RFC모델의 한국어 억양 곡선에의 적용

표 경란, 김 형순, 최 규수(부산대)

<차 례>

<Abstract>

Application of Rise/Fall/Connection(RFC) Model to Korean Intonation

Kyung Nan Pyo et al.

This is a pilot study on applying the Rise/Fall/Connection(RFC) model to Korean intonation for speech synthesis. RFC model contains successive intonation events, which can be pitch accents and intonation boundary tones. The intonation contour of RFC model is composed of piecewise linear curves of rise, fall, and connection elements, and each element can have any amplitude and duration.

In this paper, elements of RFC model is slightly modified to accommodate the characteristics of Korean intonation. Subjective preference test was conducted to compare the modified RFC model with the original one. The results show that the intonation contour produced by the modified RFC model is perceptually indistinguishable from that of the original RFC model, while the former requires less number of labels than the latter.

1. 서론

컴퓨터와 신호처리 기술의 급속한 발전에 따라, 음성은 사람과 사람 사이의 의사 전달 수단으로서 뿐만 아니라 인간과 컴퓨터 사이의 의사 소통을 위한 매개체로서의 역할이 요구되기에 이르렀고 실제로 사용되기 시작하고 있다. 문장-음성 변환(Text-to-Speech Synthesis(TTS) Conversion) 기술은 이러한 요구에 부합되는 것으로 사용자가 음성으로 바꾸고자 하는 문장을 기계에 입력하면, 기계가 그 문장을 분석한 후 미리 저장된 말소리의 기본 단위들로부터 사용자가 원하는 음성을 신호로 합성하는 것을 말한다(김형순, 1998). 이 때 요구되는 합성음은 사용자가 듣기에 명료하고 자연스러워야 한다. 신호 처리 기술의 발달에 따라 현재 합성음의 명료도는 상당 수준에 이르러 있는 상태이지만 아직은 자연스럽지 못하기 때문에, 어떻게 하면 합성음을 더 자연스럽게 만들 수 있는가에 연구의 초점이 모아지고 있다. 합성음을 자연스럽게 내기 위해서는 여러 가지 운율 정보가 필요하겠지만, 특히 운율을 조절하는 매개변수로서 억양은 합성음의 자연성에 크게 관여하므로 앞선 연구자들로부터 많은 관심의 대상이 되어 왔다(오영환, 1997).

일반적으로 음성 합성기에서 억양 곡선(intonation or pitch contour)을 합성하는 모듈은 입력 문장으로부터 억양을 예측하는 부분과 그 예측된 변수를 이용하여 원하는 억양 곡선을 생성하는 부분으로 나누어지는데(오영환, 1997), 본 논문에서 다루고 있는 것은 이미 예측된 기본주파수(fundamental frequency, F0)값을 가지고 억양을 생성하는 억양 곡선의 모델이다.

본 논문에서는 억양 합성에서 사용하기 쉽도록 구체적인 수치 형태로 표현 가능하면서, 자연음의 억양과 유사하고 자연스러운 음조를 나타내는 억양 모델로서 기존의 억양 모델 중에서 Rise/Fall/Connection(RFC)모델을 선택하여 한국어 억양에 적용해 보았다. 이 모델은 목표점/보간 모델의 하나로 따로 억양의 하락선(declination line)을 설정하지 않고, 보간 될 목표점 자체에 억양의 하락 현상을 나타내도록 하여 그 목표점들을 연결하여 억양 곡선을 모델링 하는 것이다. RFC모델은 음향학적으로 F0값을 모델링 하는 것이므로, 한국어 억양의 음향학적인 특성을 고려해서 한국어 억양에 적합하도록 RFC모델의 구성 요소를 수정하여 적용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 억양의 구조와 억양 합성 모델에 관한 앞선 연구를 정리하고, 3장에서는 영어 억양에 적용된 원래의 RFC모델에 대해서 기술하고, 4장에서는 한국어 억양 곡선을 RFC 모델로 분석하고 합성하는 과정을 통해 한국어에 적합하게 모델의 구성 요소를 수정한 RFC 모델을 제시한다. 마지막으로 4

장에서 결론을 맺는다.

2. 억양의 구조와 억양 합성 모델에 관한 앞선 연구

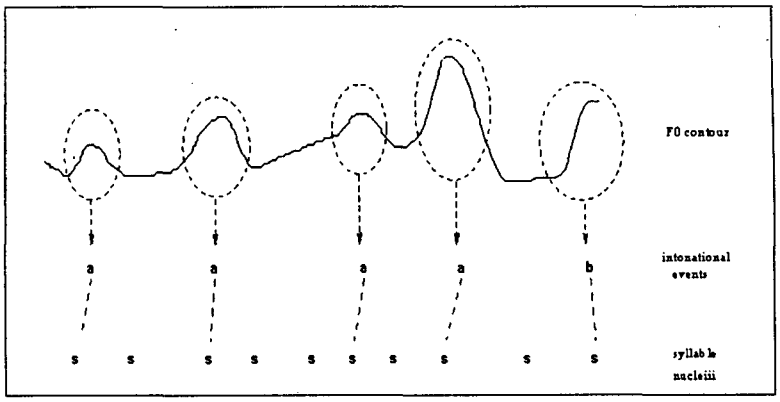
2.1 억양의 구조

억양이란 사람이 발성을 할 때 음의 높이(pitch)에서 일어나는 변화, 즉 성대의 진동에 의해서 생기는 음악적 음조의 높이의 변화를 말하며 그 물리적 의미는 F0의 변화에 해당한다(조성식 등, 1990). 그러나 언어학적으로 분석한 억양의 다양한 현상들을 규칙화하고 정량화해서 억양 합성에 필요한 규칙을 만든다는 것은 단기간 내에 되는 것이 아니라 상당한 시간과 노하우가 필요한 일이다. 따라서 억양을 합성해 내는 것을 목적으로 하는 공학적 연구에서는 음향음성학적인 관점에서 억양의 여러 현상을 단순화시켜 합성을 하고 있다(김형욱 등, 1992).

