제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

RFC 모델의 한국어 억양 곡선에의 적용

표경란, 김형순, 최규수
부산대학교 인지과학협동과정

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

Kyung Nan Pyo, Hyung Soon Kim, Kyu Soo Choi
Dept. of Interdisciplinary Research Program of Cognitive Science, Graduate School, Pusan National University
pkn@rabiner.ee.pusan.ac.kr, {kimhs,kschoi2}@hyowon.cc.pusan.ac.kr

요 약

본 논문에서는 합성음에 사용할 한국어 억양 모델을 세우기 위한 기초적 연구로서 한국어 억양 곡선에 RFC 모델을 적용해 보았다. 억양 곡선의 구조는 피치 액센트와 억양구 경계 음조의 연속으로 되어 있는데, RFC 모델은 각각의 진폭과 지속시간을 가지는 상승 음조 요소와 하강 음조 요소, 그리고 연결 요소로 이러한 억양 곡선의 모양을 모델링한다[1][2]. 본 논문에서는 한국어 억양 곡선의 특징을 잘 반영하도록 RFC 모델의 구성 요소를 수정하고, 청취 실험을 통해서 원래의 RFC 모델과 수정된 RFC 모델을 비교해 보았다. 실험 결과는 수정된 RFC 모델이 원래의 RFC 모델보다 13%정도 음조 표지 개수가 줄었음에도 불구하고 청각적으로 인지하는데 차이가 없는 것으로 나타났다.

1. 서 론

일반적으로 음성 합성기에서 억양 곡선(intonation or pitch contour)을 합성하는 모듈은 입력 문장으로부터 F0의 궤적을 예측하는 부분과 그 예측된 변수를 이용하여 원하는 F0 곡선을 생성하는 부분으로 나누어지는데[3], 본 논문에서 다루고 있는 것은 F0 곡선의 생성을 위한 억양 곡선의 모델이다. 억양 곡선을 생성하는 모델은 소스/필터 모델, 목표점/보간 모델 등과 같이 매개 변수를 사용하는 모델과 은닉 마코프 모델, 신경망 모델, 연결 합성 모델처럼 매개변수를 사용하지 않는 모델로 양분할 수 있다[4].

본 논문의 목적은 억양 합성에서 사용하기 쉽도록 구체적인 수치 형태로 표현가능하면서, 자연음의 억양과 유사하고 자연스러운 음조를 나타내는 억양 모델을 선택하여 한국어의 억양을 모델링하는 것이다. 이를 위하여 본 논문에서는 목표점/보간 모델중의 하나인 Rise/Fall/Connection(RFC)모델을 한국어 억양에 적용해 보았다. 이 모델은 하락선(declination line)의 설정없이 보간될 목표점 자체에 억양의 하락 현상을 나타내도록 하여 그 목표점들을 연결하여 억양 곡선을 모델링한다. RFC 모델은 음향학적으로 F0 값을 모델링하는 것이므로, 한국어 억양의 음향학적인 특성을 고려해서 한국어 억양에 적합하도록 RFC 모델을 적용한다면 억양 곡선을 잘 나타낼 수 있을 것이다.

본 논문의 구성은 다음과 같다. 2 절에서는 영어 억양에 적용된 원래의 RFC 모델에 대해서 기술하고, 3 절에서는 한국어 억양 곡선을 RFC 모델로 분석하고 합성하는 과정을 통해 한국어에 적합하게 모델의 구성 요소를 수정한 RFC 모델을 제시한다. 마지막으로 4 절에서 결론을 맺는다.

2. RFC 모델

RFC 모델은 하나의 억양구를 각 음절에 얹힌 피치 액센트의 연속된 모양과 억양구의 끝을 나타내는 억양

구 경계 음조로 분석하고 각각의 음조를 구성하는 요소의 값을 나타내는 변수로 모델링된다[2].

