

음성 코퍼스 구축에서 분절과 레이블링의 문제

엄용남, 이용주

원광대학교 음성정보기술산업지원센터

Segmentation and Labeling in Creation of Speech Corpus

Yongnam Um and Yong-Ju Lee

SITEC, Wonkwang University

E-mail: umyongnam, yjlee@sitec.or.kr

Abstract

In this paper it is discussed what should be taken into consideration with respect to segmentation and labeling in creation of speech corpus. What levels of annotation and what kind of contents should be included, what kind of acoustic information is checked for in segmentation, etc are discussed.

I. 서론

음성공학에서 이제 다량의 언어자원은 필수적인 것이 되었다. 그리고 음성공학에 사용되는 음성 코퍼스는 음성 파형과 더불어 여러 가지 병행 정보의 코딩이 필요한데, 분절은 음성 파일을 물리적 또는 언어적 단위들에 해당하는 부분으로 접치지 않도록 나누는 것이고 레이블링은 이런 단위들에 물리적 또는 언어적 표지를 할당하는 것이다. 그러나 음성을 경계가 있는 일련의 분절들로 취급한다는 것은 어느 정도는 음성기술의 요구에 의한 일종의 편의상의 추상적인 것이다. 예를 들면 모음과 활음 간의 경계나 모음과 다음 모음과의 경계를 정의하는 것은 매우 어려운 일이고, 또 하나의 자음의 조음 위치에 관한 정보는 보통 자음 그 자체보다 이웃하는 모음들에 들어 있기도 하다. 그러나 또한 많은 분절들이 매우 분명한 음향 경계를 가지고 있는 것 또한 사실이고, 또한 언어공학의 통계적 방법은 문제의 데이터를 다량 포함시키는 방법으로 데이터의 편향성을 피할 수가 있다. 음성 코퍼스의 구축

과정에서 필수적인 부분인 분절과 레이블링 작업을 수행함에 있어서 특별히 유의해야 하는 것들을 검토해보고 분절의 인지에 중요한 것으로 드러난 음향학적인 정보들을 살펴본다.

II. 이용 목적과 설계

표본 수집된 많은 양의 음성을 그대로 공동으로 이용한다는 것은 불가능하기 때문에 음성이 의미 있게 쓰이기 위해서는 코퍼스의 필요한 자료들의 접근을 가능하게 하기 위하여 수집된 음성에 대한 언어적 주석(분절과 표지)이 필수적이다. 따라서 수집된 음성이 음성 코퍼스로서 가치를 지니기 위해서는 수집된 음성이 언어적 주석으로 보강되어야 한다. 음성의 언어적 주석은 기호에 의한 표기를 통해서이다. 표기가 음성이 아니고 일종의 추상적인 기호에 의한 것이지만 음성학·음성공학 등에서 음성을 다룰 때는 이 기호들이 마치 음성 그 자체인 것처럼 다루어진다. 모든 형태의 음성 표기들은 결국 음성을 분석하고 분류한 결과이기 때문이다. 그러나 코퍼스의 목적에 따라서 표본된 음성의 유형이 다를 수 있고 이에 따라 어떠한 내용들이 어떠한 언어 층위에서 분석·분류되어야 하는가가 다를 수가 있다. 분절과 레이블링된 데이터베이스는 여러 가지 인식기들에서 사용되는 은닉 마르코프 모델들을 테스트하고 훈련하는 데 사용될 수 있고 텍스트에서 음성으로 합성하기 위한 규칙을 도출하는 데 필요한 데이터를 제공하기도 한다. 분절과 레이블링된 데

이더베이스는 또한 기본 언어 연구 특히 잘 연구되지 않은 언어들의 경우 기본 언어자원의 역할도 한다. 이런 데이터베이스는 언어적 문맥에 따른 지속시간 변이의 유형들 같은 것들에 관한 보다 세부적인 정보뿐만 아니라 그 언어의 언어음들의 음향 파라미터들에 관한 기본 정보를 제공한다. 따라서 먼저 코퍼스의 목적이 뚜렷이 밝혀져야 한다. 그리고 언어는 계층적 구성을 이루고 있으므로 음성이 어떠한 언어층위들에서 분석·분류되어야 되어야 할 것인가를 결정해야 한다.

