

# 한국어 낭독체 문장의 음향분석

## -바람과 햇님의 운율구 생성을 중심으로-

성철재(한국전자통신연구소)

### <차례>

1. 머리말	4.3 피실험자
2. 대화체와 대비해본 낭독체 문장의 특징	4.4 실험절차
3. ToBI 규약	4.5 실험결과 및 토의
3.1 ToBI 규약의 필요성	4.5.1 발화속도 및 운율구내부의 음절갯수
4. '바람과 햇님'의 음향분석	4.5.2 운율구 내부현상의 음향분석
4.1 실험목적	5. 토의
4.2 실험자료	6. 맺음말

### <Abstract>

#### The Acoustic Analysis of Korean Read Speech - with respect to the prosodic phrasing -

Chuljae Sung

This study aims to suggest some theoretical methodology for analysis of the prosodic patterns in Korean Read Speech. The engineering effort relevant to the phonetic study has focused to the importance of prosodic phrasing which may play a major role in analyzing the phonetic DB.

Before establishing the prosodic phrase as the prosodic unit, we should describe the features of the boundary signal in a target sentence. With this in mind, the general characteristics of Read Speech and the ToBI(tones and Break Indices), which has been currently in vogue with respect to the prosodic labelling system, were presented as the first step. The concrete analysis was carried out with the fable 'North Wind and the Sun' Korean version, where about 25 prosodic units were discriminated by perceptual approach for 5 subjects.

Establishing various informations which can be used for deciding a boundary position systematically, we can proceed to the next, viz. acoustic analysis of prosodic unit. The most important which we primarily study for improving the naturalness of synthetic speech may be, at first, detecting the boundary signals in the speech file and accordingly reestablishment it within the raw text.

## 1. 머리말

본 연구는 한국어 낭독체 음성의 운율적 특징을 파악하기 위한 연구의 방법론을 제안하는 것을 그 목적으로 한다. 일상 생활에서 의사소통의 목적으로 사용되는 대화체 음성(Dialogue Speech)과는 달리 낭독체(Read Speech) 발화는 상대적으로 안정적이며 고른 운율패턴을 보여준다. TTS(Text-to-Speech) 음성합성기에서 구현시키고자 하는 일차적인 음성형태가 낭독체이니 만큼 대화체 합성방식으로 나아가기 전에 낭독체 발화의 운율패턴을 정확하게 기술해주는 것이 당면과제이다.

운율패턴 확립과 관계하여, 현재 진행되는 공학적인 연구의 일반적인 추세는, 경계(boundary)를 특징짓는 다양한 자질(feature)들을 언어학적인 층위를 모두 고려하여 선정하고 이를 통계적인 절차를 거쳐 모델링하는 것을 목표로 한다. 이러한 자질 집합을 잘 설정하는 것이 연구를 효과적으로 진행하는 첩경임은 재론의 여지가 없다.

언어학적인 측면에서는, 실제 부려쓰인 언어에서 관찰할 수 있는 역동적인 리듬 현상과 관련지어볼 수 있다. 음악에서 볼 수 있는 것과 같은 정교한 모습은 아니지만 낱말들이 상호 결합하여 무리들. 짓는 모습은 리듬현상의 하나로 간주할 수 있기 때문이다. 이러한 무리짓기에는 여러 가지 언어학적 요소가 총체적으로 작용하고 있다. 통사적 혹은 의미적 단위와 결부된 경우도 있을 것이고, 호흡과 관련된 개인차로 인한 음성적인 무리짓기도 있을 수 있다.

낭독체음성의 연구방법론을 확립하기 위해서 우선적으로 해야할 일은 이와 같은 '무리짓기'의 언어적 본질을 파악하는 것이다. '운율구 만들기(prosodic phrasing)'라는 용어는 무리짓기 현상을 음성적 관점에서 명명해본 것이다. 통사적 단위의 구구조 분석과 마찬가지로 실제 부려쓰인 말에 있어서도 명시적인 음성표현으로 나타나는 운율구(prosodic phrase) 단위를 설정해볼 수 있다는 가정에서 출발한다. 다시말해 '운율구'란 지각적으로 명백한 끊김이 느껴지는 단위이며 음향적으로 일정 수치 이상의 지속시간의 증가나 고저의 변동으로 경계지워진 단위라 할 수 있다.

운율구 단위는 음성적 특징으로 구분되며, 경계의 지표는 여러 가지 양상으로 나타날 수 있다. 긴 묵음구간(silent interval)으로 나타날 수도 있고 경계전 음절의 장음화(preboundary lengthening)로 혹은 경계전 음절 억양의 고저 변동에 의해서 표시될 수도 있다. 요는 '경계가 있다' 라고 하는 지각적인(perceptual) 인지가 중요하며, 이의 공학적인 구현을 위해서는, 인간의 지각적인 판단을 기계가 거의 동일하게 인식할 정도의 수치화된 기준을 마련하는 것이 중요하다.

아래에서는 이러한 관점과 결부된 연구방법론을 나름대로 세워보고자 한다. 먼저 낭독체 문장의 특징을 알아보기 위해 대화체의 일반적 특징을 약술할 것이며, 현재 세계적인 추세로 진행되고 있는 ToBI(Tones and Break Indices)의 필요성, 그리고 우화 '바람과 햇님'에서 관찰될 수 있는 운율구의 음향적 특징을 기술하도록 하겠다.

## 2. 대화체와 대비해본 낭독체 문장의 특징

일반적으로 대화체 발화는 글말의 문법을 상당 부분 파괴하고 순간적인 상황의 논리에 따라 유동적인 부분이 많으며 발화의 속도(speech rate)도 일정하게 기능할 수가 없다. 낭독체 문장은 이와는 달리 어느 정도 일정한 속도가 유지되며 운율패턴도 대화체에 비해 예측가능한 면을 보인다. 낭독체 문장의 특징은 대화체의 두드러진 특징을 열거하면 자연히 파악될 것이다.