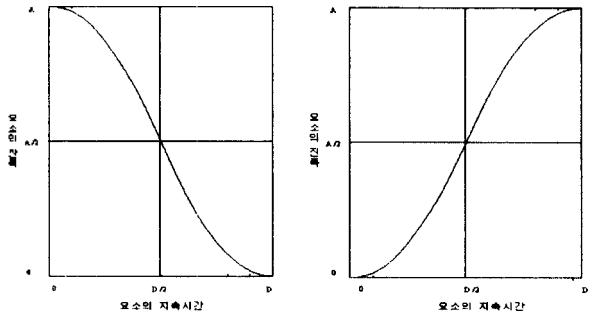
2.1 피치 액센트의 모델링

하나의 액센트는 액센트를 이루는 요소인 상승음조(rising tone)와 하강음조(falling tone)로 나누어 모델링한다. 보통 상승음조보다 하강음조의 진폭이 크게 나타나는데, 이것은 억양 곡선의 하강현상과 관계가 있다. 따라서 상승음조와 하강음조를 나누어 모델링하면, 억양 곡선의 전체적인 하락선을 따로 설정하지 않아도 되고, 다양한 피치 액센트를 나타낼 수 있다.

식(1)과 (2)는 Taylor[2]가 RFC 모델을 구현하기 위해서 제안한 식인데, 상승음조 요소와 하강음조 요소를 모델링할 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) (b)

그림 1. 하강음조 요소와 상승음조 요소의 모델링
(a) 하강음조 요소의 모델링
(b) 상승음조 요소의 모델링

식(1)은 하강음조 요소의 F0 곡선의 모양을 나타내는 식이고, 식(1)을 진폭축(f_0)으로 대칭하면 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 모델링한다. 그림 1은 식(1)과 식(2)를 그래프로 나타낸 것이다.

2.2 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)는 상승음조 요소로 모델링된다.

2.3 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 F0 곡선이 연속된 값을 가지도록 모델링한다. 이 때의 직선을 연결 요소(connection element)라고 하고, 이 요소 또한 진폭과 지속시간의 파라미터를 가진다.

3. RFC 모델의 한국어에의 적용

본 연구에서는 RFC 모델을 한국어 억양 곡선에 적용해 보기 위해 원광대학교에서 구축된 낭독 음성 데이터베이스를 사용하였다[5]. 이 데이터베이스는 16 kHz로 샘플링 되어 있고, 다양한 형식의 평서문으로 되어 있으며, 표준어의 억양으로 녹음이 되어 있다. 이 중에서 본 연구에서 분석에 사용한 데이터베이스는 50문장으로 된 공동세트 이다.

3.1 F0 값 뽑아내기와 억양 곡선 고르기

F0 값은 Entropic 사의 ESPS/Xwaves+를 사용하여 5 ms 간격으로 뽑아 텍스트 파일 형식으로 저장했다. 그리고 F0 값을 뽑아내는 알고리즘 자체의 추정오차와 인접 분절음의 영향에 의해 생긴 급격한 변화의 F0 값을 제거하기 위해서 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. 그리고 듣는이가 무성음 구간 근처의 유성음에 의해서 무성음 구간을 마스킹(masking)해서

인지한다는 연구 결과를 토대로[6], F0 값이 나타나지 않는 무성음 구간을 직선 보간(straight line interpolation)하여 연속된 값을 가지도록 하였다. 마지막으로 직선 보간을 할 때 생기는 날카롭게 꺾이는 부분을 제거하고 부드러운 곡선으로 만들기 위해서 최종적으로 7 point 메디안 필터링을 하였다.

그림 2 는 "아들 똘이 너석과 같은 유치원에 다니는 옆집 현아 엄마의 전화였습니다." 라는 문장에서 뽑아낸 F0 궤적에 전처리를 하는 과정을 나타낸 것이다.

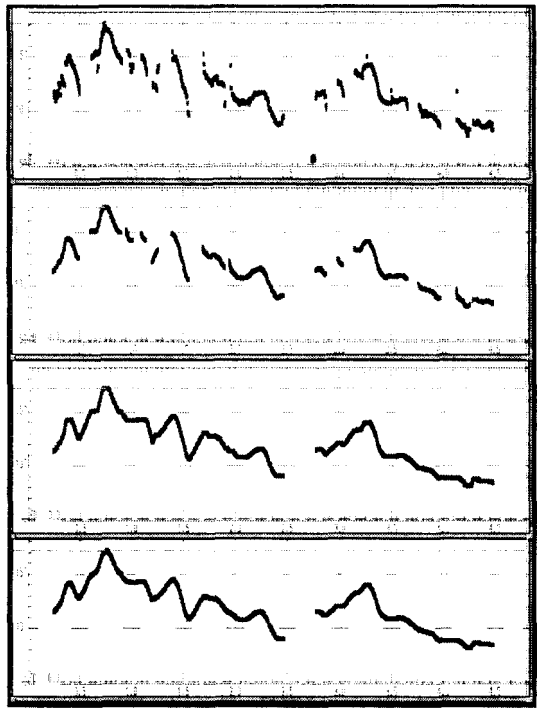


그림 2. 억양 곡선의 고르기 과정
(a) 원래의 F0 궤적
(b) 15 point 메디안 필터링 후의 F0 곡선
(c) 무성음 부분을 직선 보간한 후의 F0 곡선
(d) 7 point 메디안 필터링 후의 F0 곡선