억양은 F0 곡선을 이용해서 분석할 수 있으며, 하나의 발화는 끊김이 없는 하나의 덩어리로 발음되지 않고 몇 개의 작은 덩어리로 나뉘어 발음되는데, 여기서 그 각각의 덩어리에 대한 개념은 학자마다 다양하다(이현복, 1989; 성철재, 1995, 1996; Jun, S.-A., 1993, 1996; 김선미, 1997). 성철재(1995, 1996)는 이 각각의 덩어리를 지각 적으로 명백한 끊김이 느껴지고, 음향학적으로 일정 수치 이상의 지속시간의 증가나 고저의 변동으로 경계 지워지는 '운율구' 라고 하였으며, 구희산(Koo H.-S., 1986)은 이러한 운율구가 구나 절의 단위와 거의 일치된다고 하였다. Taylor(1994)의 RFC 모델에서 사용한 '억양구'는 이러한 운율구의 특성과 유사한 개념이다. 따라서 본 논문에서는 이러한 운율구의 마지막 음절에 경계음조를 설정하여 운율구를 경계짓고 이 단위를 Taylor의 RFC 모델에서의 표현을 따라 억양구라 이름하였다. 그리고 문장의 전체 억양 구조는 이런 억양구의 연속으로 되어있다고 본다(Taylor, 1994).

<그림 1>은 억양의 구조를 도식화한 것으로, 하나의 억양구 만을 나타낸 것이다. 점선원으로 표시된 a는 피치 액센트가 있는 음절과 연결되어 있고, b는 억양구의 경계 음조를 나타내는데 이것은 억양구의 마지막 음절과 연결되어 있다. 본 논문에서는 이와 같은 억양 곡선의 구조를 기반으로 하여 억양을 분석하고 합성하였다.

이처럼 억양을 분석하거나 원하는 억양을 합성하고자 할 때, 특히 통계적인 방법으로 억양을 연구하고자 할 때에는 운율적으로 표지(label)가 붙여진 많은 양의 음성 데이터베이스가 필요하다.



<그림 1> 억양 곡선(F0 곡선)에 피치 액센트와 억양구 경계 음조, 그리고 액센트가 얹히는 음절을 표지한 그림(Taylor, 1997)

영어의 경우에는 아직은 정해진 표준 기호가 있는 것은 아니지만 미국에서 운율을 연구하는 여러 연구자들이 ToBI(Tone and Break Indices)시스템을 사용하고 있다 (Silverman et al., 1992; Beckman & Ayers, 1994; Pitrelli et al., 1994). ToBI 시스템은 하나의 발화에 대한 운율적 요소들을 시간 축에 따라 정해진 표지를 붙이는 것으로, 단어와 그 경계를 나타내는 단어 표기 계층(orthographic tire), 기본주파수 곡선의 음조 요소들을 나타내는 음조 계층(tone tire), 단어들간의 정접의 정도를 나타내는 끊어 읽기 색인 계층(break-index tire), 그리고 헛기침 소리나 기타 잡음 등을 나타내는 기타 계층(miscellaneous tire)의 복수 계층(multiple tire) 구조로 되어 있다. 한국에서도 K-ToBI(Korean Tone and Break Indices)에 관한 연구가 진행 중에 있다(Jun. S.-A., 1990, 1993, 1995; Kim J.-J., et al., 1997; 이숙향, 1998).

그러나 ToBI 시스템은 음운론에 입각하여 억양의 분석을 하는 모델이기 때문에, 분석해 낸 다양한 억양의 양상을 억양의 합성에서 사용할 수 있도록 합성을 위한 구체적인 수치형태로 표현하는 문제가 남아 있다.

2.2 억양 합성 방식에 관한 앞선 연구

한국어에서는 아직 고유의 기본주파수 곡선을 생성하는 방법이 제안된 적이 없고 미국이나 일본 그리고 유럽에서 제안한 모델을 한국어에 적용시켜 보는 실정이어서 기존의 문장-음성 변환 시스템도 그리 자연스러운 한국어 억양을 합성해 내고 있지는 못하다.

<표 1>은 기본주파수 곡선을 생성하는 모델을 크게 둘로 나눈 것인데, 매개변수를

사용하는 모델과 그렇지 않은 모델로 분류된다(Ross, 1995). 그리고 억양 모델을 훈련하는 방법에는 규칙 기반 학습 방법과 데이터 기반 학습 방법이 있다.

<표 1> F0 곡선 생성 모델의 분류

F0 곡선 생성 모델(F0 Generation Models)	
매개변수를 사용하는 기법 (Parametric Approaches)	소스/필터 모델(Source/Filter Model)
	목표점/보간 모델(Target/Interpolation Model)
	혼합모델(Hybrid Model)
매개변수를 사용하지 않는 기법 (Non-parametric Approaches)	은닉 마코프 모델(Hidden Markov Model)
	신경망 모델(Neural Network Model)
	연결 합성 모델(Concatenative Model)

본 논문의 목적은 억양 합성에서 사용하기 쉽도록 구체적인 수치 형태로 표현 가능하면서, 자연음의 억양과 유사한 음조를 모델링 할 수 있는 모델을 선정하고, 한국어에 적용하여 보는 것이다. RFC 모델은 앞에서 언급했듯이, 목표점/보간 모델의 하나인데, 이것은 고도의 언어학적인 지식이 필요한 것이 아니라, 억양 곡선의 음향학적인 특성에 따라 음조의 오르내림을 모델링 하는 것이다. 따라서 억양 곡선의 합성에 있어서 매개변수를 조절하기에 편리하고 비교적 간단하다. 모델은 억양의 음향학적인 특성에 따라 억양 곡선의 오르내림을 모형화 하는 것이므로, 한국어 억양의 음향학적인 특성을 고려해서 한국어 억양에 적합하도록 RFC모델을 수정한다면 억양 곡선을 잘 나타낼 수 있을 것이다.