### 2.1 대화체 문장의 일반적 특징

- 1) 다양한 간투사의 삽입
- 2) 표현의 탈락
- 3) 단순화된 표현
- 4) 자기 수정(self-correction)
- 5) 머뭇거림
- 6) 의도적인 혹은 그렇지 않은 되풀이(repetition)
- 7) 화제의 전환을 나타내는 다양한 표현
- 8) 가벼운 신음 소리
- 9) 유창하지 않음
- 10) 운율의 역동적인 변화
- 11) 문법적으로 약한 문장
- 12) 주의하지 않은 발화(careless speech)
- 13) 조정되지 않은 발화 속도
- 14) 고저의 큰 변화량
- 15) 다양한 목적이 혼재
- 16) 좁은 초점을 받는 주제화(topicalization-narrowly focusing)
- 17) 필요없는 표현의 삽입

이와 같은 다양한 특징을 나타내는 대화체 음성은 기계적인 음향분석이 용이하지 않을 뿐 아니라 공학적인 모델링 작업도 힘이 드는 게 사실이다. 현재의 합성기술을 이용하여 음성서비스의 질적인 향상을 도모하는 데에는 우선적으로 낭독체 합성의 기술적 우위를 확보하는 것이 급선무이다. 대화체 문장은 음성인식 기술의 향상과 함께 점차 연구대상의 중심 영역으로 이동해가고 있다.

### 3. ToBI 규약

ToBI(Tones and Break Indices)는 영어의 억양 유형 및 운율의 다양한 제 측면을 전사(transcription)하는 체계로서 음성녹음, Fo 곡선, 그리고 총 네개의 층열(tier)로 구성되어 있다. Beckman & Hirschberg(1994)에 의해 제안된 내용을 기초로 간략하게 기술해 보면 다음과 같다.

네 개의 층열은 철자, 높낮이, 낱말의 분리도, 그리고 기타 부가적인 부분으로 구성되어 있다. 먼저 철자(orthographic tier) 층열은 정규 영어 철자를 이용하여 발화 내의 낱말을 전사하며, 높낮이 층열(tone tier)은 기본주파수 곡선에서의 높낮이 자질(tonal property)을 전사한다. 끊김 표지 층열(break-index tier)은 낱말 사이의 분리도(degree of disjuncture: pause)를 그리고 부가 층열(miscellaneous tier)은 침묵이나 웃음 등의 부가적인 부분에 대해 전사한다.

#### 1) 철자층열(The Orthographic Tier)

각 낱말의 철자 전사는 마지막 분절음의 끝부분에 표시되어야 한다.

#### 2) 끊김표지 층열(The Break Index Tier)

끊김 표지는 각 낱말 사이에 위치한, 혹은 문미 낱말과 발화 마지막의 묵음구간 사이에 놓여졌다고 생각되는 연결(juncture)의 정도를 등급매기는 표시이다. 모든 연결은 명시적인 끊김 표지가(break index value)를 할당 받아야만 한다. 즉 쉼(pause)이나 혹은 경계(boundary)의 정도(degree)를 표시하는 것이다. 0에서 4까지의 등급이 있으며 표시는 낱말의 오른쪽 끝에 위치시켜야 한다.

## Will you have marmalade, or jam ?

1 1 1 3 1 4

주저함(audible hesitation)을 나타내는 표시는 끊김표지의 오른쪽에 직접 부가적 표시(diacritic) 'p'로 나타낸다(e.g. '3p', break index 1,2,3와 함께 사용)

## Display all the flights from Baltimore to Dallas which leave after 4 p.m.

1 1 1 1 3- 2p 0 4 2p 2p 3p 2p 4

#### 3) 높낮이 층열(The Tone Tier)

두 유형의 '높낮이'가 표시된다. 하나는 억양 경계와 그리고 나머지 하나는 악센트 받는 음절과 연계되어 있다. 높낮이 표시는 High(H)와 Low(L) 그리고 경계를 의미하는 %, 악센트 받는 음절을 의미하는 \* 를 조합하여 나타낸다. 예를 들면 다음과 같다.

## will you have marmalade, or jam ?

L\* H- L\* H-H%

#### 4) 부가층열(Miscellaneous Tier)

부가 층열은 특정 침묵, 숨소리, 웃음소리, 주저함 등의 간접적인 표현을 나타내기 위해 사용한다. ToBI 규약은 그 형태의 시작과 끝에 event<...event>와 같이 표시한다.

## Will you have marmalade...

L\* . L\*  
1 1 1

cough<cough>

### 3.1 ToBI 규약의 필요성

분절음과 운율은 음성언어를 구성하는 필수요소이다. IPA를 이용한 분절음 전사 못지 않게 중요한 것이 운율을 효과적으로 표기하는 문제이다. 영어는 강세박자 언어의 대표적인 유형으로서 운율을 표기하기가 상대적으로 쉬운 언어라고 할 수 있다. 낱말 하나마다 대표적인 강세표시를 할 수 있다는 것도 큰 강점이다.

음성합성이나 인식에 이용되는 음성자료는, 대용량의 데이터베이스에서 추출된 음성정보를 효과적으로 재단하여 통계를 내고, 높은 확률로 대표값을 매긴 후 사용된다. 개개의 분절음 자체의 세밀한 음향정보와 함께 낱말, 구, 절, 문장 단위로 늘어나면서 부가되는 대단위의 운율 정보 역시 이러한 작업에서 예외가 될 수 없다.

개별음운의 대표적인 음향정보가 대단위 DB를 통해 구축되었다고 하더라도, 한 음운이 음성언어 속에서 역동적인 역할을 수행하기 위해서는, 좌우 환경(context)이 고려되고 문장 내에서의 여러 가지 언어학적 정보가 가미되어야만 한다. ToBI와 같은 운율전사 체계는 대용량의 DB를 이용하여 통계 작업을 수행할 때 절대적인 역할을 담당한다. 문장의 어떤 부분에서 경계가 형성되며 그 부분의 억양유형은 어떠한고 또 어떤 음절이 악센트를 받으며 흔들리는가 하는 문제를 효율적으로 보여주기 때문이다.

