텍스트의 의미 정보에 기반을 둔 음성컨트롤 태그에 관한 연구*

A Study of Speech Control Tags Based on Semantic Information of a Text

장 문 수** · 정 경 채*** · 강 선 미***
Moonsoo Chang · Kyeongchae Chung · Sunmee Kang

ABSTRACT

The speech synthesis technology is widely used and its application area is also being broadened to an automatic response service, a learning system for handicapped person, etc. However, the sound quality of the speech synthesizer has not yet reached to the satisfactory level of users. To make a synthesized speech, the existing synthesizer generates rhythms only by the interval information such as space and comma or by several punctuation marks such as a question mark and an exclamation mark so that it is not easy to generate natural rhythms of people even though it is based on mass speech database. To make up for the problem, there is a way to select rhythms after processing language from a higher level information. This paper proposes a method for generating tags for controling rhythms by analyzing the meaning of sentence with speech situation information. We use the Systemic Functional Grammar (SFG) [4] which analyzes the meaning of sentence with speech situation information considering the sentence prior to the given one, the situation of a conversation, the relationship among people in the conversation, etc. In this study, we generate Semantic Speech Control Tag (SSCT) by the result of SFG's meaning analysis and the voice wave analysis.

Keywords: speech synthesis, semantic analysis, semantic speech control tag(SSCT), systemic functional grammar

1. 서 론

일반적으로 합성기에서 텍스트 기반의 입력문장에 대해서는 형태소 분석[1] 결과를 사용한다. 그리고 음성 데이터베이스에서 형태소 분석 결과 단위로 음성데이터를 가져와 합성음을 만드는데 [2], 각각의 형태소 분석 결과 단위의 소리는 명확하더라도 연속되는 합성음은 운율이 단조롭기 때문에 부자연스럽고 어색한 느낌을 준다. 이를 해결하기 위해 입력문장에서 쉼표에 의한 휴지구간정

* 이 논문은 2005 년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2005-003-D00351).
** 서경대학교 소프트웨어학과
*** 서경대학교 컴퓨터과학과

보나 물음표 또는 느낌표와 같은 문장기호 성분을 추출, 운율 생성을 위한 정보로 사용하여 합성음을 생성한다. 그러나 이렇게 해서 생성된 합성음 역시 통계적 오류에서 발생하는 잘못된 운율정보 때문에, 여전히 부자연스럽고 어색하다. 그리고 휴지구간이나 문장기호 성분만으로 생성된 운율은 출력하고자 하는 합성음이 평서문인지, 의문문인지 또는 감탄문인지에 대한 정보만을 표현하므로 운율을 통해 발화자의 의도를 표현하는데 한계가 있다. 하지만 사람 사이의 대화에서는 같은 문장이라도 화자의 감정이나 주변 상황에 따라서 다르게 발음한다. 이는 발화문장의 운율이 문장 형태만으로 결정되는 것이 아니라, 이전 대화 내용이나 대화가 이루어지는 상황, 상대방과의 관계, 발화 의도 등에 의해서도 영향을 받는다는 것을 의미한다.

예를 들어, 숙제를 계속해서 하지 않는 학생에게 선생님이 꾸중을 하는 상황에서 숙제를 하지 않는 학생에게 “왜 숙제를 안했니?”라고 물어볼 때, 기존 합성기에서는 이 문장의 형태가 의문문이라는 것을 고려하여, 단순히 문장의 끝을 올리는 운율을 적용한다. 하지만 실제 대화에서는 학생을 꾸짖기 위해서나 타이르기 위해서 등 선생님의 발화 의도에 따라서 다른 운율을 문장에 적용하여 발화하게 된다. 이러한 문장의 운율은 문장의 의미를 이해함으로써 실현될 수 있는데, 이것은 자연언어처리의 의미분석[3] 과정을 통해서 처리가 가능하다.

일상생활에서 발화되는 음성은 주변 상황이나 화자의 감정, 대화 상대자와의 관계 등에 영향을 받기 때문에 형태소 분석이나 통사 분석과 같은 기초적인 자연어 처리로는 이러한 의미 정보를 얻을 수 없다. 의미분석이론에는 사회 현상과 상황의 컨텍스트를 언어체계의 일부로 다루고 있으면서 언어의 의미를 세 개의 메타기능의 조합으로 보는 시스테믹 기능문법(Systemic Functional Grammar: SFG)[4]이 있다. 본 논문에서는 이 시스테믹 기능문법에 기초하여 합성기를 통해 생성할 문장의 의미를 분석하고, 이 의미로부터 합성음의 운율을 변화시키는 의미기반 음성제어태그(Semantic Speech Control Tag: SSCT)를 추출하여 보다 자연스러운 합성음을 생성시키는 방법론을 제안한다.

