공동이용을 위한 음성DB의 구축 및 음성 자료 수집을 위한 Workbench의 구현*

김 봉완, 이 용주(원광대 컴퓨터공학과)

<차 례>

<Abstract>

Construction of Korean Speech DB for Common Use and Implementation of Workbench for Spoken Language Data Acquisition

Bong-wan Kim et al.

This study discusses Korean speech database that has been designed and constructed for common use, especially focusing on designing a list of words or sentences that covers various phonological environments. As the results, PBW(Phonetically Balanced words) and PBS(Phonetically Balanced Sentences) was selected from balanced text corpus using maximum entropy method. And, implemented workbench for spoken language data acquisition is presented in this paper. The workbench consists of grapheme to phoneme converter, utterance list selection module, speech data editing module, multi-layer labelling module, and phoneme context search module.

* 본 연구는 과기처의 지원에 의해 이루어진 것임

1. 서 론

음성은 인간끼리의 정보 전달 매체로서 옛날부터 가장 많이 이용되어 왔다. 이러한 손쉬운 통신 수단인 음성을 인간과 기계와의 대화 수단으로 이용하고자 하는 시도가 오래 전부터 있어 왔다. 이를 위해서는 그 기반이 되는 요소기술로써, 음성 인식 및 합성으로 대표되는 음성처리기술과 언어이해 및 기계번역으로 대표되는 언어처리기술에 대한 연구가 필요하다. 이러한 음성 및 언어처리기술의 연구를 위해 가장 먼저 확보되어야 할 것이 음성, 언어 및 각종 사전 데이터베이스이다.

음성연구가 진보되어감에 따라 처리 가능한 데이터의 수는 많아져 가고, 따라서 준비해야 할 데이터의 양도 대폭적으로 증가되었다. 최근에는 음성인식의 경우 통계적 수법의 발달에 따라 대량의 음성 데이터가 시스템의 학습에 필요하게 되었으며, 합성의 경우에도 대형의 음성 데이터로부터 임의길이의 음성부분을 골라내어 접속함으로써 좋은 합성품질을 얻고 있다. 또한 음성정보처리 시스템의 연구개발을 위해서는 분석, 합성, 인식의 각종 알고리즘을 적절하게 비교 평가할 필요가 있지만 이를 위한 방법으로는 현재까지는 공동음성 데이터를 이용하여 알고리즘을 수행하고 그 결과를 비교하는 방법 이외에는 알려져 있지 않다. 이러한 목적으로 이용하는 음성 데이터를 일반적으로 음성 데이터베이스, 음성 코퍼스 또는 음성 사전이라고 부른다(이하 음성DB로 표기.)

선진 각국에서는 자국어 음성DB에 대한 체계적인 구축이 음성 및 언어처리기술 확보를 위한 가장 기본적인 연구환경임을 깊이 인식하고 이에 대한 체계적인 확보가 공공연구기관을 중심으로 활발히 추진되고 있다[1]. 그러나 국내의 경우 지금까지 음성인식 및 합성연구에 필요한 음성 데이터를 각 연구자가 부분적으로 제작하여 자체 개발용으로 사용하고 있는 경우가 대부분이다[2].

다른 연구 목적간에 두루 쓰일 수 있도록 공동이용을 목적으로 음성DB를 구축하기 위해서는 가급적 다양한 종류로 대량의 것이 바람직하나 제한된 자원에서 가능한 적은 양으로 목적을 달성하기 위해서는 대상이 되는 발성목록을 어떻게 작성할 것인가 하는 점이 중요한 문제가 된다[3,4,5]. 이러한 공동 음성 데이터베이스의 구축을 위해서는 발생 가능한 모든 음운환경을 포함하며, 특정 테스크에 집중되지 않는 발성 대상 단어나 문장의 설계가 필요하다.

본 논문에서는 이와 같은 목적으로 신문, 소설, 기타 구어자료로부터 수집된 100만여 어절의 텍스트 코퍼스에서 음운 균형 발성목록(Phonetically Balanced Corpus)을 추출하고 이를 발성목록으로 음성DB를 구축한 결과와 구축된 음성DB의 특성을

제시하며 본 연구실에서 공동 음성 데이터베이스의 구축 및 활용을 위해 개발한 PC용 Workbench에 대하여 기술한다. Workbench는 음성 언어 자료의 확보를 위한 텍스트 처리 모듈들과 음성 데이터의 처리를 위한 신호처리 모듈들로 구성되어 있다. Workbench에 포함된 모듈로는 텍스트를 자동 읽기 변환하는 철자 음운 변환기, 발성목록 선정 모듈, 끝점 검출기를 이용한 음성 데이터 편집 모듈, 다단계 레이블링 시스템, 텍스트에서 원하는 음운 환경을 포함하고 있는 문자열을 다양한 조건으로 검색할 수 있는 음운 환경 검색기를 포함하고 있다.

2. 공동 이용을 위한 음성DB의 구축

2.1 발성목록의 설계

2.1.1 음운 균형 발성목록(Phonetically Balanced Corpus)

가. 음운 균형 발성목록의 정의

음성인식 및 합성 시스템의 개발 등 음성 연구에 있어서 다종 다양한 음운환경을 포함한 음성DB의 구축은 중요한 과제의 하나로 많은 시간과 노력이 필요하다. 개별적 음성DB 구축에 따른 중복투자를 줄이고 분석, 합성, 인식의 각종 알고리즘을 적절히 비교 평가하기 위해서도 공통의 음성DB는 필수적이다. 이러한 목적으로 사용 될 음성DB는 적은 발성목록으로 실제 한국어의 발성에 나타나는 음운환경을 가능한 많이 포함하며, 특정 테스크에 집중되지 않는 것이 바람직하다. 이러한 조건을 충족시키는 발성목록으로 일본을 비롯한 선진국에서는 음운 균형 발성목록이라는 단어 또는 문장의 집합을 모아왔다[6,7,8,9].

음운 균형 발성목록이란 발생 가능한 모든 종류의 음운환경을 나타내는 음소열을 포함하되 음소열간 고른 확률분포를 갖는 최소단어들의 집합을 말한다. 음소열간 고른 확률분포를 갖는 상태는 음소열을 확률사상으로 했을 때 엔트로피가 최대인 상태이다[10]. 발성목록에 나타난 총 음운환경의 종류를 N, 음운환경의 출현확률을 $P_n, (n=1, \ldots, N)$라고 할 때, 발성목록의 음운 엔트로피 S는 다음 식(1)에 의해 구할 수 있다.

$$S = -\sum_{n=1}^{N} P_n \log_2 P_n \qquad (1)$$

이때 엔트로피 S는 음운환경의 출현빈도가 모두 동일 할 때 최대치 $\log_2 N$이 된다.

