, 한국어의 「굳이」라는 음성목록에서「ㄱ,ㅜ, ㄷ, ㅣ」에 대한 개별적인 음소값을 측정하고 또한 동시에 이들이 결합되어 「 구지」 라고 구개음화되는 변화 현상에 대한 규칙성을 음향학적으로 점검하는 것이 필요하다. 또 다른 예로는 「먹는다」와 같은 철자식 발음분류 보다는 실제 발음되고 있는 「멍는다」로 변하는 자음접변이나, 고주파성분음의 진폭이 떨어지게 되는 비음화등에 대한 규칙적인 변화유형을 발성목록에 포함시켜 이를 음향학적으로 분석하여 체계적인 분류나 규칙화 알고리즘을 찾는 것이 필요하다.

덧붙여, 일상생활의 대화문도 포함되는 것이 바람직하다. 이런 대화문에서는 주로 각 개별음의 음향적 특징의 허용 범위, 다시 말해서, 분절음의 음향학적인 기본 DB에서 나온 자료가 얼마만큼 변화하면서 실제 발화에 사용되고 있는가를 확인해 볼 수 있다. 다만, 이러한 일상대화에서 나오는 음성은 상당한 공동조음(coarticulation)현상이 두드러지므로 이에 대한 레이블링이나 분석에도 상당한 어려움이 있다. 따라서, 가능하면, 특수한 분야에 대한 집중적인 연구로부터 확대하는 것이 바람직할 것이다. 예를 들어, 전

화교환원이 자주 사용하는 용어나 고객의 일반적인 질문 유형, 호텔예약, 은행 고객관리등 단순하면서도 정해진 순서로 진행되는 대화를 먼저 연구하고 이를 일상대화로 확대하는 것이 실용성이 있을 것이다. 특히, 이들 역동적인 대화에서는 합성이나 인식등에 사용할 억양의 분석은 자연스런 합성이나 인식의 문제점을 조기에 발견하는 데 중요한 단서를 줄 것이다. 덧붙여, 현재 개발되어 있는 한국어 합성기의 음질은 이러한 억양규칙을 넣어 실행시켰을 때 상당히 자연스러운 음성을 만들 수 있을 것이다.

2.2 발성 화자

다음으로는 발성 화자를 설정해야하는데 국어 음성학자의 도움이 필요하다. 인원수는 남, 여의 음성적 차이가 상당히 있으므로 이들의 차이점을 포착하기위해서라도 남, 여 각각 구분하여 DB로 입력하는 것이 좋을 것이다. 인원수는 많을수록 좋지만 실제 통계적으로 평균을 내어 처리했을 때 유의성을 갖는 일정수를 선정한다. 예를 들어, 방언적 차이가 별로 없는 음성에서는 적은 수의 발화자로서도 충분하나, 「에, 애」와 같은 방언차이가 많은 모음이나 전라도나 경상도와 같은 다른 억양형태를 고려하지않고 이를 분석하여 통계적으로 처리했을 때 그 결과는 어느 쪽도 대표하지 못할 것이다. 화자는 KBS나 MBC 등의 공인 아나운서를 활용하는 방법이나, 대학, 고등학교의 학생들의 음성을 입력하는 것도 경비를 절약하면서도 알찬 자료를 수집할 수 있을 것이다. 덧붙여, 이러한 화자의 선정시에는 설문지를 이용하여 지역적 방언을 통제하는 것이 좋다. 그렇지 않으면, 이들 분석결과가 방언 차이로 인한 통계적 극단치의 영향을 많이 받게 될 것이다. 또한, 설문조사는 각 화자에 대한 방언 및 기타 발음상의 특이점을 알 수 있도록 한다.

2.3 녹음

녹음과정은 방송국 방음장치가 갖춰진 실험환경에서 수집할 수 있도록한다. 이때 너무 과도한 입력이 되지않도록 마이크와 입의 거리를 조절하도록 한다. 실제사람의 음성은 주위환경이 시끄러우면 거기에 적응하여 피치나 성분음의 주파수가 상승하는 경향을 보인다. 반면에 조용한 방에서 대화를 나눌 때는 음성을 또박또박 크게 발음하지 않는 경향이 있다.

