대화체 음성 및 운율 DB

이호영
부산수산대학교

Spontaneous Speech and Prosody DB

Ho-Young Lee
National Fisheries University of Pusan

요 약

자연스런 대화체 발화를 합성해 낼 수 있는 음성합성기를 개발하고, 무한대 어휘의 대화체 발화를 인식할 수 있는 음성인식기를 개발하기 위해서는 정교하게 제작된 방대한 양의 대화체 음성 및 운율 DB를 필수적으로 갖춰야 한다. 이 논문에서는 대화체 음성 자료의 수집 방법과 대화체 음성 및 운율 DB 제작 방법에 관해 자세하게 논의한다.

I. 서론

자연스런 대화체 음성을 합성해 낼 수 있는 음성합성기의 개발에는 정밀한 음성 및 운율 데이타가 필요한데, 만족스런 음성 및 운율 데이타의 확보는 정교하게 제작된 대화체 음성 및 운율 DB가 있어야 가능해진다.

대어휘, 혹은 무한대 어휘를 인식할 수 있는 음성인식기의 개발에도 대화체 음성 및 운율 DB가 필수적이다. 개발하는 음성인식기의 인식 단위를 음소(혹은 변이음)로 설정하건, 다이폰이나 트라이폰으로 설정하건 인식 단위별로 정확하게 분절(segmentation)되고 표기(labeling)된 다량의 대화체 음성 및 운율 DB가 있어야 인식 시스템의 충분한 학습이 가능해진다.

정교하게 제작된 음성 및 운율 DB의 구축 없이 음성합성기와 음성인식기를 개발하는 것은 모래 위에 집을 짓는 것과 같은 일임에도 불구하고 국내에서는 음성 및 운율 DB의 구축 작업이 매우 부진하게 이루어져 왔다.

최근 들어 국내의 대학, 연구소, 기업체 들에서 음성 및 운율 DB의 구축에 많은 관심을 기울이고 있다. 그러나 국내에서는 아직 PBW(Phonetically Balanced Word List), 숫자음 DB, 전화음성 DB, 문장음성 DB 등 낭독체 음성 DB만이 구축되고 있을 뿐 대화체 음성 및 운율 DB의 구축은 매우 부진하게 이루어지고 있다[1][2][3][4].

미국, 일본, 영국 등 여러 선진국에서는 이미 많은 양의 대화체 음성 DB를 구축하고, 이를 음성합성기와 음성인식기의 개발에 이용해 오고 있다. 따라서 선진국의 음성 공학 기술을 따라잡기 위해서는 우리도 하루 바삐 대화체 음성 및 운율 DB의 구축 작업에 착수해야 할 것이다.

이 논문에서 필자는 대화체 음성 자료의 수집 방법과 대화체 음성 및 운율 DB 제작 방법에 관해 논의하도록 하겠다.

II. 대화체 음성 자료의 수집 방법

2.1. 자연스런 대화의 녹음

가장 이상적인 대화체 음성 자료는 사람들이 자연스럽게 대화하는 것을 녹음한 것일 것이다. 그러나 이러한 자료는 수집하는데 많은 어려움이 따른다. 대부분의 사람들이 녹음기 앞에서 긴장해서 부자연스럽게 말하기 때문이다. 제보자가 눈치 못채게 녹음할 때에는 녹음기와 마이크를 숨겨야 하므로 고품질의 녹음 자료를 얻을 수 없게 된다. 따라서 자연스런 대화를 녹음하려면 제보자들에게 녹음 사실을 알리고 긴장을 풀고 대화할 수 있도록 유도해야 한다. 제보자들끼리 대화가 원활하게 진행되지 않을 경우에는 녹음자가 제보자와 대화를 하고 제보자의 말을 녹음 자료로 사용하면 될 것이다.

녹음은 DAT와 같은 고성능 녹음기를 사용해야 하며, 녹음 장소로는 고품질의 음성 자료를 만들 때에는 방음 시설이 갖춰진 녹음실이 적당하고, 소음이 들어간 음성 자료를 만들 때에는 사무실이 적당한데, 큰 소음이 없는 야외도 가능할 것이다.