또한 음성DB를 위한 발성목록은 그 양이 많아지게 되면 발성자료를 수집하는 데 어려움이 따르게 된다. 따라서 음운 균형 발성목록이 음성 데이터 수집을 위한 발성목록으로 사용되기 위해서는 엔트로피를 최대화하면서도, 최소단어들의 집합에 발생 가능한 모든 음운환경들을 포함할 수 있도록 알고리즘을 구성하여야 한다.

나. 음운 균형 발성목록 추출을 위한 음운환경의 정의

실제 발화시 음성의 음향적 특징은 전후 음소의 영향을 받아서 변화하는 경우가 많다. 음운환경을 나타내는 음소열은 많을수록 좋다고 할 수 있으나 음소열의 증가에 따라 다루어야 하는 음운환경의 수와 계산량이 급격히 증가하게 된다. 본 연구에서는 한국어의 발성에 나타날 수 있는 모든 2음소열을 음운 균형 발성목록 추출의 기본 단위로 하고, 특히 3음소열로 고려해야 할 유성음화 및 자음약화, 탄설음화의 경우를 유형별로 클러스터링 하여 음운 환경으로 정의하였다.

(1) 2음소열

본 논문에서 사용하는 대상음소로는 자음 19개, 모음 21개로 구성하고 종성자음은 모두 대표자음으로 모아 다루었다. 따라서 종성자음으로 올 수 있는 음소로는 /ㄱ/, /ㄴ/, /ㄷ/, /ㄹ/, /ㅁ/, /ㅂ/, /ㅇ/의 7 가지이다. 모음의 경우 반모음은 후속모음과 합하여 하나의 음소로 취급하였다.

대상 2음소열은 위와 같이 분류된 음소의 쌍으로 구성하였으며, 실제 발성할 때 특히 자음의 경우 같은 음소라도 어두, 어말의 위치에서 음향적 특성이 서로 다를 가능성이 있기 때문에 어두, 어말의 공백소도 하나의 음소로 취급하여 공백소와 해당음소가 쌍을 이루는 형태를 2음소열의 종류에 포함하였다(즉, Bㄱ 은 ㄱ 이 어두에 초성으로 오는 경우, ㄱB는 ㄱ 이 어미에 종성으로 오는 경우를 말한다.) 다음 표 1에 음운환경을 구성하는 음소들을 나타내었다.

<표 1> 대상 음소

자 음	ㄱ ,ㄲ ,ㄴ ,ㄷ ,ㄸ ,ㄹ ,ㅁ ,ㅂ ,ㅃ ,ㅅ ,ㅆ ,ㅇ ,ㅈ ,ㅉ ,ㅊ ,ㅋ ,ㅌ ,ㅍ ,ㅎ
모 음	ㅏ, ㅐ, ㅑ, ㅒ, ㅓ, ㅔ, ㅕ, ㅖ, ㅗ, ㅘ, ㅙ, ㅚ, ㅛ, ㅜ, ㅝ, ㅞ, ㅟ, ㅠ, ㅡ, ㅢ, ㅣ
공백소	B

이와 같이 정의된 2음소열에서 반영할 수 있는 변이음 특성은 다음과 같다[11].

(가) 고모음의 무성화

(나) 목젖 소리되기
(다) 구개음화
(라) 원순음화
(마) 비음화
(바) 미파화
(사) 기타 2음소열 간 발생할 수 있는 변이음 특성

(2) 3음소열
(가) 유성음화 및 자음약화

자음 /ㅂ, ㄷ, ㄱ, ㅈ, ㅎ/는 유성음(모음, 비음 /ㅁ, ㄴ, ㅇ/, 유음 /ㄹ/) 사이에서 유성음화 되거나 약화될 수 있다. 따라서 이를 3음소열로 설정하여야 하나 그 양이 너무 많으므로(유성음 25종 * 자음 5종 * 유성음 25종 = 3125종) 유성음을 고모음과 저모음, 유음, 비음의 4종류로 클러스터링 하여 처리함으로써 단위의 수를 80종(유성음 4종 * 자음 5종 * 유성음 4종 = 80종)으로 줄였다.

(나) 유음의 탄설음화

유음 /ㄹ/은 모음과 모음 사이에서 탄설음화 된다. 이러한 경우 역시 모음을 전체 모음으로 클러스터링 하여 1종으로 취급하였다. 또한 종성 유음 /ㄹ/은 뒤에 초성 /ㅎ/가 탈락 되면서 다음 음절의 초성으로 발성되어 탄설음화 될 가능성이 있다. 이러한 경우를 반영하기 위해 /ㄹ/ 뒤에 /ㅎ/이 왔을 경우도 탄설음화의 경우에 포함하였다.

2.1.2 음운 균형 발성목록 추출을 위한 모집단

가. 발성목록 추출을 위한 모집단

본 논문에서는 단어 레벨의 음운균형 발성목록(Phonetically Balanced Words)을 추출하기 위한 모집단으로 신문, 소설, 수필 등의 자료로부터 수집한 120만여 어절의 텍스트 코퍼스(Text Corpus)에서 가급적 많은 음운환경을 포함시키기 위하여 3음절 이상의 고빈도 어절만을 선정하여 단어 레벨 발성목록 추출의 대상으로 하였다. 그리고 다음과 같은 어절은 음성DB를 위한 발성목록으로 적절하지 않다고 판단되어 모집단에서 제외하여 최종적으로 고빈도 5,000어절을 선정하였다.

(가) "클린턴", "이재필" 등 인명이 포함된 어절
(나) "루슬란"과 같이 외국어나 의미가 난해한 어절

(다) "지섭은", "채린은" 등과 같이 소설 등의 자료에서 포함되거나, 자료의 성격상 고빈도어절로 포함된 전문용어 등의 어절

(라) "그라믄", "빌어묵을" 등의 방언

(마) 그 밖의 비어, 속어 등 발성 상 적절하지 않다고 판단되는 어절

문장 레벨의 음운균형 발성목록(Phonetically Balanced Sentences)을 추출하기 위한 모집단으로는 SERI/KAIST/울산대학교에서 구축한 100만여 어절의 장르별 균형 텍스트 코퍼스를 대상으로 하였다. 신문의 제목 및 특수 기호나 외국어가 직접 사용된 문장은 대상에서 제외하였으며, 또한 문장이 너무 길 경우 발성자가 발성하기 곤란하므로 문장의 길이가 10~20어절인 것만을 모집단의 대상으로 하여 최종적으로 10,000문장을 선정하였다.