3. RFC 모델

RFC 모델은 하나의 억양구를 각 음절에 얹힌 피치 액센트(pitch accents)의 연속된 모양과 억양구의 끝을 나타내는 억양구 경계 음조(boundary tone)로 되어 있음을 기본으로 한 것이다. 그리고 이러한 음조들은 각각의 진폭과 지속시간을 가지는 상승음조(rising tone) 요소와 하강음조(falling tone) 요소, 그리고 연결(connection) 요소로 모델링 된다(Taylor, 1994).

피치 액센트는 적어도 하나의 상승음조 요소와 하강음조 요소로 모델링 되고, 억양구의 시작과 끝에서 나타나는 상승음조는 하나의 상승음조 요소로 모델링 된다. 상승음조 요소와 하강음조 요소는 피치 액센트와 억양구 경계 음조를 모델링 할 때에만

사용되고, 연결 요소는 피치 액센트와 억양구 경계 음조를 제외한 기본주파수의 나머지 부분을 모델링 할 때 사용한다.

RFC 모델은 어디까지나 기본주파수의 음향학적 특성을 정량화 하기 편리하도록 모델링 한 것이지 억양의 음운론적인 분석은 아니다. 영어의 경우는 이 모델을 사용하여 FESTIVAL 음성 합성 시스템을 구현하였다(FESTIVAL, 1997).

3.1 피치 액센트의 모델링

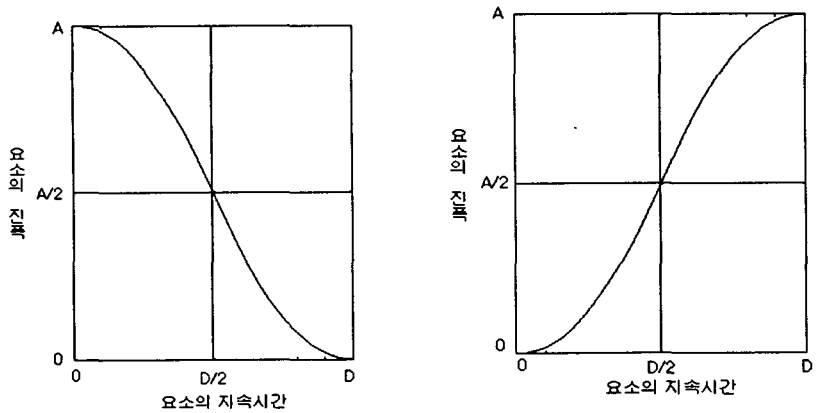
하나의 액센트는 액센트를 이루는 요소인 상승음조와 하강음조로 나누어 모델링 한다. 보통 상승음조보다 하강음조의 진폭이 크게 나타나는데, 이것은 억양 곡선의 하강현상과 관계가 있다. 따라서 상승음조와 하강음조로 나누어 모델링 하면, 억양 곡선의 전체적인 하락선을 따로 설정하지 않아도 되고, 다양한 피치 액센트를 나타낼 수 있다.

식(1)과 (2)는 Taylor(1994)가 RFC 모델을 구현하기 위해서 제안한 식인데, 상승음조 요소와 하강음조 요소를 모델링 할 때 사용된다.

$$f_0 = \begin{cases} A - 2A\,(t/D)^2 & 0 < t < D/2 \\ 2A - \,(1-t/D)^2 & D/2 < t < D \end{cases} \quad (1)$$

$$f_0 = \begin{cases} 2A - \,(t/D)^2 & 0 < t < D/2 \\ A - 2A\,(1-t/D)^2 & D/2 < t < D \end{cases} \quad (2)$$

(a) 하강음조 요소의 모델링 (b) 상승음조 요소의 모델링



<그림 2> 하강음조 요소와 상승음조 요소의 모델링

식(1)은 하강음조 요소의 F0 곡선의 모양을 나타내는 식이고, 식(1)을 진폭축 (f0)으로 대칭 하면 식(2)와 같이 상승음조 요소의 F0 곡선의 모양을 나타내는 식이 된다. 이 때, 요소의 진폭(amplitude)은 A라는 변수로, 요소의 지속시간(duration)은 D라는 변수(scaling variables)로 나타내어, 다양한 진폭과 지속시간으로 된 피치 액센트의 양상을 모델링 한다. <그림 2>는 식(1)과 식(2)를 그래프로 나타낸 것이다.

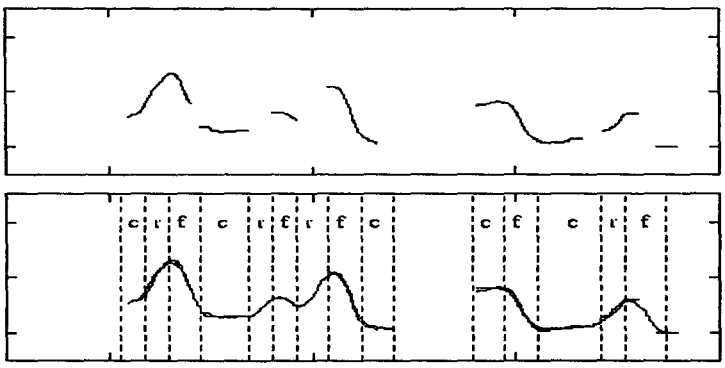
3.2 억양구 경계 음조의 모델링

영어의 경우 대개 억양구의 시작과 끝에서는 음조가 급격하게 상승(sharp rise)하는 현상이 발견되는데, 이와 같은 상승음조의 억양구 경계(rising boundary tone)는 상승음조 요소로 모델링 된다.