2.2. 방송 대담의 녹음

자연스런 대화의 녹음은 자연성이 떨어질 가능성이 있을 뿐만 아니라 다양한 내용의 대화를 이끌어 내는 데에도 어려움이 따른다. 이러한 단점을 어느 정도 극복할 수 있는 방법은 방송의 대담을 현장에서 녹음하는 것이다. 그러나 이 방법의 경우에는 방송국 당국, 담당 PD, 출연자 등의 사전 동의를 얻어야 하고, 필요한 경우 수수료를 지불해야 한다.

2.3. 테스크(task)와 관련된 대화의 녹음

자연스런 대화나 방송 대담을 녹음한 자료는 음성합성기의 개발을 위한 음성 및 운율 데이타의 수집에 유용하게 사용될 수 있지만 음성인식기의 개발을 위한 시스템 학습 자료로는 부적절할 수도 있다. 아직은 무한대 어휘 대화체 음성인식기를 개발하는 기술이 확보되어 있지 않기 때문이다.

단기간에 시연할 수 있는 음성인식기를 개발하려면 녹음할 대화의 주제를 크게 한정해야 한다. 음성인식기의 테스크(task)와 관련된 대화 주제로는 호텔이나 열차 예약, 국제 학회 참가 등록, 무역 상담, 관광 안내 등이 있다.

2.4. 주요 방언의 녹음

DB 구축을 위한 녹음 자료를 제작할 때에는 제보자의 방언에 각별한 신경을 기울여야 한다. 제보자들의 방언이 통일하지 않을 경우에는 음성합성기와 음성인식기의 개발에 많은 문제점들이 발생하기 때문이다. 우선적으로 확보해야 할 녹음 자료의 제보자는 표준말 사용자여야 한다.

표준말 이외의 방언을 합성하고 인식하는 음성합성기와 인식기를 개발하려면 주요 방언의 대화체 음성 및 운율 DB를 구축해야 한다. 방언 자료의 제보자로는 교육을 받지 못한 나이 많은 제보자보다는 교육받은 젊은 제보자가 더 적당하다.

III. 음성 단위의 분절 및 표기

3.1. 음소 단위의 분절 및 표기

음소는 대화체 음성 및 운율 DB의 가장 기본적인 분절(segmentation) 단위이다. 각 음소는 음성 환경에 따라 다른 음가의 변이음(allophone)으로 실현되므로 분절된 음소가 어느 변이음으로 실현되는지 음성기호로 표기(labeling)해 주어야 한다. 변이음의 표기는 국제음성문자(IPA)를 이용하는 것이 가장 바람직하다.

음소 단위의 분절과 변이음 표기를 위해서는 음소 단위의 분절 방법과 기준을 마련하고[5], 음성합성기와 음성인식기를 개발하는 데 이용될 수 있는 주요 변이음의 목록을 작성하고[5][6], 변이음들의 음향적 특징을 파악하고[5][7], 각 변이음을 어떤 기호로 표기할 것인가를 결정해야 한다[6][8].

다음은 대화체 음성 DB의 제작에 필요한 한국어 주요 변이음들의 목록이다.

(1) /ㅂ,ㅍ,ㅃ/
 [bʲ, pʰʲ, p͈ʲ]: i, j 앞
 [bʷ, pʰʷ, p͈ʷ]: o, u, ø, w 앞
 [bᶣ, pʰᶣ, p͈ᶣ]: y, ɥ 앞
 [b, pʰ, p]: 그 밖의 다른 모음 앞
 [p̚]: 어말이나 장순 파열음 앞
 [p]: 다른 조음 자리의 장애음 앞
 [b]: 같은 말토막 안의 유성음 사이, 수의적
 [β]: 같은 말토막 안의 모음 사이, 수의적