이러한 작업은 우선적으로는 손으로 진행되겠지만, 확보되는 DB와 컴퓨터의 훈련(training)양이 증가될수록 어느 정도는 자동적인 표지 매김이 가능하게 될 것이다. 합성음의 품질을 높이기 위한 첫번째 처방은, 현재로서는 운율구를 생성하는 메커니즘을 음향적인 관점에서 효율적으로 기술하는 일이다. 이러한 작업을 위한 전단계가 ToBI와 같은 전사체계를 마련하는 것이다.

## 4. '바람과 햇님'의 음향분석

### 4.1 실험목적

글말의 통사분석에 비유될 수 있는 소리말에서의 운율구 단위 매김은 한 나라 말의 리듬 현상을 파악함에 있어 첫 번째 작업이며 이는 음성학적으로 중요한 의의를 가진다. 한국어의 낭독체 발화에 있어서의 운율 현상을 이러한 운율구 생성과 관련지어 음향적인 관점에서 기술해 보고자 한다. 연구의 출발점은 운율구를 구분짓는 것이며 일단은 청취적 판단에 의한 지각적인 정의로 시작한다.

### 4.2 실험자료

우화 중 하나인 '바람과 햇님'을 택했다.

「바람과 햇님이 서로 힘이 더 세다고 다투고 있을 때 한 나그네가 따뜻한 외투를 입고 걸어왔습니다. 그들은 누구든지 나그네의 외투를 먼저 벗기는 이가 힘이 더 세다고 하기로 결정했습니다. 북풍은 힘껏 불었으나 불면 불수록 나그네는 외투를 단단히 여몄습니다. 그때에 햇님이 뜨거운 햇볕을 가만히 내려찍니 나그네는 외투를 얼른 벗었습니다. 이리하여 북풍은 햇님이 둘 중에 힘이 더 세다고 인정하지 않을 수 없었습니다.」

### 4.3 피실험자

서울 출신의 20 대 중반에서 후반 나이의 서울대학교에 재학중인 남자 5명(PSJ, PCW, HSH, HDH, LDW로 약칭)이 참가하였다.

### 4.4 실험절차

총 9부의 카드로 녹음을 하였는데, 피실험자 하나마다 첫번째와 마지막 부분을 제외한 7회분의 녹음이 분석대상으로 사용되었다. 녹음은 서울대 언어학과의 음성실험실에서 Inkel Digilink II 모델명 DD-3010 C 카세트테이프를 이용하였으며, Shure의 Unidyne III 545 D 다이내믹 마이크를 사용하였다. KAY ELEMETRICS의 CSL(Computerized Speech Lab) 4300 B를 이용하여 16 KHz 샘플링(sampling), 16 bit 양자화(quantization) 하여 A/D변환시켰다.

실험자료의 음향분석을 위하여 세가지 운율변수 모두가 고려되었는데 CSL을 이용하여 길이(duration), 기본주파수(Fo), 세기(dBSPL)가 측정되었다. 기본주파수의 측

정은 모음시작(vowel onset)에서부터 2/3 되는 지점을 측정하였다. 소리의 높이(pitch)를 분석하기 위해서 여러가지 방법을 이용할 수 있으나 현재까지 완전한 것은 없다고 한다. 따라서 정확한 기본주파수 검출을 위해, 오류가 발생한 부분은 CSL 화면위에서 직접 pitch marker(Impulse marker)를 수정하였다. 세기(intensity)는 분석 대상 음절의 진폭 중 가장 값이 높게 나타나는 지점을 측정하는 정상강도 측정법을 이용했다. 통계처리는 7회 반복한 자료들을 기준으로 시행되었으며 맥킨트쉬의 통계 처리 프로그램인 Statview를 이용하였다.

꾸어읽은 모습은, 음성학 수업을 받은 서울대 언어학과 대학원 및 대학에 재학 중인 21명의 학생들을 통한 청취실험의 통계와 CSL 프로그램을 통한 음향분석결과의 통계를 이용하여 유형화하였다. 청취실험의 내용은 꾸어읽은 부분의 확인과 더불어, 청취실험에 참가한 21명 각각이 결정한 운율구 내부에서 가장 돋들리는 음절을 찾으라는 것이다.

통계의 과정을 거쳐 피실험자별로 결정되는 운율구는 7회분의 자료 전체를 대상으로 하여 산출된다. 예를 들어 피실험자 PCW의 경우, 첫회에 21개, 두번째에 24개, 세번째 23개, 네번째 24개, 다섯번째 24개, 여섯번째 24개, 일곱번째 24개의 운율구를 만들어내었는데, 이 모두를 더한 164개의 운율구가 통계의 대상이 된다.

#### 4.5 실험결과 및 토의

##### 4.5.1 발화속도 및 운율구 내부의 음절갯수

피실험자 각각의 발화 속도(speech rate)는 대략 다음과 같다.

표 1. 피실험자별 전체발화시간(total-duration)과 쉼시간(pause duration), 1분당 차지하는 음절수, 1초당 차지하는 음절수, (시간의 단위는 msec(1sec =1000 msec), M=arithmetic mean, s=standard deviation, n=7)

inf.	t-dur		p-dur		n-syl/min		n-syl/sec	
	M	s	M	s	M	s	M	s
PSJ	33648	1694	6897	1211	300.2	14.9	5.0	.3
PCW	32738	1265	7931	1643	308.3	12.8	5.1	.2
HSH	32574	1221	7612	512	309.8	11.9	5.2	.2
HDH	33087	1036	6849	835	304.9	9.8	5.1	.2
LDW	29815	1096	4517	994	337.4	10.8	5.6	.3

다섯명 피실험자 모두 대략 29.8 초- 33.6 초 정도 사이에 전체 내용을 발화하였으며, 1분당 약 305 음절 정도를 말하였고, 1초 당 차지하는 음절 수는 5개 정도 그

리고 한 음절의 평균길이는 150 msec에서 변하고 있음을 알 수 있다. 발화속도는 LDW가 가장 빠르며 1분당 음절수도 337개 정도로 가장 많은 음절을 발화하였다. HDH는 피실험자 중 가장 안정된 발화속도를 유지한 것을 알 수 있다.

