

# Peak 파라미터와 피치 검색테이블을 이용한 억양 생성방식 연구

장석복, 김형순

{prime, kimhs}@hyowon.pusan.ac.kr

부산대학교 인지과학협동과정, 부산대학교 전자공학과

## A Study on Generation Method of Intonation using Peak Parameter and Pitch Lookup-Table

Seok Bok Jang, Hyung Soon Kim

Dept. of Interdisciplinary Research Program of Cognitive Science,

Pusan National University,

Dept. of Electronics Engineering, Pusan National University

### 요 약

본 논문에서는 Text-to-Speech 시스템에서 사용할 억양 모델을 위해 음성 DB에서 모델 파라미터와 피치 검색테이블(lookup-table)을 추출하여 미리 구성하고, 합성시에는 이를 추정하여 최종 F0 값을 생성하는 자료기반 접근방식(data-driven approach)을 사용한다.

어절 경계강도(break-index)는 경계강도의 특성에 따라 고정적 경계강도와 가변적 경계강도로 세분화하여 사용하였고, 예측된 경계강도를 기준으로 억양구(Intonation Phrase)와 액센트구(Accentual Phrase)를 설정하였다. 특히, 액센트구 모델은 인지적, 음향적으로 중요한 정점(peak)을 정확하게 모델링하는 것에 주안점을 두어 정점(peak)의 시간축, 주파수축 값과 이를 기준으로 한 앞뒤 기울기를 추정하여 4개의 파라미터로 설정하였고, 이 파라미터들은 CART(Classification and Regression Tree)를 이용하여 예측규칙을 만들었다. 경계음조가 나타나는 조사, 어미는 정규화된(normalized) 피치값과 key-index로 구성되는 검색테이블을 만들어 보다 정교하게 피치값을 예측하였다. 본 논문에서 제안한 억양 모델을 본 연구실에서 제작한 음성합성기를 통해 합성하여 청취실험을 거친 결과, 기존의 상용 Text-to-Speech 시스템에 비해 자연스러운 합성음을 얻을 수 있었다.

### 1. 서론

Text-to-Speech 시스템에서 명료도(intelligibility)와 자연성(naturalness)을 향상시키기 위한 최근의 노력들은 자국어의 특징을 고려한 운율(prosody)의 모델링이라는 문제로 모아지고 있다. 운율은 인간이 발화할 때 부가하는 청각적 변화요소로서 억양(fundamental frequency, F0, pitch), 길이(duration), 세기(intensity) 등으로 나타낼 수 있으며, Text-to-Speech 시스템에서는 이러한 운율 요소를 적절히 조절함으로써 지루하고 단순한 음성이 아니라 자연스러운 발화의 특성을 살릴 수 있다.

그 중에서도 억양은 높낮이의 변화를 통해 문장을 의미적 단위들로 묶거나, 돌출됨(prominence)의 상대적 차이를 통해 특정한 단어에 주의를 기울이게 만듦으로써 발화된 음성에 대한 청자의 이해를 돕는 중요한 요소이다. 억양과 관련된 선행 연구들은 규칙기반 방식(rule-based approach)[1], 신경망(neural networks)을 이용한 학습[2], 선형회귀(linear regression) 방식[3], CART[4] 등의 다양한 방법을 이용하여 억양에 대한 접근과 생성방식을 연구해 왔다.

본 논문에서는 모든 음절, 어절이 고유한 운율현상을 가지고

있는 것이 아니라 발화문장의 언어적 분석과정을 통한 운율의 계층적 구조 안에서 몇 개의 액센트 성분으로 단순화시킬 수 있으며, 그 외 음절과 어절들은 이 액센트 성분들과의 관계로 해석하여 억양을 나타낼 수 있다는 접근방식을 택하였다.

