論文 90-27-1-19

음절 단위를 이용한 한국어 음성 합성

(The Korean Text-to-Speech Using Syllable Units)

金 柄 秀* 尹 起 善** 朴 成 漢***

(Byeong Soo Kim, Gi Sun Yun, and Sung Han Park)

要 約

본 논문에서는 한국어 음성 합성시 적은 기억 용량을 차지하면서 음질을 향상시키기 위한 한국어 규칙 합성 방법을 제안한다.

이를 위하여 합성 단위를 음절(syllable)로 하며, 음절 음성 신호를 모델화 하기 위하여 12차 선형 예측(linear prediction)법을 사용한다. 또한 합성시 자연성의 향상을 위하여 음절 사이의 관계에 의한 음절연결 규칙을 개발하며, 음운 변동 규칙과 음율 변동 규칙을 적용한다.

Abstract

In this paper, a rule-based method for improving the intelligibility of synthetic speech is proposed. A 12-pole linear prediction coding method is used to model syllable speech signals. A syllable concatenation rule for pause and frame rejection between syllables is developed to improve the naturalness of the synthetic speech. In addition, phonological structure transform rule and prosody rule are applied to the synthetic speech by LPC. The illustrative results demonstrate that the synthetic speech obtained by applying these rules has better naturalness than the synthetic speech by LPC.

Ⅰ. 서 론

사회의 발달과 더불어 음성은 정보 전달의 수단으로서 인간이 가장 널리 이용하는 것의 하나가 되고 있다. 특히 정보화 시대에 있어서 기계와 사람간의 대화 수단으로서 그 중요성이 증대됨에 따라, 기존의 문자에 의한 정보 전달 방법을 대신하여 음성을 직접적으로 이용하기 위하여 많은 분야에서 과학적이며 광범위한 연구가 진행되고 있다.

음성 합성은 반도체 기술의 발달과 더불어 기억용량에 의한 제약이 점차 적어짐에 따라 음질의 향상을 중대 시키려는 방향으로 활발한 연구가 진행되고 있으며, talking computer terminal, teaching machine, 음성에 의한 경고 시스템, 일기예보 등 그 응용분야도 광범위하다. 이렇게 음성을 기계와 인간의 정보 전달의 수단으로 사용하기 위해서는 우선 text를 음성으로 나타내는 text-to-speech의 변환이 선행되어

*正會員, 金星半導體 情報通信研究所
(Lucky Goldstar R & D Complex, C & C Lab.)
**正會員, 漢陽大學校 電子工學科
(Dept. of Elec. Eng., Hanyang Univ.)
***正會員, 漢陽大學校 電子計算學科
(Dept. of Comp. Sci. & Eng., Hanyang Univ.)

接受日字 : 1989年 7月 12日
(※ 본 연구는 아산사회 복지 사업재단의 연구비 지원에 의해 수행한 것임)

야 한다.[1-4] 즉 키보드로써 입력된 문장에 대하여 사람처럼 명확하면서도 자연스러운 음을 만드는 음성 합성이 필요한 것이다.[1-4] 그러나 text-to-speech에 있어서 가장 큰 문제는 음질로서, 이는 발음의 정확도, 명료도, 자연성에 의하여 영향을 받는다.

여기서 발음의 정확도는 주어진 단어에 대하여 올바른 음소들의 연결에 의해 결정되며 명료도는 vocal tract model이 단어들을 어떻게 발음하는가에 따라 좌우된다. 자연성은 주로 음율에 영향을 받으며, 음율은 문장이나 단어수준에서의 음조 변화, 엑센트 및 소리의 지연시간등에 의해 결정되며, 이러한 요소들중 명료도는 합성시 사용되는 합성 단위 즉 음소, diphone, 반음절, 음절등에 따라 인식도를 달리하고 있으며 음질에 가장 큰 영향을 미치는 자연성의 향상을 위해서는 많은 규칙이 필요하다. 이를 실현하는 방법으로 초기에는 문장 혹은 단어 단위로 녹음/재생하는 방법을 연구하였으나 문장, 단어 단위의 경우는 자연성과 명료도는 높지만 기억용량의 제약으로 인해 사용 가능한 어휘수가 적어서 제한된 용도로만 사용되고 있다. 그러므로 text-to-speech는 합성의 기본 단위를 최소화 하되, 자연성과 명료도를 향상시킬 수 있는 규칙 합성에 대한 개발의 필요성이 강조되고 있다.[5-8]