III. 언어적 층위

1980년대 음성 및 텍스트 코퍼스 작업은 일반적으로 발화의 음향 파형에 대한 참조 없이 텍스트나 발성을 부호열로 표기하는 것이었고 이런 코퍼스들은 물리적 음향 파형과 부호 표기와의 연결시키지 않았다. 따라서 기계적으로 읽는 것이 불가능했다. 위에서 언급했듯이 이제 음향파형과 관련하여 분절과 레이블링 모두 언어 데이터베이스 작업의 중요한 부분이 되었다. 분절은 음성 파일을 물리적·언어적 단위들에 대응하는 부분으로 겹치지 않도록 나누는 것을 말하고 레이블링은 이렇게 겹치지 않도록 나누어진 단위들에 물리적 언어적 표기를 할당하는 것이다. 여기서 겹치지 않는다는 것은 하나의 언어 층위에 있어서 그러하다는 것이다. 언어 층위에 따라서 물리적·언어적 단위로서 경계를 정할 수 있는 분절들의 유형은 다양하다. 낭독 음성이든 자유발화 음성이든 독백이든 대화이든 층위에 따라서 다른 방법으로 표기될 수가 있다. 우선 실제 음성을 살펴보는 것이 필요한 전사와 이런 전사에서 도출될 수 있는 표기층위들을 구별할 수 있을 것이다. 음성 파형을 전사할 경우 인간의 귀나 컴퓨터에 의해서 이루어질 수 있다. 이런 전사에서 도출될 수 있는 표기들은 실제 음성을 참고할 필요가 없는 형태, 통사, 의미, 화용의 층위와 관련될 수 있을 것이다. (물론 음소전사도 철사전사로부터 자동으로 도출될 수도 있다.)

다음과 같은 표기 층위의 유형을 구별할 수 있을 것이다. 이런 모든 층위의 표기들이 음성 신호와 연결 시간적으로 정렬 될 수 있으나 주로 운율, 물리, 음소, 음성층위와 관련하여 사용될 것이다. 데이터베이스의 사용 목적에 따라 이러한 다양한 유형의 분절들의 층위를 결정해야 한다. 즉 몇 단계의 층위로 음성 파형을 분석하고 분류할 것인가는 데이터베이스의 사용 목적에 따라 결정해야 할 것이다. 일반적으로 음성 코퍼스와 관련하여 철자 표기, 음소 표기, 정밀 표기 세 단계가 중요한 층위가 될 것이다.

(1) 정서법 층위

이 층위에서는 발생된 음성을 철자법을 따라서 옮기는 것이다. 이것은 코퍼스가 그 언어 사용 공동체에 속하지 않는 사람들에게 의해 사용될 때에는 특히 유용하다. 발생자들의 녹음 내용을 전사를 통해서 일차적으로 텍스트로 단순하게 변환시키는 작업이다. 발생된 내용을 정서법(철자법)에 따라 단순히 옮기기만 하면 되므로 스크립트를 읽은 경우는 스크립트 자체가 쓰일 수 있다. 또 이 때는 전문적인 음운 및 음성에 대한 교육을 받지 않는 인력을 사용해서도 전사가 가능하다. 철자 전사는 해당 언어의 표준 철자법의 사용을 의미한다.