(2) /ㄷ,ㅌ,ㄸ/
 [dʲ, tʰʲ, t͈ʲ]: i, j 앞
 [dʷ, tʰʷ, t͈ʷ]: o, u, ø, w 앞
 [dᶣ, tʰᶣ, t͈ᶣ]: y, ɥ 앞
 [d, tʰ, t]: 그 밖의 다른 모음 앞
 [t̚]: 어말이나 양순 파열음 앞
 [t]: 다른 조음 자리의 장애음 앞
 [d]: 같은 말토막 안의 유성음 사이, 수의적
 [d]: 같은 말토막 안의 모음 사이, 수의적

(3) /ㄱ,ㅋ,ㄲ/
 [gʲ, kʰʲ, k͈ʲ]: i, j 앞
 [gʷ, kʰʷ, k͈ʷ]: o, u, ø, w 앞
 [gᶣ, kʰᶣ, k͈ᶣ]: y, ɥ 앞
 [g, kʰ, k]: 그 밖의 다른 모음 앞
 [k̚]: 어말이나 양순 파열음 앞
 [k]: 다른 조음 자리의 장애음 앞
 [g]: 같은 말토막 안의 유성음 사이, 수의적
 [ɣ]: 같은 말토막 안의 모음 사이, 수의적

(4) /ㅈ,ㅊ,ㅉ/
 [ʤ, ʧʰ, ʧ]: o, u, y, ø, w, ɥ 앞
 [ʥ, ʨʰ, ʨ]: 그 밖의 다른 모음 앞
 [ʥ]: 같은 말토막 안의 유성음 사이, 수의적
 [z]: 같은 말토막 안의 모음 사이, 수의적

(5) /ㅅ,ㅆ/
 [ɕ, ɕ͈]: i, j 앞
 [s, s͈]: o, u, ø, w 앞
 [ʃ, ʃ͈]: y, ɥ 앞
 [s, s͈]: 그 밖의 다른 모음 앞

(6) /ㅎ/
 [ç]: i, j 앞
 [x]: ɯ 앞
 [ɸ]: o, u, ø, w 앞
 [ɸʲ]: y, ɥ 앞
 [h]: 그 밖의 다른 모음 앞
 [ɦ]: 유성음 사이에서 수의적으로

(7) /ㄹ/
 1) [ɾ]: 모음과 모음 사이, 모음과 반모음 사이
  모음과 /ㅎ/ 사이, 외래어의 어두
  [ɾʲ]: i, j 앞
  [ɾʷ]: o, u, ø, w 앞
  [ɾᶣ]: y, ɥ 앞
  [ɾ̥]: h 앞에서 h가 약하게 발음될 때
  [ɾ]: 그 밖의 다른 모음 앞
 2) [l]: 어말이나 자음 앞, /ㄹ/ 뒤
  /ㅎ/ 앞에서 수의적으로, 외래어의 어두
  [lʲ]/[ʎ]: i, j 앞
  [lʷ]: o, u, ø, w 앞
  [lᶣ]: y, ɥ 앞
  [l]: 그 밖의 다른 모음 앞

(8) /ㅁ,ㄴ,ㅇ/
 [mʲ, nʲ/ɲ, ŋʲ]: i, j 앞
 [mʷ, nʷ, ŋʷ]: o, u, ø, w 앞
 [mᶣ, nᶣ, ŋᶣ]: y, ɥ 앞
 [m, n, ŋ]: 그 밖의 다른 모음 앞
 [m̥, n̥, ŋ̥]: h 앞에서 h가 약하게 발음될 때

(9) /j/
 [ɥ]: u, ʊ 앞
 [j]: 그 밖의 다른 모음 앞
 [ç]/[çʷ]: pʰ, tʰ, kʰ, ʨʰ, s, s͈, h 뒤

(10) /w/
 [ɥ]: i 앞
 [w]: 그 밖의 다른 모음 앞
 [ʍ]/(ʍʲ): pʰ, tʰ, kʰ, ʨʰ, s, s͈, h 뒤

(11) /ɰ/
 [ɰ]: i 앞에만 나타남. 앞에 자음이 오면 탈락함.

