시간적 분해에 기반한 F0 궤적 모델에 관한 연구

F0 Contour Model based on Temporal Decomposition

변 효 진*, 김 연 준*, 오 영 환*
(Heo Jin Byeon*, Yeon Jun Kim*, Yung Hwan Oh*)

요 약

본 논문에서는 음성합성의 억양 제어를 위한 새로운 F0 궤적 모델을 제안한다. 제안한 모델은 발성된 문장의 F0 궤적을 중첩가산되는 사건들로 분해하고, 각 사건들을 가우시안 종모양의 사건함수로 모델링한다. 그리고 제안한 모델을 위한 파라미터 추정 알고리즘을 제시한다. 제안한 모델은 특정한 음운론적 지식에 기반하지 않았으며, F0 궤적의 분석단계와 합성단계에 모두 사용 가능하다. 제안한 모델의 성능평가를 위해 다양한 장르에서 추출한 여러 형태의 500문장의 코퍼스를 구축하고, 이를 전문 아나운서에게 발성하게 하여 구축한 음성코퍼스로 실험한 결과, 원음성의 F0 궤적과 제안한 모델에 의해 합성된 F0 궤적의 평균 제곱 오류근이 7.87Hz이었다.

ABSTRACT

This paper proposes a new F0 contour model for intonation control in speech synthesis. We assume that the F0 contour of an utterance can be described using a sequence of time-overlapping events, which determine the fluctuation of a given F0 contour, described by asymmetric Gaussian functions. In addition, We propose a parameter estimation algorithm for the proposed model. The proposed model is not developed with a particular phonological theory in mind, and can be used in both F0 contour analysis and synthesis. For testing our F0 model, we collected 500 sentences from various genres and built a corresponding speech corpus uttered by a professional female announcer. As a result of F0 resynthesis experiment using the proposed model, the RMSE was 7.87Hz for given speech corpus.

I. 서 론

억양이란 문장 단위의 음성에 있어서 강세를 갖는 음절간의 상대적인 음높이를 의미한다. 한국어에 있어서 문장의 억양은 의미를 변화시키지는 않지만 문장의 형태, 구문구조, 화제의 변경, 의미, 감정에 따라 다양한 형태로 나타난다. 그러므로, 문서-음성 변환 시스템(text-to-speech system)에서 자연스러운 합성음을 생성하기 위하여 억양제어는 매우 중요한 부분이라고 할 수 있다[1]. 이러한 억양은 물리적 신호인 F0 궤적을 사용하여 표현되는데, F0 궤적이 억양의 양적인 기술에 가장 적절하다고 판단되기 때문이다. F0 궤적이란 음성의 주기적인 특성을 나타내는 기본주파수의 시간축상에서의 궤적을 의미한다. F0 궤적의 효과적인 모델은 적은 수의 파라미터로 정확한 F0 궤적을 표현하므로써, 입력되는 문장으로부터 최소한의 노력으로 파라미터를 추정하여 자연스러운 억양을 합성할 수 있게 한다. 또한 음성인식 분야에 활용되어 발성된 문장이 갖는 중의성 해소에 이용될 수도 있다[2].

* 한국과학기술원 전산학과
접수일자 : 1999년 7월 1일

궁극적으로 합성음의 자연스러운 F0 궤적을 생성하기 위한 억양 모델은 언어학적 지식을 이용한 모델(linguistic model), 억양의 지각적 특성을 이용한 모델(perceptual model) 그리고 F0 궤적 자체를 모델링하는 음향학적 모델(acoustic model) 등 크게 3가지의 형태로 나뉠 수 있다[3]. 언어학적 지식을 이용한 대표적인 방법으로는 발성된 문장의 음조(tone)의 변화를 기술하고 음조와 F0 궤적간의 관계를 모델링하는 것으로 ToBI (Tones and Break Indices)를 이용한 동적 시스템 모델[4], 선형회귀 모델[5] 등이 있다. 이러한 모델들은 발성된 문장의 억양을 ToBi 레이블링 시스템과 같이 억양의 형태를 분류한 레이블을 이용하여 표현하고, 결정된 레이블열로부터 F0 궤적을 추정하는 방식을 이용한다. 따라서 체계적인 음운론에 기반한 억양분류가 필요하며, 분류된 억양패턴의 F0 궤적으로의 변환을 위한 모델이 필요하다. 억양의 지각적 특성을 이용한 방법으로는 인간의 청각특성상 구별할 수 없는 F0 궤적의 변화를 무시하여 단순화된 F0 궤적을 생성하는 방법이 있다[6]. 이 방법으로 단순화된 F0 궤적은 몇 개의 직선으로 표현되기 때문에 비교적 적은 양의 정보로 청각적으로 차이가 적은 F0 궤적을 생성할 수 있다. 이와 달리, F0 궤적 자체를 모델링하는 음향학적 모델은 억양의 변화를 몇 개의 범주로 한정

하지 않고 파라미터의 변화에 따라 다양한 형태의 F0 궤적을 보다 정확히 표현할 수 있는 장점이 있다. 또한 모델의 접근방식에 따라 특정한 음운론적 지식에 기반하지 않고 F0 궤적 변화에 충실한 새로운 형태의 억양제어 요소를 제시할 수도 있다. 그리고 F0 궤적의 분석단계에서 추정된 파라미터를 이용하여 합성단계에서 발성된 문장의 F0 궤적을 그대로 복원할 수 있다. 음향학적 모델의 대표적인 예로는 Fujisaki 모델[7], RFC (Rise/Fall/Connection) 모델[8] 등이 있다. 본 논문에서는 이러한 음향학적 모델이 갖는 장점을 살려, 특정한 음운론적 지식에 기반하지 않고 F0 궤적 변화 특성을 잘 모델링할 수 있는 새로운 F0 궤적 모델을 제안한다. 제안된 모델은 발성된 문장의 F0 궤적을 중첩가산되는 사건들의 열로 가정하고, 각 사건들을 F0 궤적으로부터 분해하여 가우시안 종모양의 사건함수로 모델링한다. 그리고 제안된 모델의 파라미터를 추정할 수 있는 알고리즘을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 F0 궤적 모델에 대해 살펴보고, 본 연구의 배경에 대해 설명한다. 3장에서는 본 논문에서 제안한 시간적 분해에 의한 F0 궤적 모델을 자세히 설명하고, 제안한 모델의 파라미터 추정방법을 제시한다. 4장에서는 제안한 모델의 성능을 실험을 통해 보이고, 끝으로 5장에서 결론을 맺는다.