축약된 어형으로 자주 나타나는 경우 또는 음절이 줄어드는 경우 그 줄어진 어형을 철자의 표기에 반영한다. 철자표기에 축약형을 표시하기로 하는 기준이 있어야 한다. 축약형의 빈도수 및 음절수의 축소의 정도에 따라서 결정되어야 할 것이다. 이렇게 결정할 경우 사전에 실린 축약형도 있을 수 있고 사전에 실리지 않은 축약형도 있을 수 있다. 사전에 실린 축약형은 사전에 따를 필요가 있다. 따라서 일정하게 하나의 표준 사전을 사용하는 것이 바람직할 것이다. 그리고 축약형들에 대한 일괄적인 목록을 작성하는 것이 필요하다. 한국어의 경우 특별한 다른 이유가 없는 한 한글 맞춤법, 표준어규정, 외래어표기법 등을 따르면 될 것이다. 여기서 문제가 되는 것은 표준어가 아닌 방언의 경우 이 층위에서 표준어를 적고 다음 단계인 음소층에 실제 방언을 옮길 것인가 아니면 제일 층에 직접 방언형을 전사할 것인가의 문제가 있을 수 있는데 실제 방언형을 적는 것이 바람직해 보인다.

(2) 음성형 층위

이 층위에서는 제일 단계 철자표기를 기반으로 해서 화자에 의해서 발화되는 음성을 음소단위로 옮기는데, 단어들을 따로 따로 고립해서 발음할 때의 모습과 비슷하다고 할 수 있다. 주어진 단어를 온전한 분절음의 형태로 옮긴 음성형이다. 이는 각 단어의 발음을 (발음)사전에 나타내듯이 음소에 의해서 표기한 것으로, 말하자면 어떤 단어를 언급할 때 그 단어 발음의 음소형이라고 할 수 있다. 따라서 한국어의 경우 이 층위에서는 한국어의 음소 기호들만으로 필요 충분하다. 온라인 발음사전이 있을 경우 철자로부터 자동으로 도출될 수 있다. 음소이하의 보다 세부적인 내용은 다음 층위에 표기될 수 있다.

여기서는 표기에 어떠한 음소부호를 사용해야 할 것인가가 문제가 된다. 표기에 사용하는 음소 부호를 표기하는 가능한 다양한 방법들이 있을 수 있기 때문이다. 먼저 IPA부호들을 사용할 수도 있으나 어떤 플랫폼

품들은 전 영역의 IPA 부호들을 제시하는 것이 가능하지만 (예: 라텍스 폰트에 의해서 사용되는 부호) 많은 경우 IPA 부호들에 대응하는 알파벳 또는 숫자적 표기 같은 다른 방법들을 사용하지 않을 수 없다. [모든 IPA 부호들에 대한 숫자적 대응 표기는 Esling and Gaylord(1993)에서 볼 수 있다. IPA 부호들의 알파벳 대응 표기는 뉴스그룹 sci.lang에서 볼 수 있다. 주요 유럽어들을 위해서는 SAMPA 표기 시스템이 있다.] ELRA에 따르면 일반 음성학 연구에서 필요한 것처럼 모든 부호들이 모든 언어에 걸쳐서 구별되어야 할 경우가 아니라면 그리고 코퍼스가 단일 언어에 국한되고 표지가 진정한 IPA 부호가 아니라 이에 대응하는 알파벳을 사용해야 한다면 요구되는 표기상의 복잡성을 피할 수 있기 때문에 언어 고유의 특수 문자 세트를 사용하는 것이 바람직하다고 한다.

(3) 정밀 음성 층위

실제 연속적으로 발화 시에 나타나는 연속음으로 환경에 따른 음운의 변화가 표시된다. 미세 음성 전사 층위는 전사자가 녹음을 듣거나 또는 파형과 스펙트로그램을 참조해야하는 최초의 층위이다. 이것은 녹음할 때 화자가 실제로 발성하는 것을 표기하려고 하기 때문이다. 따라서 음성의 모든 부분을 수동으로 살펴야 하기 때문에 필요한 시간과 노력이 증가된다. 부호들의 목록은 해당 언어에서 음소적 지위를 갖지 않는 음들이 포함되어야 하기 때문에 음소목록보다 증가되게 된다. 이 층위에서 무성음화나 유성음화 그리고 여러 가지 이음들이 표기되기도 하고 비음화 순음화 같은 이차적 조음들이 표기되기도 한다. 이 층위의 하나의 분절음은 하위 층위의 하나 이상의 분절음에 대응할 수도 있다.