또한 음성 합성은 크게 포르만트 합성, 조음(articulatory) 합성, 분석 합성등으로 나눌 수 있다. 포르만트 합성은 성도의 전달함수를 포르만트 공진기(resonator)로 모델화한 것이며, 조음 합성은 성도를 전송선로로 모델화하여 수식화 시킨 것으로, 위의 방법들은 적절한 합성 파라미터를 구하기 위하여 많은 분석과 교정을 필요로 한다.[9]

그러나 우리나라에서는 아직 국어에 대한 과학적 연구가 부족하여, 한국어를 유창하고 자연스럽게 합성할 수 있는 규칙을 정립할 수 없는 상태이다. 따라서 본 논문은 한국어 text에 대하여 음성으로 합성하는 한국어 text-to-speech에 대한 연구로서, 명료한 음을 합성하며 명료성을 해치지 않는 범위내에서 사람의 음성과 비슷한 자연스러움을 가지는 규칙을 개발한다. 기본적인 합성방법으로 분석합성 방법인 선형예측(linear prediction) 분석 합성법을 이용하고, 단일 화자의 음성신호를 음절 단위로 분할(segmentation)하여 선형 예측 분석을 한 뒤 선형예측 계수를 구하고, 이 계수를 이용하여 음성을 합성한다. 또한 database로부터 합성된 문장 음성의 자연성을 높이기 위하여 음운학적 요소인 음운 변동규칙과 음성학적 요소인 음절 연결 규칙 및 음율규칙을 적용하여 한국어 규칙 합성을 수행한다.

Ⅱ. 음성 합성 시스템

규칙 합성은 녹음된 단어 또는 문장으로부터 필요한 부분을 편집하여 합성하는 문장 편집 합성과 달리 책이나 키보드로써 입력된 메세지에 대하여 음성을 합성하는 것으로 종류로는 음소 합성, diphone 합성, 반음절 합성, 음절 합성으로 나눌 수 있다. 이를 위하여는 크게 4가지 과정이 필요한데 첫째, 입력된 문장을 합성 처리하기 위하여 적당한 형태로 바꾸어 주며 둘째, 문장(sentence)이나 구(phrase)에 대하여 단어경계(word boundary)를 찾고 문장을 분할(segmentation) 한후 셋째, 단어의 강세(stress)나 문장의 억양(intonation)을 조정한다. 마지막으로 저장된 data와 rule에 의하여 음성 합성을 위한 제어 신호를 발생시킨다.

이러한 일반적인 규칙 합성에 의한 text-to-speech 변환 시스템의 일반적인 흐름도는 그림1과 같이 나타낼 수 있다.

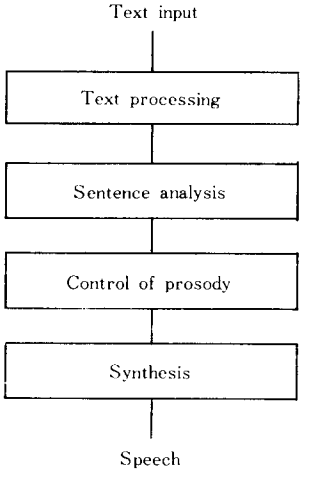


그림 1. 일반적인 text-to-speech에 대한 흐름도.
Fig. 1. The flowchart of general text-to-speech.

그러나 합성단위를 음절로하여 문장을 합성하는 경우, text를 소리나는 대로 변환시키는 음운 변동 시스템과 음절단위 합성시 문제가 되는 자연스러움을 향상시키기 위하여 음절 연결규칙, 음율 처리규칙을 포함시키는 것이 바람직하다. 그림2는 본 연구에서 제시된 음절단위에 의한 한국어 음성 합성 처리 시스템을 나타낸다.

입력단에서는, 한국어 문장을 입력받아 쉼표, 마침

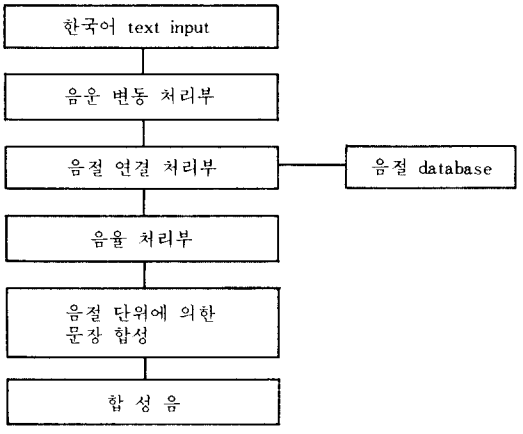


그림 2. 음성합성 처리 흐름도
Fig. 2. The flowchart of the speech synthesis processing.

