

국어 분절음 특성에 맞는 음성 데이터 베이스의 모형

김 종 미

강원대학교 영어영문학과

A Model of Speech Database in Korean in consideration of its segmental phonology

Jong-mi Kim

English Dept., Kangwon National University

요 약

본 논문에서는 국어 분절음 특성에 맞는 음성 데이터베이스의 모형을 제시하고자 한다. 음성 데이터 베이스는 i)각 음의 고유음가정보, ii)인접음 정보, iii)빈도수에 따른 확률정보를 포함해야 한다. 이 요건을 충족시키기 위하여 본 모형은 i)음운 단위별로 Labeling하여, 고유음과 인접음 정보를 편집하고, ii)음운 규칙과 제약정보에 의해 Phoneme Balanced Words를 작성하여, 허용되는 인접음을 취하고, 허용되지 않는 인접음을 탈락시키며, iii)확률정보에 의해 데이터 베이스를 구성하여 보다 더 가능한 음과 음소열을 우선적으로 인식·합성하게 한다.

본 모형의 잇점은 i)데이터의 추출시, 각 시스템의 인식 및 합성 단위별로 쉽게 편집하여 이용할 수 있다는 용이성, ii)알고리즘 작성시, 인접음정보와 확률정보를 통해 인식 및 합성의 정확을 거할 수 있다는 정확성, iii)시스템 평가시, 빈도수가 높은 음과 음소열의 우선적인 인식 및 합성을 우월하게 평가한다는 공정성, iv)데이터 집적시, 데이터의 음운기능의 중복과 편중을 피함으로써 데이터량을 감소화할 수 있다는 경제성을 들 수 있다.

1. 서론

본 논문은 인식 및 합성 연구용 공통음성 단어 데이터 베이스의 모형을 제시하고자 한다. 본 모형은 국어분절음 특성에 맞게 구성되어, i) 음운 단위별로 Labeling하고, ii) 음운규칙과 제약정보에 의해 Phoneme Balanced Words를 작성하며, iii) 확률정보에 의해 Phoneme Balanced Database로 확장한다. "분절음"이란 음소와 어음을 통칭하며, 음운과 대조되는 개념이다.

본 데이터 베이스의 범위는 "단어"로서, 연속음성의 경우는 본 데이터베이스에 허사 집가로 인한 빈도수의 제조정과, 음운정보의 첨가를 실시하여 활용할 수 있다.

본 모형의 기술순서는 i) Labeling ii) Balanced Speech의 작성, iii) Balanced Database로의 확장의 순으로 하겠다.

II. 음운 단위별 Labeling

Labeling을 하기 위한 요건은, 음성인식 시스템에 활용할 경우, i) 시스템에 따라 다른 다양한 인식단위가 모두 Label 되어야 하고, ii) 단어사전과 대비하여 의미파악을 하는 단위가 Label 되어야 하며, iii) 인접음을 예측하기 위한 단서를 줄 수 있어야 하고; 음성합성 시스템에 활용할

국어 분점음 특성에 맞는 음성 데이터 베이스의 모형

경우는, i) 시스템에 따라 다른 다양한 합성단위가 모두 Label되어야 하고, ii) Text의 관독단위가 Label 되어야 하며, iii) 음운규칙 및 제약 조건의 적용 단위가 Label 되어야 한다.

이들 요건을 충족시키기 위하여 본 모형에서 제시하는 Labeling의 단위는, 이음, 음소, 음절, 단어이며, 이 중 이음과 음절의 경우는 demi-의 형태와 di-의 형태를 편집하여 추가로 Labeling하는 것이다. "demi-"의 형태란 해당단위를 절반으로 자른 것으로서, demi-phone은 이음의 반절을, demi-syllable은 음절의 반절을 의미한다. 또한 "di-"의 형태란, 해당단위의 절반과 인접단위의 절반을 합한 것으로서 di-phone은 인접한 두음의 demi-phones의 합이고, di-syllable은 인접한 두음절의 demi-syllables의 합이다. 아래 각 Labeling 단위에 관해 약술하기로 한다.

1. 이음(Allophone)

"이음"이란 각 음소의 실제 발음되는 형태로서, 음성인식 시스템에서 흔히 일컫는 "Phoneme-like units"도 이에 해당한다. 예컨대, 국어의 /ㅂ/은 단어초에서는 기음[p^h]로, 모음간에는 유성음 [b]로, 단어말에는 무기음[p^h]로 나타나는데, 이때의 [p^h, b, p^h]를 이음이라 한다. 국어의 주요 이음의 내용은 경국의(1994)를 참조하기 바란다¹¹⁾.