표 2. 운율구 안의 음절수 빈도조사(Count Frequency(%))

inf.	count frequency- syllable number(%)
PSJ	5.6 - 6.9(33.1%), 6.9 - 8.2(26.5%)
PCW	5.4 - 6.6(30.5%), 6.6 - 7.8(28%)
HSH	5.9 - 7.2(56.5%), 4.6 - 5.9(11.9%), 7.2 - 8.5(11.3%)
HDH	5.3 - 6.4(31.3%), 6.4 - 7.5(26.3%)
LDW	5.2 - 7.4(34.6%), 7.4 - 9.6(15%)

표 2는 각 피실험자당 그들이 만들어낸 운율구 내부에서 가장 높은 빈도로 출현한 음절의 수와 백분율을 나타내본 것이다. 대부분의 경우 대략 5음절에서 8음절 사이의 음절수가 50 % 이상의 비율을 차지하고 있는 것으로 나타났다. 3.4 조로 이루어진 우리 전통시가의 울격을 생각나게 하는 대목이다.

#### 4.5.2 운율구 내부현상의 음향분석

각 피실험자들은 7회의 낭독을 통해서 매번 약간씩의 변이를 보였는데 이는 매번의 자료에서 생성되는 운율구의 수가 서로 다름을 의미한다. 아래 표 3에서 피실험자 모두의 운율구 숫자를 각 발화횟수별로 정리하였다. 총갯수와 평균값을 통해서 피실험자별 특징을 파악할 수 있을 것이다. 영어의 경우 Uldall(1971, 1972)이 이용한 'The North Wind and the Sun' 자료는 D. Abercrombie가 녹음한 것으로 약 45초 정도의 발화길이를 보여준다. 리듬단위는 56개(metric feet) 정도이며 260에서 870 msec 까지의 변이를 보인다.

표 3. 각 피실험자의 발화회수별 운율구의 수

inf. \ session	1	2	3	4	5	6	7	SUM	MEAN
PSJ	22	22	23	20	22	22	20	151	21.6
PCW	21	24	23	24	24	24	24	164	23.4
HSH	26	22	25	25	26	27	26	177	25
HDH	30	29	28	27	29	28	27	198	28.3
LDW	16	17	19	21	18	18	18	127	18.1



(1) 운율구 안에서 음향변수의 최대값을 나타내는 음절의 위치 분석  
(변수 별 파악 %)

운율구 안의 음절위치와 관련하여 첫음절과 둘째 음절 부분의 경우는 기본주파수와 세기의 최대값이 집중되는 경향을 보였으며 마지막 부분은 압도적으로 지속시간이 길어진다. 여기서는 지속시간의 최대값 자체에 의미를 두었으므로 마지막 음절부분의 통계적 가치도 동등하게 생각한다. 듣는이의 청취적 지각판단과는 상관없는 냉정한 물리적 수치의 산출과 관계되기 때문이다(그림 1-3).

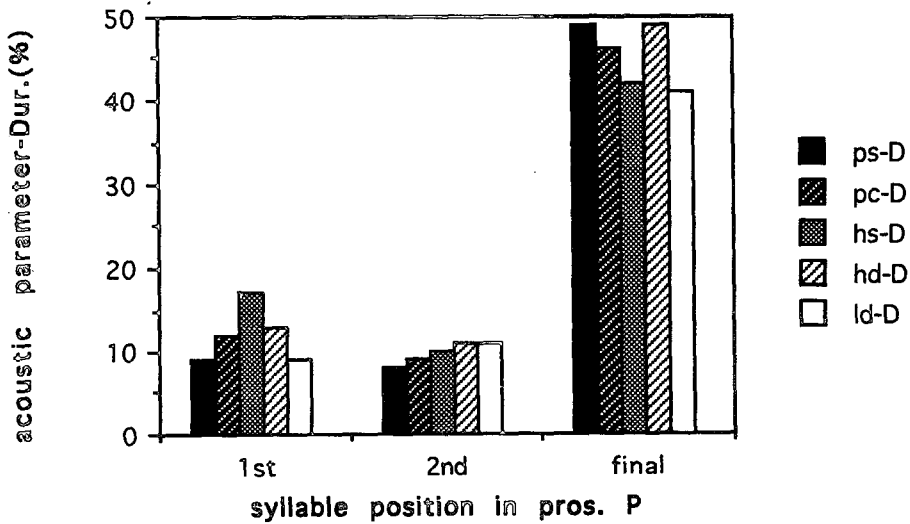


그림 1. 운율구 안에서 지속시간의 최대값을 나타내는 음절의 위치 분석(%)