표 등을 포함한 무제한 한국어 문자로 구성된 text를 음운 변동 처리부에서 필요한, 적절한 symbol로 바꾸어 준다. 음운 변동 처리부에서는 음성합성 system에서 음운학적인 면과 음성학적인 면과의 차이를 해결한다. 즉 합성음의 명료성을 위하여 한국어의 음성학적 발음 특성을 토대로 발음규칙을 설정하여, 입력된 한글 text를 정확한 발음 표기로 바꾸어 주며, 평문, 의문문, 느낌문, 감탄문에 대한 문의 종류와 문장내에서 단어의 수등 음율처리 필요한 정보를 추출한다. 음절 연결 처리부에서는 음절 database 내에서 해당 음절을 가져와 음절 연결 규칙에 따라 연결하고, 음율 처리부에서는 음운 변동 처리부에서 추출한 음율 처리 정보를 이용 합성시 자연스러움을 향상시켜주는 규칙을 적용해서 음절 단위에 의한 문장 합성을 수행한다.

Ⅲ. Database 작성

음성 합성을 위한 전단계로서 합성시 음질의 향상을 위하여 정확한 분석이 필요하다.

본 연구에서는 파워 부호화 방식인 LPC 분석 방법을 이용하며, pitch period를 구하기 위하여 분석단위 프레임의 길이를 10 KHz sampling시 일반적으로 사용되고 있는 25.6ms(256 samples)로 하여 12.8ms (128 samples) 간격으로 분석한다. 그러나 무성음 영역에서도 유성음 영역과 같이 한 프레임을 25.6ms로 하였을 경우 무성음 영역에서도 스펙트럼 포락선 (spectrum envelope)의 불규칙적인 변화에 대응하기가 어려우므로 무성음 영역의 프레임 단위를 줄여 이것의 문제를 해소시킨다. 따라서 무성음 영역에서의 한 프레임은 12.8ms로 하고 6.4ms 간격으로 분석을 하며, 유성음 영역에서는 한 프레임을 25.6ms로 하고 12.8ms 간격으로 분석을 한다.

1. 시스템의 구성

그림3은 음성 database를 작성하기 위한 전체 시스템의 개략도이다. 마이크를 이용하여 입력된 음성 신호는 증폭기를 통하여 증폭되고, 증폭된 신호는 4kHz 저역 여파기를 거쳐서 표본화 주파수 10kHz로 12비트 A/D 변환한 뒤, 변환된 결과를 모니터상에 graphic display하여 음절단위로 분리하여 각각의 파일로 음성 데이타를 hard disk에 저장한다. 또한 음절 데이타의 수집은 무반향실이 아닌 실험실에서 수행하되 소음을 최대한 제거하고 진행하였다.

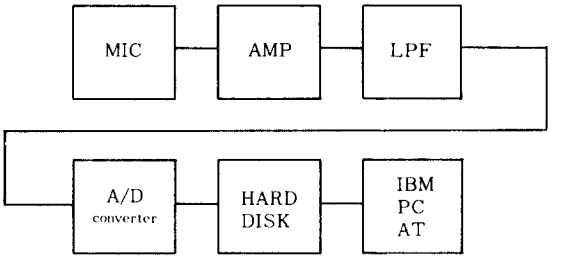


그림 3. 음성 database 작성을 위한 block diagram
Fig. 3. The block diagram to obtain speech database.

2. Database의 구성

음절 단위를 기본으로 하는 합성이므로, 단음절 단위로 발음한 음성에 대하여 분석한다. 그러나 많은 경우에 있어서 음절 단위로 발음한 경우 음성의 시작점에서 터짐 소리가 생기므로 정확한 분할이 어려울 뿐만 아니라 분석 합성시 정확한 합성이 되지 않으므로, 이것을 해소하기 위하여 음소간의 상호작용이 덜한, 두음절로 구성된 의미 없는 단어를 이용하여 그 중 두번째 음절을 선택하는 방법을 취한다. 예로 '양'이라는 데이터를 얻기 위해서 "한양" 이라는 의미있는 단어보다는 "우양"이라는 의미없는 단어를 선택하여 발음한후, '양'을 분리 데이터 베이스를 만든다. 그리고 음절 단위에서 정확한 음을 선택하기 위하여 음절을 분리할 때 음절의 시작점에서 약간의 휴지를 포함하도록 한다.