나. 철자-음운 표기 변환

이상에서 선정된 모집단은 글자 음운 변환기에 의해 소리나는 대로의 형태로 자동 변환하고 잘못 변환된 어절은 직접 수정을 통하여 올바르게 표기하였다. 글자 음운 변환기는 한글을 소리나는 대로 바꿔주는 읽기 규칙을 구현한 것이다[12]. 읽기 규칙의 기본 원칙은 교육부에서 고시한 표준어 규정의 표준 발음 규칙을 따르고 있다. 읽기 규칙은 크게 모음의 발음, 장음 처리, 음운 변동 규칙 및 음의 첨가의 4가지로 나눌 수 있으며, 본 논문에서는 장음 처리를 제외한 3가지의 규칙을 적용하고, 두 가지 이상으로 적용 가능한 경우는 보다 일반적인 경우를 적용하였다. 그리고 규칙으로 구현할 수 없는 경우는 예외 발음 사전에 추가하여 처리하도록 하였다.

2.1.3 음운 균형 발성목록 추출 알고리즘

모집단에서 발생한 음운환경 n의 출현빈도를 N_{Pn}, 후보세트에서의 출현빈도를 N_n, 모집단에서 발생한 음운환경의 총 종류의 수를 T_P, 후보세트에서의 음운환경의 총 종류의 수를 T라고 한다면 본 논문에서 사용한 발성목록 추출 알고리즘은 다음과 같다.

가. 초기 음운 균형 발성목록 후보세트의 구성

단계 1. 모집단의 모든 어절들을 조사하여 $N_{Pn}=1$인 음운환경을 갖는 모든 어절을 모두 후보세트에 포함시키면서 모집단에서 제거한다.

나. 추가 과정

단계 2. 모집단의 나머지 어절들을 모두 조사하여, 각 어절에 포함된 음운환경 중 $N_n=0$인 음운환경을 가장 많이 갖는 어절 또는 문장을 임시 선택한다.

단계 3. 선택된 대상이 복수일 경우, 하나씩 후보세트에 추가하여 후보세트의 엔트로피를 계산하였을 때 엔트로피가 최대화되는 대상을 후보세트에 포함시키면서 모집단에서 제거한다.

단계 4. 단계 2~단계 3의 과정을 T가 T_P와 같아질 때까지 반복한다.

단계 5. 후보세트에 포함되지 않은 모집단의 나머지 어절 또는 문장들을 모두 한번에 하나씩 후보세트에 추가해서 후보세트의 엔트로피를 계산해 보고, 추가되었을 때 엔트로피를 최대화시켰던 대상을 후보세트에 추가하고 모집단에서 삭제한다.

단계 6. 단계 5의 과정을 후보세트의 엔트로피를 증가시키는 어절이 더 이상 없을 때까지 반복한다.

다. 삭제과정

단계 7. 후보세트를 구성하고 있는 어절들 중, $N_n \geq 2$인 음운환경으로만 이루어진 어절 또는 문장들을 모두 임시 선택한다.

단계 8. 선택한 대상들을 모두 한번에 하나씩 후보세트에서 삭제해서 후보세트의 엔트로피를 계산해 보고, 삭제되었을 경우 후보세트의 엔트로피를 최대화시켰던 대상을 후보세트에서 삭제한다.

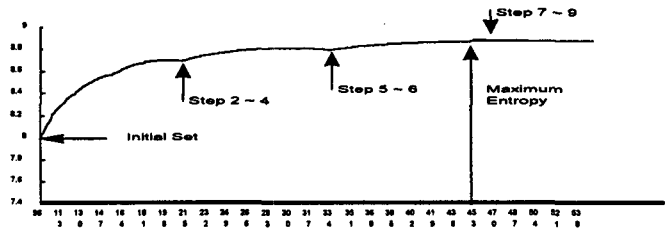
단계 9. 단계 7~단계 8의 과정을 $N_n \geq 2$인 음운환경으로만 이루어진 대상이 더 이상 존재하지 않거나, 존재하더라도 삭제했을 경우 후보세트의 엔트로피를 증가시키는 어절이 더 이상 없을 때까지 반복한다.

2.1.4 추출된 음운 균형 발성목록

가. 단어 레벨 음운 균형 발성목록(Phonetically Balanced Words)

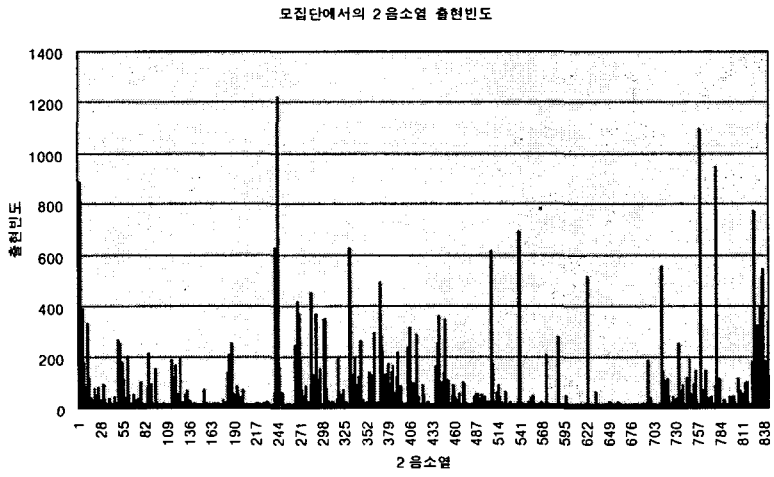
이상과 같은 알고리즘에 의하여 본 논문에서 PBW를 추출하는 과정에서 초기 후보세트로 95개의 어절이 추출되었으며 단계 2~단계 3의 과정을 통하여 238어절, 단계 5~단계 6의 과정을 통하여 167어절이 추출되었다. 이렇게 해서 추출된 총 500어절 중 단계 7~단계 9의 삭제과정을 통하여 총 48개의 어절이 삭제되어 최종적으로 452어절의 PBW를 추출하였다. 다음 그림에 PBW를 추출하는 과정에서의 엔트

로피 변화를 나타내었다.