따라서, 입력문장을 분석하여 어절 경계강도를 부여하고 이에 기반한 억양구와 액센트구조로 운율 구조를 설정하였다. 그리고 비경계 액센트로 볼 수 있으며 인지적, 음향적으로 중요한 요소인 돌돌림을 나타내는 정점에 대한 모델 파라미터를 설정하고, 경계 액센트로 볼 수 있는 구 경계 음절들에는 피치테이블을 구성하여 추정하는 방법을 사용한다.

이러한 파라미터와 피치테이블은 운율 음성 DB에서 모델 생성에 필요한 파라미터들을 미리 추출하여 구성하고 이를 합성시에 추정하여 합성음을 만들어내는 자료기반 접근방식을 사용하여 설정하였다.

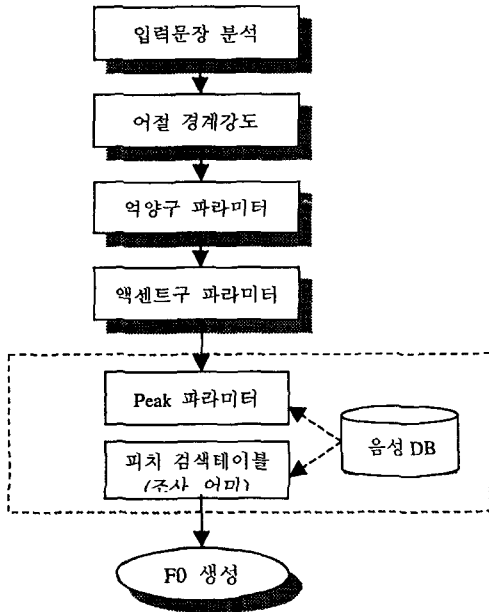


그림 1. F0 생성의 흐름도

## 2. 어절 경계강도 생성

### 2.1. 음성 DB 및 문장 분석

운율모델 생성을 위한 데이터베이스는 현직 여성 아나운서

를 통해 방송 부스에서 녹음하였으며, 발생 목록은 뉴스, 일기 예보, 교통안내, 신문사설, 소설, 교과서 등에서 골고루 추출하였다. 문장의 길이는 최소 1 어절에서, 최대 20 어절의 길이를 가지고 있으며 총 어절 수는 문장부호를 제외하고 2786 개였다. 전체 문장은 400 문장으로서 이중 훈련문장은 250 문장을 사용했으며 검증문장은 150 문장을 사용하였다. 음성의 음소 단위 레이블링은 본 연구실에서 개발한 자동음성 분할시스템을 통해 레이블링 한 후 수동으로 수정하였다.

F0 생성은 다음과 같은 과정을 거쳤다. 먼저 5 ms 간격으로 F0을 추출하여 15 point median filtering을 거친 후 다시 7 point median filtering을 통해 피치값을 smoothing하였다. 그리고, 무성음 구간은 선형 연결(linear interpolation)하였다. 그 후 10 ms 간격으로 피치값을 추출하였다.

본 논문에서 품사는 어절단위로 할당하는 것이 아니라 각 어절에 대하여 어근, 어간의 실질형태소를 머리품사(Head Part-of-Speech)로, 조사, 어미의 문법형태소는 꼬리품사(Tail Part-of-Speech)로 할당하는 방식을 취한다. 분석에 사용한 품사집합은 운율적 행동들이 비슷한 품사들을 묶어서 단순화시켰으며, 머리 품사로 쓰이는 것은 체언, 의존명사, 용언, 보조용언, 관형사, 부사, 접속부사, 감탄사, 삽입표와 마침표의 문장부호를 합해서 10 가지이며, 꼬리 품사는 용언, 보조용언을 제외한 머리품사 8 개와 조사 8 개, 어미 5 개를 사용하였다. 구문구조는 어절간의 지배, 의존관계로 나타내었다.