또한 시스템에서 검출한 피치와 목측에 의한 결과

와 비교하여 유성음과 무성음 영역의 구분에 오차가 있는 경우는 유성음과 무성음을 재구분하여 다시 선형 예측 계수와 피치값을 구한다.

그림4는 database 작성을 위한 흐름도이다.

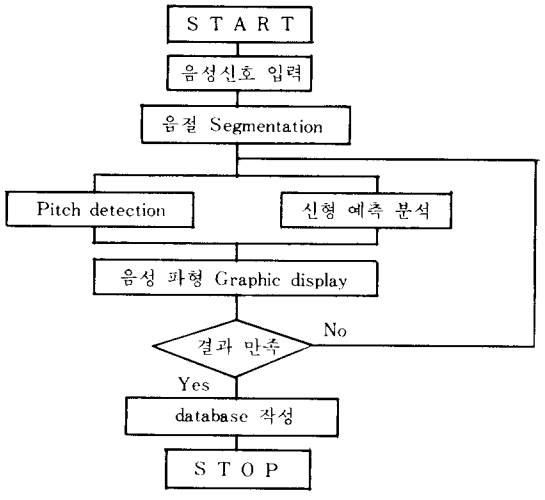


그림 4. Database 작성을 위한 흐름도
Fig. 4. The flowchart for the making of database.

Ⅳ. 음운 변동 처리기

한국어에 있어서 text의 문장과 text를 발음한 경우의 문장과는 음소들이 차이를 나타내고 있다. 즉 음절이 모여 낱말로써 발음될때 음절과 음절이 잇달아 소리나거나 단어와 단어가 잇달아 소리나면 서로 영향을 주고 받아서 여러가지 다른 소리로 발음된다. 이러한 현상을 고려하여 text로 표기된 것을 발음나는 대로 바꾸어 주는 한국어 음운 변동 처리 시스템이 필요하다.

본 연구에서 이용된 합성을 위한 한국어 음운 변동처리 시스템은 크게 입력부, 변동 규칙 처리부, 출력부 3가지로 나눌 수 있다. 첫째로 입력부에서는 키보드로써 입력된 문장의 음절을 2바이트 코드에서 초성, 중성, 종성의 코드를 분리한다. 둘째로 변동 규칙 처리부 에서는 분리된 코드에 변동 규칙을 적용하여 초성, 중성, 종성을 조음 현상에 맞게 변환하다. 세째로 출력에서는 변환된 초성, 중성, 종성을 음절 정합을 하기 위하여 2바이트 한글 코드로 변환한다.

한국어에서 음운 변동을 적용할 때 음운 변동 규칙의 순서에 맞게 처리하는 것이 필요하다. 그 예로서 "많이"는 "마니"로 발음되는데, 여기에는 두 규칙이 적용된다. 그런데 이 규칙을 반대로 적용하려고 하면 규칙이 적용되지 않는다. 따라서 먼저 "많이"에 ㅎ"탈락을 적용하여 "만이"를 이끌어 내고, 다음으로 연음 법칙을 적용하여 "마니"를 이끌어 내어야 순조롭게 풀려 나간다.[8] 음운 변동 현상들은 자음접변, 경음화, 구개음화, 격음화, 종성법칙, 연음법칙, 절음법칙, 음운첨가, 음운탈락 등의 9개로 나누어 처리하며, 이러한 규칙들에서 처리되지 않는 단어들은 예로 "삵괭이"같은 단어는 종성 법칙에 의하면 "삭괭이"이나 실제로는 "살괭이"로 된다. 따라서 이러한 예외 규칙들은 사전을 구성하여 처리하여 거의 대부분의 규칙을 처리하도록 하였다.

또한 음운 규칙들의 충돌을 방지하기 위한 음운변동 처리 순서에는 다음과 같은 규칙이 있다.

i) 연음법칙과 격음화 현상 다음에 구개음화 현상이 발생한다.
ii) 연음 법칙보다 먼저 자음탈락(ㅎ 탈락) 현상을 먼저 적용한다.
iii) 자음 접변은 대표음 처리 다음에 일어난다.

Ⅴ. 음절 연결 규칙

합성 단위가 음절이고, 음절은 초성, 중성, 종성으로 구성되며 이들 음소간의 상호 작용에 의하여 음향적인 특성이 변형되어진다. 또한 음절은 음소들의 결합으로 구성된 가장 기본적인 음소론적 단위이므로 음절사이에서 결합의 방법과 결합의 제약이 나타나게 된다. 이러한 음절간의 특징을 음향적인 면에서 관찰하여 음절 단위의 합성시 음절간의 자연성을 향상시키는 것을 목적으로 한다.