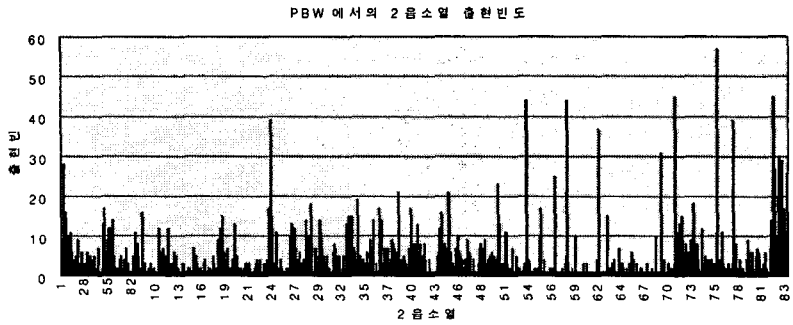


<그림 1> PBW추출 과정의 엔트로피 변화

모집단 고빈도 5,000어절과 PBW set에는 모두 842종류의 음운환경을 포함하고 있으며, 출현한 음운환경의 총 빈도 수는 모집단에서 42,043, PBW set에서 3,504이다. 모집단에는 VC, CV, VV, CC가 각각 40.78%, 46.13%, 3.12%, 9.97%를 보이고 있으나 PBW에서는 37.50%, 43.81%, 9.54%, 9.15%의 분포를 이루고 있다. 모집단에서의 평균 음절수는 3.357, 평균 음소 수는 7408음소로 나타났고, PBW에서의 평균 음절수는 3.398음절, 평균 음소 수는 6752음소로 나타났다. 또한 추출을 위한 모집단에서의 엔트로피는 7.98993에서 추출된 PBW의 엔트로피는 8.88397로 음운의 균형이 모집단에 비해 좋아졌음을 알 수 있다. 그림 2와 그림 3에 모집단과 PBW에서의 2음소열 출현빈도를 나타내었다. 그림에서도 PBW에서의 2음소열 출현빈도가 고른 분포를 보이고 있음을 알 수 있다.



<그림 2> 모집단에서의 2음소열 출현빈도

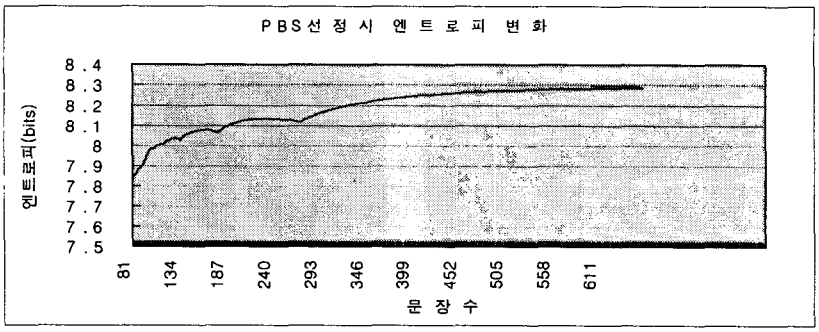


<그림 3> PBW에서의 2음소열 출현빈도

나. 문장 레벨 음운 균형 발성목록(Phonetically Balanced Sentences)

앞에서 기술한 알고리즘에 의거하여 본 연구에서 PBS를 추출하는 과정에서 초기 후보 세트로 81문장이 추출되었으며, 단계 2~단계 4의 과정을 통하여 190문장, 단계 5~단계 6의 과정을 통해서는 390문장이 추출되었다. 이렇게 해서 추출된 총 661문장 중 단계 7~단계 9의 삭제과정을 통하여 총 59문장이 삭제되어 602문장의 PBS를 추출하였다.

추출된 602문장 중 발성 목록으로 부적절 하다고 판단되는 20문장을 삭제하고 단계 2~단계 9의 과정을 다시 수행하여 최종적으로 PBS 589문장을 추출하였다. <그림 4>에 602문장을 추출하는 과정의 엔트로피 변화를 나타내었다.



<그림 4> PBS 추출 과정의 엔트로피 변화

모집단 10,000문장과 PBS에는 모두 서로 다른 1,091종의 음운 환경을 포함하고 있으며, 출현한 음운환경의 총 빈도 수는 모집단에서 692,454, PBS에서 40,129이다. 또한 모집단에서의 음운 엔트로피는 7.786559인데 반해 8.295044로 음운 환경의 균형이 모집단에 비해 좋아졌다. 추출된 PBS의 음절수 평균은 37.83, 어절 수 평균은

12.96으로 평균 1 어절은 2.92음절을 갖고 있는 것으로 나타났다.

2.1.5 기타 발성목록

무제한 어휘의 인식기 및 합성기를 위한 음성DB의 발성목록으로 테스크에 종속적이지 않으면서, 한국어의 음운환경을 충분히 고려한 PBW외에도, 음성 연구자들이 꾸준히 요구하고 있는 것이 숫자음 음성이다. 관련연구에 의하면 필요로 하는 음성 데이터에 대한 설문 등에서 숫자음 음성은 우선 순위가 높게 나타나고 있다[5]. 따라서 공동이용을 위한 음성DB의 기본적인 내용으로 포함되는 것이 바람직하다.

가. 숫자음

숫자음의 발성 목록으로는 우리말에서 흔히 쓰이는 단독숫자 37종과 지시어 4종, 그리고 숫자간 결합이 모두 포함되도록 구성한 4연 숫자음 35종을 포함하여 구성하였다.

나. 단문

한국어의 자연스런 합성과 인식률 향상을 위해서는 한국어의 운율(Prosody)에 대한 연구가 이루어져야 한다. 이를 위해서는 한국어의 운율에 대한 정보를 추출하고 연구할 수 있는 운율음성DB가 구축되어야 한다. 그러나 한국어의 특징을 잘 나타낼 수 있는 정형화된 운율 문장은 아직 국내에 설계된 바가 없다. 운율추출을 위한 발성목록의 설계는 문장수준의 음성DB 구축 시 고려해야 할 사항이지만, 단어수준 음성DB에서도 운율 추출을 위한 기본적인 데이터를 제공하기 위하여 단문을 포함하였다. 포함된 단문은 국제음성학회 한국어 발음예문인 "바람과 햇님"을 발성목록으로 하였다.

다. 고빈도 2,000어절

고빈도 2,000어절은 PBW를 추출하기 위한 모집단인 고빈도 5,000어절 중 음소중복을 최소화하고, PBW 452어절이 포함되도록 고빈도 2,000어절을 선정하였다.