### 2.2. 어절경계강도 설정

어절 경계강도는 발화된 음성을 청취할 때 느끼는 어절간의 운율적 이질감으로서 객관적이고 물리적인 현상이기 보다는 심리 음향학적 현상이라고 할 수 있으며, ‘끊어읽기 색인’, ‘운율 경계강도’ 이라고도 한다. 이 경계강도에 따라서 휴지기의 유무와 길이, 경계음절의 길이 및 억양 변화가 특징적으로 나타나기 때문에 어절 경계강도를 설정하고 예측하는 것은 합성음의 자연스러움에 매우 큰 영향을 미친다.

본 논문에서의 어절 경계강도에 대한 기본적인 접근방식은 어절간의 관계에서 화자의 차이, 문장의 구조, 발화속도, 문맥 상황에 비교적 덜 영향을 받는 관계로서 어떤 환경에 오더라도 휴지기가 와야 한다든지, 이어서 읽어야 한다든지 하는 운율

구조가 지켜져야 자연스러운 것과 그렇지 않은 것을 분리하는 것이다.

표 2는 어절 경계강도를 음성 DB의 청취실험과 품사정보를 통하여 나타낸 것이다.

표 1. 어절 경계강도

	고정적 경계강도	가변적 경계강도
변이 요소 (화자, 발화속도, 문맥 상황)	비교적 영향을 덜 받음	영향을 많이 받음
위반 시	의미전달에 지장을 주어 어색함	의미전달에 영향 없음
기호 (BI)	BI-0, BI-4, BI-5	BI-1, BI-2, BI-3
예	용언의 관형형 + 의존명사	목적격 조사 + 용언

고정적 경계강도는 어절과 어절 사이의 결합강도가 비교적 정해져 있어, 이를 어졌을 경우 의미전달에 지장이 있거나 어색하게 들리는 어절간의 관계를 일컫는다. 'BI-0'은 섬이 없는 어절 경계이며, 'BI-4'는 문장 내에서 큰 휴지를 가지고 끊어 읽어야 할 부분으로 역양구를 가리킨다. 'BI-5'는 문장 마지막 어절에 부여하는 어절 경계강도이다.

가변적 경계강도는 화자, 발화속도, 고정적 경계강도와 관계, 문장 내에서의 위치에 따라 경계강도의 크기가 달라질 수 있다. 예를 들면, '시원한 물을 먹다'에서 발화속도를 빨리 할 때는 '물'과 '먹다' 사이에 휴지가 없을 수도 있고, 발화속도를 천천히 할 때는 휴지가 있을 수도 있다. 즉, 꼭 붙여서 읽어야 할 어절 관계를 제외하고 어떠한 어절이라도 경계음절로서 기능할 수 있다는 것을 보여준다. 이러한 어절 관계들을 한 부류의 어절관계로 묶을 수도 있으나, 품사 관계와 경계음절의 특성인 휴지길이, 경계음절 장음화 등의 경향을 분석하여 추가적인 변별적 자질들을 부여함으로써 운용제어에 보다 융통성을 갖게 되는 장점이 있다.

결국, 운용구조에 있어 고정적이며 의미전달에 중요한 어절 경계강도는 통계적 방법 등을 통해 정확하게 먼저 모델링한 후, 상대적으로 가변적이며 덜 중요한 어절 경계강도는 먼저 정해진 경계강도와 관계와 품사 bigram 등을 이용하여 간단한 방법으로 처리할 수 있다.

표 2. 음성 DB에서의 어절 경계강도의 분포

어절경계강도	훈련데이터	검증데이터
BI-0	522 (30.35)	333 (31.24)
BI-1	328 (19.07)	211 (19.79)
BI-2	266 (15.47)	171 (16.04)
BI-3	141 (8.20)	69 (6.47)
BI-4	213 (12.38)	132 (12.38)
BI-5	250 (14.53)	150 (14.07)
합	1720	1066

### 2.2.1. 고정적 어절경계강도 예측

고정적 어절경계강도는 가변적 어절 경계강도인 'BI-1', 'BI-2', 'BI-3'을 하나의 부류로 묶어서 'BI-123'로 나타내고 'BI-0', 'BI-123', 'BI-4', 'BI-5'로 나누어서 예측하였다. 예측방법은 Tree기반 모델링기법 중 하나인 CART를 이용하였다. CART는 패턴분석이나 회귀분석을 통해 운율구조의 설정, 휴지길이 추정 등의 운율구조 생성에 매우 유용하게 사용되고 있는 통계적 방법이다[4].