1. 음절의 구성

한국어 음절의 됨됨이를 공식으로 보면 다음과 같다.[7]

$$\text{음절} = C^{I} + V + C^{F}$$

여기서 C^{I} : 초성 자음
V : 중성 모음
C^{F} : 종성 자음

한국어 음절 연결을 공식화하면 다음과 같다.[8]

$$C^{11}\ V^{1}\ C^{f1} * C^{12}\ V^{2}\ C^{f2}$$

여기서 가능한 음절의 연결은

i) $C^{11}\ V^{1} * C^{12}\ C^{f2}$: 두음절의 종성이 모두 없는 경우
ii) $C^{11}\ V^{1} * C^{12}\ V^{2}\ C^{f2}$: 첫음절의 종성이 없는 경우

iii) $C^{11}V^{1}C^{\wedge}*V^{2}C^{\wedge 2}$: 두번째 음절의 종성이 없는 경우

iv) $C^{11}V^{1}C^{\wedge}*C^{12}V^{2}C^{\wedge 2}$: 두음절의 초, 중, 종성이 모두 있는 경우

의 4가지 경우이다.

2. 음절 연결 규칙의 구현

본 연구에서는 우선 남자 아나운서가 발음한 문장과 단어 데이터에 대하여 10kHz로 표본화 하고 A/D 변환시켜 그래픽 터미널을 이용하여 음성파형을 관찰, 분석하여 음절 연결 규칙을 만든다.

1) 음절 연결 규칙을 위한 발음표기

다음은 음절 연결 규칙을 만들기 위해 분류한 한국어 자음의 발음기호 분류이다.[8]

1. /p, t, k/　　6. /c'/
2. /s, h/　　7. /p^h, t^h, k^h/
3. /c/　　8. /c^h/
4. /p', t', k'/　　9. /m, n, l,/
5. /s'/　　10. Empty

2) 음절간의 휴지(pause)

다음의 결과는 단어만을 발음하여 두 음절사이의 pause을 관찰하여 얻어진 결과이다.

A/B → pause 길이

여기서 A : $C^{\wedge}$의 발음 표기 번호

B : C^{12}의 발음 표기 번호

a. 10/1, 3, /s/ → 10-30ms
b. 10/4, 5, 6 → 40-80ms
c. 10/7, 8 → 30-60ms
d. 9/4 → 70-90ms
e. 9/5, 6 → 10-20ms
f. 9/7, 8 → 20-50ms
g. 1/1, 2, 3 → 70-90ms
h. 1/4 → 100-130ms
i. 1/5 → 80-120ms
j. 1/6 → 70-100ms
k. 1/7 → 10-30ms
l. 1/8 → 40-60ms

그러나 문장을 발음한 경우 단어에 대하여 발음한 경우보다 두 음절사이의 pause가 상대적으로 짧으며, 또한 실제의 음성 합성시 음절사이가 너무 떨어져서 발음이 나는 경향이 있으므로 이러한 현상을 줄이기 위하여 다음과 같이 최대 pause를 조정한다. 또한 단어에 대한 음절수에 따라 음절간의 pause의 길이가 달라지므로 본 연구에서는 음절수가 길수록 상대적으로 음절간의 pause 길이를 줄이는 방향으로 한다.

(최대) (최소, 5음절 이상)

i) a, e, k. : 10ms - 4 ms
ii) c, f. : 15ms - 7 ms
iii) b, l. : 25ms - 10ms
iv) d, j, g. : 40ms - 15ms
v) h, i. : 50ms - 20ms

예로 두음절 이하로 구성된 단어들에 대해서, i)의 경우 첫번째 음절의 종성이 없고 두번째 음절의 발음이/p, t, k, c/인 경우, 첫음절의 발음이 유성비음이고 두번째 발음이 /s' c'/인 경우는 최대 10ms로 두음절간의 pause를 조정한다. 이러한 음절 연결규칙을 사용하여 음성 합성을 한 결과 음절이 끊어지는 부자연스러운 합성음이 출력되어, 이러한 문제를 해소하기 위하여 reference 데이타를 분석하여 자연스러움이 더해지도록 크게 다음의 4가지 패턴의 연결로 규칙을 만든다.

규칙 1) $C^{11}V^{1}*V^{2}C^{\wedge 2}$인 경우

V^{1}의 마지막 5프레임과 V^{2}의 처음 3 프레임을 제거하고, 4-5프레임에 걸쳐 선형 보간후 연결한다.