II. 관련 연구

F0 궤적 모델의 목적은 발성된 문장의 F0 궤적을 가능한 적은 수의 파라미터로 표현하여 효과적으로 문장의 억양을 분석 또는 합성하는 것이다. 특히 F0 궤적을 몇 개의 정해진 패턴으로 분류하여 분석 또는 합성하는 방법(prototypical model)보다는 연속적인 값을 갖는 파라미터로 표현하여 다양한 F0 궤적의 변이를 표현할 수 있는 방법이 더욱 효과적이다[8]. 이러한 모델들의 예로는 Fujisaki 모델, RFC 모델, MBD (maximum-based description) 모델[9], PaIntE (Parametric intonation event) 모델[10] 등이 있다.

Fujisaki 모델

Fujisaki 모델은 음원을 운율구의 기저패턴을 나타내는 요소(phrase command)와 단어의 액센트를 표현하는 요소(accent command)로 구분하고, 이 두 요소를 성문의 진동 메카니즘(glottal oscillation mechanism)을 모델링한 필터를 통과시켜 중첩가산함으로써 발성된 문장의 F0 궤적을 구한다. 따라서 이 모델은 음원을 표현하는 두 가지 요소를 파라미터화하여 F0 궤적을 분석 또는 합성할 수 있다. 운율구의 기저패턴을 나타내는 요소는 시간축상에서의 발생위치와 크기를 갖는 펄스로 파라미터화되고, 단어의 액센트를 표현하는 요소는 시간축상에서의 발생위치 및 지속시간과 크기를 갖는 펄스로 파라미터화된다. 식 2.1은 Fujisaki 모델을 수식화하여 보인다.

$$\ln F0(t) = \ln F_{min} + \sum_{i=1}^{I} A_{pi} G_{pi}(t - T_{0i}) + \sum_{j=1}^{J} A_{aj}\{G_{aj}(t - T_{1j}) - G_{aj}(t - T_{2j})\}$$

단,

$$G_{pi}(t) = \begin{cases} \alpha_i^2 t \exp(-\alpha_i t) & , \text{ for } t \geq 0 \\ 0 & , \text{ for } t < 0 \end{cases}$$

$$G_{aj}(t) = \begin{cases} Min[1-(1+\beta_j t)\exp(-\beta_j t), \theta_j]) & , \text{ for } t \geq 0 \\ 0 & , \text{ for } t < 0 \end{cases}$$

(식 2.1)

식 2.1에서 I, J는 각각 운율구의 기저패턴을 나타내는 요소와 단어의 액센트를 나타내는 요소의 개수를 나타내며, A_{pi}와 A_{aj}는 각 요소의 크기, T_{0i}과 T_{1j}, T_{2j}는 각 요소의 발생위치와 지속시간을 나타낸다. 그리고 α_i, β_j, θ_j는 필터의 특성을 나타내는 파라미터이다. 이러한 Fujisaki 모델은 일본어의 평서문 억양특성을 기반으로 만들어 졌다. 따라서 문장의 끝에서 F0 궤적이 증가형태가 되는 의문문이나 F0 궤적의 하강현상(declination)이 뚜렷하지 않은 억양의 분석 및 합성에는 적합하지가 않다. 또한 모델의 파라미터는 분석-합성 방식(analysis-by-synthesis)에 의해 추출하는데, 각 요소의 개수와 파라미터를 자동으로 추출하기 어려운 문제점이 있다. 그러나 F0 궤적을 몇 개의 운율구와 단어의 액센트로 분리하는 비교적 간단명료한 모델의 특성 때문에 여러 언어권에서 변형된 방법으로 널리 이용되고 있다[11].

RFC 모델

RFC 모델은 F0 궤적의 증가부분을 R, 감소부분을 F로 표현하고 각각을 2차 곡선으로 모델링한다. 또, 증감의 기울기가 억양의 인지에 큰 영향을 주지 못하는 부분을 단순한 직선형태로 잇는 C로 표현한다. 식 2.2는 F로 표현된 F0 궤적을 모델링한 식을 보이며, A는 F0 궤적의 크기, D는 지속시간을 나타낸다.

$$F0(t) = \begin{cases} A - 2A(\frac{t}{D})^2 & , \ 0 < t \leq \frac{D}{2} \\ 2A(1-\frac{t}{D})^2 & , \ \frac{D}{2} < t < D \end{cases} \quad \text{(식 2.2)}$$

RFC 모델은 식 2.2에서와 같이 A와 D의 값에 따라 다양한 형태의 기울기와 지속시간을 갖는 F0 궤적의 증감형태를 표현할 수 있으며, 단순히 F0 궤적의 증감 형태에 따라 파라미터를 추출할 수 있으므로 특정한 음운론적 지식이 요구되지 않는다. 따라서 RFC 모델은 영어권의 억양을 표현하기 위해 개발된 모델이나 다른 언어권에도 쉽게 적용할 수 있는 장점이 있다. 실제로 F0 궤적의 증감을 표현하는 레이블을 확장하여 한국어 대화체 음성의 F0 궤적을 표현하는데 적용된 바 있다[12]. 또한 F0 궤적

의 증감 기울기를 크기와 지속시간으로 파라미터화하는 것이므로 정형화된 방법으로 쉽게 파라미터를 자동추출할 수 있는 장점이 있다. 그러나 이 모델은 R, F와 C의 구분을 위하여 억양의 인지에 영향을 주는 기울기에 대한 임계값을 구해야 한다. 이 임계값을 구하는 방법은 전문가가 실제 음성을 듣고 억양의 인지에 영향을 주는 부분과 그렇지 못한 부분을 구분하여 정한다. 따라서 임계값에 의해 파라미터의 개수와 원음성의 F0 궤적과 RFC 모델에 의해 재합성된 F0 궤적간의 오류가 큰 영향을 받는다. 임계값을 높게 설정하면 파라미터 개수가 적어지는 반면 오류가 커지고, 반대로 임계값을 낮게 설정할 경우 오류가 줄어드는 반면 추출해야 할 파라미터 개수가 늘어나는 단점이 있다.