3.3 피치 액센트와 억양구 경계 음조 이외의 구간 모델링

억양 곡선의 전체적인 양상이 피치 액센트의 연속된 모습이라고는 하지만 억양 곡선의 모든 부분이 피치 액센트나 억양구 경계 음조와 연결되는 것은 아니다. 이렇게 피치 액센트에 영향을 끼치지 않는 부분은 직선으로 연결하여 F0곡선이 연속된 값을 가지도록 모델링 한다. 이 때의 직선을 연결 요소(connection element)라고 하고, 이 요소 또한 진폭과 지속시간의 파라미터를 가진다.

<그림 3>은 지금까지 기술한 RFC 모델의 구성 요소로 영어의 억양 곡선을 모델링 한 예이다(Taylor, 1992).



"The large window stays closed: the small window you can open"

<그림 3> RFC 모델의 적용 예(Taylor, 1992)

<그림 3>에서 위의 그림은 자연음에서 뽑아낸 기본주파수 궤적을 F0값을 고르는

과정을 거쳐서 부드러운 억양 곡선으로 나타낸 것이고, 아래의 그림은 RFC 모델의 요소로 억양 곡선에 음조를 표지 한 것이다. 상승음조 요소는 r, 하강음조 요소는 f, 연결 요소는 c, 그리고 억양구 경계 음조는 rb로 표지 된다. 그런데 <그림 3>은 rb가 없는 경우이므로 억양구 경계 음조를 나타내는 표지가 나타나 있지 않다.

4. RFC 모델의 한국어에의 적용

한국어를 대상으로 하여 공개적으로 사용할 수 있도록 구축된 음성 데이터베이스는 많지 않으며, 얼마 안 되는 공통 데이터베이스도 대부분 음성인식 분야를 위한 것이어서 음성합성 분야 특히 운율 연구에 사용할 수 있는 음성 데이터베이스는 거의 없는 실정이다(이용주, 1996).

본 연구에서는 원광대학교에서 구축한 낭독 음성 데이터베이스를 사용하였다(Kim B.W. et al., 1997). 이것은 시스템공학연구소(SERI)에서 구축한 100만 어절의 한국어 텍스트 코퍼스에 의해 선정된 10,000 문장 중 처리를 거쳐 선정된 589문장으로 되어 있다. 그리고 이 589개의 문장들은 50문장으로 이루어진 공동세트와 나머지 539문장을 5개의 개별세트로 나누어져 있다. 이 데이터베이스는 음성학적으로 음소들이 잘 배분되어 있고, 다양한 형식의 평서문으로 되어 있으며, 표준어의 억양으로 녹음이 되어 있다. 음성은 방음 부스에서 Senheizer HMD224X 마이크를 사용하여 녹음되었으며, 발성된 음성 데이터는 디지털 오디오 테이프에 저장된 다음 16 kHz로 표본화(sampling)하고 16 bits로 양자화 된 것이다.

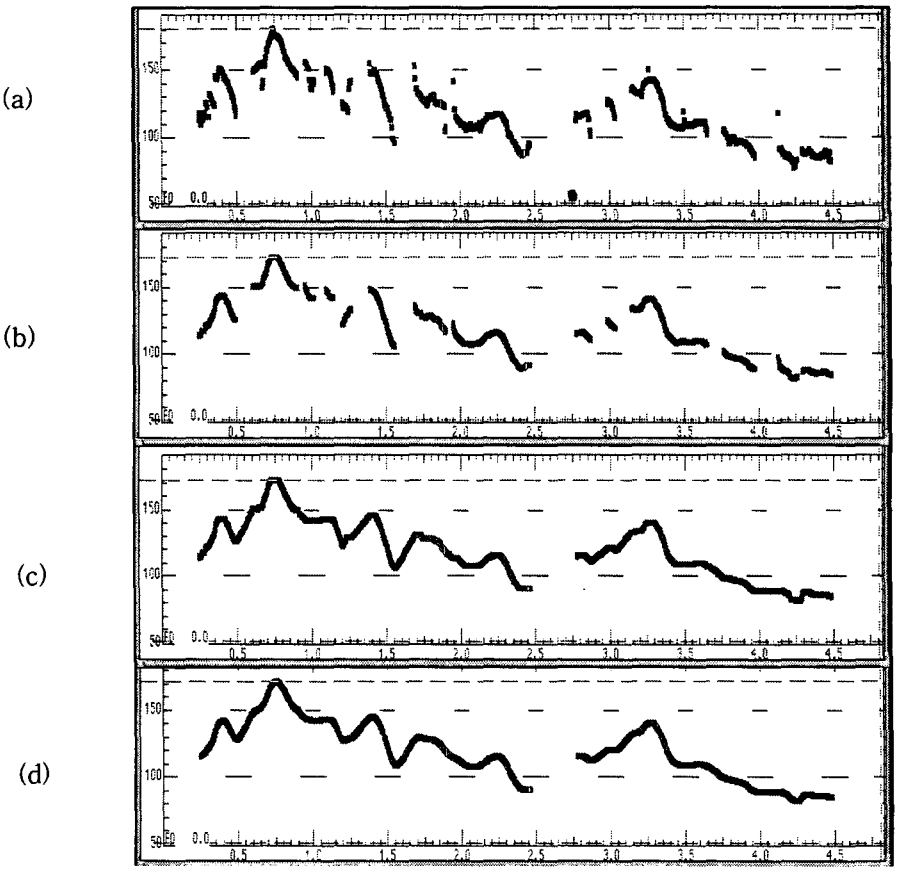
본 연구에서는 이 데이터베이스 중에서 50문장으로 된 공동세트를 사용하였다. 화자는 23세의 남성이고, 부모님의 고향이 서울이며, 서울에서 20년간 거주하였다.

4.1 F0값 뽑아내기와 억양 곡선 고르기

F0값은 Entropic사의 ESPS/Xwaves+를 사용하여 5 ms간격으로 뽑아 텍스트 파일 형식으로 저장했다.