규칙 2) $C^{11}V^{1}C^{\wedge}*V^{2}C^{\wedge 2}$인 경우

(a) $C^{\wedge}$이 유성자음인 경우

$C^{\wedge}$의 마지막 3프레임과 V^{2}의 처음 3프레임을 제거하고, 3-4프레임에 걸쳐 선형 보간후 연결한다.

(b) $C^{\wedge}$이 무성자음인 경우

$C^{\wedge}$의 마지막 1프레임과 V^{2}의 처음 3프레임을 제거한 후 연결한다.

규칙 3) $C^{11}V^{1}*C^{12}V^{2}C^{\wedge 2}$인 경우

(a) C^{12}이 유성자음인 경우

V^{1}의 마지막 4프레임과 C^{12}의 처음 2프레임을 제거하고, 3-4프레임의 선형 보간후 연결한다.

(b) C^{12}이 무성자음인 경우

V^{1}의 마지막 4프레임과 C^{12}의 처음 2프레임을 제거 후 연결한다.

규칙 4) $C^{11}V^{1}C^{\wedge}*C^{12}V^{2}C^{\wedge 2}$인 경우

(a) $C^{\wedge}$와 C^{12}가 유성자음인 경우

$C^{\wedge}$의 마지막 2프레임과 C^{12}의 처음 2프레임을 제거하고, 2프레임의 선형 보간후 연결한다.

(b) $C^{\wedge}$이 유성자음이고 C^{12}이 무성자음인 경우 $C^{\wedge}$의 마지막 2프레임과 C^{12}의 처음 0프레임을 제거 후 연결한다.

(c) $C^{\wedge}$와 C^{12}가 무성자음인 경우

프레임 제거없이 연결한다.

(d) C^{n}이 무성자음이고 C^{12}이 유성자음인 경우 C^{n}의 마지막 1프레임과 C^{12}의 처음 1프레임을 제거 후 연결한다.

위의 4가지 규칙들의 연결시 합성에 중요한 요소로 작용하는 에너지의 연결을 자연스럽게 하기 위하여 음절 연결 규칙인 pause의 삽입의 크기에 따라 pause 길이가 길수록 3프레임, pause 길이가 짧을수록 2프레임의 에너지 레벨을 감소시키며, 연결부위의 계수와 에너지, 피치를 smoothing 해서 이들 파라미터의 갑작스러운 변화를 줄인다.

Ⅵ. 음율 변동 규칙

음성의 음율은 기능면에서 듣는 사람으로 하여금 메세지에 대한 구조적 정보를 제공한다. 이러한 음율은 메세지의 템포, 리듬, 멜로디를 형성하는 몇가지 음향학적인 요소인 intensity, stress, pitch contour등에 의해 표시된다. 즉 강세가 있는 음절은 음율적 요소가 있다고 할 수 있다.[8-10]

1. 강세(stress)

한국어에서 강세는 pitch, duration, intensity 중에서 pitch, intensity의 순서로 영향을 받는다. 그러나 강세는 그 절대적인 위치에 대하여 많은 연구가 되어 있지 못한 실정이며, 또한 단어에 있어서 단어자체에 억양을 가지는 경우가 있으므로 강세의 정확한 모델링이 어렵다. 그러므로 본 연구에서는 한 음절로 구성된 단어에 대하여서는 적은 강세가 가해지며, 두 음절로 구성된 단어에 대해서는 두번째에 강한강세, 첫번째 음절에 약한 강세가 주어지며, 세 음절 이상으로 구성된 단어에 대해서는 두번째 음절에 강세가 가해지며 음절이 갈수록 점차 강세가 작아지는 방향으로 모델링 하며, 악센트의 구현은 pitch와 intensity로 구현한다.

2. 억양(intonation)

억양의 통사적 기능은 문중에서 나타나는 경우인 문중 억양과 문말에서 나타나는 경우인 문말 억양으로 나눌 수 있으며, 문중 억양은 단어의 연결에서 실현되고, 문말 억양은 문의 결합 또는 문의 연결에서 실현된다.[7]