전사자로 하여금 음성 자체를 점검할 것을 요구하고 이것은 시간과 노력의 면에서 요하는 자원을 증가시키게 된다. 따라서 충분한 이유가 있지 않다면 이 층위의 전사는 시작하지 않는 것이 좋을 수도 있다고 권장하고 있다. 이것은 위에서 언급했듯이 음성코퍼스의 목적에 따라서 결정될 문제이다. 이 층위에서는 분절의 경계를 결정하는 데 잠재적인 문제가 있을 수 있다. 그러나 이 표기 층위가 말해진 것에 대한 정확한 기록으로서 이 층위에서 정확한 전사를 하게 된다면 더 이상 자세한 정보가 필요하지 않을 정도로 충분할 것이다. 이 층위는 하나 이상의 세부적인 층위로 나누어 질 수가 있기 때문에 이 층위를 어떻게 정의하느냐에 따라서 그리고 얼마나 많은 하위 층위를 두느냐에 따라서 정밀도에서 차이가 있게 될 것이다. 따라서 일정한 규정이 필요하게 된다.

음성 코퍼스들은 여러 층위들 중에서 하나 이상을

선택해야 할 것이다. 철자 전사는 큰 규모의 음성 코퍼스들과 단어의 발음에 관한 자세한 것이 중요하지 않는 연구를 위한 코퍼스들에서 사용된다. 대 규모 코퍼스의 경우 적어도 그 코퍼스의 일부는 자세한 전사가 필요할 수도 있겠지만 작업의 규모 때문에 이것이 항상 가능하지가 않는 것이다. 그리고 단어의 발음에 관한 자세한 것이 중요하지 않다면 정확한 전사는 필요하지 않을 것이다.

IV. 분절

음성을 경계가 있는 일련의 분절들로 다루는 것은 음성기술의 요구에 따라 필요하게 되어진 어느 정도의 추상화이다. 예를 들면 모음과 활음간의 경계나 모음과 모음과의 경계를 정의하는 것은 쉬운 일이 아니다. 게다가 하나의 자음의 조음 위치에 관한 정보는 보통 그 자음 그 자체보다 이웃하는 모음들에 들어 있다. 연구들은 또 위치 동화의 경우에도 자주 기저음의 제스처가 남아 있는 것을 보여준다. 따라서 음성 신호를 이산적인 방식으로 음들의 단순한 열로 기술한다는 것은 불가능한 일이겠지만 많은 분절들이 분명한 음향 경계를 가지고 있는 것 또한 사실이어서 음성을 분절하려고 하는 것은 타당한 것이라고 Roach et al. (1990)은 주장한다. 또 음성 공학의 관점에서 보면 음성 신호에 분명한 음향 경계가 존재하지 않을 때는 전혀 분절을 하지 않는 것보다는 상당히 임의적인 지점이라도 선택하는 것이 바람직할 수 있다. 예를 들면 분절된 코퍼스는 은닉 마르코프 모델 기반의 인식기들을 훈련하는데 유용하게 사용될 수 있는데 관련되는 다량의 데이터를 포함시킴으로서 경계 위치에 관한 견해의 편향을 피할 수 있기 때문이다. 분절은 수동 분절, 자동 분절, 반자동 분절이 있을 수 있는데 수동 분절은 전문 전사자가 스펙트로그램과 음성 파형을 참고하여 음성 파일을 수동으로 분절하고 표기하는 과정을 말한다. 수동적인 방법은 보다 정확하다고 믿어지고 있다. 또한 결국 음성이 인간의 인식과 관련되어 사용될 것이므로 인간 전사자에 의한 분절 경계와 표지들은 인식적인 면에 있어서도 타당성을 확보할 수 있는 장점이 있다. 그러나 어떤 형태로든 점검 절차와 함께 전사자간 또는 전사자 개인의 일관성을 보장하기 위한 명시적인 분절 기준을 마련할 필요성이 제기된다. 자동 분절은 프로그램을 통해서 자동적으로 분절 경계를 결정하는 것을 말한다. 이런 프로그램으로는 입력으로 정확한 부호열이 주어지는 은닉 마르코프 모델 기반의 인식기가 쓰일 수 있다. 그러나 훈련 데이터가 충분하지 않을 경우 경계들의 배정이 만족스러울 만큼 정확