이음별 Labeling의 용도는 인식 시스템에서 i) 인식단위로 흔히 쓰여지고, ii) 인접음 예측의 단서가 된다는 점과, 합성시스템에서는 iii) 합성의 단위로서 이음, demi-phone, di-phone의 여러형태를 사용한다는 점이다¹²⁾. 이음을 인식 및 합성의 단위로 사용하는 이유는 음향분류가 쉽고, 조음결합에 따른 특성이 반영되기 때문이다.

이음 Labeling System의 요건은 다음과 같다. i) 침착, 변경이 용이해야 하는데, 그 이유는 "이음"은 다양한 음향을 표시하므로 그 수효를 한정지을 수 없고 분류가 부정확하기 때문이다. 따라서 System은 우선 비교적 예측 가능한 것을 수록한 후, 추가 변경을 수시로 해야 한다. ii) 또한, Labeling시 먼저 해당 이음을 Label한 후, 절반을 잘라 demi-phone형태를 Label하고, 또 인접음의 demi-phones을 합하여 di-phone의 형태로도 Label해야 한다.

2. 음소(Phoneme)

"음소"란 각 언어의 의미를 변별하는 최소의 소리단위이다. 예컨대, 국어의 단어중 '지각'과 '지갑'의 의미를 변별하는 최소의 소리는 'ㄱ'과 'ㅂ'이므로, 이들을 가리켜 음소라 한다. 국어의 음소는 19개의 자음, 10개의 모음, 2개의 받모음으로 되어 있다. /p/, /t/, /s/, /c/, /k/, /p', /t', /s', /c', /k', /ph/, /th/, /ch/, /kh/, /h/, /m/, /n/, /ŋ/, /l/이며, 모음은 /i/, /e/, /ɛ/, /u/, /y/, /ɔ/, /a/, /a', /o/이다. 또한 받모음은 /y/와 /w/로서, /y/는 '예, 애, 의, 여, 야, 유, 요'의 7개의 형태로, /w/는 '위, 세, 왜, 워, 와'의 5개의 형태로 나타난다.

음소별 Labeling의 용도는 인식 시스템에서는 i) 의미 파악의 단위와, ii) 인접음 예측의 단위가 되고, 합성시스템에서는 iii) Text관독 단위와 iv) 규칙 및 제약조건의 적용단위가 된다. 이 중 i)의 이유는 앞서 '지각'과 '지갑'의 예에서와 같이 'ㄱ'과 'ㅂ'의 음소차이에 의해 의미가 달라지기 때문이고, ii)와 iv)는 뒷절 "III. 음운규칙과 제약 정보에 의한 Phoneme Balanced Words의 작성"에서 후술할 것이다. 그리고 iii)음소가 합성시 Text 관독 단위가 되는 이유는, 국어가 일본어나 중국어와는 달리 음소에 기반한 '음소문자'이기 때문이며, 여러 음소문자 언어 중에서도 특히 문자와 소리의 일대일 대응이 정확하며 Text-To-Speech 합성이 용이하기 때문이다. (cf. 영어, 독일, 불어등)

음소별 Labeling시에 필요한 시스템의 요건은 다음과 같다. i) 음소의 수는 위에 기술한 21개로 고정적이므로 침착 및 변경의 기능이 필요없다. ii) demi-나 di-로의 재편집이 필요없는데, 그 이유는 음소가 실제 발음형태가 아니어서 Transition정보가 무의미하기 때문이다. iii) Pause는 하나의 음소로 등록하여 인식 및 합성시 단어 경계로 사용하게 한다.

3. 음절(Syllable)

국어의 음절은 초성, 중성, 종성으로 구성되어 있는데, 이 세 위치에 /p/, /t/, /s/, /c/, /k/, /p', /t', /s', /c', /k', /ph/, /th/, /ch/, /kh/, /h/, /m/, /n/, /l/ 등의 자음을 합하여 19종류가 될 수 있고, 종성에는 국어의 모든 받모음과 북모음을 총괄하여 22종류가 될 수 있으며, 종성에는 북

음과 /p/, /t/, /k/, /m/, /n/, /ŋ/, /l/ 등의 자음을 합하여 8 종류가 올 수 있다.