2. 관련연구

2.1 합성기의 음성 태깅

기존 합성기의 경우 휴지구간, 음의 높낮이, 발화 속도 등을 나타내는 기본적인 태그를 사용하여 운율을 생성하고 있다. L&H(Lernout & Hauspie)코리아의 L&H PCMM Korean RealSpeak 합성기[5]를 예로 들어 합성기의 기본 특징과 태그의 종류를 살펴본다. L&H 코리아의 합성기에서 사용되는 기본태그의 형태는 <ESC><char><value>와 <ESC>\<char>=<value>\이다. 이 때, <char>는 태그의 종류를 나타내고, <value>는 해당 태그의 빠르기나 강도와 같은 수치 값을 나타낸다. 태그를 붙일 때에는 입력문장의 앞부분에 삽입하여야 하며, 삽입하는 태그문자열에는 공백이 존재할 수 없다. 태그문자열에서 <char>부분에 들어가는 문자는 대문자와 소문자 모두 사용할 수 있다. 그리고 특정 태그에 의해 설정된 값은 같은 태그의 값을 수정함으로써 변경시킬 수 있으며, <ESC>F 태그에 의해서 초기화 시킬 수 있다. <표 1>은 L&H PCMM Korean RealSpeak 합성기에서 사용하고 있는 기본 태그의 종류이다.

표 1. L&H PCMM Korean RealSpeak 기본 태그

태그 종류	태그 기능	기 타
<ESC>Vx	합성음의 출력 볼륨 값 결정	Range(x) : 0 - 9
<ESC>Rx	합성음의 출력 속도 값 결정	Range(x) : 1 - 9
<ESC>Wx	모든 단어 사이에 휴지구간 삽입	Range(x) : 0 - 9
<ESC>Px	태그 삽입 부분만 휴지구간 삽입	Range(x) : 1 - 9
<ESC>Bx	합성음을 발음하는 톤의 높낮이 결정	Range(x) : 0 - 9
<ESC>@x	단어의 문법상의 분류 표현	Range(x) : N, J, A, v, R, V
<ESC>Hx	900ms 이상의 휴지구간을 표현할 때 삽입	Range(x) : 1-10,000,000
<ESC>"	문장의 강세를 표현	단어의 앞에 위치
<ESC>F	다른 태그의 설정값을 기본값으로 복원	
<ESC>\Mrk=x\	입력문장에서 특정 위치를 가리킴	Range(x) :0 - 2147483648
<ESC>\pron=x\	출력 방식 결정	Range(x): addr: 연설 방식 std : 표준 방식

<표 1>에 설명한 기본 태그는 합성음 출력의 속도 조절, 출력 볼륨의 조절, 발음하는 톤의 높낮이 조절, 휴지구간의 삽입과 길이 설정, 강세의 표현 등 합성음을 출력하는데 있어 기본적인 역할을 하고 있다. 본 연구에서는 이러한 기본 태그들의 조합을 통하여 음성 제어 태그를 생성하고, 제공되지 않는 태그는 새롭게 정의하여 적용시킨다.

2.2 시스테믹 기능문법 이론

Chomsky의 변형 생성문법(Transformational generative grammar : 1965년)[1]은 문장의 구조를 심층적으로 분석하여 문의 구조를 규명하는 것을 주된 목적으로 한다. 의미 분석 단계에서는 통사부문에서 분석한 문의 구조를 바탕으로 기본적인 의미만을 분석하며, 문장이 실세계와 가지는 연관 관계를 분석하는 화용분석 단계를 거치지 않는다. 반면 Halliday에 의해 고안된 시스테믹 기능문법(Systemic Functional Grammar)은 절(clause)의 기능적(functional) 의미를 분석하는 것을 주된 목적으로 한다. 그래서 문의 기본적인 의미 분석 이외에 화용분석을 함께하여 문맥, 또는 상황에 주목하여 의미를 해석한다.

2.2.1 시스테믹 기능문법의 언어체계

<그림 1>은 시스테믹 기능문법의 언어체계[4]를 나타내고 있다. 언어체계는 음운층(phonology), 문법층(lexico-grammar), 의미층(semantics)이라는 세 개의 층으로 구성되고, 언어체계를 둘러싸는 컨텍스트(context)는 특정의 사회집단의 가치관이나 판단 기준인 문화의 컨텍스트와 사회활동의 장면에 상당하는 상황의 컨텍스트에 의해 구성된다. 상황의 컨텍스트는 실현되는 텍스트의 형태에 영향을 주는 세 개의 사회적 요소, 즉 필드(field: 언어사용영역), 테너(tenor: 화자와 청자 사이의 사회

적 역할), 모드(mode: 커뮤티케이션의 채널, 혹은 수단)로 구성된다. 그리고 상황의 컨텍스트를 구성하는 필드, 테너, 모드를 반영하여 어휘문법적 구성요소의 선택을 수행하는 것은 언어체계의 세 개의 메타기능(meta function)으로써 관념구성적(ideational) 기능, 대인관계적(interpersonal) 기능, 텍스트형성적(textual) 기능이 있다. 이와 같이 다층(multi-stratified) 구조에 의해 언어체계가 구성되어, 각 계층은 텍스트로 실현(realize)되는 잠재성(potential)을 가진 선택가지에 의해 구성되는 체계로서 표현된다. 이 체계를 표현하는 방법으로서 시스테믹 네트워크(systemic network)[6]라는 선택가지의 네트워크가 이용되고 있다.

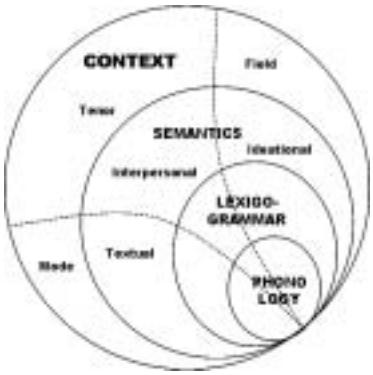

그림 1. 시스테믹 기능문법

2.2.2 시스테믹 기능문법의 메타기능

시스테믹 기능문법의 가장 큰 특징은 세 개의 메타기능의 조합으로 의미를 표현하고 있다[4]는 것이다. 각 메타기능을 기능문법의 언어체계에서 의미론 층의 관점에서 서술하면 다음과 같다.