(12) /ㅓ/
 [ə:]: 장모음으로 발음될 때
 [ʌ]: 단모음으로 발음될 때

(13) /ㅚ/
 [ø]: 어두의 초성 자음과 받침 사이에서, 어중에서
 [we]: 받침 없는 어두의 첫 음절에서, 어말에서

(14) /ㅟ/
 [y]: 어두의 초성 자음과 받침 사이에서, 어중의 자음으로 시작하는 음절에서
 [wi]: 받침 없는 어두의 첫 음절에서, /ㅟ/로 시작하는 어중의 음절에서, 어말에서

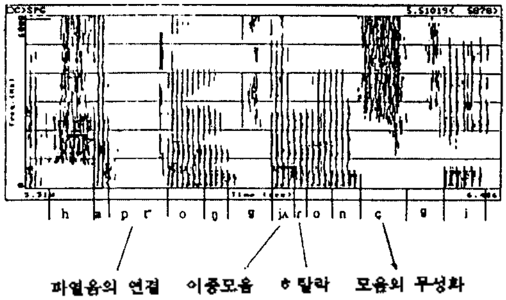
(15) /ㅢ/
 [ɰi]: 어두에서, 둘째 음절 이하에서 수의적으로
 [i]: 자음 뒤에서, 둘째 음절 이하에서
 [ɯ]: 뒤에 /ㅣ/가 올 때
 [e]: 조사 /의/

(16) /ㅣ/: [i], /ㅔ/: [e], /ㅐ/: [ɛ], /ㅏ/: [a]
 /ㅗ/: [o], /ㅜ/: [u], /ㅡ/: [ɯ]

대화체 발화를 음소 단위로 분절할 때는 다음과 같은 여러가지 문제점들이 발생한다.

첫째로, 고모음 /i, u, y, ɯ/는 유기음 /ㅍ,ㅌ,ㅋ,ㅊ/과 마찰음 /ㅅ,ㅆ,ㅎ/ 다음에서 완전히, 혹은 부분적으로 무성음화되는데, 완전히 무성음화될 때는 자음만 분절하고 표기해야 한다[9][7].

(17) 합동 결혼식이 [hapt͈oŋ gjʌɾonɕgi]



둘째로, 이중모음은 반모음의 시작 부분에서부터 모음의 안정 구간까지 지속적인 전이(transition)를 보이므로 반모음과 모음 사이의 경계를 잡을 수 없다. 이 때에는 이중모음은 하나의 음성 단위로 처리하거나 편의상 전이의 중간 부분에 경계를 부과해야 한다.((17) 참조)

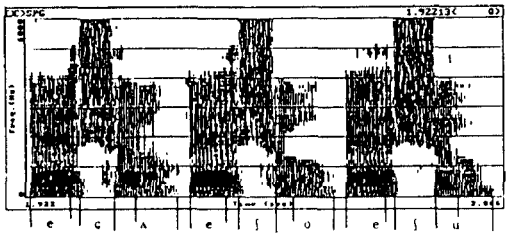
셋째로, 두 개의 파열음이 이어날 때에는 스펙트로그램상에 음소 경계가 나타나지 않는데, 이 때에도 두 파열음을 하나의 단위처럼 표기하던가, 편의상 중간 부분에 경계를 부과해야 한다.((17) 참조)

넷째로, 성문 마찰음 /ㅎ/은 모음 사이에서 약화되어 유성 마찰음 [ɦ]로 발음되거나 탈락하기도 하는데, 탈락할 경우에는 앞 자음만 분절하고 표기해야 한다[7].((17) 참조)

다섯째로, 반모음 /j, w/는 마찰음 /ㅅ,ㅆ,ㅎ/, 파찰음/ㅈ,ㅊ,ㅉ/, 유기 파열음 /ㅍ,ㅌ,ㅋ/ 뒤에서 마찰음이나 기(aspiration)와 융합되어 탈락하게 되는데, 이 경우에는 마찰음만을 분절하고 표기해 주어야 한다[9][7].