발음 속도는 느리고 또박또박 발음하는 방식과 보통 대화 속도의 두가지로 나누는 것이 바람직하다. 음성입력은 일반 음성 녹음 테이프 또는 CD음질의 고성능 테이프를 사용하거나 컴퓨터에 바로 입력하는 것이 좋다. 고주파부분의 간섭을 피하기 위해서는 8 kHz이하의 저주파만 통과시키도록 한다. 실제 음성은 0 -4 kHz에서도 거의 의미나 음질에서 차이가 많지 않으므로 이것의 두배인 10 kHz 전후의 샘플링 주파수를 사용한다면 보다 많은 정보를 디스크에 담을 수 있다. 또한, 각종 컴퓨터에서 이 소리를 들을 수 있도록 표준 방식의 텍스트나 이진 파일, 또는 압축파일로 디스켓이나 CD에 약 60분 분량으로 저장하여 각 연구원 및 연구기관에 배포하는 것도 보다 많은 연구자들의 연구를 조장하는 방안이 될 것이다. 예를 들어 IBM의 사운드 블래스터나 Macintosh의 SoundWave, Signalyze 등에서 열어볼 수 있는 파일이면 바로 분석과 합성 실험을 통해 어떤 특정한 분야나 종합적인 자료 검토를 할 수 있게 된다. 또한 Internet이나, Hitel의 음성자료란에 올려줌으로써 보다 많은 학자들이 연구할 수 있게 한다.

3. 분석

이렇게 수집된 자료는 여러 각도에서 많은 학자들이 동시에 연구해볼 만하다. 이들의 연구 결과를 종합하게되면 바로 복잡한 음성의 비밀을 밝혀줄 수 있을 것이다. 수집된 자료에 각 음소와 변화형을 표기해야 하는데 이런 표기 방식은 반드시 통일된 방식을 취해야 하며, 한국어 특유의 발음을 포착할 수 있도록 한다. 표기상의 정밀화 정도는 청각적인 실험을 통해 어느 정도의 표기가 적당한지 결정한 뒤 이뤄져야한다. 이미 출판되어 있는 한국어 발음사전등을 조사하여 실용적인 면을 부각시켜서 조정하는 방안도 필요하다.

분석 내용으로는 우선 자음과 모음으로 나누어 몇가지만 지적해 보기로 한다. 모음의 음향적 특성으로는 성도의 특징인 제 1,2,3 포먼트와 각각의 진폭값이 될 것이다. 특히, 음성합성에 활용하기위해서는 제4,5,6 포먼트도 조사하여 자료화하는 것이 더 자연스럽고 정밀한 합성에 필요하다. 덧붙여, 모음의 구간에서 나타나는 피치나 진폭 윤곽선 (amplitude envelope)에 대한 값도 필요하다. 이들은 자연스런 음성 합성에 상당한 영향을 미치는 파라메터이기 때문이다. 남여의 포먼트 값에도 일정한 비율의 차이를 보여주기 때문에 이를 고려해야한다. 예를 들어, 성도의 총길이가 다르기 때문에 발생하는 차이와 모음 「이」와 같은 음성에서 발생하는 해부구조상의 차이때문에 발생하는 요소등을 염두에 두어야한다. 또한, 통제되지 않는 환경에서 발성이 되었을 때는 이들자료가 주위 환경에 영향을 받아 측정값이 증가하거나 축소되므로 이를 적절히 정규화시키는 작업도 필요하다. 그

방법으로는 음향학적인 수치는 주관적인 청각 기준과는 비선형적인 관계를 갖고 있으므로 mel이나 bark scale로 변환하여 조사하거나, 한국인의 귀에 알맞는 청각적인 잣대를 실험을 통해 확정하여 사용하는 것도 필요하다.