1) 관념구성적(ideational) 기능

관념구성적 의미는 컨텍스트의 대화구조와 어휘문법적 특징을 연결하는 것으로 정의된다. 어휘문법적 특징은 타동성(transitivity)을 표현하는 언어자원으로 구성되는데, 프로세스(process), 참여자(participant), 상황어(circumstance)로 구성된다. 프로세서는 절(clause)의 기능을 나타내는 것으로 물질(material), 정신(mental), 관계(relational), 존재(existential), 행동(behavioral), 언동(verbal) 등으로 크게 6가지 프로세스가 있다. 참여자는 절의 기능을 수행하는데 참여하는 참여자로써 행위자(actor), 목표(goal), 감지자(sensor), 현상(phenomenon) 등이 있다. 상황어는 프로세스의 기능을 보조하는 역할로 주로 부사나 부사구에 해당하는 시간, 공간, 감정 등을 나타내는 문장요소들이다.

2) 대인관계적(interpersonal) 기능

대인관계적 의미는 기본적으로 발화 기능(speech function)을 이용한다. 발화에는 교환의 역할로써 “제공(giving)”과 “요구(demanding)”이 있다. 또, 교환의 대상으로써 “물건이나 서비스”와 “정보”가 있어, 그 조합으로 구성되는 4 개의 요소, 제안(offer), 명령(command), 진술(statement), 질문

(question)이 발화기능이 된다. 세부적으로는 진술의 경우 응답, 보증, 진술, 희망 등으로 나눌 수 있다.

3) 텍스트형성적(textual) 기능

텍스트형성적 의미는 주제(theme), 결속관계(cohesive relation), 정보의 신구(new/old information), 그리고, 초점(focus) 등에 의해 파악된다. 주제는 발화의 문두에 출현하여 발화의 의도를 나타내는 것으로, 대화의 흐름을 읽는데 중요한 역할을 담당하는 것이다. 주제에는 3개의 메타기능적 의미를 가진 주제로 분류된다. 우선, 일반적으로 주제라고 말하는 경험적 주제는 전하고 싶은 메세지를 나타내는 것이고, 대인관계적 주제는 상대방과의 관계를 나타내는 것이 문장 앞에 나올 때이다. 그리고 텍스트형성적 주제는 텍스트의 논리적 결속관계를 나타내는 것으로, 접속사 등이 여기에 속한다. 이들 주제는 하나의 어휘요소에 중복되어 나타날 수도 있다.

3. 연구내용

3.1 시스테믹 기능문법을 통한 의미분석

시스테믹 기능문법에 적용하여 문의 의미를 분석하면 상황 정보의 틀 속에서 문법 해석을 통한 메타기능의 의미로 결과를 도출하게 된다[7]. 다음은 하나의 문장을 예로 들어 시스테믹 기능문법적 의미해석 과정과 결과를 설명한다. 예제 문장은 특정한 상황 속에서 발화되는 대화형 문장이다. <표 2>는 예제 문장과 이 문장을 발화하는 시점의 상황 컨텍스트를 나타내고 있다.

표 2. 예제 문장과 상황의 컨텍스트표

예문		"뒤에 절벽 있어!"
사회적 요소	필드	바위산에서 등반 도중 바람이 많이 부는 언덕임. 위급상황에 대한 알림. 상대방은 위급상황을 인지하지 못함.
	테너	친한 친구 사이
	모드	대화체 음성 채널. 간결한 문장 사용.

1) 형태소 분석

예문의 형태소 분석은 (주)미디어젠의 형태소 분석기를 사용하였으며, 분석 결과의 신뢰도는 95% 이상이다. 형태소 분석은 어절 단위로 나누어 하는데, 한국어의 경우 어절의 단위와 띄어쓰기 단위가 거의 일치하므로 띄어쓰기 단위로 문을 나누어서 분석하였다. <표 3>은 예문에 대한 형태소 분석 결과이다. 시스테믹 기능문법의 어휘문법적 해석은 형태소 분석 결과를 토대로 수행한다.

표 3. 예문에 대한 형태소 분석 결과

어절단위	실험문장을 발화하는 상황
뒤에	(NX 뒤/nc) + (JCX 에/jca)
절벽	(NX 절벽/nc)
있어	(PX 있/pa) + (EFX 어/ef)
.(마침표)	(SYX ./s.)

2) 어휘문법적 해석

시스테믹 기능문법에서 어휘문법적 해석은 복잡한 언어학적 분석을 통해서 이루어진다. 본 논문에서는 합성음의 운율에 영향을 줄 수 있는 의미를 생성하는데 필요한 범위 내에서 분석을 시도한다. <표 4>는 예문에 대한 어휘문법적 해석을 세 개의 메타기능에 따라 분석한 결과를 나타낸다.