(18) ㄱ. 에햐[eça], 에혀[eçʌ], 에효[eʍʰo]

ㄴ. 에셔[eʃʌ], 에쇼[eʃo], 에슈[eʃu]



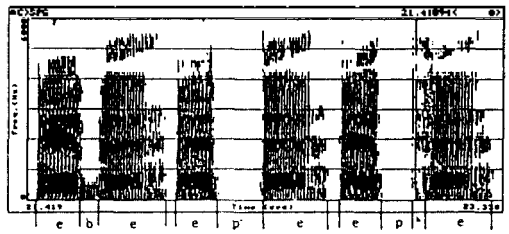
여섯째로, 대화체 발화의 스펙트로그램에서는 유음 /ㄹ/과 비음 /ㅁ,ㄴ,ㅇ/의 경계를 찾기가 어려워지는데, 이 때에는 유음과 비음이 나타나는 부분을 확대하여 음파와 비교하면서 분절해야 한다.

3.2. 음향 event 단위의 분절 및 표기

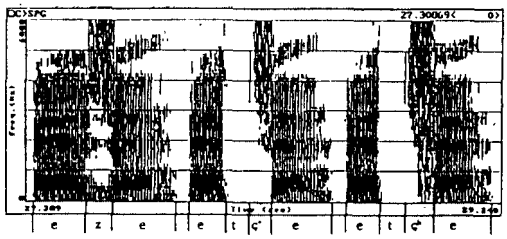
포먼트 합성 방식의 음성합성기의 개발에 필요한 음성 데이타의 수집을 위해서는 음향 event 단위로 분절되고 표기된 음성 DB를 확보해야 한다. PSOLA 방식의 음성합성기의 개발을 위해 다이폰 목록을 작성할 때에도 음향 event 단위로 분절되고 표기된 음성 DB가 되어 있어야 다이폰 단위의 분절이 수월해진다. 뿐만 아니라 음성인식기의 개발을 위한 대화체 발화의 트라이폰 단위의 분절을 위해서도 음향 event 단위의 분절과 표기가 필요하다.

자음의 경우 유기 파열음 /ㅍ,ㅌ,ㅋ/는 막을 지속 단계와 개방 후 마찰 단계(기의 생성 단계)를 나누어 분절하고 표기할 필요가 있고, 파찰음 /ㅈ,ㅊ,ㅉ/은 막음 지속 단계와 개방후 마찰음 생성 단계를 나누어 분절하고 표기할 필요가 있다[7].

(19) ㄱ. 에베[ebe], 에뻬[ep'e], 에페[ep^{h}e]



ㄴ. 에제[eze], 에쩨[et'e], 에체[et^{h}e]



모음의 경우에는 모음의 전이 구간(transition)과 안정 구간(steady part)을 나누어서 표기해야 하는데, 경우에 따라서는 안정구간이 나타나지 않기도 한다.((18) 참조)

3.3. 다이폰과 트라이폰 단위의 분절 및 표기

국내 여러 기관에서 개발하는 PSOLA 방식의 음성합성기는 다이폰을 단위로 하고 있기 때문에 대화체 발화를 다이폰 단위로 분절하고 표기할 필요가 있다. 뿐만 아니라 트라이폰을 단위로 하는 음성인식기를 개발할 때에도 대화체 발화를 트라이폰 단위로 분절하고 표기한 DB가 있어야 시스템을 충분히 학습시킬 수 있다.

3.5. 간투사, 반복 발화, 교정 발화의 경계 표기

대화체 발화에서는 말의 속도가 빠르기 때문에 음소의 탈락과 융합이 많이 나타날 뿐만 아니라 간투사[10], 반복 발화, 교정 발화(speech repair) 등이 많이 나타난다[4]. 음성인식기가 자연스런 대화체 발화를 인식하게 하려면 간투사, 반복 발화, 교정 발화 등을 찾아서 제거시키는 알고리즘을 만들어야 한다. 따라서 대화체 음성 DB에는 간투사, 반복 발화, 교정 발화 등의 경계를 반드시 표기해 주어야 한다.