따라서 본 연구에서는 문의 종결양식인 문말 고저곡선(pitch contour)으로 문의 양식을 표시한다. 그림 5에서 보여주는 바와같이 문의 종결양식중 가장 일반적인 형태는 점차 하강하는 형태이며, 상승-하강, 평탄-하강 형태들도 볼 수 있다. 문장의 종류에 따라 평서문에서는 하강형태를, 의문문에서는 하강-상승 형태를, 명령문에서는 하강 형태로 나누어 처리하며, 기본 주파수 곡선(fundamental frequency countour)은 기본 주파수의 범위를 표시하는 포락선(envelop)내에서 taget들의 연결로서 표시된다. 즉 taget들 사이의 기본 주파수 곡선을 transition rule로서 표시할 수 있다. 이러한 FO contour를 나타내는 방법으로 본 연구에서는 target들 사이의 선형 보간에 의해서 문장 전체의 기본 주파수 형태를 구현하는 schematic algorithm을 사용하며, 알고리듬의 강세에 있는 음절의 Fo 값은 다음과 같은 공식에 의해 표시된다.[4]

Fo=baseline+accent taget(topline-baseline)

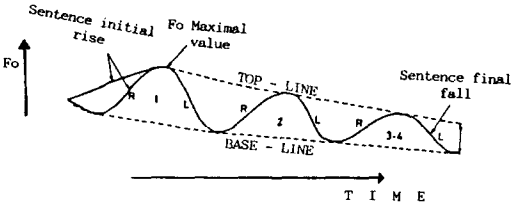


그림 5. FO의 전형적인 pattern
Fig. 5. Typical pattern of FO.

Ⅶ. 실험 및 결과

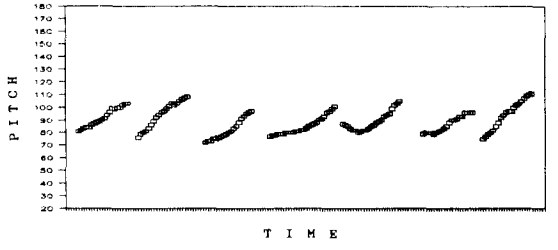
본 연구에서는 합성 단위로 음절을 사용하므로 단음절에 대하여 분석합성(anaysis-synthesis)을 한 경우, 단음절의 database을 이용하여 문장을 합성할때, 문장의 자연스러움을 향상시키기 위하여 음운 변동 규칙, 음절 연결 규칙, 음율 변동 규칙을 적용한다. 특히 음운 변동 규칙이 적용되지 않는 합성음에 대해서는 문장의 이해가 힘든 경우도 있다. 또한 음절 연결 규칙에서 음절사이의 pause만을 첨가한 경우 명료도에는 별 문제가 없으나, 합성음이 매우 부자연스럽기 때문에 음절 연결 규칙에서 연결되는 두음절의 음소의 종류에 따라 프레임의 제거규칙을 이용하면 명료도에는 해가 가지 않는 범위내에서 자연음에 가까운 진행속도와 유사한 속도를 나타내어 음질을 향상시킬 수 있다. 음율변동 규칙을 적용하므로 문의 종류에 따른 억양의 변화에 pitch 패턴의 형태가 변화함으로 합성음이 자연스러워진다. 음율 변동 규칙을 적용할때 음율 변동 규칙에 의해 구해지는 pitch 값만으로 pitch 패턴을 모델링하여 합성한 경우 분석한 data의 원래 음질을 상실하여 명료한 음

성을 얻을 수 없으므로 규칙에 의해 구해진 pitch 값과 database에 저장된 pitch의 값을 50%로 공통적으로 이용하므로 합성음의 명료도를 향상 시킬 수 있다.

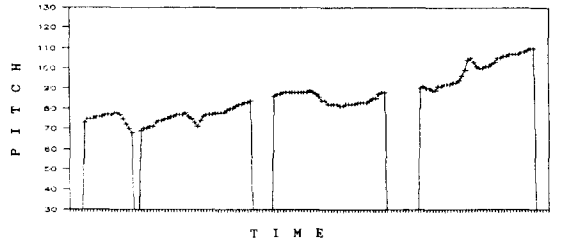
그림6-a는 pause규칙만 적용된 경우의 "바다에 눈이 온다"의 pitch pattern이고, 6-b.는 pause 규칙만의 단점을 보완하기 위해 본 논문에서 제안한 규칙이 모두 적용된 경우의 pitch pattern이다. 그러나 음절 단위를 이용한 한국어 음성합성에서 가장 큰 문제는 database의 기억용량 문제로 다른 규칙 합성보다 많은 기억용량을 필요로 하며, 또한 batabase의 작성시 한 음절의 시간이나 강약으로 인하여 분석합성시 합성음이 다른 음으로 들리는 경우가 있다.