표 4. 예문의 어휘문법적 해석

예문		뒤에	절벽(이)	있어!
메타기능	관념구성적기능	상황어	참여자	프로세스
		position	존재자(existent)	existential
	대인관계적기능	Residue		Mood
				현재형, 평서문
	텍스트형성적 기능	Theme	Rheme	

3) 의미층

시스테믹 기능문법의 의미는 의미층의 메타기능으로 표현되며, 이것은 어휘문법적 해석과 상황의 컨텍스트로부터 정보를 받아서 분석된다. <표 5>는 의미층의 메타기능을 나타낸 것이다

표 5. 예문의 의미층의 메타기능

메타기능	의미분석 결과
관념구성적기능	위급상황의 위험에 대한 경고
대인관계적기능	진술(statement)
텍스트형성적 기능	Topical Theme: 뒤에

합성 발화를 생성할 경우에는 위의 해석 결과와는 반대로 특정 상황 속에서 발화해야할 의미가 도출되고, 이 의미로부터 어휘문법의 언어자원이 선택된다. 본 논문에서는 이러한 분석과정 속에서 나오는 의미 요소와 언어자원을 바탕으로 음성의 운율을 제어하는 태그를 생성한다.

3.2 육성발화와 합성발화의 비교

3.2.1 실험자료

본 실험의 목적은 문장부호 이외에는 어떤 언어정보도 없이 발화하는 합성기 발화와 여러 가지 상황이나 감정을 설정하여 발화하는 육성발화를 비교하여 그 차이점을 살펴보는 것이다. 이 비교실험을 통해 일상생활언어를 표현하기 위해서는 상황이나 감정이 포함된 언어정보가 필요함을 보이고자 한다. 실험에 사용하기 위한 실험 문장은 <표 6>과 같다. 실험문장목록을 만들 때에는 육성발화와 합성 발화를 비교해서 확연한 차이점이 나타날 수 있도록 하기 위해 화자의 감정이 많이 포함된 문장을 기준으로 선정하였다. 각 음성파일은 샘플링 주파수가 16 kHz이고 채널은 모노, 샘플당 비트크기는 16 bit로 하여, 실험문장을 각각 3회씩 발화하도록 하였다. 그리고 객관적이고 정확한 실험자료를 얻기 위해 피실험자에게는 발음에 대한 지시를 주지 않고, 설정된 상황에 대해서 충

분히 숙지하도록 하였으며, 자연스러운 발화가 나올 수 있도록 연습을 한 후 녹음을 진행하였다. 자연스러운 발화를 위해 핀마이크를 착용하여 녹음을 하였으며, 녹음과 분석을 위해서 Cool Edit (Version, Pro 2.1)와 Wavesurfer(Version, 1.8.5)를 사용하였다. 본 연구에서 파형을 나타내는 그림의 시간 스케일은 0.1 초 단위로 하였다.

3.2.2 피실험자 선정

본 실험에서 사용하는 합성기의 발화가 여성화자의 목소리로 이루어지기 때문에 20 대 초반의 여성 7 명을 피실험자로 선정하였다.

표 6. 실험문장 목록

실험문장	실험문장을 발화하는 상황
“뒤에 절벽 있어!”	위험한 상황에 빠질 수 있음을 알리고자 하는 상황
“바닥이 미끄럽다”	넘어지지 않도록 조심하라고 경고하는 상황
“내일 비 온대”	내일 날씨를 상대방에게 전달하려는 상황
“짐 빨리 옮겨”	1) 짐을 옮겨 달라고 명령형으로 요청하는 상황 2) 강압적으로 짐을 옮기도록 하기위해 명령하는 상황
“야! 안돼”	위험이 발생하고 있는 상황
“볼펜 달라고”	시끄러운 환경에서 볼펜을 빌리고자 하는 상황

3.2.3 육성발화와 합성발화의 파형 비교

합성기의 발화를 개선시킬 수 있는 음성제어태그를 생성하기 위하여 합성기 발화와 육성 발화를 비교하여 차이점을 찾는다. 본 논문에서는 두 발화의 차이점을 찾기 위해 동일한 문장을 발화하게 하고, 동일한 시간대역으로 음성파일의 파형을 비교하면서 실험을 진행하고자 한다. 단, 음성의 크기(Amplitude)는 녹음 상황에 따라 다르게 나오므로 파형분석을 위하여 일부 조정하여 나타낸다.

가) “뒤에 절벽 있어!”

두 발화의 파형에서 “a”부분을 비교해보면, 육성발화의 경우 톤(tone)이 자주 변화하는 것을 볼 수 있고, 합성기 발화의 경우 비슷한 톤으로 발화 하고 있으며 중간에 짧은 휴지구간이 존재하는 것을 볼 수 있다. 그리고 “b”부분을 비교해보면 육성발화의 경우 길게 늘여서 발화하지만 합성음의 경우에는 짧게 발화하는 것을 볼 수 있다.

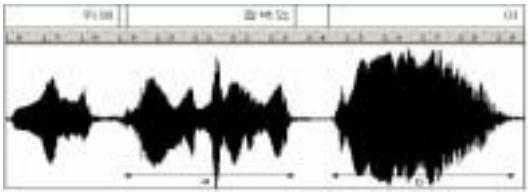

그림 2a. KJH(20세) 육성발화

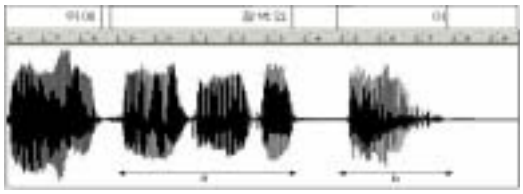

그림 2b. L&H Korea 합성발화

나) “바닥이 미끄럽다”

두 발화의 파형에서 “a”부분을 비교해보면 육성발화의 경우 ‘다’발음이 높은 톤으로 길게 발화되며, 합성 발화의 경우 낮은 톤으로 짧게 발화하는 것을 볼 수 있다.