MBD 모델과 PaIntE 모델

MBD 모델은 F0 궤적의 증감 변화의 중심이 되는 국부 최대값이 발생하는 위치를 중심으로 좌우에 나타나는 F0 궤적을 $\cos^2$ 함수를 이용하여 근사시킨다. 이 모델로 F0 궤적을 기술하기 위해서는 하나의 국부 최대값에 대해 발생 위치를 나타내는 delay, 크기를 나타내는 amplitude, 좌측에 나타나는 국부 최소값까지의 $\cos^2$함수 근사를 위한 파라미터인 rise와 우측에 나타나는 국부 최소값까지의 근사를 위한 파라미터인 fall 등 모두 4개의 파라미터가 필요하다. 식 2.3은 국부최대값 좌측에 나타나는 F0 궤적을 근사한 식을 보인다.

$$F0(t) = amplitude \cdot \cos^2\left(\frac{t \cdot rise}{4amplitude}\right) \qquad \text{(식 2.3)}$$

식 2.3에서 t는 국부 최대값의 발생위치에 대한 상대적인 시간을 나타내는 변수이다. MBD 모델은 국부 최대값 발생 위치에서 좌우 국부 최소값까지의 F0 궤적을 모델링한 것이므로 RFC 모델에서의 C 요소가 생략된다. 따라서 임계값에 의한 구분이 없어 임계값 설정에 따른 문제가 없는 장점이 있으나, 근사 구간을 하나의 함수 형태로 표현하기 때문에 기울기 변화를 표현하는 RFC 모델의 C 요소를 근본적으로 해결한 것은 아니다.

PaIntE 모델은 MBD 모델과 같이 국부 최대값이 발생하는 위치를 중심으로 좌우에 나타나는 F0 궤적을 모델링한다. 이 모델이 MBD 모델과 다른점은 $\cos^2$ 함수로 근사하는 것이 아니라 시그모이드(sigmoid)함수로 근사한다는 점이다. 따라서 이 모델도 MBD 모델과 같이 RFC 모델에서의 C 요소의 문제를 근본적으로 해결하지 못한다.

본 논문에서는 앞에서 기술한 여러 F0 궤적 모델들의 장점을 취하여 정형화된 방법으로 파라미터를 쉽게 추출할 수 있는 시간적 분해에 기반한 F0 궤적 모델을 제안한다. 제안한 모델은 F0 궤적을 중첩가산되는 사건들의 열로 가정한다. 각 사건들은 MBD 모델이나 PaIntE 모델과 같이 국부 최대값을 중심으로 발생한다고 가정함으로써 사건의 발생위치와 개수를 쉽게 추출할 수 있도록 하였다. 또한 각 사건들을 단조증감 형태인 가우시안 종 모양의 함수로 근사함으로써 사건함수를 이루는 파라미터들의 추출도 용이하게 하였다. 그리고 각 사건들을 Fujisaki 모델과 같이 서로 중첩가산시키므로써 두 사건 사이에 기울기의 변화가 일어나는 RFC 모델의 C 요소를 포함할 수 있도록 하였다.

III. 시간적 분해에 기반한 F0 궤적 모델

본 논문에서 제안한 모델은 F0 궤적을 시간적으로 떨어져 있는 사건들의 열로 표현한다. 즉, F0 궤적의 변화를 결정하는 몇 개의 사건들이 서로 중첩가산되어 발성된 문장의 F0 궤적을 형성한다고 가정한다. 이러한 가정은 기존의 F0 궤적 모델들의 표현 방법을 일반화한 것이다. 예를 들면, Fujisaki 모델에서 각 운율구의 기저패턴을 나타내는 요소와 단어의 액센트를 나타내는 요소를 발생시간에 따라 나열하면, 이를 F0 궤적을 형성하는 사건들의 열로 볼 수 있다. RFC 모델의 경우에서도 F0 궤적 표현 방법은 R, F 또는 C로 표현되는 사건들의 열로 설명할 수 있다. 이와 마찬가지로 MBD 모델이나 PaIntE 모델에서도 각 모델의 표현방법은 국부 최대값을 중심으로 나타나는 사건들의 열로 해석 가능하다. 따라서 제안한 모델은 이러한 가정에 기반하여 F0 궤적을 몇 개의 사건들로 분해하고 각 사건들을 정형화하여 표현할 수 있는 방법을 제시한다.

3.1 F0 궤적의 시간적 분해

F0 궤적을 몇 개의 사건들의 중첩가산으로 표현하기 위해서는 각 사건의 시간축에서의 변화하는 모습과 발생위치를 결정해야 한다. F0 궤적을 이루는 사건들 중 j번째 사건의 시간축에서의 변화 형태를 시간 t에 대한 함수로 표현한 사건함수를 $event_j(t)$라 하면, F0 궤적의 시간 t에 대한 함수 $F0(t)$는 식 3.1과 같다.

$$F0(t) = \sum_{j=1}^{J} event_j(t) \qquad \text{(식 3.1)}$$

식 3.1에서 J는 F0 궤적을 이루는 사건의 개수를 의미한다. 이 때, F0 궤적을 이루는 사건들은 그 수가 적으면서 단순한 형태일수록 효과적으로 F0 궤적을 표현할 수 있다. 일반적으로 F0 궤적을 모델링 할 때, F0 궤적의 국부 최대값을 중심으로 표현하는 방법이 널리 쓰인다. 이는 발성된 문장의 억양을 지각하는 것은 F0 궤적의 절대값보다는 높낮이의 변화를 지각한다는 가정을 바탕으로 한 것이다. 예를 들어, RFC 모델이나 MBD 모델, PaIntE 모델 등은 모두 F0 궤적의 국부 최대값의 좌우에 나타나는 F0 궤적의 형태를 모델링한 것이며, Fujisaki 모델 또한 운율구의 기저패턴으로부터 단어 액센트에 의한 F0 궤적의 변화가 얼마만큼 증가후 감소하느냐는 것을 모델링한다. 따라서 본 논문에서 제안한 시간적 분해 모델에서도 F0 궤적을 이루는 사건들은 F0 궤적의 국부 최대값을 중심으로 발생한다고 가정한다.