2.2 음성DB의 구축

음성DB를 구축하기 위해서는 음성시료를 수집하고 이를 편집하여 공통으로 사용할 수 있는 포멧으로 저장하여야 한다. 이러한 과정을 다음에 기술하였다.

2.2.1 음성 데이터의 녹음

작성된 발성목록을 대상으로 표준어를 사용하는 서울지역 일반인 및 아나운서 72

명분의 음성을 수집하였다. 음성은 방음 부스에서 Senheizer HMD 224X Headset 마이크를 사용하여 녹음하였으며, 발성된 데이터는 디지털 오디오 테이프에 저장하였다. 16kHz로 샘플링하고 16Bits로 양자화 하였다. 각 데이터베이스 종류별 발성횟수를 표 2에 나타내었다.

<표 2> 종류별 발성횟수

발성목록	발성 횟수
단독숫자 및 지시어, 단독 숫자	일반인 70명 X 4회, 어나운서 2명 X 2회
단문	일반인 70명 X 2회, 어나운서 2명 X 2회
PBW	일반인 70명 X 2회
고빈도 2,000어절	어나운서 2명 X 2회
PBS	일반인 20명 - 공통셋 50문장, 일반셋 539문장

2.2.2 음성 데이터의 편집

A/D하여 컴퓨터의 하드디스크에 저장된 음성 데이터는 그 양이 많으므로 각기 필요 없는 부분은 제외하고 사용할 수 있는 부분만으로 다시 편집을 하여야 한다. 그 과정은 우선 양자화 된 음성 데이터를 대상으로 자동 끝점 추출 알고리즘을 사용한 음성 데이터 편집 시스템 구현하고 이를 이용하여 필요한 음성 데이터를 찾아내고, 불량 데이터의 제거, 수정 데이터의 삽입, 음성 데이터의 앞뒤에 일정한 길이의 무음 구간의 확보 등 전반적인 사항을 수정한다.(사용된 편집 시스템은 본 논문의 음성 언어 자료 확보를 위한 Workbench 부분을 참조할 것.)

이와 같은 과정을 거친 각각의 데이터파일을 디렉토리 구조로 저장하여 CD-ROM 형태로 제작되었을 때 효율적으로 검색하고 활용할 수 있도록 하였다.

2.2.3 음성 데이터 파일의 구조

음성 데이터를 보관하는 형태로는 음성 파형으로 저장하는 방법과 분석처리를 한 후 파라미터의 형으로 저장하는 방법이 있으며, 후자의 경우에 저장하기 위한 정보량의 압축 및 이용 시에 계산량 절감 등의 장점이 있으나 이용할 분석법에 대한 선택의 어려움과 특정 분석법에 의한 이용상의 제약이 있어 배포를 목적으로 하는 경우에는 음성 파형의 형태로 저장하는 것이 바람직하다[4].

음성 데이터를 저장하고 있는 각각의 파일들은 음성 데이터에 대한 정보를 유지하기 위해 헤더를 갖는다. 본 논문에서 구성한 헤더 구조는 미국의 DARPA-TIMIT Speech Database의 헤더 구조를 참조하여 상호 호환성을 유지하도록 구성하였다

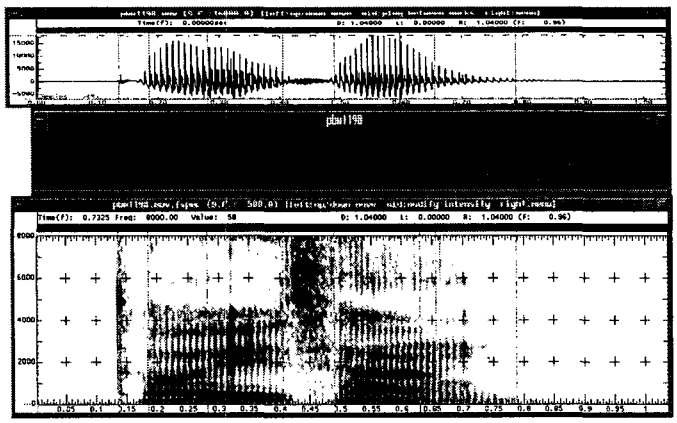
[13]. 헤더의 크기는 1024바이트로 고정되어 있으며, 헤더 첫 부분의 고정영역을 제외하고는 필드의 추가, 삭제, 주석의 첨가 등이 자유롭게 이루어질 수 있도록 되어있다.

2.2.4 음성 데이터 파일의 레이블링

레이블링된 음성 데이터베이스는 음성 인식 및 합성기의 훈련 및 평가, 성능 개선을 위한 중요한 자료로서 활용된다. 그러나 음성 데이터베이스를 레이블링하는 작업은 대부분 수작업으로 이루어지고 있으며 또한 그 기준마저 통일되지 않고, 각 연구자의 필요에 따라 작성하여 사용하고 있으므로, 기준 및 결과가 공동 이용을 목적으로 사용하기에는 부적합하거나 객관적 검증이 이루어지지 못하고 있는 실정이다.

이에 본 연구에서는 공통적으로 이용할 수 있는 레이블링된 음성 데이터베이스를 구축하기 위한 한국어 레이블링 기준안 초안[14]을 작성하여 PBW 60명분을 레이블하였다.

이 초안은 향후 여러 분야에서 공통으로 활용할 수 있는 기준안이 되기 위해 다양한 사용자 요구사항을 조사, 대상으로 하는 음성 시료의 확대, 관련 분야 전문가들의 객관적의 평가 및 검증작업을 거치고 있다.



<그림 5> 음소 레이블링

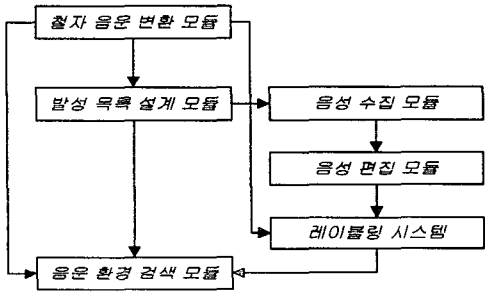
3. 음성 언어 자료 확보를 위한 Workbench

음성 언어 자료의 확보를 위해서는 발성 목록의 설계에서부터 음성 시료의 수집

및 편집, 레이블링, 검색 환경의 제공 등 다양한 소프트웨어의 도움이 필요하다. 본 장에서는 음성 언어 자료의 확보를 위한 Workbench를 설계하고, 구현한 결과에 대하여 기술한다.