자음의 음향적 특징으로는 주파수 및 진폭 그리고 영교차율에 의한 유무성음 구분, 파열음일 경우에는 유성음 개시의 특징인 막음지속시간(Hold Onset Time)과 성대진동개시시간(Voice Onset Time)을 모두 고려해야 한다. 또한 자음의 연음에 의한 포먼트의 특징인 절단주파수(Cutoff Frequency)와 지속시간도 중요한 음향적 단서이므로 이들에 대한 적절한 측정치를 수집해야한다. 이들 값에 대한 측정치를 모아야 한다. 그러나 이들 값에 대한 음향적인 중요도는 결국 청각실험을 통해서 검증되어야 한다. 예를 들어, 컴퓨터에서 측정한 지속시간의 값을 소숫점이하 두자리 이상 표시한다는 것은 별로 의미가 없다. 인간의 청각체계는 특별한 경우를 제외하고는 4 ms정도의 차이를 동일하다고 여기기 때문이다.

다음으로는 자모음이 결합되어 생기는 공동조음효과에 대해 기본값을 중심으로 이것이 어떻게 변화되어 나가는가를 조사할 필요가 있다. 일종의 시간왜곡방법(Time Warping)을 사용하여 추적하는 것이 바람직하다. 예를 들어, 느린 속도와 빠른 속도의 음성을 수집하여 소리의 시작과 끝을 기준으로 삼아 비교해 보면, 일정한 비율로 변화되고 있음을 알 수 있다. 이러한 모든 음성 분석의 중요성은 결국 합성에 의해 인간의 청각적인 판단을 통해 수정 보완하는 방식으로 진행되어야 할 것이다.

4. 합성

이렇게 분석된 파라메터의 중요도는 청각실험을 통해 검증되어야 한다. 음성 합성 방식에는 여러가지가 있는데 임의로 두가지로 분류해 보면 음절단위로 녹음하여 이를 조합하는 단순한 방식과 음의 특성을 추출하여 이를 재조합하는 포먼트 합성이나 LPC 계수를 활용한 정밀하고 복잡한 방식으로 나눌 수 있다. 단순방식에 의한 방법은 한음절씩 녹음하여 결합하기 때문에 매우 빨리 비용을 들이지 않고 시간안내나 자동전화번호안내 등의 단순 반복적인 응용장치에 활용된다. 그러나 그만큼 많은 정보를 저장해야 하고 또한 속도나 자연성에서 일정 수준이상을 넘지 못하고 만다. 특히 피치를 조정하지 않았을 때는 매우 어색한 느낌을 준다.

이에 반해 Klatt가 제시한 포먼트 합성 방식은 여러가지 파라메터에 대한 기본적인 연구와 시행착오가 있겠지만 음성은 여러가지 원하는 형태로 마음대로 변화시킬 수 있는 장점이 있으므로 한국어 DB만 확보되면 이를 응용한 합성 방식이 더 좋을 것이다. 기존 연구에서도 실제음성과 거의 차이를 느끼지 못할 정도로 개개인의 음성 특성을 살릴 수 있다. 이들 합성에 필요한 파라메터도 가능하면 자동 분석 방식에 의해 추출하고 이를 바로 합성하여 지각 실험을 할 수 있도록 개발하는 것이 필요하다. 특히, 청각실험환경은 Macintosh 의 Hypercard나 IBM계열의 Toolbook과 같은 멀티미디어 저작도구를 활용하여 외부함수등을 이용하여 조립하는 방식을 택하는 것이 더욱더 실험에 몰두할 수 있는 시간을 벌 수 있을 것이다. 가능하면 개발된 함수나 도구를 공개하여 보다 많은 사람들이 이용할 수 있도록 하는 것이 바람직하다.