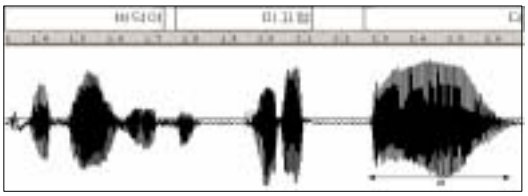

그림 3a. JMY(24세) 육성발화

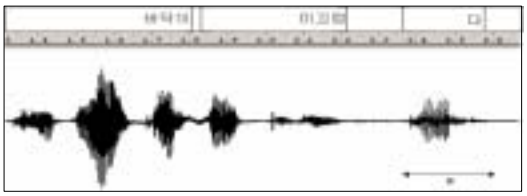

그림 3b. L&H Korea 합성발화

다) “짐 빨리 옮겨”

두 발화의 파형을 비교하면 “a”부분에서 큰 차이를 보인다. 육성발화에서는 목소리 톤이 일정하게 올라가서 어느 정도 수평을 유지하다가 다시 일정하게 떨어지는 특징이 나타나지만, 합성발화의 경우에는 목소리 톤이 높아졌다가 바로 낮아지고 있다.

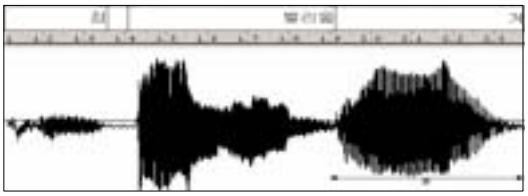

그림 4a. JMY(24세) 육성발화

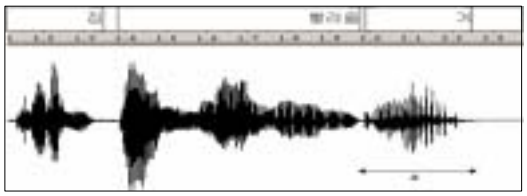

그림 4b. L&H Korea 합성발화

라) “야! 안돼”

두 발화의 파형을 비교하면, 육성발화에서는 “a”부분의 톤이 긴 시간동안 점점 높아졌다가 거의 끝 부분에서 낮아지는 것을 볼 수 있다. 하지만 합성기 발화에서는 중간 부분까지 높아졌다가 그 이후로 다시 낮아지고 있다.

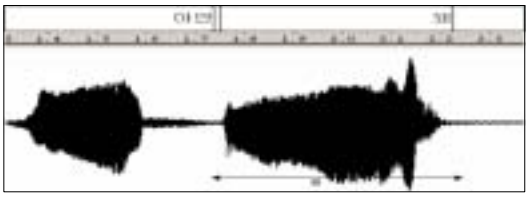

그림 5a. LJY(21세) 육성발화

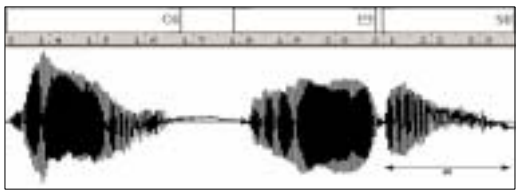

그림 5b. L&H Korea 합성발화

마) “볼펜 달라고”

두 발화의 파형에서 “a”부분을 비교해보면, 육성발화의 경우 파형 중간에 긴 휴지 구간이 존재한다. 그러나 합성기 발화의 경우에는 매우 짧은 휴지구간이 존재한다.

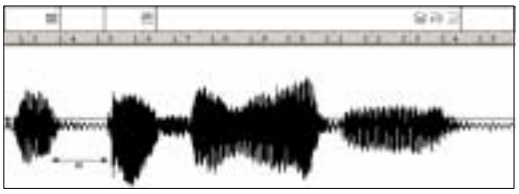

그림 6a. LJY(21세) 육성발화

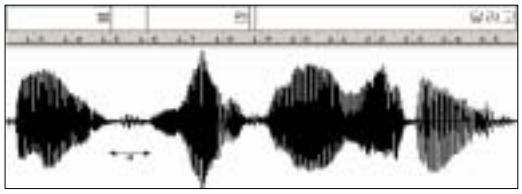

그림 6b. L&H Korea 합성발화

합성기 발화와 육성발화에서 발생하는 차이점들은 육성발화에서 감정이나 상황을 표현하기 위해 기본운율을 변화시키는 과정에서 발생한 것으로 보인다. 따라서 합성기의 무미건조한 발화를 일상생활 음성과 유사하게 생성하기 위해서는 이러한 차이점들을 극복하는 방안을 마련해야 한다.

3.3 음성제어태그의 생성

합성기 발화에 자연스러운 운율을 추가하기 위하여 3.2 절에서 비교 분석한 자료를 바탕으로 음성을 제어하는 태그(Speech Control Tag: SCT)들을 정의한다. <표 7>은 기호체계로 표현한 SCT와 태그의 의미를 기술하고 있다. 태그명은 기능 표현의 영문 이니셜을 사용하고 있으며, 음성제어태그의 포맷은 L&H PCMM Korean RealSpeak 기본 태그형태를 참조하였다.