사건함수 정의

사건함수는 F0 궤적을 이루는 가장 기본적인 단위를 표현해야 한다. 즉, 화자가 문장을 발성할 때, F0 궤적의 변화를 주기 위해 취하는 한번의 행위를 사건함수로 가정한다. 따라서 시간적 분해 모델에서의 사건함수는 단조증가후 단조감소하는 형태로 가정한다. 만약 단조증감이 아닌 두번 이상의 굴곡을 갖는 사건이 있다면, 이 사건을 화자가 취하기 위해서는 굴곡이 변하는 부분에서 또 한번의 행위를 취해야 하므로 한번의 행위로 사건함수가 이루어진다는 가정에 위배된다. 또한 단조증감은 비교적 간단한 형태이므로 사건함수를 파라미터화하기가 쉽다. 이러한 조건을 만족하는 사건함수로 시간적 분해 모델에서는 식 3.2와 같은 가우시안 종모양의 함수를 이용한다.

$$event(t) = a + b\exp(-0.5(\frac{t-c}{d})^2) \qquad \text{(식 3.2)}$$

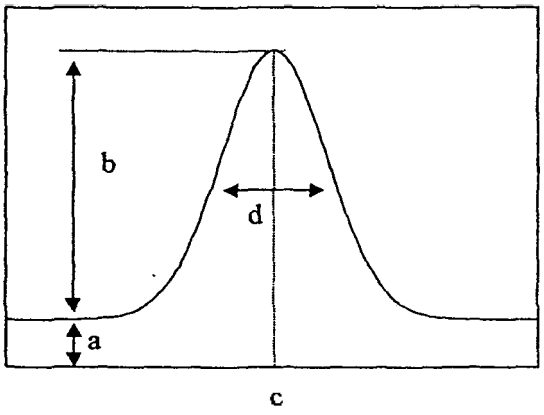

그림 3.1. 가우시안 종모양의 사건함수
Fig. 3.1. Gaussian bell shape event function.

식 3.2에서 a는 사건의 기저값, b는 사건의 크기를 나타내며, c와 d는 각각 사건의 중심위치와 사건의 영향도(degree of effect)를 나타낸다. 그림 3.1은 식 3.2를 도식화한 것이다. 가우시안 종모양의 사건함수는 d의 값을 변화시키므로써 다양한 기울기 즉, 사건함수가 좌우 F0 궤적에 미치는 영향도를 반영할 수 있다. 그러나, 그림 3.1에서 알 수 있듯이 식 3.2의 경우는 사건의 중심인 c를 중심으로 좌우대칭형이므로 다양한 형태의 사건함수를 표현하기에 어려움이 따른다. 따라서 본 논문에서 제안한 시간적 분해 모델에서는 사건함수가 좌우에 미치는 서로 다른 영향도에 변화를 줄 수 있도록, 영향도를 나타내는 d를 사건의 중심 c의 좌우에 식 3.3과 같이 각각 dl과 dr로 나누어 표현한다.

$$event(t) = \begin{cases} a + b\exp(-0.5(\frac{t-c}{dl})^2) & , \ t < c \\ a+b & , \ t = c \\ a + b\exp(-0.5(\frac{t-c}{dr})^2) & , \ t > c \end{cases}$$

(식3.3)

식 3.3에서 좌우함수 모두 $t=c$일 때 함수값은 $a+b$가 되며, 미분값 또한 0이 되므로 정의된 사건함수는 자연스럽게 연속되는 함수형태를 취한다. 그림 3.2는 사건함수의 좌측 영향도가 우측영향도의 $\frac{1}{2}$이 되는 경우의 예를 보인다.

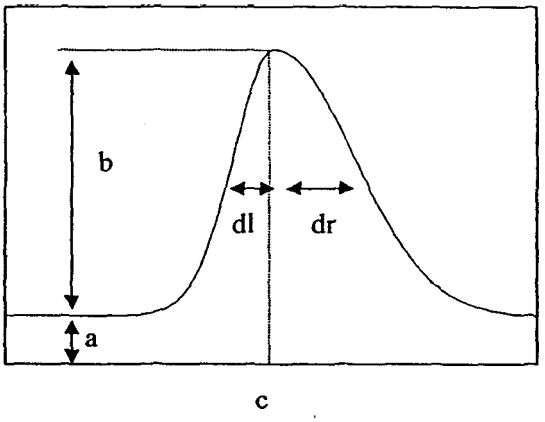

그림 3.2. 가우시안 종모양에 기반한 좌우 비대칭형 사건함수 (dr = 2dl)
Fig. 3.2. Asymmetric Gaussian bell shape event function (dr = 2dl).

운율구 단위의 F0 궤적

운율구란 사람이 문장을 읽을 때 띄어 읽는 것을 모사하기 위한 것으로 지각적으로 띄어 읽는 단위를 의미한다. 그림 3.3은 발성된 문장의 F0 궤적을 운율구별로 나눈 것을 보인다. 운율구가 F0 궤적 모델에서 중요한 의미를 갖는 이유는 그림 3.3에서와 같이 운율구의 F0 궤적의 연속성과 형태를 전체 F0 궤적에서 분리해 낼 수 있기 때문이다. 일반적으로 F0 궤적은 각 운율구마다 서로 다른 기저선(baseline)을 갖고, 운율구 경계부분에서의 변화형태도 다양한 것으로 알려져 있다. F0 궤적을 운율구별로 분리할 수 있다는 것은 F0 궤적을 이루는 각 사건함수들의 영향범위가 운율구별로 제한될 수 있음을 의미한다. 즉, 임의의 사건함수는 F0 궤적 전체에 영향을 미치는 것이 아니라 그 사건함수가 포함되어 있는 운율구 경계내에서만 영향을 미치는 것으로 볼 수 있다. 따라서 각 운율구는 서로 중첩가산되지 않으며, 각 사건함수는 자신이 포함되어 있는 운율구 범위내에서만 중첩가산됨을 의미한다. 이러한 운율구의 특성을 반영하면 식 3.3의 사건함수는 식 3.4와 같이 전개 된다.