하지 않을 수 있다. 이런 경우를 위해서 자동 분절이 수행된 이후 수동적인 점검과 편집을 할 수 있는데 이는 반자동 분절이라고 할 수 있다. 자동 분절된 데이터베이스는 언제나 그 정확성을 판별할 필요가 있을 것이고 정확성의 척도로는 데이터베이스의 일부에 대하여 수동분절과 자동분절 간의 일관성이 될 것이다. 그리고 분절과 표지의 정확성의 척도의 하나는 전사자들간의 일관성일 것이다.

V. 전사와 관련한 음성 인식의 정보

위에서 언급했듯이 자동전사보다는 수동전사가 더 정확하고 자동전사 프로그램의 훈련데이터로 수동전사의 자료가 쓰이고 자동전사의 결과를 수동 점검할 필요가 있다는 면에서 수동전사가 일차적이고 또 수동전사는 인간의 인식에 기반한다는 의미에서 타당성이 인정된다. 이런 관점에서 스펙트로그램과 음성 파형과 관련하여 인간의 음성의 인식 연구에서 밝혀진 인간의 음성 인식의 중요한 정보로 이용되는 음성의 특성들을 살펴본다.

(1) 포먼트 전이

음절과 음절의 연결 그리고 음과 음의 연결에서는 일반적으로 자음과 모음의 연결로 이루어지고 자음에서 모음 그리고 모음에서 자음으로의 전이는 음향신호의 모음 부분에 포먼트 전이로서 나타난다. 제2 포먼트의 전이 그리고 적게는 제3 포먼트의 전이는 구강의 협착의 위치를 알려주는 정보를 제공한다고 알려져 있다. 따라서 폐쇄음, 파찰음, 마찰음 비음 등의 조음 위치를 나타내는 정보를 제공한다. 자음들 중에서도 활음과 유음은 자체의 포먼트 구조를 갖는다. 활음(/j/, /w/)들은 (/i/와 /u/처럼) 협착의 중심에서의 제2 포먼트와 제3 포먼트의 값의 차에 의해서 구별되고 포먼트의 빠른 이동과 안정기간의 지속시간에 의해서 다른 모음들과 구별된다. 접근음들은 동적인 성격 때문에 전이 정보를 이웃하는 모음에 크게 의존한다.

(2) 마찰 소음

마찰음의 좁은 협착에 의해서 생기는 상당히 긴 시간의 비주기파 소음을 특징으로 갖는다. 마찰 소음의 스펙트럼은 치찰음의 조음 위치를 구별하는 데 충분하다. 그러나 낮은 진폭과 보다 확산된 스펙트럼을 가진 비치찰음들의 경우는 조음 위치를 구별하는 데 제2 포먼트의 전이 정보가 필요한 것으로 알려진다. 마찰 소음도 어느 정도는 모음처럼 인접 자음에 관한 정보를 제공하는 전이를 보여줄 수 있다.

(3) 폐쇄음의 개방 파열

폐쇄음의 조음에서는 구강의 완전한 폐쇄와 이에 따른 폐쇄 공간의 기압의 상승과 갑작스런 개방으로 짧은 고진폭의 소음을 생성한다. 개방 파열은 짧은 비주기성 파열음이다. 이들은 폐쇄음의 조음 위치의 인식에 중요함이 알려졌다. 폐쇄음의 개방 파열과 포먼트 전이가 조음위치 정보를 보여주지만 폐쇄음의 개방 파열 스펙트럼과 F2 전이가 상승되면 F2 전이가 더 중요하다. 연구개음을 구별할 때 전이에 가장 의존하게 된다고 알려진다. 주변 소음이 없는 소음이 없는 환경에서는 폐쇄음의 개방 파열도 충실한 정보를 제공해주지만 사소한 소음에도 가려질 수 있다. 파찰음은 폐쇄음과 마찰음 모두와 유사한 점을 지니고 있다. 따라서 개방 파열에 이은 짧은 기간의 마찰이 나타난다. 개방 파열과 짧은 기간의 마찰 모두 위치 정보를 제공해 준다.