음성 파형에서 추출한 실제 F0값을 보면(<그림 4>의 (a)), F0값을 뽑아내는 알고리즘 자체의 추정오차와 인접 분절음의 영향에 의해 생긴 급격한 변화의 F0값이 많이 나타나는 것을 볼 수 있다. 억양 곡선을 상승음조 요소와 하강음조 요소로 모델링 하기 위해서는, 음조의 변화에 영향을 끼치지 않는 범위에 한에서 이러한 값들을 제거하고 억양 곡선을 고르는(smoothing) 것이 구현하는 관점에서 편리하다. 그래서 음성

파형에서 뽑아낸 F0값에 다음과 같은 전처리를 하였다. 먼저 파형의 순간적 흐트러짐(jitter)이나 피치 흔들림(perturbation)의 영향을 없애기 위해서 15 point 메디안 필터링(median filtering)을 하였다. 그렇게 하면 <그림 5>의 (b)와 같이 갑작스럽게 값이 변화하는 부분이 거의 제거된다. 그리고 듣는 이가 무성음 구간 근처의 유성음에 의해서 무성음 구간을 마스킹(masking)해서 인지한다는 연구 결과를 토대로(Kohler, 1991), F0값이 나타나지 않는 무성음 구간에 대해서는 직선 보간(straight line interpolation)하여 연속된 값을 가지도록 하였다. 이것은 억양 곡선이 연속된 값을 가지도록 하기 위함이다. 마지막으로 직선 보간을 할 때 생기는 날카롭게 꺾이는 부분을 제거하고 <그림 4>의 (d)와 같은 부드러운 곡선으로 만들기 위해서 최종적으로 7 point 메디안 필터링을 하였다.



<그림 4> 억양 곡선의 고르기 과정
(a) 원래의 F0값들의 모습 (b) 15 point메디안 필터링 후의 F0 곡선
(c) 무성음 부분을 직선 보간한 후의 F0 곡선
(d) 7 point 메디안 필터링 후의 F0 곡선

<그림 4>는 "아들 똘이 녀석과 같은 유치원에 다니는 옆집 현아 엄마의 전화였습니다." 라는 문장에서 뽑아낸 F0값을 위에서 기술한 방법으로 전처리를 하여 억양 곡선으로 만드는 과정을 나타낸 것이다.

4.2 RFC 모델을 이용한 음조 표지 붙이기

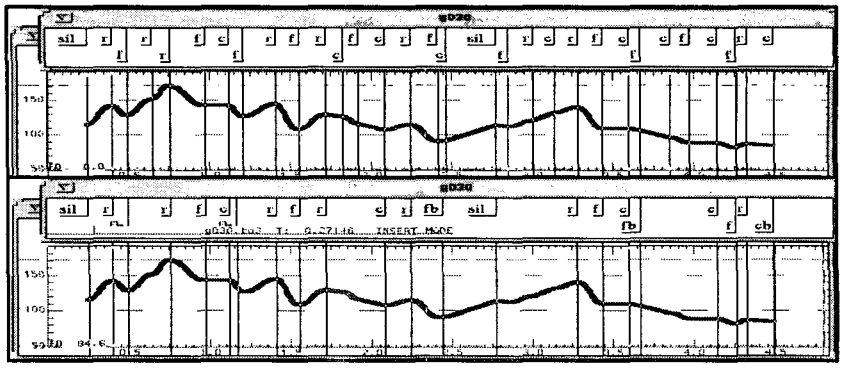
RFC 모델을 이용한 음조 표지 붙이기 작업은 ESPS의 xlabel을 사용하였다. xlabel은 텍스트 양식으로 된 음조 표지 파일을 생성할 수 있도록 하는 프로그램인데, 음조 표지의 시작 시점을 시간 축에 따라 표시하는 시간 축 필드(time-series field)와 그 구간에 해당하는 음조를 표시하는 음조 표지 필드(label field)를 가진다. 이때 바로 뒤의 음조 표지의 시작 시점에서 현재 음조 표지의 시작 시점을 뺀 값이 현재 음조 표지의 지속시간이 된다. 각 음조 표지의 진폭은 음조 표지 파일과 앞 절에서 전처리를 통해 얻어진 억양 곡선의 파일을 입력으로 받아서 두 파일을 시간 축에 맞게 정렬(align)한 다음 현재 음조 표지가 끝나는 시점의 기본주파수 값에서 현재 음조 표지 시작점의 기본주파수 값을 뺌으로써 얻을 수 있다. 본 논문에서는 PC에서 C언어로 구현한 프로그램을 통해 각 음조 표지의 지속 시간과 진폭을 구한 다음 음조 표지 파일의 다른 필드에 저장하였다.

본 연구에서는 상승음조에는 r, 하강음조에는 f, 피치 액센트가 없는 부분에는 c, 억양구 경계 음조에는 기존의 상승음조로 된 억양구 경계인 rb외에 하강음조로 된 억양구 경계 음조(falling boundary tone)에는 fb, 연결 요소로 된 억양구 경계 음조(continuation boundary tone)에는 cb라는 음조 표지를 붙였다. 이들 2가지의 억양구 경계 음조를 더 설정한 이유는, 영어 억양 곡선에 적용했던 원래의 RFC모델에는 rb만 있었는데, 한국어 억양 곡선의 경우에는 억양구 경계 음조에 하강음조 요소나 연결 요소로 모델링 해야 하는 음조가 나타나기 때문이다. 그리고 표준어 억양에서 흔히 보이는 계단식 음조는 청각적으로 그 단계가 느껴지지 않으므로 하나의 음조로 표지 하였다.