3.1 Workbench 개요

음성 언어 자료의 확보를 위한 Workbench는 발성 목록의 설계 및 음성 수집 모듈, 음성 편집 모듈, 레이블링 시스템, 음운 환경 검색 모듈로 구성되어 있으며 각 모듈들간의 관계는 다음 그림과 같다.



<그림 6> Workbench의 개요

3.2 철자 음운 변환 모듈

철자 음운 변환기는 한글을 소리나는 대로 바꿔주는 읽기 규칙을 구현한 것이다. 읽기 규칙의 기본 원칙은 정부에서 고시한 표준어 규정의 표준 발음법을 따르고 있다 [12]. 읽기 규칙은 크게 4가지로 구별되며, 모음의 발음, 장음처리, 음운 변동 규칙, 음의 첨가로 나눌 수 있다. 구현된 모듈은 표준어 규정에 있는 원칙 중 장음 처리를 제외한 3가지의 규칙을 적용하고, 두 가지 이상으로 적용 가능한 경우에는 보다 일반적인 경우를 적용하도록 하였다. 또한 규칙으로 구현할 수 없는 경우에는 예외 발음 사전에 추가하여 처리하도록 하였다.

3.3 음운 균형 발성 목록 선정 모듈

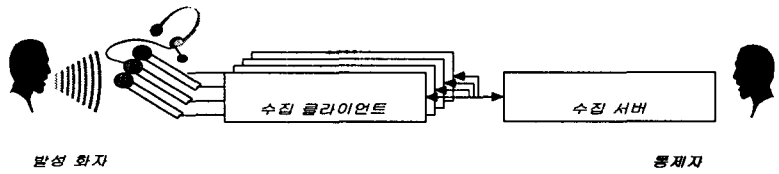
본 모듈은 대량의 텍스트 코퍼스를 입력으로 받아 앞에서 기술한 철자 음운 변환 모듈을 통해 음소열로 변환한 후, 이로부터 음운 현상들이 균형 있게 분포하는 (Phonetically balanced) 발성 목록을 선정해 주는 모듈이다.

음운 균형 발성목록에 관해서는 2.1절을 참조할 것.

3.4 클라이언트/서버 음성 수집 모듈

음성 데이터베이스를 구축하기 위해 음성 시료를 수집할 경우 사용자에게 발성 목록을 어떻게 제시해 줄 것인가 하는 점도 중요한 요소가 된다.

본 모듈은 사용자에게 자동으로 발성 목록을 제시해 주고, 또한 동일한 발성 내용을 다양한 환경(다른 PC기종-예를 들면 노트북과 데스크탑 PC, 다른 종류의 사운드 카드, 다른 종류의 마이크 등)에서 수집하기 위해 개발된 것이다.



<그림 7> 클라이언트/서버 음성 수집 모듈

이를 위해 음성 수집 클라이언트에서는 각각의 환경(사운드 카드, 마이크 등)을 설정하고 네트워크를 통해 수집 서버에 접속한다. 접속 후 클라이언트가 서버의 통제에 의해 수행하는 기능은 다음과 같다.

- 발성 화자 별 음성 데이터의 관리
- 발성 목록의 제시
- 발성 시간 정보의 표시

수집 서버는 접속된 수집 클라이언트들에 대하여 다음과 같은 기능을 수행한다.

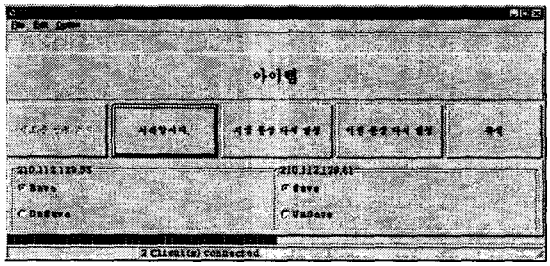
- 클라이언트 별 음성 수집 통제
- 클라이언트 간 동기 유지
- 발성 목록의 전송
- 잘못 발성된 발성 목록의 재발성
- 발성 화자의 통제(발성 시작, 휴식 등)
- 발성 화자 정보의 관리
- 샘플링 주파수의 설정
- 목록별 발성 시간의 설정

따라서 발성 화자는 복수의 마이크 앞에서 한번만 발성하면 각각의 클라이언트에

서는 네트워크를 통해 서버의 통제를 받아 음성 시료를 수집함으로써 음성 시료 수집의 효율성을 높일 수 있다. 또한, 동일한 음성 내용에 대한 다양한 환경의 음성 시료를 확보할 수 있으므로 마이크, 사운드 카드 등 환경에 의한 영향을 분석할 수 있는 시료를 얻을 수 있다.

다음 그림은 음성 수집 클라이언트와 2개의 클라이언트가 접속되어 있는 상태에서 음성 수집을 통제하고 있는 음성 수집 서버를 나타낸 것이다.

<그림 8> 음성 수집 클라이언트



<그림 9> 음성 수집 서버

3.5 음성 편집 모듈

음성 편집 모듈은 수집된 음성 데이터가 연속적으로 저장되어 있을 경우, 데이터 중 무음 구간, 잘못 발성된 구간을 제외하고 원하는 음성 데이터만을 파일로 저장하기 위한 모듈이다.

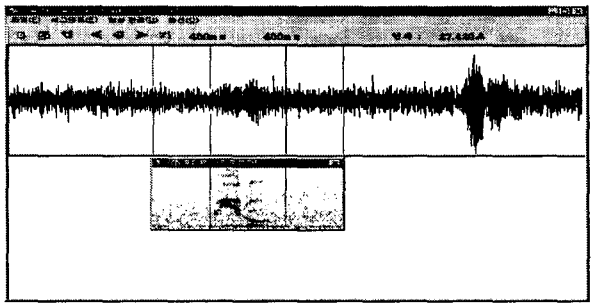
연속된 음성 데이터에서 음성 구간을 검출하기 위해서는 끝점 검출 알고리즘이 필요하며, 특히 음성 인식기의 경우 끝점 검출이 정확하지 않을 경우 성능이 저하된다는 사실은 잘 알려져 있다[15].