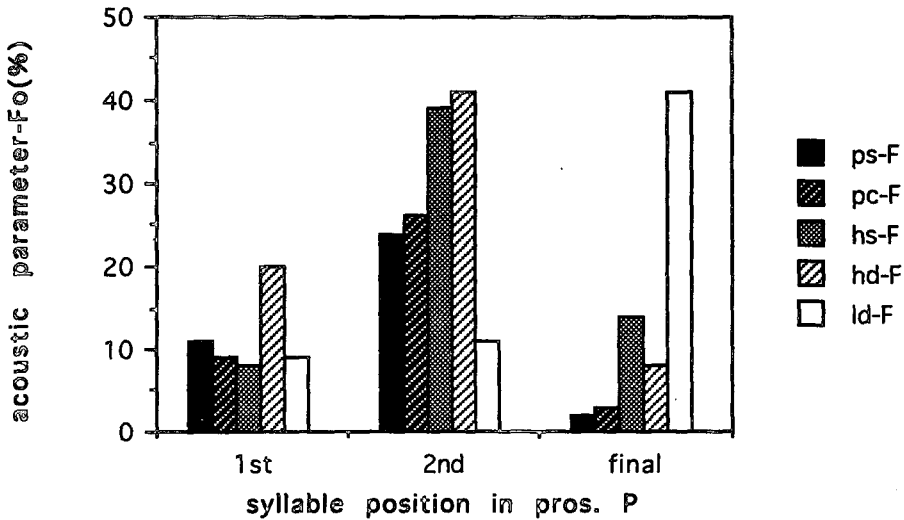


그림 2. 운율구 안에서 Fo의 최대값을 나타내는 음절의 위치 분석(%)

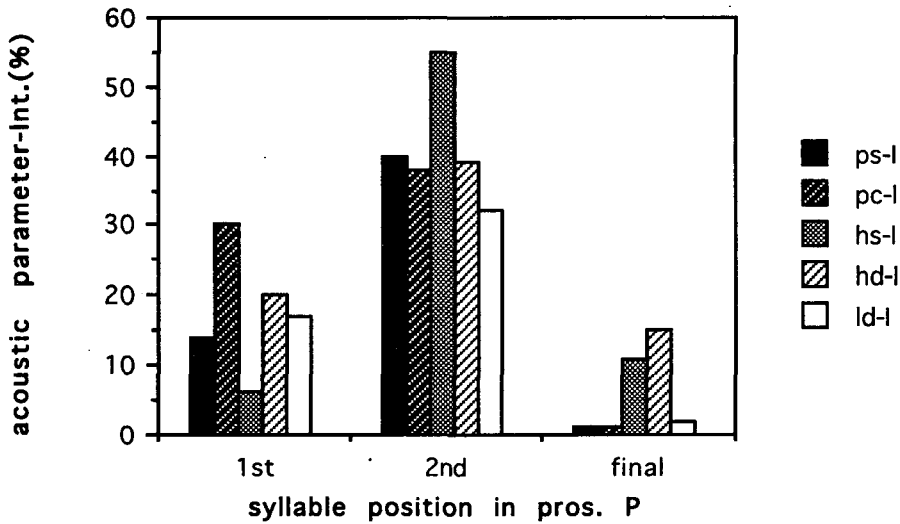


그림 3. 운율구 안에서 세기의 최대값을 나타내는 음절의 위치 분석(%)

(2) 음향 변수의 최대값이 내용어 및 조사(토씨)에 붙는 경우의 빈도(%)

가장 높은 수치를 갖는 음향변수가, 어절의 앞부분을 차지하는 내용어에 일치될 때 그리고 마지막 부분의 문법적 형태소에 나타날 때 어떻게 실현되는지가 문제의 핵심이다. 결과, 내용어 부분에선 기본주파수와 세기의 최대값이 집중되는 경향을 보였고 토씨 부분에선 지속시간이 압도적으로 많이 등장한다. 이는 한 운율구 내부에서 지속시간은 마지막 부분에서 거의 규칙적으로 그 최고수치를 실현시킴을 의미한다. 흔히 예상하는 것처럼 기본주파수의 최대값이 규칙적으로 나타날 것이라는 생각은 이러한 결과를 토대로 할 때 무리한 주장임을 알 수 있다.

음향변수의 최대값이 자리잡는 위치는, 운율구 내를 기준으로 한다고 할 때 기본주파수의 최대값이 운율구 앞쪽의 둘째음절부분에 약 35% 정도 집중되는 것을 그림 2를 통해서 파악할 수 있다. 어말 문법형태소 부분의 억양상승에 대한 지각은 그 앞음절(penultimate)의 하강과 연이은 상승의 인상 때문으로 풀이할 수 있겠다(윤일승 1992:96)(그림 4-5).

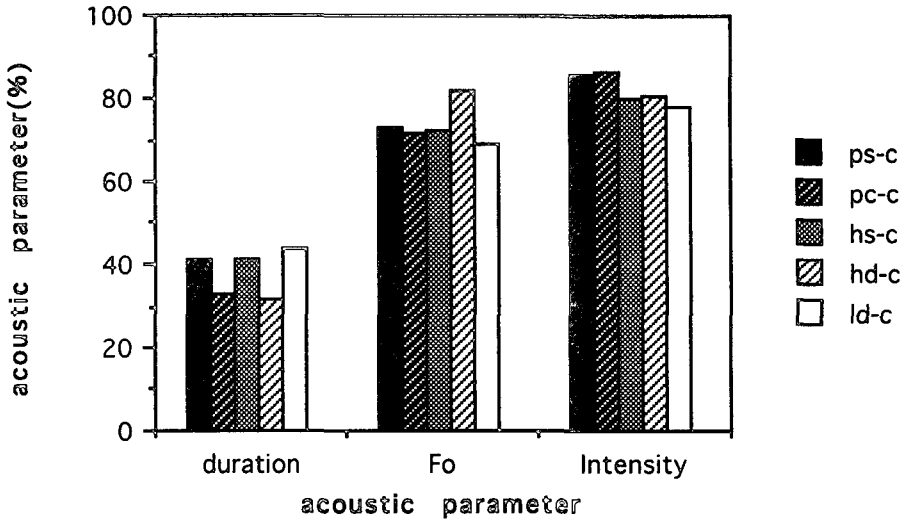


그림 4. 음향 변수의 최대값이 내용어에 붙는 경우의 빈도(%)