예측결과를 보면 종결 어절인 'BI-5'를 제외하고, 훈련 데이터에서는 11.02%의 예측 오류율을 나타내었으며, 검증 데이터에서는 16.20%의 예측 오류율을 나타내었다.

표 3. 고정적 어절경계강도 예측

훈련데이터	Predicted Class			
	BI-0	BI-123	BI-4	BI-5
Actual Class				
BI-0	473 (90.61)	41 (7.85)	8 (1.53)	0 (0)
BI-123	61 (8.31)	612 (83.38)	61 (8.31)	0 (0)
BI-4	4 (1.88)	11 (5.16)	198 (92.96)	0 (0)
BI-5	0 (0)	0 (0)	0 (0)	250 (100)

검증데이터	Predicted Class			
	BI-0	BI-123	BI-4	BI-5
Actual Class BI-0	291 (87.39)	37 (11.11)	5 (1.50)	0 (0)
Actual Class BI-123	35 (7.74)	375 (82.96)	42 (9.29)	0 (0)
Actual Class BI-4	7 (5.30)	18 (13.64)	107 (81.06)	0 (0)
Actual Class BI-5	0 (0)	0 (0)	0 (0)	150 (100)

그림 2를 볼 때, 관측어절 꼬리품사(RPos), 관측어절 다음어절의 품사정보(NLPos, NRPos), 의존 문장 트리에서의 어절 간 Depth의 차이(ReNext), 끝어절과의 어절거리(WordToLast), 음절거리(SylToLast) 등이 고정적 어절 경계강도 예측에서 중요도가 높은 특징 변수이었다. 결정트리에서의 종결노드의 개수는 19개였다.

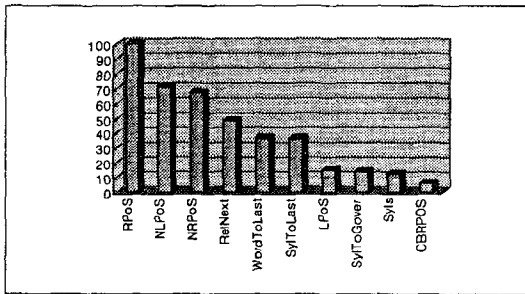


그림 2. 고정적 어절경계강도 예측에서의 변수 중요도

### 2.2.2. 가변적 어절경계강도 예측

가변적 어절 경계강도는 고정적 어절경계강도의 예측결과에서 'BI-123'로 예측된 어절에 대하여 'BI-1', 'BI-2', 'BI-3'으로 경계정보를 확장하는 단계이다. 주로 품사 bigram을 이용하여 휴지 길이와 피치 변화 등의 어절 경계적 특성이 큰 것에 높은 값을 할당한 후, 고정적 어절경계강도의 개수나 어절 거리 등의 요소를 가지고 추가적으로 경계강도를 변화시켰다. 예를 들면, 'BI-0'이 연이어 두 개가 나오는 경우에는 그 다음 어절의 가변적 경계강도를 한 단계 올리고, 종결어절 앞에 나오는 어절의 가변적 경계강도는 한 단계 낮추었다.

## 3. 억양구, 액센트구 파라미터

### 3.1. 억양구 파라미터

억양구는 낱숨의 단위이며 하나의 의미단락을 이루고 있는 단위로써 구나 절, 짧은 문장이 보통 억양구의 기본 단위가 된다. 억양구 내에서 피치의 변화는 일반적으로 문장의 앞쪽에 있는 억양구가 피치 영역이 넓고 높으며, 끝으로 갈수록 억양구의 피치 영역이 낮고 좁아지는 경향이 있다. 그러나, 절이 끝나는 경우나 새로운 의미단락이 시작하는 경우에는 피치영역이 재설정된다.