일반적으로 고립 단어의 끝점 검출을 위해서는 단구간 에너지와 영교차율을 파라미터로 사용한다[16]. 그러나 이러한 끝점 검출 과정은 소음이 존재하지 않는 경우에는 좋은 성능을 나타내지만, 자동차 소음과 같이 SNR이 0dB에 가까운 소음 환경에서는 끝점 검출이 거의 불가능하다. 이러한 경우 주파수 영역의 에너지를 파라미터로

이용하면 보다 좋은 성능을 얻을 수 있다.

따라서 본 모듈에서는 단구간 영교차율, 에너지, 주파수 에너지의 파라미터를 사용자가 선택하여 사용할 수 있도록 구현하였다.

다음 그림은 본 모듈을 이용하여 자동차 소음 환경에서 끝점을 검출하고 음성 데이터의 편집을 수행하는 그림이다.



<그림 10> 음성 편집 모듈

3.6 다단계 음성 레이블링 시스템

레이블링 된 음성 데이터베이스는 음성 인식 및 합성기의 훈련 및 평가, 성능 개선을 위한 중요한 자료로서 활용된다. 그러나 현재 음성 데이터베이스를 레이블링 하는 작업은 대부분 수작업으로 이루어지고 있으며, 자동 레이블링을 위한 연구가 활발하게 진행되고 있으나 자동 레이블링 시스템의 훈련을 위해서도 수동 레이블링 된 데이터는 필요한 실정이다.

수동 레이블링을 위해서는 레이블링 시스템이 필요하나 현재 대부분의 소프트웨어가 유닉스(Unix)기반으로 되어 있거나 전용 하드웨어를 요구하므로 보급이 미흡한 실정이고, 그나마 대부분이 고가이므로 연구자들이 손쉽게 접하기 힘든 면이 있다.

본 절에서는 PC에서 사운드카드를 이용하여 손쉽게 레이블링을 할 수 있도록 다단계 음성 레이블링 시스템을 구현한 결과에 대하여 기술한다.

구현된 레이블링 시스템은 경계 분할 정보, 음성 정보, 언어 정보 등의 다른 계층을 두고 레이블링을 수행할 수 있도록 되어 있다. 레이블링 시스템에서 제공하는 개략적 기능은 다음과 같다.

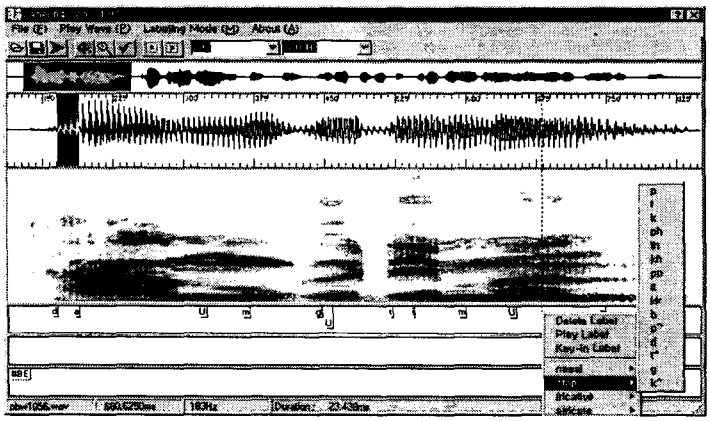
- 웨이브 폼 보기
- 스펙트로그램 보기

- 스펙트로그램 명암 조절
- FFT 분석 포인트의 조절
- 전체 구간 듣기
- 확대 구간 듣기
- 선택 구간 듣기
- 레이블 구간 듣기
- 다단계 레이블
- 레이블 심볼의 사용자 정의

본 레이블링 시스템에 사용된 스펙트로그램의 특징은 다음과 같다.

- Frame size : 8ms(125Hz bandwidth)
- Frame advance rate : 1ms
- Window : Hanning window
- 기본 FFT 포인트 : 512 포인트

다음 그림은 구현된 레이블링 시스템을 이용하여 레이블링을 수행하는 화면을 나타낸 것이다.



<그림 11> 다단계 레이블링 시스템

3.7 음운 환경 검색 모듈

특정 음운 현상을 분석하기 위해서는 원하는 조건을 갖춘 발성 목록을 설계하고 음성 시료를 수집하여 이를 분석하는 방법이 사용된다.

그러나 최근 공통으로 사용하기 위한 다종 다양한 환경을 포함하고 있는 대량의 음성 데이터베이스[2, 3, 17]가 배포됨에 따라 이러한 음성 데이터베이스로부터 원하는 음운 환경을 포함하고 있는 발성 목록을 추출하고 이를 분석할 수 있는 방안도 필요하다.

예를 들어, 한국어의 유성음화에 대한 분석을 수행할 경우, 음성 데이터베이스의 발성 목록에서 유성음화가 발생할 수 있는 조건인 "모음 + 연자음 + 모음"의 음운 환경을 포함하고 있는 발성목록을 선별하고 분석을 수행할 수 있다면 음성 데이터베이스의 활용도를 높일 수 있다.

본 모듈은 이러한 목적을 달성하기 위하여 발성 목록을 입력으로 받아, 원하는 조건을 만족하는 발성 목록을 포함하고 있는 문장 또는 단어를 선별해 주는 모듈이다.

본 모듈에서 지원하는 검색 방법은 문자열 검색, 음소열 검색 및 음운 환경 검색의 3가지이다. 문자열 검색은 발성 목록에서 원하는 문자열을 포함하고 있는 목록을 선정하여 출력하는 방법이고, 음소열 검색은 발성 목록에서 원하는 음소열을 포함하고 있는 목록을 선정하여 출력하는 방법을 말하며, 이러한 경우 발성 목록을 위에서 기술한 철자 음운 변환 모듈을 거쳐 변환한 음소열에서 검색을 수행하게 된다.

음운 환경 검색은 특정 음소열을 지정하지 않고 위의 예와 같이 "모음 + 연자음 + 모음"처럼 음운 환경을 만족하는 유형을 지정하여 검색을 수행하는 방법을 말한다. 음운 환경을 정의하는 방법은 다음과 같다.

(1) 자음 : 자음(<조음방법>,<조음위치>,<강세>)

<조음방법>

조음방법으로는 비음, 파열음, 마찰음, 파찰음, 유음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 조음방법 유형은 어떠한 조음방법도 다 포함할 것(Don't care)을 나타낸다.

<조음위치>

조음위치로는 양순음, 치음, 경구개음, 성문음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 조음위치 유형은 어떠한 조음위치도 다 포함할 것(Don't care)을 나타낸다.

<강세>

강세로는 연음, 격음, 경음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 강세유형은 어떠한 강세도 다 포함할 것(Don't care)을 나타낸다.