3.2 RFC 모델을 이용한 음조 표지 붙이기

ESPS 의 xlabel 을 사용하여 상승음조에는 r, 하강음조에는 f, 피치 액센트가 없는 무분에는 c, 억양구 경계 음조에는 기존의 상승음조로 된 억양구 경계인 rb 외에 하강음조로 된 억양구 경계 음조(falling boundary tone)에는 fb, 연결 요소로 된 억양구 경계 음조(continuation boundary tone)에는 cb 라는 음조 표지를 붙였다.

이들 2 가지의 억양구 경계 음조를 더 설정한 이유는, 영어 억양 곡선에 적용했던 원래의 RFC 모델에는 rb 만 있었는데, 한국어 억양 곡선의 경우에는 억양구 경계 음조에 하강음조 요소나 연결 요소로 모델링해야 하는 음조가 나타나기 때문이다. 그리고 표준어 억양에서 흔히 보이는 계단식 음조는 청각적으로 그 단계가 느껴지지 않으므로 하나의 음조로 표지하였다.

그림 3 은 RFC 모델로 한국어 억양에 음조를 표지한 예이고, 표 1 은 50 문장을 분석한 결과이다.

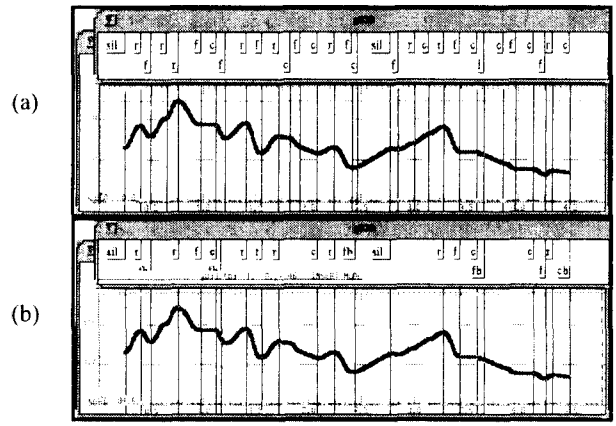


그림 3. RFC 모델을 이용한 한국어 억양의 음조 표지 예(그림 2 와 같은 문장)
(a) 원래의 RFC 모델로 음조 표지한 경우
(b) 계단식 음조를 하나의 음조로 단순화하고, 3 가지 억양구 경계 음조로 된 수정한 RFC 모델로 음조 표지한 경우

표 1. 50 문장의 음조 표지의 총 개수

	r	f	c	rb	fb	cb	합계
(a) 원래의 RFC 모델	557	704	476	80	0	0	1817
(b) 수정한 RFC 모델	463	422	380	80	199	39	1538
(a) - (b)	94	282	96	0	-199	-39	234

표 1 을 보면 수정된 RFC 모델의 전체 음조 표지의 개수가 원래 RFC 모델의 음조 표지 개수보다 13%가 줄었다. 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 RFC 모델을 억양의 합성 모델로 사용하려 할 때에는, 음조 표지들 사이의 패턴을 찾아서 억양이 나타나는 양상을 좀더 간단

허 하면 음조 표지의 개수를 더 줄일 수도 있을 것이다.

3.3 RFC 모델을 이용한 억양 곡선의 합성과 청취 실험

수정된 RFC 모델로 억양 곡선을 합성하여도 원래 억양의 음조와 차이가 나지 않을 것이라는 가정 하에, 청취 실험을 통해 가정을 검증하기 위해서 앞 절에서 사용한 50 문장 중에서 원래의 RFC 모델로 음조 표지한 음조의 개수와 차이가 많이 나고, 합성음이 명료한 문장 10 개를 골라서 청취실험에 사용하였다. 이때 합성음은 ESPS 의 LPC 합성 방법을 사용하여 생성하였다.