if $t, c \in$ *givenphrase*,

$$event(t) = \begin{cases} a + b\exp(-0.5(\frac{t-c}{dl})^2) & , \ t < c \\ a+b & , \ t = c \\ a + b\exp(-0.5(\frac{t-c}{dr})^2) & , \ t > c \end{cases}$$

oterwise,

$$event(t) = 0$$

(식 3.4)

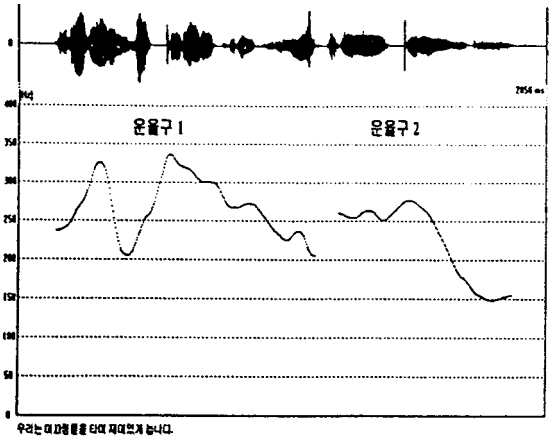

그림 3.3. F0 궤적의 운율구
Fig. 3.3. Prosody phrase of F0 contour.

시간적 분해 모델은 식 3.4와 같은 사건함수들이 자신이 포함된 운율구내에서 중첩가산되므로 사건함수의 기저값을 나타내는 a는 운율구의 기저값보다 클 수가 없다. 따라서 a를 주어진 운율구의 기저값이라 하면, 두면 주어진 운율구의 F0 궤적함수 $phrase(t)$는 식 3.5와 같이 표현할 수 있다.

$$phrase(t) = a + \sum_{i=0}^{I} event_i(t)$$

단,

$$event_i(t) = \begin{cases} b_i \exp(-0.5(\frac{t-c_i}{dl_i})^2) & , \ t < c_i \\ b_i & , \ t = c_i \\ b_i \exp(-0.5(\frac{t-c_i}{dr_i})^2) & , \ t > c_i \end{cases}$$

(식 3.5)

식 3.5에서 I는 주어진 운율구에 포함되어 있는 사건의 개수이며, $event_i(t)$는 주어진 운율구에 포함된 i번째 사건함수를 나타낸다.

시간적 분해 모델

이상에서 기술한 사건함수와 운율구 단위의 F0 궤적을 이용하여 최종적으로 제안된 F0 궤적의 시간적 분해 모델은 다음과 같다. 발성된 문장의 F0 궤적을 표현하는 시간축 변수 t에 대한 함수 $F0(t)$는 식 3.6과 같이 I개의 연속된 운율구의 F0 궤적 함수 $phrase_i(t)$로 분리된다.

$$F0(t) = \sum_{i=1}^{I} phrase_i(t) \qquad \text{(식 3.6)}$$

또한 $phrase_i(t)$는 중첩가산되는 사건함수들의 합으로 이루어져 있으며, 식 3.7과 같이 표현된다.

$$phrase_i(t) = \begin{cases} a_i + \sum_{j \in phrase_i} event_j(t) & , \ t \in phrase_i \\ 0 & , \ otherwise \end{cases}$$

(식 3.7)

식 3.7에서 a_i는 i번째 운율구의 기저값을 나타내며, 사건함수 $event_j(t)$는 식 3.8과 같이 표현된다.

$$event_j(t) = \begin{cases} b_j \exp(-0.5(\frac{t-c_j}{dl_j})^2) & , \ t < c_j \\ b_j & , \ t = c_j \\ b_j \exp(-0.5(\frac{t-c_j}{dr_j})^2) & , \ t > c_j \end{cases}$$

(식 3.8)

식 3.8에서 b_j는 j번째 사건의 크기, c_j는 j번째 사건의 중심위치를 나타내며, dl_j와 dr_j는 각각 j번째 사건의 좌우 형태를 결정하는 파라미터이다.

3.2 파라미터 추정 방법

본절에서는 제안된 F0 궤적의 시간적 분해 모델의 파라미터 추정방법에 대하여 기술한다. 발성된 문장의 F0 궤적이 I개의 운율구로 분리되고 모두 J개의 사건으로 이루어진다면, 시간적 분해 모델은 각 운율구의 기저값을 나타내는 파라미터인 I개의 a와 각 사건함수의 형태를 결정하는 파라미터인 b, c, dl, dr이 각각 J개씩 필요하므로 모두 $I+4J$개의 파라미터가 추정되어야 한다. 이 때, F0 궤적은 서로 중첩가산되지 않는 운율구의 열로 이루어져 있다는 가정에 따라 파라미터 추정은 각 운율구별로 나누어 실행한다.

운율구의 기저값 a 추정

운율구의 기저값 a는 주어진 운율구의 F0 궤적이 a미만의 값은 취하지 않는다는 의미다. 즉, 주어진 운율구의 F0 궤적을 이루는 사건들은 a를 기본값으로 그 위에 중첩가산된다. 따라서 a는 식 3.9와 같이 주어진 운율구의 F0 궤적의 최소값 이하의 값으로 표현된다.

$$a = \min(phrase(t)) - \epsilon, \quad \epsilon \geq 0 \qquad \text{(식 3.9)}$$

사건의 중심위치 c 추정

운율구 F0 궤적을 이루는 각 사건들은 3.1절에서 설명된 것과 같이 운율구의 F0 궤적의 국부 최대값이 나타나는 시간을 중심으로 발생한다. 그러나 사건의 중심위치 c가 국부 최대값이 나타나는 시점이 아닌 경우도 가능하다. 다시 말하면, 제안된 모델은 사건들의 중첩가산에 의해 일어나므로 사건의 중심이 국부 최대값이 나타나는 시간과 항상 일치하지는 않는다. 이는 제안한 모델의 표현방법에 의하면 구하고자하는 현재 사건 이전에 발생하는 사건의

영향에 의해 현재 사건의 중심이 국부 최대값이 나타나는 시간보다 조금 늦게 나타나는 지연현상이 발생하기 때문이다. 이전 사건이 현재 사건에 영향을 주는 부분은 이전 사건의 우측부분 즉, 단조감소형태의 영향을 받는 것이므로 이전 사건의 영향을 F0 궤적에서 제외하고 나면 현재 국부 최대값이 나타나는 시기가 시간적으로 지연될 수는 있어도 앞당겨 질 수는 없다. 따라서 사건의 중심위치 c는 운율구 F0 궤적의 국부 최대값이 나타나는 시간으로 초기화한 후, 순차적으로 구한 사건함수를 운율구 F0 궤적에서 제외시켜나가면서, 초기화된 c의 위치 이후에서의 새로운 국부 최대값이 나타나는 시간으로 추정할 수 있다.