(4) 비음

구강 폐쇄음과 마찬가지로 비음들은 구강의 협착이 있어서 인접모음에 포먼트 전이를 남긴다. 그러나 반공명 효과와 500Hz 이하의 낮은 주파수 공명 때문에 상위 포먼트들에 현저한 약화를 보인다. 이러한 모습이 조음위치 정보를 제공해준다. /n/과 /m/의 구별에서 /ŋ/의 구별에서보다 현저한 역할을 한다. 청자들은 신호의 비음의 내면적 부분보다는 포먼트 전이로부터 조음위치를 찾아내는 정보를 찾아낸다. 따라서 F2 전이가 가장 중요한 정보라고 여겨진다.

(5) 조음 방식에 대한 정보

모든 구강의 협착은 신호의 약화를 가져오게 되는데 특히 상위 주파수들에서 더 잘 드러난다. 상대적인 주파수 약화의 정도가 조음 방식에 대한 정보를 제공한다. 모든 주파수에 있어서 갑작스런 약화는 폐쇄음의 존재를 나타낸다. 유성 폐쇄음은 F0의 주파수는 유지된다. 모음과 모음사이의 신호에서 에너지가 없는 정지상태는 폐쇄음의 조음 방식에 대한 정보를 제공한다. 배음 주파수의 완전한 약화와 마찰 소음은 마찰음을 나타낸다. 마찰 소음의 강도가 높으면 높을수록 폐쇄음과 마찰음의 구별이 가능하게 한다. 보다 심하지 않은 진폭의 저하와 콧소리 그리고 비음의 특징인 상위 포먼트들의 현저한 약화와 하위 주파수로의 치우침 등이 비음 정보를 제공한다. 상위 포먼트의 약화, 포먼트 대역폭의 확대, 비음 포먼트의 등장과 같은 모음의 비음화는 앞으로 나타날 비음의 존재를 예견하게 하여 준다. 활음과 유음은 조음기관의 접근 중 포먼트 구조를 유지하고 있다. 포먼트 전이중 급속한 변화를 보이는 폐쇄음 등의 자음들과는 전이의 상대적 점진성으로

구별된다.

(6) 자음에서의 유성/무성

신호에서의 주기성이 유성음에 대한 정보를 나타낸다. 이 이외에도 VOT, 기음의 존재와 진폭, 유성의 시작시의 피치의 교란상태, 지속시간 등이 있다. 마찰음에서의 마찰 소음 중에 주기성의 존재가 유성음을 확인하게 한다. 언어에 따라서는 음절말 등에서 폐쇄음 같은 유성 저해음이 성대의 활성과는 관계가 없을 수도 있는데 이런 경우는 다른 정보에 의존하게 된다. 음절초 폐쇄음의 경우 VOT가 주요 정보가 된다. 유성음의 경우 짧은 또는 마이너스 VOT, 무성음의 경우는 긴 VOT, 기음의 경우는 아주 긴 VOT가 구별 정보이다. 또 음절초 폐쇄음의 경우 개방 파열의 상대적 진폭도 유성음에 대한 정보를 제공한다. 낮은 진폭의 개방 파열은 유성을 높은 진폭은 무성음을 나타낸다. 모음의 지속시간과 스펙트럼의 성격 등은 다음에 오는 폐쇄음이나 마찰음의 유성성을 나타낸다. 모음이 짧고 포먼트의 안정상태의 지속시간의 전이에 대한 비율이 상대적으로 낮을 때는 유성음을 나타낸다. 자음 폐쇄의 협착의 지속시간도 폐쇄음이나 마찰음의 유성상태를 나타내는 정보이다. 보다 긴 경우가 무성음을 나타낸다.