음절별 Labeling의 용도는 인식시스템에서 i) 인식의 단위, ii) 인접음 예측단위, iii) 확률 추정 단위가 되고, 합성 시스템에서는 iv) 합성의 단위, v) 규칙 및 제약의 적용단위가 된다. 이 중 i)과 iv)에 언급한 바, 음절이 인식 및 합성의 단위로 쓰이는 이유는 국어에 쓰이는 음절의 종류가 다른 언어와는 달리 간단하고¹³⁾, 각 음절에는 음절내 음소끼리의 transition 정보가 포함되어 있으며, 모음당 음절 하나라는 일대일 대응관계가 있어 음절의 위치 파악이 쉽기 때문이다. 이처럼 모음에 따라 음절의 위치를 쉽게 찾을 수 있다는 특징과, transition 정보가 포함되어 있다는 특징 때문에 demi-syllable과 di-syllable의 형태를 인식과 합성의 단위로 삼기도 한다. 그리고 ii), iii), iv)에 언급한 바, 인접음 예측, 확률 추정, 규칙 및 제약의 적용을 위한 음절의 용도는 다음 III절과 IV절에서 후술할 것이다.

음절 Labeling system의 요건은 다음과 같다. i) 추가는 가능하나, 변경은 불필요하다. 그 이유는 우선 나타나는 음절을 대상으로 데이터 베이스를 작성한 후, 추후 발견되는 음절을 추가할 것이기 때문이다. 또한 ii) demi-syllable과 di-syllable의 형태도 수록해야 한다. 이들 형태를 인식 및 합성단위로 사용하기도 하기 때문이다.

4. 단어

국어의 단어는 조사나 체언이 붙지 않은 상태를 일컫고, 이들 허사가 붙은 뒤의 형태는 어절이라 한다.

국어 단어의 용도는 인식시스템에서 i) 의미파악의 단위와, ii) 인접음 예측의 단위가 되고, 합성시스템에서는 iii) Text 판독 단위와 iv) 규칙 및 제약조건의 적용단위가 된다. 이 중 i)과 iii)의 이유는 국어의 어휘사전이 단어별로 구성되어 있기 때문이고, ii)와 iv)의 이유는 뒷절 "III. 규칙 및 제약 정보에 의한 Phoneme Balanced Words의 작성"에서 후술 하겠다.

단어 Labeling System의 요건은 다음과 같다. i) 단어 경계를 북음으로 표시하여 모든 단어는 '북음 + 음소열 + 북음'의 형태로 수록한다. 그 이유는 한 단어를 구성하는 음소의 수효가 일정치 않아 인접단어를 합하여 한 단어로

오인식하는 것을 막기 위해서이다. 또한, ii) 연구영역별 재편집을 위해 추가, 삭제가 용이해야 한다. 그리고 iii) demi- 나 di-의 형태는 불필요하다. 이는 각 단어의 의미를 잃게하고, 규칙적용의 경계를 잃게 하기 때문이다.

이상 4가지의 음운단위별 Labeling을 통하면 인식 및 합성을 위한 모든 요건을 충족 시킬 수 있다. 우선 i) 시스템에 따라 다양한 인식 및 합성단위로써 이음, demi-phone, di-phone, 음절, demi-syllable, di-syllable이 모두 Label되어 활용가능하고, ii) 의미파악의 단위와 text 판독단위로써 음소와 단어 Labeling이 활용되며, iii) 이음, 음소, 음절, 단어의 Labeling을 통해 인접음예측단위, 음운규칙 및 제약 조건의 적용단위로 쓰인다.

III. 음운 규칙과 제약정보에 의한 Phoneme Balanced Words의 작성

Phoneme Balanced Words (PBW)는 시스템 Train용으로 사용할 수 있는 커스 음성 데이터 베이스의 통칭이다. PBW의 요건은 인식 시스템에서 i) 국어에 불가능한 이음열, 음소열, 음절, 단어로 오인식 하는 것을 막아야 하고, ii) 인접음 예측정보를 제공하여야 하며, iii) 의미 인식형태인, 발음 규칙적용 이전의 기저음소열을 파악하여야 한다는 것이다. 또한 합성시스템에서는 i) 국어에 불가능한 이음열, 음소열, 음절의 생성을 막아야 하고, ii) 음운환경에 따른 음가변화를 반영하여야 한다.

이들 요건을 충족시키기 위하여 본 문헌에서 제시하는 각 규칙 및 제약정보는 이음별 음소별, 음절별, 단어별로 아래와 같이 예시되며, 전체적인 정보는 정국의(1994)를 참조하기 바란다¹⁴⁾.