이상과 같은 방법으로 주어진 운율구에서 사건 발생의 중심위치 c를 추정하는 알고리즘은 다음과 같다.

단계 1 : 주어진 운율구의 국부 최대값들을 사건 중심 위치로 초기화 한다.
단계 2 : 주어진 시점의 사건함수를 구한다.
단계 3 : 구해진 사건함수는 원 F0 궤적에서 감산하여 그 영향을 제외시킨다.
단계 4 : 이전 사건함수의 영향을 제외한 F0 궤적에서 현재 사건의 초기화된 사건 발생 중심위치 이후에 나타나는 국부 최대값 위치를 구하고자 하는 사건의 중심위치로 추정한다.
단계 5 : 주어진 운율구의 사건함수가 모두 구해질 때까지 2-4단계를 반복 수행한다.

사건의 좌우 영향도 dl_i, dr_i 추정

운율구 F0 궤적 $phrase(t)$를 이루는 I개의 사건은 순차적으로 발생한다. 따라서 사건도 순차적으로 추정하며, 이미 추정된 사건을 운율구 F0 궤적에서 감산하여 그 영향을 제외한 후 다음 사건을 구한다. 즉, j번째 사건함수 $event_j(t)$의 좌우 영향도를 나타내는 파라미터 dl_j, dr_j를 추정할 때는 $phrase(t)-\sum_{i=0}^{j-1} event_i(t)$와의 평균제곱오류(mean square error)가 최소가 되도록 추정한다. 이 때, 시간 t의 범위는 현재 사건의 중심위치 c_j의 좌우에 나타나는 국부 최소값 사이로 한정한다. 이는 현재 사건의 중심을 기준으로 나타나는 피크(peak)를 가장 잘 표현할 수 있도록 사건의 영향도를 추정하기 위함이다. dl_j, dr_j를 추정하기 위한 대상이 되는 운율구 궤적함수 $y_j(t)$는 식 3.10과 같다.

$$y_j(t) = phrase(t) - \sum_{i=0}^{j-1} event_i(t) \approx event_j(t) \quad \text{(식 3.10)}$$

계산상의 편이를 위해 식 3.10에서 사건함수 $event_j(t)$를 0과 1사이의 값으로 평활하여 로그값을 취한 후의 평균제곱오류(MSE)는 식 3.11과 같이 표현된다.

$$MSE = \sum_{t=lmin_j}^{c_j} (y_j(t)-(-0.5(\frac{t-c_j}{dl_j})^2))^2 + \sum_{t=c_j+1}^{rmin_j} (y_j(t)-(-0.5(\frac{t-c_j}{dr_j})^2))^2 \quad \text{(식 3.11)}$$

식 3.11에서 $lmin_j$는 c_j 왼쪽에 나타나는 국부 최소값의 위치를, $rmin_j$는 c_j 오른쪽에 나타나는 국부 최소값의 위치를 나타낸다. 평균제곱오류가 최소가 되는 dl_j, dr_j를 추정하기 위해서는 식 3.11을 dl_j, dr_j에 대해 각각 편미분하여 그 값이 0이 되는 dl_j, dr_j를 구한다. 식 3.12는 평균제곱오류가 최소가 되는 dl_j, dr_j의 추정치를 보인다.

$$dl_j = \sqrt{-\frac{\sum_{i=lmin_j}^{c_j-1}(t-c_j)^4}{2\sum_{i=lmin_j}^{c_j-1}(t-c_j)^2\log(\frac{y_j(t)-a}{b_j})}} \quad \text{(식 3.12)}$$

$$dr_j = \sqrt{-\frac{\sum_{j=c_j+1}^{rmin_j}(t-c_j)^4}{2\sum_{j=c_j+1}^{rmin_j}(t-c_j)^2\log(\frac{y_j(t)-a}{b_j})}}$$

식 3.12에서 a는 운율구 F0 궤적의 기저값이며, b_j는 사건의 크기를 나타내는 값으로 dl_j, dr_j를 추정하기 위해 $y_j(c_j)-a$로 초기화하여 사용한다.

사건의 크기 b 추정

앞에서 구한 파라미터들로부터 사건의 크기 b를 추정하는 방법 역시 평균제곱오류가 최소가 되도록 구한다. 운율구 F0궤적을 0과 1사이 값으로 평활화된 사건 $event_j(t)$들로 표현하면 식 3.13과 같다.

$$phrase(t) = a + \sum_{j=0}^{I} b_j event_j(t)$$

단,

$$event_j(t) = \begin{cases} \exp(-0.5(\frac{t-c_j}{dl_j})^2), & t<c_j \\ 1, & t=c_j \\ \exp(-0.5(\frac{t-c_j}{dr_j})^2), & t>c_j \end{cases} \quad \text{(식 3.13)}$$

식 3.13을 이용하여 평균제곱오류를 구하면 식 3.14와 같다.

$$MSE = \sum_{t=0}^{T}(phrase(t)-(a+\sum_{j=0}^{I} b_j event_j(t)))^2 \quad \text{(식 3.14)}$$

식 3.14에서 T는 운율구를 이루는 F0 궤적의 샘플 개수를 나타내며, I는 운율구를 이루는 사건들의 개수를 나타낸다. 평균제곱오류가 최소가 되는 b_j를 추정하기 위해서는 식 3.14를 b_j에 대해 편미분하여 그 값이 0이 되는 b_j를