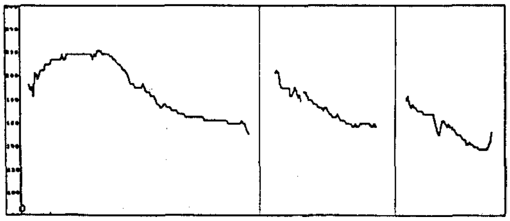
Ⅳ. 운율 단위의 분절 및 억양 패턴 표기

4.1. 말토막(리듬구, 억양구) 단위의 분절

문장이 여러 어절로 이루어져 있는 때에는 문장 안에서 한번 이상 끊기를 한다. 끊기의 단위를 '말토막(rhythm unit)'이라 하는데, 말토막은 하나의 강세음절과 영개 이상의 비강세 음절(들)로 이루어진 리듬구일 뿐만 아니라 '말토막억양(phrasal tone)'이라는 억양 단위가 부과되는 억양구이다[11]. 대화체 음성 및 운율 DB에는 말토막 경계를 표기해야 하는데, 말토막 경계는 '/'로 표기한다.

다음의 그림에서 보듯이 말토막과 말토막 사이에는 억양 곡선의 끊김이 생기며, 잠재적 휴지(tentative pause)에 의해 앞 말토막의 끝음절이 길게 발음된다[12][13].

(20) 영만이 엄마가 그려는 그림
[^j^ŋmani ^mmaga / `gɯrj^non / `gɯrim]



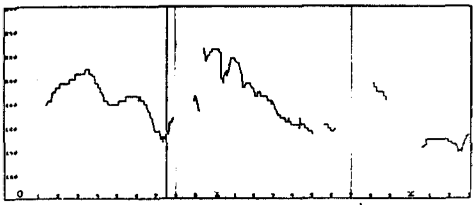
4.2. 말마디(억양절) 단위의 분절

하나의 문장은 하나의 숨단위(breath group)로 발음될 수도 있지만 둘 이상의 숨단위로 발음될 수도 있다. 이 때 하나의 숨단위를 '말마디(intonation group)'라 한다[11]. 말마디는 하나 이상의 말토막억양이 긴밀하게 연결된 억양 단위이기 때문에 '억양절'이라고도 부를 수 있다. 대화체 음성 및 운율 DB에는 말마디의 경계를 표기해야 하는데, 말마디 경계는 '//'로 표기한다.

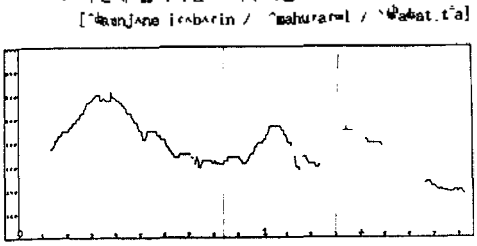
말마디의 끝음절에는 '핵억양(nuclear tone)'이라는 중요한 억양패턴이 얹히는데[11], 핵억양의 존재로 말마디의 경계를 파악할 수 있다. 핵억양 외에도 말마디 경계에는 휴지(pause)에 의한 에너지 공백(silence)이 나타나며, 핵억양이 얹히는 음절은 길게 발음되는 경향이 있다.

다음의 그림은 중의적인 문장인 '작년에 잃어버린 마후라를 찾았어'의 의미가 말마디 경계의 부과 유무에 의해 구별되는 것을 보여 준다. (21ㄱ)은 '잃어버린 마후라를 작년에 찾았다'는 의미를 전달하고, (21ㄴ)은 '작년에 잃어버린 마후라를 올해 찾았다'는 의미를 전달한다[13].