5. 인식

다음으로는 이렇게 합성한 음성을 HMM이나 Neural Network등을 통해 컴퓨터에 의한 통계적인 인식실험을 실시할 수 있는데 이들 방식을 개선하기 위해서는 인간의 지각실험에서 나온 결과를 활용하여 전처리하는 것도 필요하다. 예를 들어, 음향학적인 주파수인 Hertz를 log나 mel등으로 바꾼다면 남여의 차이와 같은 비언어적 요소들이 상당히 극복될 것이다. 또한, 이들 표준자료를 분석하고 이들의 파라메터를 모두 응용하여 음성을 합성한 뒤 한국인에게 들려 주었을 때 이들의 청각적인 반응은 바로 음성합성에 필요한 파라메터 설정을 하는데 중요한 자료가 될 것이다. 예를 들어, 한국어 단모음의 합성 실험에 따르면, 모음의 중요한 파라메터인 포먼트의 값을 상당히 변화시켜도 동일한 음으로 감지하고 있음을 발견했다(양병곤, 1995). 이들 화자들이 동일하다고 반응하는 포먼트의 범위를 보면 제 1포먼트는 150 Hz, 제2포먼트는 300 Hz, 제3포먼트는 800 Hz정도의 변화에도 동일한 음성으로 들렸다. 이러한 청각실험결과는 바로 음성합성용 파라메터 가운데 하나인 포먼트를 얼마만큼 정밀하게 해야하는가 알 수 있게 해주며 동시에 음성 인식방법에서도 이러한 허용치를 바탕으로 인식의 범위를 정하므로써 실제 사람의 인식과 같은 과정을 프로그램화할 수 있을 것이다.

6. 활용

일단 한국어에 대한 DB가 구축되면 이들을 활용하여 한국어에 대한 음성 음운체계를 확립하는데 기여할 것이다. 기존의 음성학적인 논의나 용어등은 주로 영어나 기존의 연구된 많은 서양언어의 자료를

활용하고 있는데 이들로 부터 추출한 결과는 별로 유용하지 않을 때가 많다. 예를 들어, 음성합성용 파라메터만 하더라도 영어나 서양어에 바탕을 두고 음성을 합성했을 때 이들에 대한 한국인의 청각적인 인상은 어색할 수 밖에 없는 것이다. 따라서, 한국어에 대한 DB 구축은 합성에 가장 적절한 파라메터설정에도 기여할 것이다.

이러한 DB분석을 통한 합성과 인식실험을 바탕으로 한국어의 음성평가 기준도 마련될 수 있다. 다시 말해서, 어떤 비교하고자 하는 음성자료를 수집한 뒤 이것을 분석하여 자음 모음 및 기타 초분절음에 대한 구체적인 음향학적 특성을 추출한뒤 이를 원 음성 DB의 자료와 수치적으로 비교함으로써 적절한 평가를 할 수 있게 된다. 물론 평가기준을 설정하고 이를 비교하고자 할 때는 성별이나 나이차이와 같은 비언어적인 요소를 고려해야 할 것이다. 그러나 이러한 비언어학적인 차이는 음원인 성대의 떨림의 차이나 여파기인 성도의 길이차이에서 오는 경향이 많다. 따라서, 이러한 성도의 길이차이를 극복한 뒤에 비교해야만 공정한 결과를 얻을 수 있을 것이다.

지금까지의 모든 분석과 합성인식에 의한 기계적인 실험과 주관적인 청각실험들을 별개의 과정으로 진행하기보다는 전국적으로 이를 통합 조정하는 기구가 필요하다. 이런 토론은 Internet이나 기타 하이텔 천리안 등과 같은 상용 통신망을 이용하여 자료를 교환하거나 규칙적인 토론회 등을 통해 각 분야의 연구결과를 발표해서 이를 즉각적으로 응용하여 실제 실험 및 수집과정에 활용될 수있도록 한다. 특히 각 분야의 연구내용은 서로 발표하여 공개하는 것이 본인의 연구발전을 위해서 뿐만아니라 한국 음향 음성학의 발전을 위해서도 바람직하다는 사실을 깊이 인식하고 이를 실천해야 할 것이다.

사람의 음성은 매우 정교한 장치에의해 생성되는 예술작품과 같다. 한국어 음성 DB의 개발과 그 특성에 대한 집중적인 연구와 활용이 잘되어서 외국의 기초연구에 대한 비싼 로얄티를 지급하지 않고서도 그 자료를 응용한 여러 가지 음성처리및 인식기기가 개발되기를 기대해 본다.

참고 문헌

양 병곤(1992). 음성학 입문. 부산:진영문화사.
양 병곤(1995). 한국어 단모음의 합성에 의한 청각 실험 연구. '94 동의대 자체학술연구비논문.