(7) 모음의 질

모음은 자음과는 달리 비교적 열린 구강의 통로를 통해서 생성되기 때문에 모음의 구별은 구강의 공명도에 달려 있다. 모음의 구별은 일반적으로 기본 주파수(F0)와 구강의 처음 세 공명도(F1, F2, F3)의 상대적 거리에 따라서 인식된다. 그러나 포먼트들은 항상 일정 구간을 유지하지는 않고 주위의 자음 때문에 보다 낮은 포먼트도 가질 수 있다. 따라서 포먼트의 전이로부터도 모음에 관한 정보를 얻을 수 있다.

VI. 운율 및 기타

지금까지는 분절음의 레이블링을 다루었으나 음성 데이터는 운율 층위에서도 주석을 할 수가 있다. 운율 주석은 최근에야 음성 및 언어 기술 분야에서 선호되기 시작하였다. 이제 음성 분절들의 합성과 인식에 관하여 상당한 숙련의 단계에 이르자 연구자들은 음성 기술의 운율적인 면에 훨씬 더 많은 작업이 필요하다는 것을 인식하게 되었다. 그러나 운율 현상들에 대한 음향적 정보들이 훨씬 적기 때문에 분절음 표기보다는 훨씬 덜 분명하지만 F0 곡선과 피치 곡선이 중요한 지표가 될 수 있다. 파형이 음성의 현 위치를 보여주

고 보통 F0 곡선과 함께 보여진다. 분절된 단위들은 주어진 연구 프로그램의 기반이 되는 이론적 편향에 달려 있을 것이다. 대체로 단위들의 경계를 표기하는 운율 표지 시스템과 F0 최고점처럼 고립된 운율 사건들을 표기하는 시스템과 근본적인 구별이 이루어질 수 있다.

첫 번째 유형의 방법은 아마 억양구, 음운구, 음운단어, 음보, 음절 같은 Nespor and Vogel(1986)에 의해서 제안된 억양 범주들을 사용하는 것이다. 대안으로 마썸 데이터베이스에서처럼 주요 성조 단위와 부차적 성조 단위의 보다 전통적인 단위들을 표시할 수도 있다. 또는 타이밍에서 초기, 후기, 아주 후기인 상승과 하강, 변화율의 빠름과 느림, 전 크기와 반 크기 같은 화란의 억양 연구에서 사용되는 음운 인식 범주들을 주석할 수도 있다. 이런 유형의 주석은 형태·통사 층위의 주석과 함께 사용되어 지속시간, 휴지 등의 면에서 통사층과 운율층의 관계에 관한 정보를 생성할 수 있을 것이다.

두 번째 유형의 방법은 기본 이론에서는 위에서 언급된 단위들을 참조할 수 있지만 실제로는 그것들을 주석하지 않고 여러 가지 종류의 고저 성조들의 발생을 표시한다. 최근에 완성된 ToBI 전사 시스템은 이런 종류의 시스템으로 여기서는 운율 단위들이 성조 층위에서보다는 휴지 색인 층위에서 주석이 이루어진다.

운율이외에도 자연발화 음성에는 음성과 함께 여러 현상들이 나타나는데 이러한 현상들을 위해서 또 다른 추가적인 표기를 사용할 수 있다. 음성 중에는 개 짖는 소리, 문을 닫는 소리 등 외적 소음과 기침, 웃음, 입다시는 소리 같은 화자가 생성하는 내재적 소음들이 들어갈 수 있다. 이런 것들을 위해서 하나의 독립된 표기 층을 마련할 수 있다. 이 층위는 또 화자의 비유연성들과 휴지사이에 들어간 군말 같은 것들을 표기하는 데 사용될 수 있다. 이런 주석들을 위해서 사용될 표기의 유형은 데이터베이스의 성격에 달려 있게 될 것이다. 이 표기 층위에서 생각할 수 있는 주석 가능한 현상들의 유형에는 다음의 것들을 생각할 수 있다.