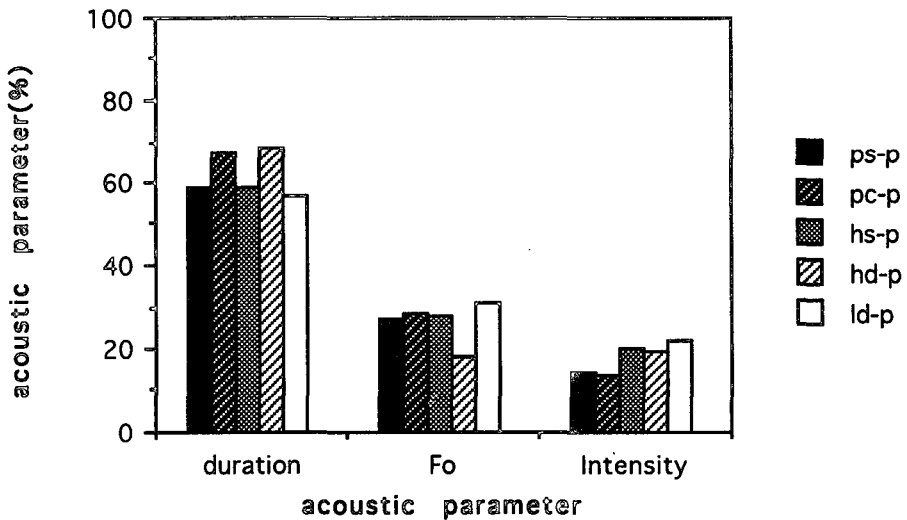


그림 5. 음향 변수의 최대값이 토씨에 붙는 경우의 빈도(%)

(3) 음향 변수의 최대값이 토씨에 붙은 경우, 그 토씨가 실현하는 문법관계의 빈도 (%)

음향 변수의 최대값이 토씨에 붙은 경우, 그 토씨로 인해 실현되는 문법관계(흔히 격의 의미로 파악하는 주격, 목적격, 속격 등의 의미)의 빈도를 조사해보았다. 혹시라도 음향변수의 최대값 실현과 문법관계가 어떤 상관관계를 가지고 있지는 않나하는 물음에서 출발한 것인데, 결론을 내리기에는 미흡하나 나름대로의 시도로 생각하

면 좋겠다. 이러한 시도에 대한 결론적인 해결책은 역시 '다양한 자료'의 코퍼스로 구축된 데이터베이스가 있어야만 되는 것이다.

그림 6 - 8을 통해 알 수 있는 것은 피실험자 모두 지속시간과 관련하여서는 주격을 나타내는 '은,는,이,가' > 목적격을 나타내는 '을,를' > 인용의 특수토씨 '고' 등의 순으로 최대값의 집중빈도를 나타내었고, 다른 운을변수와 관련하여서는, 기본주파수와 세기 둘다 주격의 '은,는,이,가'와 '와,과' 부분에 집중됨을 알 수 있다. 물론 언급한 것처럼 결론적인 얘기가 아니기 때문에, 데이터베이스의 크기에 따라 다른 결과가 나올 수도 있는 가능성을 인정할 수 있다(그림 6 - 8).

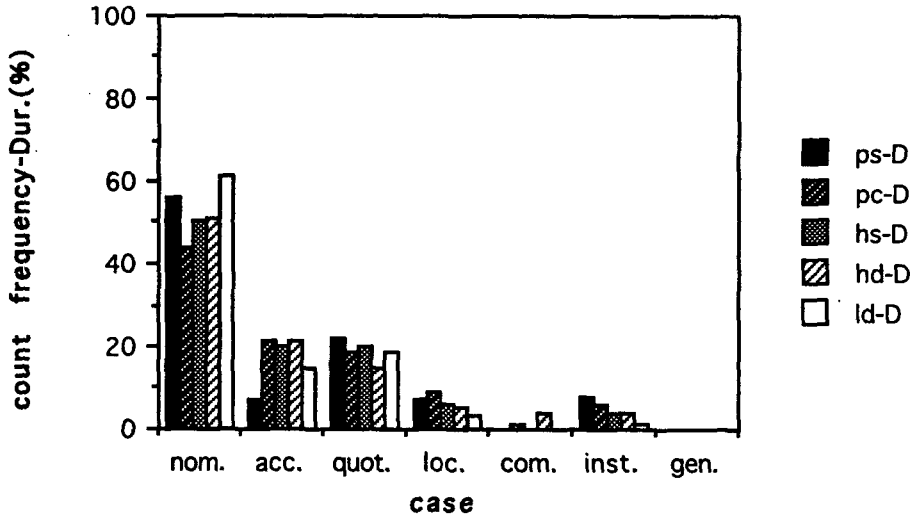


그림 6. 지속시간의 최대값이 토씨에 붙은 경우, 그 토씨가 실현하는 문법관계의 빈도

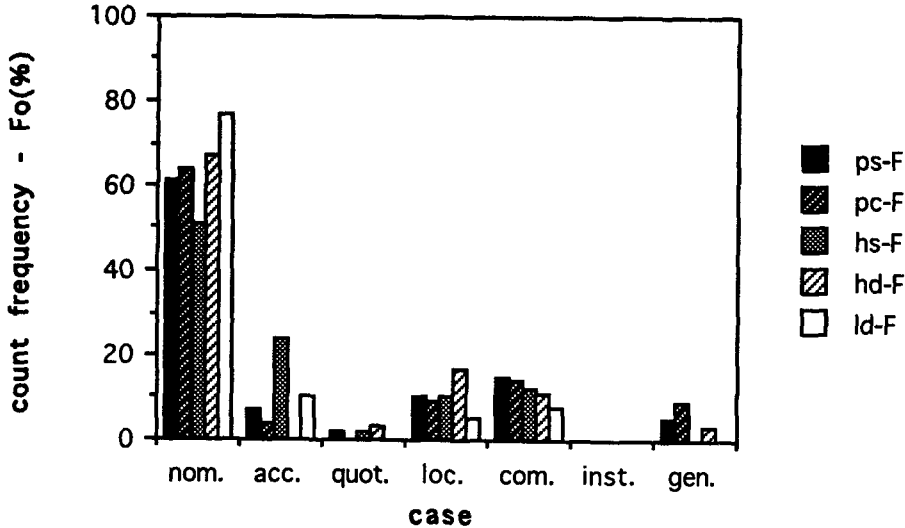


그림 7. 기본주파수의 최대값이 토씨에 붙은 경우, 그 토씨가 실현하는 문법관계의 빈도(%)