표 7. 음성제어태그(Speech Control Tag : SCT) 리스트

태그종류	태그기능
<ESC>IE	확장된 시간영역에서 음성 크기를 선형적으로 증가시킴
<ESC>IR	축소된 시간영역에서 음성 크기를 선형적으로 증가시킴
<ESC>DE	확장된 시간영역에서 음성 크기를 선형적으로 감소시킴
<ESC>DR	축소된 시간영역에서 음성 크기를 선형적으로 감소시킴
<ESC>FE	확장된 시간영역에서 음성 크기를 일정하게 유지시킴
<ESC>FR	축소된 시간영역에서 음성 크기를 일정하게 유지시킴
<ESC>AI	고정된 시간영역에서 음성 크기를 선형적으로 증가시킴
<ESC>AD	고정된 시간영역에서 음성 크기를 선형적으로 감소시킴
<ESC>SA	임의의 시간영역에서 음성 크기를 천천히 감소시킴
<ESC>FA	임의의 시간영역에서 음성 크기를 빠르게 감소시킴
<ESC>SC	스타카토 형식으로 발화된 음을 자연스럽게 연결시킴
<ESC>SE	임의 시간에 해당하는 위치의 음을 강조시킴
<ESC>LSU	임의의 구간에서 발화의 속도를 증가시킴
<ESC>LSD	임의의 구간에서 발화의 속도를 감소시킴
<ESC>PI	임의의 시간에 해당하는 위치에 휴지구간을 삽입시킴
<ESC>PD	임의의 시간에 해당하는 위치의 휴지구간을 삭제시킴

<표 7>에서 정의한 SCT는 몇 개의 그룹으로 나누어 보면, 시간 영역에서 같은 음의 길이를 일정하게 변화시키는 태그군(IE, IR, DE, DR, FE, FR)과 음성의 크기를 변화시키는 태그군(AI, AD), 발화 속도와 관련된 태그군(LSU, LSD, PI, PD), 그리고 특수한 경우에 적용되는 태그군(SA, FA, SC, SE) 등으로 나누어 볼 수 있다. 이러한 태그들은 독립적으로 적용될 수도 있고 필요한 경우에 태그의 조합으로 적용될 수 있다. 그리고 파라미터를 받아서 운율 변동폭을 결정할 수 있다. 예를 들어 음성 톤을 일정하게 늘여주는 "FE" 태그의 경우, "FE(d)"와 같은 형태로 파라미터를 받아서 그 길이만큼 발화를 길게 늘여줄 수 있다.

3.4 의미기반 음성제어 태그(SSCT)

의미 분석을 통하여 제공되는 의미 정보에 의해서 발생하는 운율의 변화는 다양한 양상으로 나타난다. 따라서 이것을 표현하기 위해서는 앞 절에서 정의한 음성제어태그를 복합적으로 적용해야 한다. 본 논문에서는 문장의 의미 분석을 통하여 제공되는 의미 정보를 앞서 정의한 음성제어태그의 조합으로 나타내는 의미기반 음성제어태그(Semantic Speech Control Tag)를 제안한다. 본 절에서는 앞 절에서 사용한 예문에 나타나는 의미정보를 통하여 SSCT의 예들을 제시한다.

3.4.1 정보전달 의지, 경고, 알림, 부탁: InfoNoti

화자는 정보를 전달하려는 의지를 강하게 표현하거나, 앞으로 일어날지도 모르는 위험 상황에 대한 경고나 주의, 알림 등을 나타낼 때 발화의 끝부분을 길게 늘이는 경향을 보인다. 또한 부탁이나 요청을 할 경우에도 같은 경향을 나타내며, 특히 연인 사이나 친한 남녀 친구 사이에서 여성화자가 부탁의 의미를 전달할 때 더욱 강하게 이러한 경향을 보이고 있다. 본 논문에서는 이와 같은 의미정보에 대한 음성제어태그를 InfoNoti(Information Notification)로 나타낸다. <표 8>은 "뒤에 절벽 있어!"와 "바닥이 미끄럽다" 그리고 "짐 빨리 옮겨"라는 문장에 대한 의미 분석 결과 중에서 발화에 영향을 미치는 부분만 나타낸 것이다.

표 8. 문장 끝말 늘이기의 예문 분석

예문1	"뒤에 절벽 있어!"		
	의미층	관념구성적 기능	추락 위험에 대한 경고
예문2	"바닥이 미끄럽다"		
	컨텍스트층	테너	화자:어른, 청자:어린아이
		필드	길을 가다가 바닥이 미끄러움을 알았을 때
	의미층	관념구성적 기능	바닥 미끄러움에 대한 조심, 경고
예문3	"짐 빨리 옮겨"		
	컨텍스트층	테너	연인사이, 화자:여자, 청자:남자
	의미층	관념구성적 기능	무거운 물건 운반 요구
		대인관계적 기능	명령형 요청

이와 같은 경우의 운율 변화(InfoNoti)는 3.3 절에서 제시한 SCT의 FE와 SA를 조합함으로써 표현할 수 있다. 3.2절의 <그림 2a>의 "b"부분과 <그림 3a>의 "a"부분 그리고 <그림 4a>의 "a"부

분에서 이러한 현상을 볼 수 있다. 이때 파형은 오른쪽이 긴 사다리꼴로 나타나는데 화자의 감정이 강하게 표현될수록 평평한 부분이 길어지게 된다.