<그림 5>는 RFC 모델로 한국어 억양에 음조를 표지 한 예인데, <그림 5>의 (a)는 Taylor가 제안한 RFC 모델로 억양을 음조 표지 한 것이고, (b)는 억양 곡선 상으로는 음조의 변화가 있어 보이지만 청각적으로는 액센트가 느껴지지 않는 부분을 직선으로 음조를 표지 하거나, 표준어의 억양에서 흔히 보이는 계단식 음조를 하나의 음조로 표지 하여 음조 표지의 개수를 줄인 것이다.

(a)

(b)



<그림 5> RFC 모델을 이용한 한국어 억양의 음조 표지 예(<그림 4>와 같은 문장)
(a) 원래의 RFC모델로 음조 표지 한 경우
(b) 계단식 음조를 하나의 음조로 단순화하고, 3가지 억양구 경계 음조로 된 수정한 RFC모델로 음조 표지 한 경우

다음 <표 2>는 50문장을 분석한 결과이다.

<표 2> 50문장 음조 표지의 총 개수

	r	f	c	rb	fb	cb	합계
(a) 원래의 RFC 모델	557	704	476	80	0	0	1817
(b) 수정된 RFC 모델	463	422	380	80	199	39	1583
(a) - (b)	94	282	96	0	-199	-39	234

<표 2>를 보면 수정된 RFC 모델의 전체 음조 표지의 개수가 원래 RFC모델의 음조 표지 개수보다 13% 정도 줄어들었음을 알 수 있다. 그러나 이것은 RFC 모델로 억양 곡선을 합성했을 때 자연음의 억양 곡선의 음조와 거의 차이가 없을 정도까지만 음조 표지의 개수를 줄이기 위해서 최소로 줄인 것이다. 따라서 합성기에서 수정된 RFC모델을 억양의 합성 모델로 사용하려 할 때에는, 음조 표지들 사이의 패턴을 찾아서 억양이 나타나는 양상을 좀더 간단히 하면 음조 표지의 개수를 더 줄일 수도 있을 것이다. 물론 이렇게 하면 자연음의 억양과는 다소 차이가 나겠지만, 자연스러운 억양 곡선을 생성하는 데에는 무리가 없을 것이라고 생각된다.

4.3 RFC 모델을 이용한 억양 곡선의 합성과 청취 실험

RFC 모델로 분석한 억양 곡선이 한국어 억양 곡선을 잘 표현하는지 실험하기 위

해서 Linear Predictive Coding(LPC) 합성 방법으로 합성음을 생성하였다. LPC합성은 ESPS 합성 프로그램중의 하나를 이용하였으며, 2개의 입력 파일을 사용하여 하나의 합성음 출력 파일을 만들 수 있도록 되어 있다. 이들 두 파일은 자연음의 유성음 구간에서 피치 동기적(pitch-synchronous)으로 얻어진 반사 계수(reflection coefficient)가 저장된 파일과 RFC 모델로 분석한 표지로 합성한 억양 곡선 파일이다. 억양 곡선은 음조 표지마다 저장되어 있는 음조 각각의 지속시간과 진폭을 입력 값으로 하여 PC상에서 C언어로 구현하여 생성하였다.

수정된 RFC 모델로 억양 곡선을 합성하여도 원래 억양의 음조와 차이가 나지 않을 것이라는 가정 하에, 청취 실험을 통해 가정을 검증하기 위해서 앞 절에서 사용한 50문장 중에서 원래의 RFC 모델로 음조 표지한 음조의 개수와 차이가 6개 이상 나는 문장 19개를 고른 후, 이 중에서 비교적 합성음이 명료한 문장 10개를 간추려 청취실험에 사용하였다. <표 3>은 합성과 청취실험에 사용한 문장의 목록이다. 이 때 억양구의 경계는 연구자의 지각에 의존하여, 명백한 끊김이 느껴지고 음향적으로 고저의 변동이 나타나는 경계로 결정하였다.

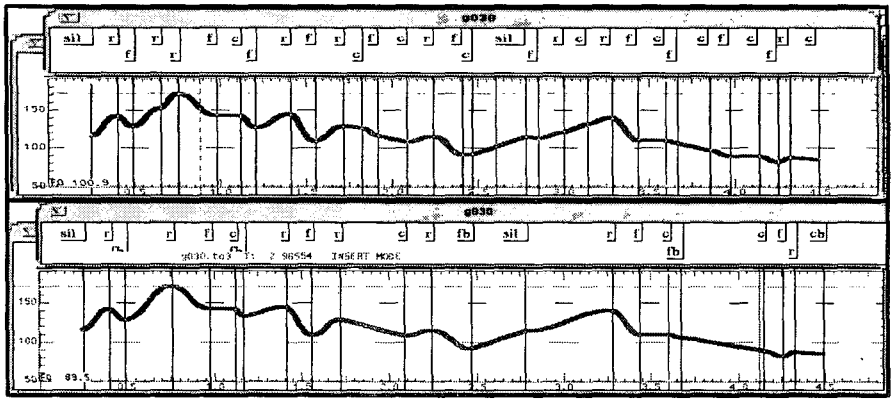
<표 3> 합성과 청취 실험에 사용한 문장 목록(/ : 억양구)