1. 이음별 규칙

이음별 규칙은 아래와 같이 예시된다.

국어 본절음 특성에 맞는 음성 데이터 베이스의 모형

	가능한 이음열	불가능한 이음열	PBW
(a)	r+모음	자음+r, 단어끝r, r+자음	오라[ora]
(b)	f+i, f+j	s+i, s+j	시계[figiel]

위 (a)는 '라'의 탈설음화라는 규칙으로 /r/음소는 초성에서 반드시 [r]로 소리가 나며, 기타의 환경에서는 [l]로 소리가 난다는 규칙을 말한다. 이 규칙을 반영하는 PBW는 '오라'이다.

위 (b)는 '시'의 구개음화라는 규칙으로 /s/음소는 /i/나 /j/모음 앞에서 구개음 [ʃ]로 소리가 난다. 이 규칙을 반영하는 PBW는 '시계'이다.

2. 음소별 규칙

음소별 규칙은 아래와 같이 예시된다.

	가능한 음소열	불가능한 음소열	PBW
(a)	평음 + 경음	평음 + 평음	각시 [kaksʰi]
(b)	각음	h + 각음	농고 [nokho]
(c)	비음 + 비음	성애음 + 비음	검문사 [cɛmmusa]
(d)	리리	리리, 리리	천리 [tʰɛlli]

위 (a)는 평음과 평음이 겹치면 첫 평음은 경음화 된다는 규칙으로서 이에 해당하는 PBW는 '각시'이다. 위(b)는 /h/와 평음이 합하면 각음이 되는 규칙으로서 이에 해당하는 PBW는 '농고'이다. 위(c)는 비음앞의 성애음은 비음화된다는 규칙으로 이에 해당하는 PBW는 '검문사'이다. 위(d)는 /r/과 /l/이 인접하면 '리리'로 치고, 규칙으로서 이에 해당하는 PBW는 '천리, 천리'이다.

3. 음절별 규칙 및 제약조건

음절별 규칙 및 제약조건은 아래와 같이 예시된다.

2. 상국, 구희산, 이찬도, 김종미, 한선희, "음성 인식/합성을 위한 국어의 음성-음운론적 특성연구," 한국음향학회지 제 13권 5호 (예정), 1994.

	가능한 음절	불가능한 음절	PBW
(a)	평과열 종성	각음 종성 경음 종성 치찰음 종성	부엌 [puɛk] 박 [pak] 갓 [gat]
(b)	단자음 종성	복자음 종성	값 [kap] 딛 [tak]
(c)	비과음 종성	과열종성	옥 [ok]

위 (a)는 종성에서는 경음, 경음, 치찰음이 허용이 안되고 평과열을 많이 허용된다는 제약조건으로, '부엌, 박, 갓' 등의 종성이 각각 [k, ɸ, t]으로 소리난다. 위 (b)는 종성에서는 단 한개의 자음만이 허용된다는 제약조건으로 '값, 딛'등의 복자음 종성이 [b, ɸ]처럼 단자음으로 소리난다. 위(c)는 종성 과열음은 비과화 된다는 규칙으로 '옥'의 경우 영어처럼 [ok]로 소리나지 않고 단협소리 [ok]으로 소리난다.

4. 단어별 규칙 및 제약조건

국어의 어두에는 '리'이 올 수 없다는 제약조건이 있다. 따라서, '리'이나 '이'으로 시작하는 단어는 있을 수 없어도 '리'로 시작하는 단어는 외래어의 경우가 아닌 한 없다.

위에 예시한 이음열, 음소별, 음절별, 단어별 규칙 및 제약조건이 인식 및 합성 시스템을 위한 PBW의 요건을 충족시킨다는 점은 다음과 같다. 인식 시스템에서 위 PBW의 사용은 i) 국어에 불가능한 이음열, 음소열, 음절, 단어로 오인식하는 것을 막을 수 있다. 불가능한 소리의 연속은 이미 PBW의 Training시 List되어야 할 것이기 때문이다. ii) 인접음 예측정보를 제공한다. 예컨대, [r]이 인식되었다면 그 후속음은 반드시 모음이 있다는 예측을 할 수 있다. iii) 의미 인식 형태인 발음규칙 적용 이전의 지저음소열을 파악하게 해준다. Training시 수록되었던 지식정보에 의해 변형된 소리를 파악해야 단어사건과 태즈할 수 있기 때문이다. 예컨대, [tʰɛlli]를 /tʰɛlli/로 파악해야 오해석을 막을 수 있다.