<조음방법>, <조음위치>, <강세>에 해당하는 토큰들의 어떠한 조합도 사용 가능하다. 단, 각 토큰들의 위치는 지켜져야 하며, 각 토큰은 쉼표에 의해 구분되어야 한

다.

(2) 모음 : 모음(<종류>,<수평위치>,<수직위치>)

<종류>

종류로는 모음, 이중모음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 모음과 이중모음을 구분하지 않겠다는 것(Don't care)을 나타낸다.

<수평위치>

수평위치로는 전설음, 중설음, 후설음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 수평위치 유형은 어떠한 수평위치도 다 포함할 것(Don't care)을 나타낸다.

<수직위치>

수직위치로는 고모음, 중모음, 저모음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 수직위치 유형은 어떠한 수직위치도 다 포함할 것(Don't care)을 나타낸다.

<종류>, <수평위치>, <수직위치>에 해당하는 토큰들의 어떠한 조합도 사용 가능하다. 단 각 토큰들의 위치는 지켜져야 하며, 각 토큰들은 쉼표에 의해 구분되어야 한다.

위와 같이 정의된 음운 환경을 이용하여 본 모듈에서 "고모음 + 파열 연자음 + 고모음"을 검색하는 경우를 다음 그림에 나타내었다.

<그림 12> 음운 환경 검색 모듈

4. 결 론

공동이용을 위한 음성DB를 구축하기 위해서는 그 발성목록이 한국어의 다양한 음운환경을 포함하고 있으며, 일정한 도메인의 테스크에 종속되지 않는 것이 바람직하다. 본 논문에서는 이를 위하여 100만여 어절의 텍스트 코퍼스에서 음운 균형 발성목록을 추출하고 제시하였다. 추출된 음운 균형 발성목록은 다음과 같은 조건을 만족한다. 첫째, 모집단에 나타난 모든 음운환경을 포함한다. 둘째, 최소한의 어절로 구성되어 있다. 셋째, 발성목록을 구성하고 있는 음운환경들 간의 확률분포가 최대한 고르게 분포한다.

그 외 발성목록으로 임의단위 합성 시스템을 위한 고빈도 2,000어절을 작성하고 숫자음으로는 단독숫자음 및 4연숫자음 76종을 작성하고, 단문은 1종을 선정하였다.

이와 같이 구성된 발성목록을 표준어를 사용하는 72명의 일반인 및 어나운서를 대상으로 음성시료를 수집하고 이를 양자화 하여 편집하고 CD-ROM에 저장하여 공동으로 이용이 가능한 형태로 구축하였다.

또한 본 논문에서는 음성 언어 자료의 확보를 위한 Workbench를 설계하고, 구현한 결과에 대하여 소개하였다. 구현된 모듈로는 텍스트를 자동 읽기 변환하는 철자음운 변환기, 발성 목록 선정 모듈, 끝점 검출기를 이용한 음성 데이터 편집 모듈, 다단계 레이블링 시스템, 텍스트에서 원하는 음운 환경을 포함하고 있는 문자열을 다양한 조건으로 검색할 수 있는 음운 환경 검색기를 포함하고 있다. 향후 다양한 분석 기능을 포함하여 종합적인 음성 언어 처리 Workbench로 발전시켜 나갈 계획이다.

<참고문헌>

[1] 이용주, 김봉완 외(1995.6), 해외 음성 DB 구축 동향, 「한국음향학회 제12회 음성통신 및 신호처리 워크샵 논문집」

[2] 이용주(1996. 8), 음성데이터베이스의 현황및 과제, 「한국음향학회 제13회 음성통신 및 신호처리 워크샵 논문집」

[3] 이용주(1994. 10), 한국어 음성언어정보처리와 음성 데이터베이스, 「한국어정보처리 소식 제2권 특별기고」

[4] 이용주, 정유현 외(1992. 7), 보급형 음성 데이터베이스 구축에 관한 연구, 「1차년도 최종보고서」, 과학기술처.

[5] 정유현, 최준혁 외(1992. 6), 공통 음성 데이터베이스 구축을 위한 사전 조사연구, 「전자공학회 하계 학술대회」

[6] K. Shikano(1984), Phonetically balanced word list based on information entropy, *proceedings of Acoustical society of Japan.*

[7] Yeonja Lim, et al,(1995), Implementation of the POW algorithm for Speech Database, *ICASSP 95, Detroit.*

[8] Hayamizu(1986), Generation of VCV/CVC balanced word sets for speech data base, *Bulletin of Electrotechnical Laboratory.*

[9] K. Iso, et al,(1988), Design of a Japanese Sentence List for a Speech Database, *Proc. ASJ., 2-2-19, March.*

[10] Jianhua Lin(1991), Divergence Measure Based on the Shannon Entropy, *IEEE Transactions on information theory, Vol. 37,* No. 1,

[11] 이호영(1996), 「국어 음성학」, 태학사.

[12] 최운천, 지민제, 이용주(1992. 10), 문장음성 변환시스템 글소리II를 위한 읽기 규칙, 「1992년도 제4회 한글 및 한국어 정보처리 학술발표 논문집」

[13] NIST(1991), Speech Corpora Produced on CD-ROM Media by The National Institute of Standards and Technology(NIST), April, 1

[14] 이용주, 이숙향 외(1996), 한국어 음성DB 구축을 위한 한국어 레이블링 기준에 관한 연구, 「제13회 음성통신 및 신호처리 워크샵 논문집」

[15] J.G. Wilpon, et al.(1984. 3), An improved word-detection algorithm for telephone-quality speech incorporating both syntactic and semantic constraints, *AT&T Tech. J., 63(3),* 479-493.

[16] L.R. Rabiner, et al(1975. 2), An algorithm for determining the endpoint of isolated utterance, *Bell syst. Tech. J., 54(2),* 297-315,

[17] 이용주(1998,2), 음성언어코퍼스, 「정보과학회지」, *16(2),* 41-48.

접수일자: 1998년 11월 26일
게재결정: 1998년 12월 21일

▶ 김봉완(Bong-wan Kim)
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 컴퓨터공학과 휴먼인터페이스 연구실
전화: 0653) 850-6885
e-mail: joe@speech.wonkwang.ac.kr

▶ 이용주(Yong-Ju Lee)
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 컴퓨터공학과
전화: 0653) 850-6885
e-mail: yjlee@wonnms.wonkwang.ac.kr