3.4.2 긴급상황, 절규, 분노: EmEnd

긴급한 상황을 표현하는 발화에서는 발화 끝부분의 톤을 점차 높여가며 길게 늘여서 발화하는 경향을 보이다가 끝부분의 종단에서 빠르게 감쇠한다. 이러한 특징은 분노나 절규와 같은 자신의 심리적 상황을 표출하는 상황에서 더욱 잘 나타난다. 본 논문에서는 이와 같은 의미정보에 대한 음성제어태그를 EmEnd(Emergency End)로 나타낸다. <표 9>는 “야 안돼!”라는 문장에 대한 의미 분석 결과를 나타낸 것이다.

표 9. 크게 말하면서 문장 끝말 늘이기의 예문 분석

	“야 안돼!”		
예문1	컨텍스트층	필드	도미노 게임 도중 친구가 도미노 하나를 쓰러뜨림
		테너	화자:도미노를 쌓는 사람, 청자:친구
	의미층	관념구성적 기능	긴급상황. 상황에 대한 부정
		대인관계적 기능	명령:외침

<표 9>의 예문에 대한 운율변화(EmEnd)는 SCT의 IR, IE, FA를 조합함으로써 표현할 수 있다. 3.2절의 <그림 5a>의 “a”부분에서 이러한 현상을 볼 수 있다. 이때 파형은 화자의 감정이 강하게 표현될수록 IE의 영향을 크게 받아 발화길이가 길어지게 된다.

3.4.3 명령: ComEnd

상대방의 행동을 제어하려는 명령형 형태의 발화에서, 화자는 발화 끝부분의 시작부분을 강하게 발화하고 그 톤을 약간 유지하다가 종단부분에서 톤이 줄어드는 경향을 보인다. 이 경우에 대한 SSCT는 ComEnd(Command Ending)로 정의한다. <표 10>은 “짐 빨리 옮겨”라는 문장에 대한 의미 분석 결과 중에서 발화에 영향을 미치는 부분만 나타낸 것이다.

표 10. 크게 말하면서 끝말 줄이기의 예문 분석

	“짐 빨리 옮겨”		
예문1	컨텍스트층	테너	직장상사 - 부하
		필드	서둘러 짐을 운반해야하는 상황에서 지시를 내림
	의미층	관념구성적 기능	짐 운반 지시, 신속한 명령
		대인관계적 기능	강압적 명령

이와 같은 경우 운율변화(ComEnd)는 SCT의 IR, FR, FA의 조합으로 표현할 수 있다. 아래 <그림 7a>는 “짐 빨리 옮겨”라는 문장을 명령형으로 발화 한 것이고, <그림 7b>는 같은 문장을 명령형 요청으로 발화 한 것이다. 두 그림을 비교 했을 때, 큰 차이점은 “a”부분 중에서 평평한 구간의 길이의 차이(FE와 FR의 차이)이다. 명령형 요청의 경우 평평한 구간이 길게 지속되지만 명령형의

경우에는 짧게 지속된다. 명령형 발화에 대한 파형은 왼쪽 변이 높게 올라가고 윗변 부분이 짧으며, 오른쪽 변이 빠르게 떨어지는 직사각형에 가까운 사다리꼴의 형태로 나타난다.

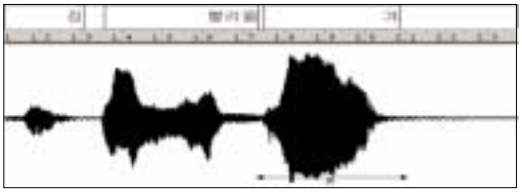

그림 7a. “짐 빨리 옮겨” - 명령형

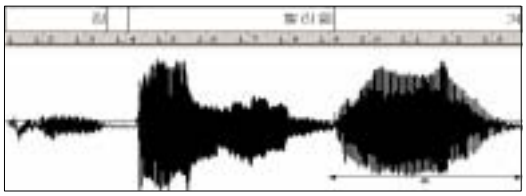

그림 7b. “짐 빨리 옮겨” - 명령형 요청

3.4.4 정보의 선택, 강조: InfoSel

화자가 여러 가능성 있는 정보들 중에서 한 가지를 선택하여 그 결과를 강조하여 발화해야하는 경우, 해당 정보 부분을 강조하여 발화한다. 이 경우에 대한 SSCT는 InfoSel(Information Selection)로 정의한다. <표 11>은 “내일 비 온대”라는 문장에 대한 의미 분석 결과 중에서 발화에 영향을 미치는 부분만 나타낸 것이다.

표 11. 강조하여 말하기의 예문 분석

	“내일 비 온대”		
예문1	컨텍스트층	필드	날씨가 흐리고 기온이 내려가서 첫눈이 올 가능성이 높은데 상대방이 날씨를 물어보는 상황 기상예보에서 “내일은 남쪽에서 올라오는 따뜻한 기압골의 영향으로 비가 올 것임” 이라고 발표
	의미층	관념구성적 기능	비가 온다는 정보를 선택 강조함
		대인관계적 기능	진술: 선택정보 전달

<표 11>의 경우 운율변화(InfoSel)는 SCT중에서 SE 하나로 표현할 수 있다. <그림 8a>는 <표 11>에 대한 음성 파형을 나타낸 것이고, <그림 8b>는 날씨가 바뀌어서 “비가 온다”는 일반적인 정보를 전달하기 위해 발화한 것이다. 그래서 <그림 8a>의 발화는 일반적인 발화인 <그림 8b>와 비교하여 “비” 부분의 파형이 강조되어 나타나게 된다.