본 논문에서는 앞에서 정해진 어절 경계강도를 기준으로 억양구를 설정하였는데 어절 경계강도가 'BI-4', 'BI-5'인 어절에 할당하였다. 억양구의 피치영역을 제어하기 위해 상위선(Pitch Topline)과 기저선(Pitch Baseline)을 정하였으며 피치의 상위선은 하나의 억양구를 거칠 때마다 7 Hz 하강하는 것으로 설정하였으며, 기저선은 하강값을 두지 않았다. 이는 문장을 발화할 때 문장 끝으로 갈수록 발화의 높이는 낮아지지만 억양구의 어디에서나 피치영역의 기저까지 피치값이 떨어질 수가 있기 때문이다. 피치 영역을 설정하면 피치값을 피치 영역안의 비율값으로서 설정할 수가 있기 때문에, 피치영역의 변화를 통해서 남성과 여성의 차이를 나타내거나, 절과 문장 종류에 따른 차이를 나타낼 수 있어서 유리하다.

표 4. 억양구 파라미터 (여성화자)

	의미	설정
상위선 (Topline)	피치의 상위 제한 주파수값	기준(Default) - 270 Hz 하강(decline) - 7 Hz
기저선 (Baseline)	피치의 기저 제한 주파수값	기준(Default) - 140 Hz
재설정 (reseting)	기준값으로 재설정	새로운 절의 시작 의미단락이 끝나는 쉼표

### 3.2. 액센트구 파라미터

억양 연구를 위해서는 발화된 문장의 이해에 영향을 많이 미치는 요소를 선정하고 이 요소를 될 수 있는 한 적은 파라미터를 가지고 자동적으로 생성할 수 있어야 하며, 생성된 파라미터를 가지고 모든 가능한 F0 곡선을 만들어내는데 문제가 없는

것일수록 좋은 모델 파라미터라 할 수 있다.

액센트구는 하나의 피치 액센트를 중심으로 주변의 여러 음절이 모여 있는 단위인데 본 논문에서는 액센트구를 피치 생성의 기본 모델단위로 설정하였다. 그 이유는 액센트구의 특징과 구조에 따라 돌출됨에 주요한 영향을 미치는 피치액센트의 값이 결정되며 액센트구의 경계음절에는 경계 액센트가 나타나기 때문에 액센트구를 기본단위로 파라미터를 설정하여 모델을 구성하기가 용이하기 때문이다.

따라서, 액센트구의 피치 액센트 음절에 F0 peak 값을 설정하고 peak 와 주변 음절들과의 관계를 통해 상승과 하강의 흐름을 만들어 내고, 추가적으로 액센트구 경계음절과 억양구 경계음절을 더 상세하게 모델링한다면 문장 전체의 억양곡선을 생성하는 것을 쉽게 할 수 있다.

본 논문에서 액센트구는 기본적으로는 어절 경계강도가 2 이상 되는 어절에 설정하였고 음절수와 어절수를 제한을 두어 그 제한을 넘었을 경우도 액센트구로 설정하였다.

### 3.2.1. Peak 파라미터

데이터에 기반한 F0 파라미터화와 관련된 연구에는 Tilt parameter[5], parametric intonation event (PaintE) parameter[6], maximum-based model[7] 등의 방법이 있다.

본 논문에서는 maximum-based model 을 이용하여 F0 의 파라미터 모델을 설정하였다. 이 모델은 억양 곡선에서 인지적, 음향적으로 가장 큰 영향을 미치는 것은 돌출됨(prominence)이며, F0 곡선에서 돌출됨을 나타내는 peak 을 정의할 수 있는 파라미터들을 정하여 모델로 설정하는 것이다.

따라서, 액센트구의 Peak 파라미터를 maximum-based model 에 의거하여 다음과 같은 네 가지로 구성하였다.