구한다. 식 3.15는 계산상의 편이를 위해 행렬연산에 의해 계산된 평균제곱오류가 최소가 되는 b_j의 추정치를 보인다.

$$(b_1, b_2, \ldots, b_J)^T = H^{-1}(F_1^T, F_2^T, \ldots, F_J^T)^T D$$

단,

$$F_j = (event_j(1), event_j(2), \ldots, event_j(T))$$

$$H = (h_{ij}), \quad h_{ij} = F_i^T F_j^T, \quad i,j = 1,2, \ldots, J$$

$$D = (phrase(1) - a, phrase(2) - a, \ldots, phrase(T) - a)^T$$

(식 3.15)

F0 궤적의 시간적 분해 모델의 파라미터 추정 알고리즘

이상에서 기술한 방법을 이용한 F0궤적의 시간적 분해 모델의 파라미터 추정 알고리즘은 다음과 같다.

단계 1 : 주어진 F0 궤적을 운율구 단위로 분리한다.

단계 2 : 현재 파라미터를 추정하고 있는 운율구의 기저값 a를 식 3.9와 같이 추정한다 (ε = 1).

단계 3 : 주어진 운율구의 초기 사건 개수는 운율구에 나타나는 국부 최대값이 나타나는 위치 개수로 초기화한다.

단계 4 : 현재의 사건 발생 중심위치에서 식 3.12를 이용하여 구하고자 하는 사건의 좌우 영향도 dl, dr을 추정한다.

단계 5 : 단계 4에서 구해진 사건을 운율구 F0궤적에서 감산하여 사건의 영향을 제외시킨다.

단계 6 : 단계 5에 의해 새로 변형된 운율구 F0궤적에서 다음 국부 최대값이 나타나는 위치를 선정하여 다음 사건의 중심위치 c를 추정한다.

단계 7 : 운율구를 이루는 사건들의 파라미터들이 모두 추정될 때까지 단계 4부터 단계6까지의 과정을 반복 수행한다.

단계 8 : 주어진 운율구내에서 구해진 사건들의 크기를 식 3.15를 이용하여 추정한다.

단계 9 : 모든 운율구의 사건들이 모두 추정될때까지 단계 2부터 단계 8까지의 과정을 반복 수행한다.

IV. 실험 및 결과

본 연구에서 제안한 F0 궤적의 시간적 분해 모델의 성능 평가를 위해 사용한 코퍼스는 초등학교 교과서, 논문 요약, 소설, 영화 해설 등에서 발췌한 것으로 단문, 복문, 평서문, 의문문, 감탄문 등 다양한 형태의 문장들로 이루어진 500문장이다. 이를 전문 여성 아나운서가 무향실에서 발성하였고, 이를 녹음하여 약 40분 분량의 음성코퍼스를 구축하여 실험에 이용하였다. 음성코퍼스의 F0 궤적은 SHS(Sub-Harmonic Summation) 방법[13]을 이용하여 5ms 단위로 추출하였으며, 이를 보정하여 정확한 F0 궤적을 구했다. 이렇게 구해진 F0 궤적에서 억양변화의 인지에 영향을 미치지 못하는 F0 궤적의 미세변화(micro-segmental effects)의 영향을 제외시키기 위해 15 포인트 중간값 필터(median filter)로 평탄화하였다. 그리고 F0 궤적의 무성음 구간을 선형 보간(linear interpolation)하고, 이를 다시 7 포인트 중간값 필터를 이용해 보간에 의한 무성음 구간 경계에서의 급격한 변화를 완화하여 최종적으로 실험에 이용할 F0 궤적 코퍼스를 구축하였다. 표 4.1은 시간적 분해 모델에 의해 재합성된 F0 궤적 실험 결과를 보인다.

표 4.1. 시간적 분해 모델의 실험 결과
Table 4.1. Experimental results of temporal decomposition model.

내 용	실험 결과
평균 제곱 오류근 (Hz)	7.87
상관 계수	0.96
문장당 평균 운율구 개수	3.53
운율구당 평균 사건 개수	3.18

표 4.1에서 알 수 있듯이 실험 결과, 시간적 분해 모델에 의해 재합성된 F0 궤적과 원음성의 F0궤적의 평균 제곱 오류근 7.87Hz를 얻을 수 있었다. 그리고 비교 대상이 되는 두 궤적간의 전체적인 상관관계를 나타내는 상관계수도 0.96으로 높은 수치를 나타냈다. 또한 실험 자료에서 시간적 분해 모델이 한 문장당 3.53개의 운율구를 운율구당 3.18개의 사건들로 분해하였다. 따라서 본 논문에서 제안한 F0 궤적의 시간적 분해 모델이 F0 궤적을 비교적 정확하고 효과적으로 표현할 수 있음을 알 수 있다. 그림 4.1과 4.2는 시간적 분해 방법으로 재합성된 F0 궤적과 원 음성의 F0 궤적을 비교하여 보인다.

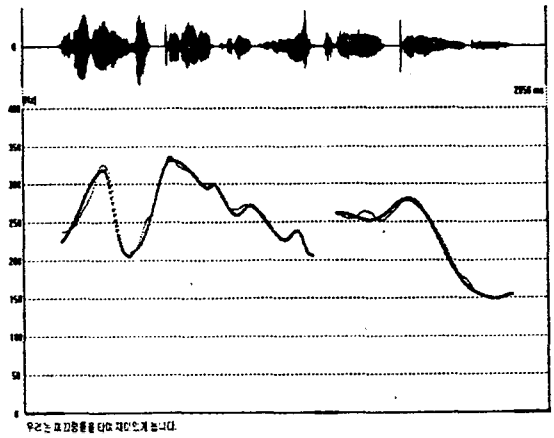

그림 4.1. 시간적 분해 모델의 F0 궤적 생성 실험 결과 - 평서문
(+ : 원음성의 F0 궤적, □ : 합성음의 F0 궤적)
Fig. 4.1. Resynthesized F0 contour by temporal decomposition model - declarative.
(+ : original F0 contour, □ : Resynthesized F0 contour)