청취 실험은 실험실 환경에서 부산대학교 전자공학과 대학원생 15 명에게 헤드폰을 착용하게 한 후 실시하였다. 실험 방법은 자연음의 억양을 그대로 사용하여 합성한 문장을 기준음으로 들려주고, 원래의 RFC 모델로 합성한 문장(A)과 수정된 RFC 모델로 합성한 문장(B)을 피험자에게 어떤 합성음이 어떤 방법으로 생성한 것인지 알 수 없도록 무작위(random)로 들려주고 A 와 B 중 어느 것의 억양이 기준음의 억양에 가까운지를 선택하게 하였다. 이 때 A 와 B 중 어느 것이 기준음의 억양과 더 가까운지 우열을 결정할 수 없는 경우에는 ? 표를 선택하도록 하였다. 청취 실험 결과는 그림 4 와 같았다.

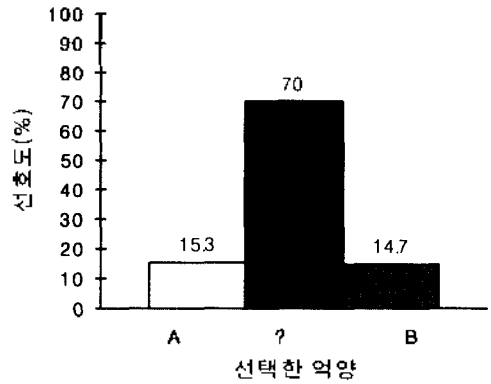


그림 4. 합성한 억양의 선호도에 관한 청취실험 결과
□ 원래의 RFC 모델로 합성한 억양 A 의 선택
■ 두 억양의 우열을 구별할 수 없는 경우 ? 선택
▨ 수정된 RFC 모델로 합성한 억양 B 의 선택

실험 결과는 그림 4 에서 보는 바와 같이 수정된 RFC 모델로 억양 곡선을 합성하여도 자연음 억양의 음조와 차이가 나지 않음을 알 수 있었다. 즉, 2 종류의 억양구 경계 음조를 더 설정하여 전체 억양구 경계 음조의 개수는 늘었고, 피치 변화에 덜 민감한 부분 하나의 음조로 합쳐서 전체 r, f, c 의 개수가 줄었지만 자연음의 억양과 별로 차이가 나지 않았다. 따라서 수정된 RFC 모델로 분석/합성한 억양 곡선의 패턴을 유형화하면, 자연스러운 억양을 생성하는데 도움을 줄 수 있을 것으로 생각된다.

4. 결 론

합성음에 사용할 억양 모형을 생성하기 위해서는 자연음의 억양 곡선의 특징을 잘 반영하면서 구현하기도 쉽도록 정량적인 값으로 억양 곡선을 모형화 하는 모델이 필요하다. RFC 모델은 이러한 면에서 한국어 억양에도 충분히 적용할 가능성이 큰 모형이므로, 합성기의 억양 합성 모듈 중에서 억양 곡선을 생성하는 부분에서 사용될 수 있을 것이라고 생각한다.

그리고 본 논문에서는 수동으로 음조 표지를 하였지만 시간과 노력이 많이 소요되므로, 억양 합성에 필요한 많은 데이터베이스를 확보하기 위해서는 기계가 자동으로 음조를 표지할 수 있도록 되어야 할 것이며, 입력 문장에서 억양의 구조를 예측하는 모듈을 세우는 연구가 앞으로 더 진행되어야 할 것이다.

< 참 고 문 헌 >

[1] Paul Taylor, *A Phonetic Model of English Intonation*, Ph. D. dissertation, Univ. of Edinburgh, 1992.

[2] Paul Taylor, "The Rise/Fall/Connection Model of Intonation", *Speech communication 15*, 1994.

[3] 오영환, "*음성언어정보처리* ", 홍릉과학출판사, 1997.

[4] K. N. Ross, *Modeling of Intonation for Speech Synthesis*, Ph.D. dissertation, Boston Univ., 1995.

[5] KKWON 한국어 음성 데이터베이스 CD-ROM

[6] K. J. Kohler "A Model of German Intonation", In K. J. Kohler, editor, *Studies in German Intonation*, Universitat Kiel, 1991.