3. 조붓한 허리와 함께/ 얄깃얄깃 노는/ 그 몸놀림은/ 웬만큼 흉내를/ 낼 수도 없었다.
12. 이상이/ 전력 계통이나/ 동력 계통이 아니라/ 전자 뇌와/ 위상 동기 장치에 있음은/ 확실했다.
14. 학교 당국의/ 지나친 관료주의적/ 교육 방법과/ 뚜렷한 차별 교육은/ 없는 집 아이들을/ 소외시키고/ 미래의 꿈을/ 잃게 만들었다.
19. 네이비/ 블루 빛깔의/ 시트가 씌워진/ 트윈 베드와/ 커다란/ 창문이/ 바다 쪽을 향해/ 시원스레 틔어 있어/ 분위기가/ 좋은/ 방이었다.
20. 공기 중의/ 먼지와 땀으로/ 더러워지기 쉬운 피부에는/ 무엇보다/ 깨끗한 세안이/ 중요하다.
28. 그리고/ 바로/ 이런 절차를 통해/ 왼쪽 뇌 우위의 상태에서/ 오른쪽 뇌 우위의/ 상태로/ 바뀐다고/ 말할 수 있겠다.
33. 저 멀리/ 희양의/ 넓고 넓은/ 눈 덮인/ 광야가/ 햇빛을 받아/ 찬란하게/ 펼쳐져 있었다.
39. 적도지역은/ 태양 에너지의/ 순흡수 지역이며/ 극 지역은/ 태양 에너지의/ 순방출 지역이다.
41. 이 집/ 저 집으로부터/ 방아 찧는 소리가/ 쿵덕쿵 / 쿵더쿵/ 들려 왔습니다.
47. 해양오염 문제에/ 대한 연구와/ 바다에서/ 생명의 기초에/ 대한 연구는/ 공히/ 밀접하고 복잡하게/ 얽혀 있다.

<그림 6>의 (a)는 원래의 RFC 모델로 억양 곡선을 합성한 것이고, (b)는 계단식 음조를 하나의 음조로 단순화하고, 3가지 억양구 경계 음조로 된 수정한 RFC모델로 억양 곡선을 합성한 것이다.

(a)

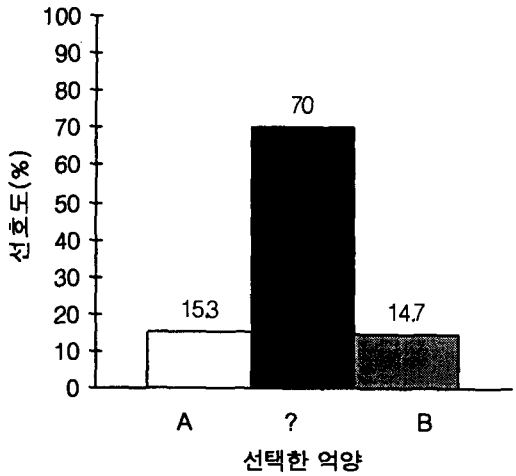
(b)



<그림 6> RFC 모델을 이용한 억양 곡선의 합성
(a) 원래의 RFC 모델로 표지한 음조로 합성한 억양 곡선
(b) 수정된 RFC 모델로 표지한 음조로 합성한 억양 곡선

청취실험은 실험실 환경에서 부산대학교 전자공학과 대학원생 15명에게 헤드폰을 착용하게 한 후 실시하였다. 실험 방법은 자연음의 억양을 그대로 사용하여 합성한 문장을 기준음으로 들려주고, 원래의 RFC모델로 합성한 문장(A)과 수정된 RFC모델로 합성한 문장(B)을 피험자에게 어떤 합성음이 어떤 방법으로 생성한 것인지 알 수 없도록 무작위(random)로 들려주고 A와 B중 어느 것의 억양이 기준음의 억양에 가까운지를 선택하게 하였다. 이 때 A와 B중 어느 것이 기준음의 억양과 더 가까운지 우열을 결정할 수 없는 경우에는 ? 표를 선택하도록 하였다.

실험 결과는 <그림 7>에서 보는 바와 같이 수정된 RFC 모델로 억양 곡선을 합성하여도 자연음 억양의 음조와 차이가 나지 않음을 알 수 있었다. 즉, 2종류의 억양구 경계 음조를 더 설정하여 전체 억양구 경계 음조의 개수는 늘었고, 피치 변화에 덜 민감한 부분 하나의 음조로 합쳐서 전체 r, f, c의 개수가 줄었지만 자연음의 억양과 별로 차이가 나지 않았다. 따라서 수정된 RFC모델로 분석/합성한 억양 곡선의 패턴을 유형화하면, 자연스러운 억양을 생성하는데 도움을 줄 수 있을 것으로 생각된다.



<그림 7> 합성한 억양의 선호도에 관한 청취실험 결과
□ 원래의 RFC 모델로 합성한 억양 A의 선택
■ 두 억양의 우열을 구별할 수 없는 경우 ? 선택
▒ 수정된 RFC 모델로 합성한 억양 B의 선택

4. 결론

본 논문에서는 합성음에 사용할 억양 모형을 생성하기 위한 기초적 연구로서 자연음의 억양 곡선의 특징을 잘 반영하면서 구현하기도 쉽도록 정량적인 값으로 억양 곡선을 모형화 하는 RFC 모델을 이용하여 한국어 억양 곡선을 분석하고, 분석한 표지들을 입력으로 하여 억양 곡선을 합성해 보았다.

RFC 모델의 요소는 영어의 억양 곡선에 적합하도록 되어 있으므로 한국어에 적합하도록 RFC 모델의 구성 요소를 수정하였다. 먼저 피치 액센트의 경우는 다음과 같이 수정하였다. 한국어 억양 곡선에서는 계단식 음조가 많이 나타나는데, 이것은 음조의 변화가 거의 느껴지지 않으므로 계단식 음조를 하나의 음조로 표지하였다. 억양구 경계 음조의 경우에는 원래의 RFC 모델에서는 표지가 rb 하나 뿐이었지만, 한국어 억양 곡선에서는 fb나 cb와 같은 억양구 경계 음조가 나타나므로 이 2가지를 추가 설정하였다.

본 논문에서 수정된 RFC 모델을 50 문장의 한국어 억양에 적용시킨 결과, 원래의 RFC 모델에 비해 음조 표지 개수가 약 13% 줄어들었다. 그리고 수정된 RFC 모델이

원래의 RFC 모델보다 전체 음조 표지의 개수가 줄어들어 단순화 되었지만, 청취 실험을 해 본 결과 원래의 RFC 모델로 합성한 억양과의 차이가 거의 없음을 알 수 있었다.