1. 낭독 텍스트에서 빠뜨린 것
2. 발성 중 탈락되거나 정정된 사항
3. 불완전한 단어
4. 이해하기 어려운 단어들
5. 머뭇거림과 휴지시간의 군말
6. 침묵, 들리는 숨소리, 웃음, 머뭇거림 등과 같은 현상들의 지속시간
7. 비음성적 음향
8. 끼어 든 동시 음성
9. 이야기 차례

이런 내용의 전사에 관하여도 프로젝트들에서 표기

방식들로 제안된 것들이 있다.

VII. 결론

음성 코퍼스의 구축과정에서 필수적인 부분인 분절과 레이블링 작업을 수행함에 있어서 분절과 레이블링에서 다루어지는 사항을 살펴보았는데 중요한 것은 코퍼스의 목적과 성격에 따라서 필요한 주석의 내용이 결정될 수가 있다는 것이고 주석의 내용이 결정되면 여러 사람들이 이런 작업에 관여할 때 생길 수 있는 일 개인 또는 개인간의 변이와 일관성의 문제를 해결하기 위한 분절 및 레이블링 규약을 정하는 것이 필요하다. 그리고 이러한 규약을 정할 때는 표기해야 할 내용, 따르게 될 표기 방식, 부호 등을 명확히 해야 한다. 이러한 규약들은 다른 기획들에서 제안된 방식들을 검토할 필요가 있다. 분절에 있어서는 음향의 인지에 관여하는 중요한 정보가 어디에 있는가를 파악하는데 중점을 두어야 하지만 분절의 어려움은 항상 존재한다. 그렇지만 가능한 한 임의의 분절을 하는 것이 분절을 하지 않는 것보다는 낫다는 사실을 염두에 둘 필요가 있다. 그것은 정보의 잉여성 때문이기도 하다.

참고문헌

- [1] R. Daniloff, G. Schuckers, and L. Feth, *The Physiology of Speech and Hearing*, New Jersey, Prentic Hall, Inc, Englewood Cliffs, 1980.
- [2] J. Esling and H. Gaylord, "Computer codes for phonetic symbols," *Journal of International Phonetic Association* 23(2), 83-97, 1993.
- [3] D. Gibbon, R. Moore, and R Winski, *Handbook of Standards and Resources for Spoken Language Systems*, New York, Mouton de Gruyter, 1997.
- [4] D. Gibbon, I. Mertinology, and R. Moore (eds.), *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Production Evaluation*, Dordrecht, Kluwer Academic Publishers, 2000.
- [5] D. Kewley-Port, D. B. Pisoni, and M. Studdert-Kennedy, "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants." *Journal of the Acoustical Society of America*, 73(5), 1779-1793, 1983.
- [6] K. M. Kurowski, and S. E. Blumstein, "Acoustic properties for place of articulation in nasals,"

Journal of the Acoustical Society of America, 1917-1927, 1987.

- [7] A. Malécot, "Acoustic cues for nasal consonants: An experimental study involving tape-splicing techniques," *Language*, 32, 274-284. 1956.
- [8] M. Nespore and I. Vogel, *Prosodic Phonology*, Dordrecht, Foris, 1986.
- [9] P. Roach, H. Roach, A. Dew and P. Rowlands, "Phonetic analysis and the automatic segmentation and labelling of speech sounds," *Journal of International Phonetic Association* 20(1), 15-21, 1993.
- [10] S. D. Soli, "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," *Journal of the Acoustical Society of America*, 70, 976-984, 1981.
- [11] K. N. Stevens, *Acoustic Phonetics*, Cambridge, MA, MIT Press. 1998.
- [12] R. Wright, "Perceptual cues in contrast maintenance," In E. Hume and K. Johnson (eds.), *The Role of Speech Perception Phenomena in Phonology*, New York, Academic Press, 2001.
- [13] R. Wright, S. Frisch,, and D. B. Pisoni, "Speech Perception," In J. G. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 20, pp. 175-195, 1999.