첫번째 파라미터는 peak delay 로서 액센트음절 모음 시작에서 peak 까지의 상대적 시간으로 정의한다. 즉, peak 가 모음의 onset 이후에 나오면 양(+의 값을 가지며, 모음 onset 이전에 나오면 음(-)의 값을 가진다. 이 값은 액센트 음절 전체 길이에 대한 상대값으로 정규화하였다.

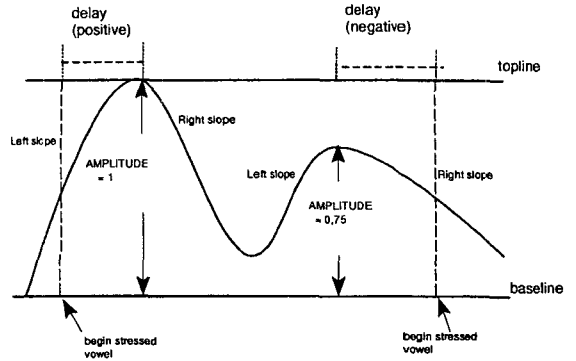


그림 3. maximum-based model 에서의 peak 파라미터[7]

두 번째 파라미터는 Peak 의 크기(amplitude)로서 상위선과 기저선 사이의 비율(%)값으로 나타낸다. 상위선과 기저선은 화자의 피치영역을 나타내는 것으로서 억양곡선의 임계값 역할을 한다. 상위선과 기저선은 억양구 단위로 재설정된다.

나머지 두 개는 앞 기울기(left slope)와 뒤 기울기(right slope)로서 peak 에서의 억양구의 앞뒤 음절들에 대한 기울기 값이다. 이 값들은 음성 DB 를 통해 파라미터를 추정할 때에는 cos2 함수를 사용하였으며 최적화된 값은 원음의 F0 와 최소제곱오류 (least square error)가 가장 작은 값으로 정하였다.

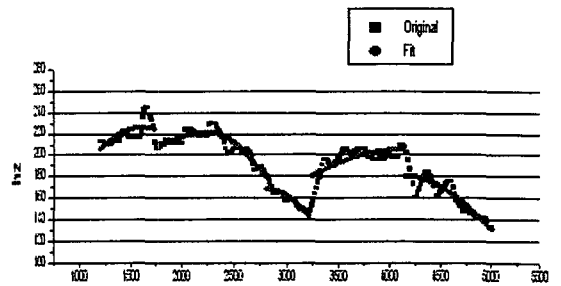


그림 4. 원음의 F0 와 peak 파라미터로 추정한 F0 의 비교  
(문장 : “사람은 모두 집에서 살고 있다”)

음성 DB 에서 자동적으로 위의 파라미터를 추정한 후, 추정

된 값으로 다시 엑센트구를 생성한 결과, 평균 오류값은 10.79Hz, 표준편차는 53.24Hz 였다. 엑센트구의 수는 1076 개였다.

위의 네가지 파라미터는 CART를 이용해 회귀 분석을 수행하여 각각 예측규칙을 만들었다. 각 파라미터에 대한 추정 성능은 표 5와 같다.

표 5. peak 파라미터의 추정성능

	상관계수 (correlation coefficient)
peak delay	0.74
peak amplitude	0.77
left slope	0.65
right slope	0.59

### 3.2.2 조사, 어미 피치 검색 테이블

엑센트구의 마지막 경계음조는 다양한 형태를 지닌다. 특히, 억양구의 경계와 일치할 경우는 F0가 상승과 하강이 같이 나타나는 경우도 있으며 그 폭도 다양하다. 이런 엑센트구 경계음조 특히 문장 마지막의 경계음조는 언어적, 화용적 유형의 정보를 전달하기 위해서 운율적 변화가 많이 일어나는 곳이다. 청자들이 합성음을 끊어서 의미적으로 이해할 때도 이런 엑센트구 경계의 휴지, 피치변화를 통해 파악한다. 한국어의 경우 주로 엑센트구 끝 조사, 어미의 억양 변화가 이런 경계음조를 주로 지닌다. 한국어는 기본적으로 관계언이 발달한 언어이며 따라서 조사, 어미가 문장의 기능, 화자의 의도 등을 나타내는 주요한 역할을 담당한다. 조사, 어미의 수는 제한되어 있으며 같은 조사나 어미는 피치 모양이 크게 다르지 않다.