합성음의 청취는 과학적인 인식율 test를 사용하여 음성 합성 시스템의 성능을 평가하여야 하나 본 연구에서는 아직 과학적인 인식율 test를 하지 못하였다. 20대의 7-8명의 성인을 대상으로 하여 듣기 평가를 수행한 결과를 종합해 보면, 'ㅅ', 'ㅈ' 같은 몇몇 변이음에 대해서는 예외가 있었으나 원음성과 비교시 거의 인식할 수 있었다. 또한 미리 합성음에대해 알려준 경우는 완벽하게 인식하였으나, 그렇치 않은 경우는 합성음에 대한 기대가 커서 완벽하게는 인식할 수가 없었다.

또한 LPC를 사용하여 무성음을 모델링 할 때, sampling 주파수가 10kHz이고, cutoff주파수가 4kHz이므로 고주파가 많은 무성음 부분이 정확히 모델링 되지 않아 무성음의 음질이 떨어진다. 그러므로 분석합성시 음질의 향상을 위하여 무성음의 정확한 모델링과 정확한 단음절의 녹음이 필요하며, 한국어에 대한 정확한 음율 규칙등의 연구가 있어야 할 것이다.



(a)



(b)

그림 6. "바다에 눈이 온다"의 pitch pattern
(a) pause 규칙만 적용된 경우
(b) 본 논문에서 제안한 규칙이 모두 적용된 경우

Fig. 6. Pitch pattern of "바다에 눈이 온다"
(a) the case of applience only pause rule.
(b) the case of appliense all rules proposed in this paper.

Ⅷ. 결　　론

본 연구에서는 명료도의 향상을 위하여 합성 단위를 음절로 하고, 문장을 소리나는 대로의 문장으로 바꾸어주는 음운 변동 규칙을 적용하며, 자연성의 향상을 위하여 음절과 음절 사이의 연결시 고려되는 모든 경우에 대한 음절 연결 규칙과 문장의 합성음을 문맥시 맞게 하여 자연스러움을 향상시키는 음율 변동 규칙을 적용 하므로서, 합성음이 끊어지는 듯이 발음나는 것을 자연스럽게 하였다.

그러나 합성음이 뛰어난 한국어 text-to-speech system을 만들기 위하여서는 우선 정확한 한국어 음절 음성 데이타의 수집과 각 음소의 특징에 관한 연구가 필요하며, 자연성의 향상을 위하여 한국어의 음율에 관한 과학적인 연구가 있어야 할 것이다.

參 考 文 獻

[1] F. Fallside and W.A. Woods, "Computer speech processing," Prentice Hall, 1985.
[2] G. Bristow "Electronic speech synthesis," McGraw Hill, 1984.
[3] I.H. Witten "Principles of computer speech," *Academic Press*, 1985.
[4] N. Umeda "Liguistic rule for Text-to-speech synthesis," *IEEE, Proceeding*, vol. 64, no. 4, April 1976.
[5] J.L. Flanagan "Voice of man and machine," JASA, May 1972.
[6] J. Allen "Linguistic-Based algorithm offer pratical Text-to-Speech systems," *Speech Technology*, vol. 1, no. 1, 1981.
[7] 허웅 "국어 음운학"정음사, 1982.
[8] 이철수 "한국어 음운학,"인하대학교 출판부, 1985.

[9] G. Akers and M. Lenning "Intonation in Text-to-speech Synthesis: evaluation of algorithm" JASA, vol. 77, Jun 1985.

[10] R.A. Cole "Perception and production of fluent speech," *Lawrence Erlbaum Associates,* 1980.

著 者 紹 介

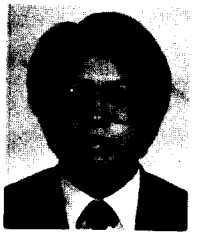
金 柄 秀(正會員)

1965年 10月 24日生. 1987年 2月 한양대학교 전자공학과 졸업. 학사학위 취득. 1989年 2月 한양대학교 대학원 전자공학과 졸업. 공학석사 학위 취득. 1989年~현재 금성 반도체 통신기기 연구소. 주관심분야는 신호처리, data 통신 등임.

●

朴 成 漢 (正會員) 第25卷 第12號 參照

현재 한양대학교 전자계산학과 부교수



尹 起 善(正會員)

1966年 3月 25日生. 1988年 2月 한양대학교 전자공학과 졸업. 학사 학위 취득. 1989年 12月~현재 한양대학교 대학원 전자공학과 석사과정 재학중. 주관심분야는 Speech signal processing등임.