합성 시스템에서 위 PBW의 사용은 i) 국어에 불가능한 이음열, 음소열, 음절, 단어의 생성을 막아준다. 불가능한 소리의 연속은 Training시에 조정된 것이기 때문이다. 그리고 ii) 음운환경에 따른 음가변화를 반영한다. PBW 자체가 음운규칙 적용 대상 단어들을 포괄하고 있기 때문에

다.

그러면, PBW를 실제로 작성하는 방법을 아래와 같이 제시하겠다. 우선 i) 모든 이음, 음소의 표준환경 발음을 망라한다. 이는 실제 언어음은 인접음에 수시로 영향 받므로 기준음의 형태가 필요하기 때문이다. 예컨대, 자음의 경우는 '아' 모음을 인접시켜 표준환경 발음을 만들 수 있고, 모음의 경우는 인접 자음이 없는 상태의 표준환경 발음을 작성한다. ii) 위에 나타난 규칙 및 제약정보에 민감한 단어들을 수록한다. iii) PBW의 규모에 따라 빈도수 높은 음절을 우선적으로 수록한다. 한국어에 가능한 음절의 형태는 3,344개(초성 19×중성 22×종성 8)이지만, 한국어 전체 나타나지 않는 음절(예: 썩, 썩)과 빈도수 낮은 음절(찍, 켜)을 제외하고, 빈도수 높은 음절을 우선적으로 수록한다. PBW의 규모가 작으면 물론 음절의 빈도수까지 고려할 여지는 없을 것이다. 음절의 빈도수 정보는 표준한국어 발음 대사전(1993)을 참조하기 바란다^[4]. 그리고, 마지막으로 iv) 본 PBW를 Task-domain이 한정되어 있는 인식 및 합성시스템에 사용할 때는, 영역별 고빈도 단어를 추가시킨다. 예컨대 "항공안내시스템"이라면, 항공안내서 필요한 특수용어들(외래어포함)을 추가해야 할 것이다.

IV. 확률정보에 의한 Balanced 음성 데이터베이스의 구성

데이터베이스는 기본적으로 방대한 양이어야 하나, Balanced 음성 데이터베이스의 목적은 최소의 데이터로 최대한 정확한 음성인식 및 합성시스템을 구현함으로써 노력, 경비, 시간을 절약하는데 있다.

최선의 음성데이터베이스의 요건은, 인식 및 합성 시스템에서 i) 알고리즘 작성시 보다 더 빈도수 높은 음과 음소열을 우선적으로 해결하게 해야하고, ii) Test Data로 사용시 빈도수 높은 음과 음소열을 우선적으로 인식 및 합성하는 System에 평가상의 Credit를 주어야 한다. 따라서 Balanced 음성 데이터베이스에는 빈도수의 반영이 필요하다.

빈도수에 의한 확률정보는 크게 출현확률과 결합확률로 나뉠 수 있다. 출현확률은 각음소나 음절이 출현할 수 있는

확률을 인건고, 결합확률은 각음소끼리의 결합하는 확률을 인건는다.

1. 출현확률

국어 음소 및 음절의 출현 빈도수는 다음과 같다. 음소별로는 /아/, /어/, /이/가 모음 전체의 절반을 차지하고, /ㄱ/, /ㄴ/, /ㅇ/, /ㄹ/, /ㄷ/ 이 자음 전체의 절반을 차지한다^[5]. 또한 음절 별로는 우리말에 실제 쓰이는 음절 수는 1,453개로서 전체 가능한 음절 수인 3,344개의 1/3정도 밖에 되지 않는다^[4]. 실제 쓰이는 음절간의 출현 빈도수 서열은 모음의 종류와 밀접한 관계가 있어, 역사적으로 오래된 모음인 [이, 으, 어, 아, 우, 오]로 구성되는 음절은 빈도수가 가장 높고, 역사적으로 늦게 생성된 모음 [에, 예]의 경우나 이중모음 [여, 야, 웨, 의]의 경우 빈도수가 낮다.