한국어 낭독체 문장의 음향분석

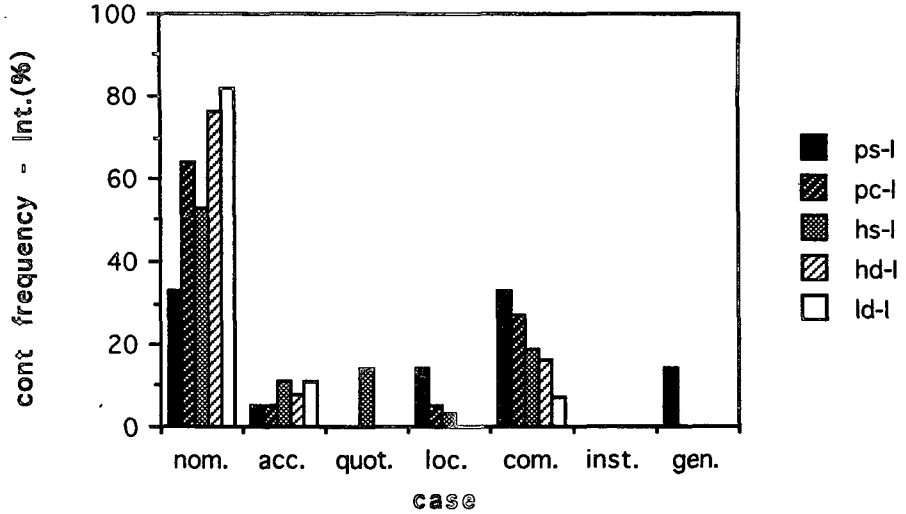


그림 8. 세기의 최대값이 토씨에 붙은 경우, 그 토씨가 실현하는 문법관계의 빈도 (%)

### 5. 토의

글말에서 통사적인 분석을 시도할 때의 구단위 생성(syntactic phrasing)에 착안하여 소리말에서의 이와 같은 작업을 '운율구 만들기(prosodic phrasing)'라 부르기로 하였다. Kozhevnikov & Chistovich(1965)에서 설정한 'syntagma'는 조음프로그램이 작동되는 조음의 단위로서 대략 7개의 음절로 구성되어 있다고 했다. 아울러 쉼(pause)으로 그 경계가 구분되는 의미의 단위이고 문장 혹은 문장의 하위부분이라는 정의를 하고 있다. 그들 논의의 핵심은 이러한 7음절 길이의 단위가 작동기억(operational memory)의 한계와 관련되어 있다는 주장이다. 낭독체에서의 '운율구'의 길이가 대략 그 정도이고 보면 생각해볼만한 내용이다.

이러한 논의와 더불어 운율구 내부에서의 음향변수의 추이를 운율구 내에서의 위치, 문법관계, 문장성분 등과 관련해서 파악해보았다. 지속시간은 어떤 논의와 관련해서도 중요한 성분으로 파악되었으며 위치에 따라서 상당히 규칙적인 실현을 보였다.

지속시간이 가장 긴 부분은 어떤 단위든 대체로 마지막 음절부분에서 나타나며 세기와 기본주파수의 경우는 그 최대값이 집중되는 부분이, 둘째 음절 위치와의 연관성을 많이 보여주고 있다. 지속시간과 관련하여 Wenk & Wioland(1982)의 불어에 대한 기술과 비슷한 면모를 보인다. 또한 이러한 관찰은 구 회산(1986)과 성 철재(1991)의 무의미 반복 낭독형 낱말을 사용한 실험과 어느 정도 같은 모습을 보여주는 부분이라 하겠다. 대개 기본주파수의 상승은 세기의 증가도 유도하므로 이러한 결과가 나왔을 것이다. 문법성분과 관련하여 토씨부분의 음향적인 특징을 파악해보았는데 대용량의 데이터 속에서는 다른 결과가 나올 수도 있다.

이러한 시도 및 관찰은 머리말에서 밝혔던 것처럼 낭독체 문장의 운율을 음향적으로 분석하는 데 있어서의 방법론 확립과 관계되어 있다. 첫번째 작업은 경계를 확정하는 것이며 둘째는 확정지어진 경계 단위 내부의 현상을 효과적으로 기술하는 작업이었다.

경계 확정에 사용되는 요소는 이미 밝혔듯이 지속시간과 고저 유형의 변동이며, 향후 과제는 자동적인 경계 확정을 위해서, 경계 유형을 ToBI의 끊김표지(break index)와 같이 몇가지로 등급매기는 작업이 필요하다. ToBI는 어느 정도 주관적인 판단이 가미되는 언어학적 체계인데 공학적인 구현을 위해서는 수치화시킨 후 통계적인 확정을 해야만 한다.

문장성분이나 형태소 분석 등도 운율구 확정에 유용한 정보라 할 수 있다. 특히 낭독체 문장에서는 통사적 위계나 의미적 단위가 상호 밀접히 연관되어 머리 속에서 배열될 것이므로, 어느 정도의 일탈은 있다하더라도 많은 부분 문법적인 장치에 따라갈 것이기 때문이다. 현재 세계적인 연구의 추세는 경계부분에서의 다양한 특질을 추출하여 기술하는 것과 경계 확정을 위해 음향적 정보 뿐만 아니라 문법적 정보도 컴퓨팅 파워를 이용하여 통계를 내는 것이다.