(21) ㄱ. 작년에 // 잃어버린 마후라를 / 찾았다.
[`ʨaŋnj^ne // `iɾ^b^ɾin mahuraɾɯl / `ʨʰaʔat.tʼa]



ㄴ. 작년에 잃어버린 / 마후라를 / 찾았다.
[^ʨaŋnj^ne iɾ^b^ɾin / ^mahuraɾɯl / `ʨʰaʔat.tʼa]



4.3. 말토막억양의 표기

한국어에는 네 개의 말토막억양—수평조, 오름조, 오르내림조, 내림조—이 있는데, 말토막 억양은 말토막의 첫 음소 왼쪽에 표기한다[11]. 수평조는 '¯'로 표기하고, 오름조, 오르내림조, 내림조는 각각 '´', '^', '`'로 표기한다.((21) 참조)

4.4. 핵억양의 표기

한국어에는 아홉 개의 핵억양—낮은수평조, 가운데수평조, 높은수평조, 높내림조, 낮내림조, 온오름조, 낮오름조, 내리오름조, 오르내림조—가 있는데, 핵억양은 말마디 마지막 음절의 첫 음소 왼쪽에 표기한다[11]. 낮은수평조, 가운데수평조, 높은수평조는 각각 '_', '-', '¯'로 표기하고, 높내림조와 낮내림조는 '`'와 '.'로 표기한다. 그리고 온오름조와 낮오름조는 '/'와 ','로 표기하며, 내리오름조와 오르내림조는 'ˇ'와 'ˆ'로 표기한다.((21) 참조)

V. 결론

이상에서 대화체 음성 자료를 수집하는 방법에 대해 논의한 다음 대화체 음성 및 운율 DB를 제작하는 방법에 관해서 논의했다.

대화체 발화를 합성하고 인식하는 음성합성기와 인식기를 개발하려면 대화체 음성 및 운율 DB를 가능한 한 많이 구축해야 할 뿐만 아니라 최대한 정밀하게 제작해야 한다. 이를 위해서는 충분한 수의 전문 인력을 양성해야 하고, 충분한 예산을 확보해야 한다. 그리고 인력과 예산, 그리고 시간을 절약하기 위해서는 DB의 구축 초기 단계에서부터 치밀하게 기획해야 한다.

참 고 문 헌

[1] 이용주 외, "ETRI의 음성 및 텍스트 데이타베이스의 구축 현황", 제1회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, 1993
[2] 최인정 외, "자동통역용 한국어 음성 데이타베이스", 제11회 음성통신 및 신호처리 워크샵 논문집, 1994
[3] 최승호 외, "전화음성데이타 수집에 관한 연구", 제11회 음성통신 및 신호처리 워크샵 논문집, 1994
[4] 이용주 외, "자유발화음성 및 텍스트코퍼스 구축에 관한 검토", 제11회 음성통신 및 신호처리 워크샵 논문집, 1994
[5] 이현복 외, 한국어의 운율분석 및 음운의 분절표기에 관한 연구, 한국전자통신연구소 위탁과제 최종보고서, 1993
[6] 이호영, "한국어 자음 변이음들의 조음적 특성", 어문교육 제2집, 부산수산대학교 어학연구소, 1993
[7] 이호영 외, "동시조음에 의한 변이음들의 음향적 특성", 한글 제220호, 한글학회, 1993
[8] Lee, H.B., "Illustrations of the IPA: Korean", Journal of the IPA 23, 1993
[9] 이호영, "한국어의 변이음 규칙과 변이음의 결정요인들", 말소리 21~24호, 대한음성학회, 1992
[10] 장태엽 외, "자연발화상에 나타난 단음절 단일 간투사의 길이 특성 분석", 제11회 음성통신 및 신호처리 워크샵 논문집, 1994
[11] 이호영, "한국어의 억양체계", 언어학 제13호, 한국언어학회, 1991
[12] 이호영, "서울말과 경상도 방언의 운율 유형론", 언어학 제15호, 한국언어학회, 1993
[13] Jun, S.A., The Phonetics and Phonology of Korean Prosody, Ph.D. Dissertation, The Ohio State University, 1993