2. 결합확률

국어음소간의 결합빈도수는 유재원(1993)에서 상세히 조사되었으며, 그 몇몇 사항은 다음과 같다^[6]. i) 국어의 "자음 + 모음"의 결합에는 장애음이 83.48%로 국어의 초성은 평음장애음의 확률이 높음을 보이고 있고, ii) "모음 + 자음"의 결합에는 ㅇ, ㄴ, ㄹ, ㄷ 순으로 높은 빈도수를 보여 국어의 종성은 공명음이 장애음에 비해 안도적인 우세를 보인다. 따라서 국어는 초성은 장애음, 종성은 공명음이 선호됨을 알 수 있다. 또한 iii) 국어 반모음의 결합확률은 j 계열 > w계열 > '의'의 순이며, iv) '모음 - 모음'의 결합에서 절대 나타나지 않는 쌍들은, '이+에, 예+에, 에+에, 에+에, 의+에' 등으로서, 진성모음의 중복을 피하는 성향을 뚜렷이 보인다.

위에서 예시한 출현확률 및 결합확률을 이용하여 Phoneme Balanced Database를 작성할 때의 유의사항은 다음과 같다. 즉, 현재 보고된 확률정보는 국어계량언어학의 연구가 초기 단계라는 점에 미루어 아직 불완전하다 할 수 있으므로, 끊임없는 침착과 수정이 가능한 프로그램을 사용하여야 한다는 것이다.

마지막으로 본 Phoneme Balanced Database를 활용한 대용량 인식 및 합성의 과정을 순차적으로 아래와 같이 제시하고자 한다. 인식의 과정을 순차적으로 기술하자면, 우

국어 분절음 특성에 맞는 음성 데이터 베이스의 모형

선 i) 입력음을, ii) 자모음으로 분할한뒤, iii) 표준 어음을 먼저 인식하고, iv) 규칙 및 제약정보를 통해 인접음을 예측한뒤, v) 음소별로 정리하고, vi) 음절별로 재배열한 뒤, vii) 확률 정보를 통해 인식의 정확성을 향상하고, viii) 음운규칙 및 제약에 따라 기저음소를 확인하여, ix) 단어사건과 대조를 통한 의미 파악을 한다.

합성의 과정을 순차적으로 기술하자면, 우선 i) Text상의 문자를, ii)음소로 변환하여, iii) 음절별로 정리한 후, iv) 음운규칙을 적용하고, v) 이음을 생성한후, vi) 확률정보를 통한 파형대조를 하고, vii) 정취 평가를 한다.

V. 결론

본 논문은 국어 분절음 특성에 맞는 음성데이터 베이스의 모형으로서, i) 음운 단위별 Labeling, ii) 음운 규칙과 제약정보에 의한 Phoneme Balanced Words의 작성, iii) 확률정보에 의한 Phoneme Balanced Database로의 확장을 제시하였다.

본 데이터베이스의 잇점으로는 용이성, 정확성, 긍정성, 경제성을 들수있다. 첫째, 용이성의 측면에서 볼때, 본 모형은 다양한 인식 및 합성단위가 세분화되어 label되어 있으므로 각 시스템의 필요한 단위별로 쉽게 편집하여 부분적으로 이용할 수도 있고, 검색할 수도 있다. 둘째, 정확성의 측면에서 볼 때, 규칙 및 제약정보의 확률 정보를 통해 인식 및 합성의 정확을 기할 수있다. 셋째, 긍정성의 측면에서 볼 때, 빈도수가 높은 음과 음소영의 우선적인 인식 및 합성을 우월하게 평가한다. 넷째, 경제성의 측면에서 볼 때, 데이터의 음운기능의 중복과 편중을 피함으로써 데이터량을 최소화하여 시간, 노력, 비용을 절약할 수 있다.

참고문헌

1. 정국, 구희산, 이찬도, 김종미, 한선희, "음성 인식/합성을 위한 국어의 음성-음운론적 특성연구," 한국음향학회지, 제 13권 5호 (예정), 1994.
2. 이종락, "반음소 : 새로운 음성합성 및 인식단위," 제 10 회 음성통신 및 신호처리 워크샵 논문집, PP. 208-212,

1993년 8월.

3. 김종미, "음소와 변이음과 음절의 상호관계," 한국어 음성인식 시스템 개발연구. 과학기술원 통신연구실, CRL-P-8801. PP. 160-168, 1988.
4. 표준 한국어 발음 대사전. 한국방송공사/KBS편, 1993.
5. 이용주의, 연속음성인식기술 개발에 관한 연구(II). 한국 전자통신 연구소. 1988년 6월.
6. 유계원, "음운통계(국어의 음소결합계약)에 대한 연구," 윤용단위 음운론 및 음운 통계에 관한 연구. 한국 전자통신연구소 연구보고서, PP. 77-95. 1993년 12월.