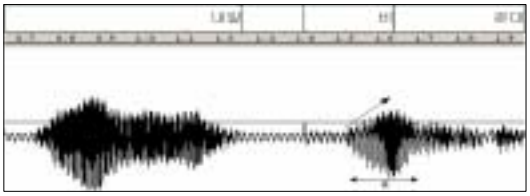

그림 8a. “내일 비 온대” - 정보 선택적 전달

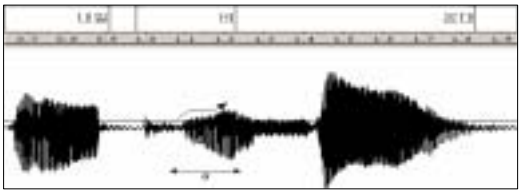

그림 8b. “내일 비 온대” - 정보의 전달

이상 살펴본 예를 비롯하여 본 논문에서 정의하는 SSCT를 <표 12>에 나타내었다. <표 12>는 각 태그들이 어떤 SCT들의 조합으로 이루어졌는지, 어떤 변위가 가능한지를 나타내고 있다. 예를 들어 EmEnd에서 IE(d)는 감정 표현의 정도를 파라미터로 받아서 발화의 시간축 상의 길이로 나타냄으로써 의미 변화를 음성 파형 변화에 적용할 수 있다. 감정 표현의 정도는 의미 분석의 어휘문법분석에서 “더욱”, “매우”, “절대” 등과 같은 감정 요소가 발견되는 대로 반영하게 된다.

표 12. SSCT의 종류

SSCT	SCT조합	파형변화
InfoNoti	FE(d)+SA	
EmEnd	IR+IE(d)+FA	
ComEnd	IR(d)+FR(d)+FA	
InfoSel	SE(d)	
AmpPau	PI(d)	

4. 요약 및 결론

현재 합성음의 음질은 대용량 음성 DB가 구축되어 상당히 개선되고 있지만 아직 자연스러운 발화와는 거리가 있으며, 일상생활에서 사용하는 발화에서 나타나는 운율의 변화는 거의 반영되지 못하여 무미건조한 음성에 머무르고 있다.

본 논문에서는 합성음의 운율을 육성 발화에 가깝게 변화시키는 한 방법으로 발화문장에 대한 의미 분석을 통하여 제공되는 의미 정보를 운율을 제어하는 태그(SSCT)로 변환하는 방법을 제안하였다. 이를 위하여 본 논문에서는 상황이 설정된 음성 발화와 합성음의 발화를 비교 분석하여 태그 생성의 정확성을 높였다. 본 논문에서 제안하는 의미정보는 상황이나 감정 등의 컨텍스트 정보와 문법적 해석 정보를 조합하여 생성하며, 또한 의미를 기능별로 나누어 제공함으로써 태그 생성으로의 활용이 용이하였다.

향후 연구로는 크게 세 가지 방향을 제시할 수 있다. 첫째, 언어적 분석을 보다 정교하게 하여 활용 가능한 언어정보를 확대하는 것이다. 둘째, 본 논문에서는 음성 파형 분석에서 시간 축 분석만 시도하였으나, 음성 변화는 주파수 대역에서도 발생하므로 이 부분에 대한 연구가 진행되어야 한다. 셋째, 본 논문의 아이디어를 시스템으로 실현하기 위해서는 제안하는 태그를 구현하는 기술이 개발되어야 한다. 그 방법은 신호처리를 통하여 음성 변조 기술을 개발하는 것과 음성 DB를 확충하는 방법이 있을 수 있으며 이들 방법은 상호보완적으로 적용되어야 할 것으로 본다.

참 고 문 헌

[1] 김영택 외 공저. 2001. *자연언어처리*, 서울: 생능출판사.
[2] 미디어젠. 2001), *음성합성기 보고서*.
[3] 나가오마코토 외 공저. 2000. *문자와 소리의 정보처리*, 한국학술정보(주).
[4] Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*, London: Edward Arnold.
[5] Lernout & Hauspie. 2000. *PCMM RealSpeak for Windows V1 Software Development Kit version1.1 User's Guide and Programmer's Reference*.
[6] Halliday, M.A.K. & Matthiessen, C. 1999. *Construing Experience Through Meaning : A Language-Based Approach to Cognition*, London: Cassell Academic.
[7] Kobayashi, I., Chang, M. S. & Sugeno, M. 2002. "A study on meaning processing of dialogue with an example of development of travel consultation system." *Information Sciences, 144*, 45-74.

접수일자: 2006. 11. 5
게재결정: 2006. 11. 30

▲ 장문수
서울특별시 성북구 정릉 4동 (우: 136-704)
서경대학교 소프트웨어학과
Tel: +82-2-940-7509 Fax: +82-2-919-5075
E-mail: cosmos@skuniv.ac.kr

▲ 정경채
서울특별시 성북구 정릉 4동 (우: 136-704)
서경대학교 컴퓨터과학과
Tel: +82-2-940-7291 Fax: +82-2-919-5075
E-mail: cherish1492@ihci.skuniv.ac.kr

▲ 강선미
서울특별시 성북구 정릉 4동 (우: 136-704)
서경대학교 컴퓨터과학과
Tel: +82-2-940-7291 Fax: +82-2-919-5075
E-mail: smkang@skuniv.ac.kr