본 논문에서의 운율구 단위는 청취적 끊김을 이용한 지각적 단위였지만 이의 객관적 이용을 위해서는 수치적인 기준을 설정해야한다고 하였다. 예를 들면 묵음구간(silent interval)이 50 msec 이상일 때는 끊김표지 '3'을 주고, 묵음구간이 없이 경계 앞음절 장음화만으로 경계가 인지될 경우 그 음절 지속시간이 평균값의 15 % 정도를 초과했을 경우, 그렇지 않을 경우를 나누어 표지를 할당하는 식이다. 이런 방식의 접근은, 경계 인지에 관련된 다양한 특질을 자국어 정보내에서 추출하여 실질적으로 이용가능한 단위설정으로 귀결되어야 한다. 경계부분에서 추출해 볼 수 있는 특질을 들면 다음과 같다.

- 1) 경계 앞 분절음, 음절, 낱말의 평균 지속시간, 고저
- 2) 경계 앞 분절음, 음절, 낱말의 최대값
- 3) 경계 앞 분절음, 음절, 낱말의 최소값
- 4) 경계 뒤 분절음, 음절, 낱말의 평균지속시간, 고저
- 5) 경계 뒤 분절음, 음절, 낱말의 최대값
- 6) 경계 뒤 분절음, 음절, 낱말의 최소값
- 7) 후속 묵음구간의 지속시간
- 8) 경계 뒤 음절의 평균 Fo에 대한 경계 앞 음절의 최대값 비율
- 9) 경계 앞 음절의 평균 Fo에 대한 그 음절 자체의 최대값/최소값 비율
- 10) 정규화(normalize)된 음절 지속시간에 대한 경계 앞 음절의 지속시간 비율
- 11) 문장 평균 Fo에 대한 어말 Fo의 비율
- 12) 경계 전/후 문장성분의 결합도
- 13) 토씨의 유형과 관련된 음향적 유동성

운율구 내부의 정보와 관련하여 가장 유용한 것은 돌출립(prominence) 정보라 할

수 있다. 돌출성(prominence)을 파악하는 자질로 다음을 열거할 수 있다.

- 1) 위치와 관련된 정보(positioning effect)
- 2) 모음의 종류
- 3) 낱말의 종류- 음절구성, 분절음 구성과 관련
- 4) 품사정보
- 5) 낱말의 음절수
- 6) 문장 성분
- 7) 통사적 위계
- 8) 의미적 고려
- 9) 다음 운율구의 음절수

운율구 경계라든지 돌출성 확정에 필요한 이와 같은 내용은 자연스러운 합성음을 만들어내는 데 있어 중요한 역할을 담당할 수 있는 요소이다. 이러한 내용들을 잘 조직화하여 데이터베이스를 효과적으로 가공할 수 있다면 보다 나은 품질의 합성음을 만들어낼 수 있을 것이다.

## 6. 뺏음말

지금까지 ‘바람과 햇님’ 자료를 이용한 한국어 낭독체 발화의 음향분석을 ‘운율구 만들기’ 작업에 비중을 두어 시도해 보았다. 머리말에서 밝혔듯이 본 논문의 목적은 문장의 효율적 음성분석을 위한 방법론을 제시하는 것이다. 이를 위해서 낭독체 및 대화체 문장의 특징과 운율표기 체계로서의 ToBI를 살펴보고 실질적인 작업의 한 예제로 ‘바람과 햇님’을 분석해 보았다.

다섯명의 결과를 통해 소리말의 실제 발화에서의 경계가 설정되는 모습은, 통사경계 및 의미 단위로의 분할도 중요한 요소이긴 하지만 말하는 사람의 생리적 호흡능력 그리고 어느 정도 무의식적인 음절수 조정노력도 무시 못할 요소임을 알 수 있다.

운율구 경계 짓기는 현 상황에서 가장 효율적으로 음성자료를 가공하는 수단임을 지금까지의 논의를 통해 살펴보았는데, 지속시간이든 억양이든 운율적 요소의 공학적 구현을 위해서는 일차적 선결작업으로 시도되어야 할 부분임을 강조하였다. 구미에서 확정되어 어느 정도 본래도에 올라 있는 ToBI는, 운율구 경계 자체와 그 내부의 조직적인 유형화를 위해서 한국어에 알맞은 운율표기 체계로 알맞게 변형되어야 할 것이다.

이후의 과제는 이러한 정보의 확보를 통해서, 언어학적으로 보다 설득력있는 음성/음운 이론의 검증에 이용하며, 공학적으로는 합성 및 인식에 사용되는 언어 단위의 자질 설정에 효과적인 도움을 줄 수 있도록 하는 것이다

<참고문헌>

- 성철재(1995). <한국어 리듬의 실험음성학적 연구> 서울대학교 언어학과 박사학위 논문.
- 이현복(1986). 한국어 음성의 합성과 인식에 관한 음성언어학적 고찰. <한글 194>. 한글학회. 55-72
- 이호영(1990). The Structure of Korean Prosody. 런던대학교 박사학위 논문.
- 지민제, 이용주(1990). 한국어 Pause Pattern의 음향음성학적 분석. <통신처리를 위한 음성정보 변환기술 개발> 한국전자통신연구소. 86-88.
- Beckman, M. & Hirschberg, J.(1994). "The ToBI Annotation Conventions", manuscript. Ohio State University.
- Kozhevnikov, V.A. & Chistovich, L.A.(1965). Speech: Articulation and Perception. Joint Publications Research Service 30. US Department of Commerce. Washington D.C.
- Ross, K.N.(1995). Modelings of Intonation for Speech Synthesis. Ph.D. dissertation. Boston University.
- Wightman, C.W. & Ostendorf, M.(1994). "Automatic Labeling of Prosodic Patterns", IEEE Transactions on Speech and Audio Processing vol.2 - no 4. 469-481.