본 논문에서는 위와 같은 운율적 특징을 뚜렷하게 가지는 조사, 어미를 보다 상세하게 모델링하기 위하여 음성 DB에서 조사, 어미에 해당하는 피치 패턴을 저장하고 이와 관련된 요소를 key-index로 저장하여 조사, 어미 피치 검색테이블을 구성하였다.

훈련단계에서 먼저 음성 DB에서 문장 전체에 대한 피치를

추출한 후 필터링(filtering)과 보간(Interpolation)을 통해 피치 곡선을 smoothing 한다. 여기서 조사, 어미에 해당하는 피치만을 추출하게 되며 이 때, key-index로 사용할 정보들을 추가적으로 처리하여 테이블을 구성하게 된다.

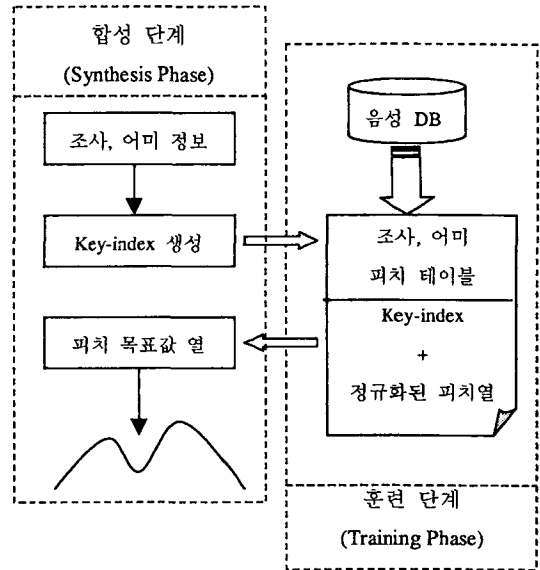


그림 5. 조사, 어미 피치 검색테이블의 작성 및 사용

key-index로 사용하는 정보는 조사, 어미의 품사 코드, 음절 개수, 발음표기 변환된 열, 어절 경계강도, 어절의 음절길이, 추출한 어절의 다음 어절 초성의 고음조 여부 등을 사용하였다.

정규화된 피치열은 조사, 어미의 전체 지속시간이 msec 단위로 저장되고, 그 이후는 전체 길이에 대해 5% 간격으로 상위선과 기저선인 피치 영역내의 F0 비율값으로 저장하였다. 합성 시에는 상위선, 기저선 내의 비율값을 실제 F0 값으로 변환하고 테이블에서 검색된 조사, 어미의 길이와 예측된 조사, 어미 길이의 비율을 가중치로 사용하였다.

최종 억양곡선은 억양구 파라미터와 엑센트구 파라미터, 검색된 조사, 어미 피치 목표값 열을 결합하여 만들어 낸다.

## 5. 청취실험 및 평가

억양요소의 평가방법은 여러 가지 방법들이 시도되고 있지만

정해진 표준이나 기준은 아직 없다. 이전에는 주로 RMSE(root mean square error)나 원음과 합성음의 F0 값을 대비하여 나타내는 방법 등이 사용되었으나 이들은 억양에 대한 정확한 평가라고 보기 힘들다.

본 논문에서는 합성음의 운율 요소만을 별개로 평가하기 위하여 PURR (Prosodic Unveiling through Restricted Representation)[9] 의 방법을 이용하였다. 이 방법은 합성음을 운율 요소만을 나타내는 제한적 형식으로 변경하는 것으로 변경된 합성음은 억양, 리듬 구조, 세기 등만 나타난다. 본 논문에서는 PURR의 방법 중 sinusoidal 표현방법을 통해 합성음을 변경하였다.