RFC 모델은 영어의 억양을 나타내기 위한 억양 모델이지만, 한국어 억양에 적합하도록 수정하여 적용해 본 모델이 한국어 억양을 비교적 잘 모델링 하였으므로, 합성기의 억양 합성 모듈 중에서 억양 곡선을 생성하는 부분에서 사용될 수 있을 것이라고 생각한다. 그러기 위해서는 입력 문장에서 억양의 구조를 예측하는 모듈을 세우는 연구가 앞으로 더 진행되어야 할 것이다.

그리고 본 논문에서는 수동으로 음조 표지를 하였지만 시간과 노력이 많이 소요되므로, 억양 합성에 필요한 많은 데이터베이스를 확보하기 위해서는 기계가 자동으로 음조를 표지 할 수 있도록 되어야 할 것이다.

<참고문헌>

김선미(1997), 한국어의 리듬 단위와 문법 구조 -음성 합성에서 리듬 구현의 자연성 향상을 위한 음성언어학적 연구-, 서울대학교 언어학과 박사학위 논문.

김형순(1998), PC용 TEXT-TO-SPEECH 시스템 개발, 산업자원부 1차년도 중간보고서.

김형욱 외 3명(1992), 한국어 문장-음성 변환기의 운율 제어용 구문분석기, 「제9회 음성통신 및 신호처리 워크샵 논문집」, 218-223.

성철재(1995), 한국어 리듬의 실험음성학적 연구, 서울대학교 언어학과 박사학위 논문.

성철재(1996), 한국어 낭독체 문장의 음향분석, 「제2회 음성학 학술대회 논문집」, 대한음성학회, 157-172.

이숙향(1998), 한국어 성조 이벤트와 음향적 길이, 「한국음향학회 학술발표대회 논문집」, 제 17권 제 1(s)호, 383-386.

오영환(1997), 「음성언어정보처리」, 홍릉과학출판사.

이용주(1996), 음성 데이터베이스의 현황 및 과제, 「제13회 음성통신 및 신호처리 워크샵 논문집」, 한국음향학회, 279-284.

이현복(1989), 「한국어의 표준발음」, 교육과학사.

조성식 등 편저(1990), 「영어학 사전」*(A Dictionary of English Linguistics)*, 신아사.

Beckman, M. E., Ayers, G. M.(1994), *Guidelines for ToBI labelling* (version 2.0,

Feb. 1994), Linguistics Dept., Ohio State Univ. [ftp site -- ftp 129.138.1.82]

Jun, S.-A.(1990), The prosodic structure of Korean-in terms of voicing, In E-J. Back, ed., *Proceeding of the 7th International Conference on Korean Linguistics,* 87-104, University of Toronto Press.

Jun, S.-A.(1993), The Phonetics and Phonology of Korean Prosody, Ph D. dissertation, Linguistics, Ohio State Univ.

Jun, S.-A.(1995), Asymmetrical prosodic effects on the laryngeal gesture in Korean. In Bruce Connell and Amalia Arvanit, eds., *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV,* 235-253, Cambridge Univ. Press.

Kim B. W. et al.(1997), Design and construction of Korean speech database for common use, *Proceedings of ICSP '97,* 759-763.

Kim J.-J., et al.(1997), An analysis of some prosodic aspects of Korean utterances using K-ToBI labelling system, *Proceedings of ICSP '97,* 87-91.
[K-ToBI site - http://speech.wonkwang.ac.kr/~jjkim/ktobi.html]

Kohler, K. J.(1991), *A Model of German Intonation, In K. J. Kohler, editor, Studies in German Intonation,* Universitat Kiel.

Koo H.-S.(1986), *An Experimental Acoustic Study of the Phonetics of Intonation in Standard Korean,* Ph. D. dissertation, Univ. of Texas at Austin, 한신문화사.

Pitrelli, John et al.(1994), Evaluation of prosodic transcription labeling reliability in the ToBI framework, *Proceedings of ICSLP '94, vol 2,* 867-870.

Ross, K. N.(1995), Modeling of Intonation for Speech Synthesis, Ph. D. dissertation, Boston Univ.

Silverman, Kím et al.(1992), "ToBI : a standard for labeling English prosody", *Proceedings of ICSLP '92, vol 2,* 867-870.

Taylor, P.(1992), A Phonetic Model of English Intonation, Ph. D. dissertation, Univ.of Edinburgh.
[Homepage of Taylor - http://www.cstr.ed.ac.uk/~pualt]

Taylor, P.(1994), The Rise/Fall/Connection Model of Intonation, *Speech Communication 15,* 169-186,
[http://www.cstr.ed.ac.uk/~pualt]

Taylor, P.(1997), Analysis and Synthesis of Intonation using the Tilt Model, Draft Journal paper on Tilt model.
[http://www.cstr.ed.ac.uk/~pualt]

KKWON 한국어 음성 데이터베이스 CD-ROM.

FESTIVAL Speech Synthesis System (version 1.2.0, Sep. 1997), Center for Speech Technology Research (CSTR), Univ. of Edinburgh.
[Homepage of FESTIVAL- http://www.cstr.ed.ac.uk/projects/festival]
waves+ Manual (version 5.1, Mar. 1996), Entropic Research Laboratory, Inc.

접수일자: 1998년 10월 14일
게재결정: 1998년 11월 1일

▶ 표경란(Kyung Nan Pyo)
주소: 부산광역시 금정구 장전동 산 30
소속: 부산대학교 인지과학 협동과정

▶ 김형순
주소: 부산광역시 금정구 장전동 산 30
소속: 부산대학교 공과대학 전자공학과
전화: 051) 510-2452
e-mail: kimhs@hyowon.cc.pusan.ac.kr

▶ 최규수(Kyu Soo Choi)
주소: 부산광역시 금정구 장전동 산 30
소속: 부산대학교 인지과학 협동과정