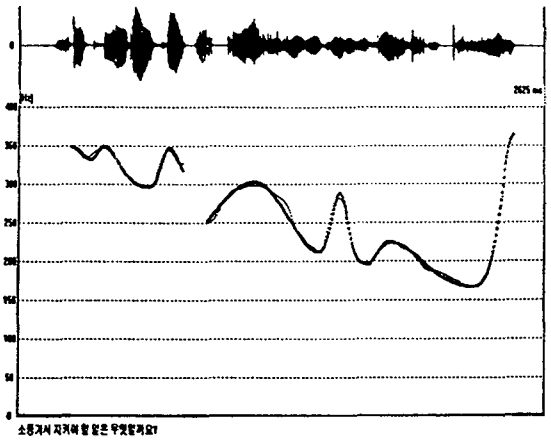

그림 4.2. 시간적 분해 모델의 F0 궤적 생성 실험 결과 - 의문문
(+ : 원음성의 F0 궤적, □ : 합성음의 F0 궤적)

Fig. 4.2. Resynthesized F0 contour by temporal decomposition model - interrogative.
(+ : original F0 contour, □ : Resynthesized F0 contour)

V. 결 론

본 논문에서는 발성된 문장의 F0 궤적을 가능한 적은 수의 파라미터로 표현하여 효과적으로 문장의 억양을 분석 또는 합성하기 위하여 시간적 분해에 기반한 F0 궤적 모델을 제안하였다. 제안한 모델은 F0 궤적을 국부 최대값을 중심으로 발생하며 서로 중첩가산되는 사건들의 열로 표현하고, 각 사건들을 가우시안 종 모양의 사건함수로 근사하였다. 또한 모델의 파라미터를 자동으로 추출할 수 있는 알고리즘을 제시하였다. 제안한 모델은 특정한 음운론적 지식에 기반하지 않았으며, F0 궤적 표현에 있어서 임계값 설정이 필요 없다. 따라서 임계값 설정에 따른 문제점을 해결하였으며, 모델 파라미터 추출이 용이하다.

실험 결과, 제안한 시간적 분해모델에 의해 재합성된 F0 궤적과 원 음성의 F0 궤적 사이의 평균 제곱 오류근이 7.87Hz였으며, 비교 대상이 되는 두 F0 궤적 사이의 상관계수가 0.96로 비교적 우수한 성능을 보임을 알 수 있었다. 앞으로는 제안한 F0 궤적의 시간적 분해 모델의 결과를 이용하여, 주어진 문장으로부터 모델 파라미터를 예측하여 문장의 억양을 제어하는 연구를 계획중이다.

참 고 문 헌

1. 김연준, 구문분석에 의한 운율조절을 이용한 한국어 문서-음성 변환 시스템의 구현, 석사학위 논문, 한국과학기술원 전산학과, 1993
2. P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, C. Fong, "The use of prosody in syntactic disambiguation," JASA Vol.90, No.6, pp01.2956-2970, 1991
3. Thierry Dutoit, *An introduction to Text-to-Speech synthesis* (Kluwer Academic Publishers, Dordrecht, 1997), Chap. 6, pp.133-145
4. K. N. Ross, *Modeling of intonation for speech synthesis*, PhD thesis, Boston University, 1995
5. A. W. Black, A. J. Hunt, "Generation F0 contours form ToBI labels using linear regression," Proceedings ICSLP'96, pp.1385-1388, 1996
6. C. d'Alessandro, P. Mertens, "Automatic pitch contour stylization using a model of tonal perception," Computer Speech and Language, Vol.9, pp.257-288, 1995
7. H. Fujisaki, H. Kawai, "Realization of linguistic information in the voice fundamental frequency contour," Proceedings ICASSP'88, pp.663-666, 1988
8. P Taylor, "The rise/fall/connection model of intonation," Speech Communication, Vol.15, pp.169-186, 1994
9. T. Portele, B. Heuft, "The maximum-based description of F0 contours and its application to English," Proceedings ICSLP'98, pp.663-666, 1998
10. G. Mohler, A. Conkie, "Parametric modeling of intonation using vector quantization," Proceedings of the 3rd ESCA/-COCOSDA International Workshop on Speech Synthesis, pp.311-316, 1998
11. H. Fujisaki, M. Ljugqvist, H. Murata, "Analysis and modeling of word accent and sentence intonation in Swedish," Proceedings ICASSP'93, pp.211-214, 1993
12. H. J. Byeon, Y. J. Kim, Y. H. Oh, "Generation of F0 contour using stochastic mapping and vector quantization control parameters," Proceedings ICASSP'97, pp.939-942, 1977
13. D. J. Hermes, "Measurement of pitch by subharmonic summation," JASA Vol.83 No.1, pp.257-264, 1988

▲변 효 진(Heo-Jin Byeon)

1994년 2월 : 한국과학기술원 전산학과 (학사)

1996년 8월 : 한국과학기술원 전산학과 (석사)

1996년 9월 ~ 현재 : 한국과학기술원 전산학과 박사 과정 재학중

※주관심분야 : 음성합성, 운율제어, 패턴 인식

▲김 연 준(Yeon-Jun Kim)

1991년 2월 : 한국과학기술원 전산학과 (학사)
1993년 8월 : 한국과학기술원 전산학과 (석사)
1994년 3월 ~ 현재 : 한국과학기술원 전산학과 박사 과정 재학중
※주관심분야 : 음성합성, 운율제어, 패턴 인식

▲오 영 환(Yung-Hwan Oh)

1972년 : 서울대학교 공과대학 (학사)
1974년 : 서울대학교 교육대학원 (석사)
1980년 : Tokyo Institute of Technology 정보공학전공 (박사)
1983년 ~ 1985년 : 충북대학교 컴퓨터공학과 조교수
1983년 ~ 1984년 : University of California, Davis 연구교수
1995년 ~ 1996년 : Carnegie-Mellon University 연구교수
1985년 ~ 현재 : 한국과학기술원 전산학과 교수
※주관심분야 : 음성인식, 음성합성, 음성코딩, 화자인식, 대화관리, 신경회로망, 전문가 시스템