실험대상은 상용 시스템으로 사용하고 있는 A사와 B사 그리고 본 실험실에서 구현한 Text-to-Speech 시스템으로 합성한 여성음이며 모두 16bit, 16kHz로 저장하였다. 평가항목은 평서문 2개, 의문문 2개, 방송뉴스, 일기예보, 교통안내 각 1개의 문장으로 했으며, 총 15명의 피험자를 대상으로 하였다. 평가방법으로는 MOS(mean opinion score) Test를 시행하였다.

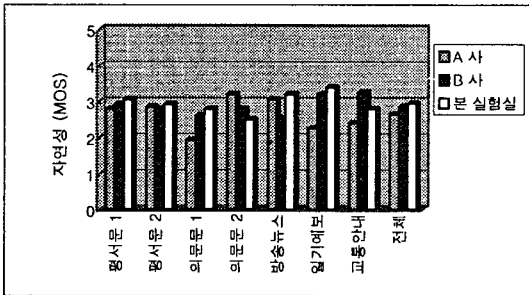


그림 6. 3개 Text-to-Speech 시스템의 자연성 MOS 평가

청취평가 결과, 본 실험실의 합성음이 다른 합성음에 비해 약간 자연스러움을 알 수 있다. 그러나, 그 차이가 그다지 크지 않고 전 항목에 걸쳐 고른 성능의 우수성이 보이지 않는 것으로 볼 때, 테스트 문장을 늘리는 등의 추가적인 실험이 필요할 것으로 생각된다.

## 6. 결론

본 논문에서는 Text-to-Speech 시스템에서 사용할 억양 모델을 위해 자료기반 추정방식을 이용하였다. 액센트 성분인 peak와

경계음조를 파라미터와 피치검색테이블을 구성하여 예측하는 방법을 통해 합성 시스템을 구성하였다. 억양 모델의 PURR 평가 결과, 기존의 상용 음성 합성기에 비해 자연스러운 합성음을 얻을 수 있었다.

## [참고 문헌]

- [1] M. Jilka, G. Möhler and G. Dogil, "Rules for the Generation of ToBI-based American English Intonation," *Speech Communication* 28, pp. 83 - 108. 1999.
- [2] C. Traber, "F0 generation with a database of natural F0 patterns and with a neural network," in *Talking Machines: Theories, Models, Designs* (G. Bailly, C. Benoit, and T.R. Sawallis, eds.), pp.287-304, Elsevier Science, 1992.
- [3] A.W. Black and A.J. Hunt, "Generating F0 contours from ToBI labels using linear regression," in *Proc. of ICSLP'96*, pp.1385-1388, 1996.
- [4] 이상호, 오영환 "CART를 이용한 운율구 추출 및 휴지 기간 모델링," 제15회 음성통신 및 신호처리 워크샵 15권 1호, pp. 81-86, 1998.
- [5] P. Taylor and A.W. Black, "Synthesizing conversational intonation from a linguistically rich input," in *Proc. ESCA Workshop on Speech Synthesis*, pp.175-178, 1994.
- [6] G. Möhler and A. Conkie, "Parametric modeling of intonation using vector quantization," in *Proc. of 3rd ESCA Workshop on Speech Synthesis*, 1998.
- [7] B. Heuft, T. Portele, "Synthesizing prosody: a prominence-based approach," in *Proc. ICSLP'96*, pp.1361-1364, 1996.
- [8] Xuedong Huang, et al., "WHISTLER: A Trainable Text-to-Speech System," in *Proc. ICSLP'96*, pp.2387-2390, 1996.
- [9] G.P.Sonntag, T. Portele, "PURR- a method for prosody evaluation and investigation," *Computer Speech and Language, Special Issue on Evaluation*, Vol.12(3), 1998.