제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

연결형 합성시스템을 위한 비정형 합성단위 추출 및 F0 모델링에 관한 검토

김 영일*, 김 종진*, 이 용주*, 이 숙향**
* 원광대학교 컴퓨터공학과, ** 원광대학교 영어영문학과

Study on the Non-uniform synthesis unit selection and F0 modeling for concatenative speech synthesis system

Young-Il Kim*, Jong-Jin Kim*, Yong-Ju Lee*, Suk-Hyang Lee**
* Dept. of Computer Engineering, Wonkwang Univ.
** Dept. of English Language and Literature, Wonkwang Univ.
yjlee@wonnms.wonkwang.ac.kr

요 약

본 연구에서는 자연스러운 한국어 음성을 합성할 수 있는 비정형 합성단위 선택기술 및 접합을 이용한 한국어 합성 시스템의 개발을 최종 목표로 하고 있다. 이러한 최종 목표에 도달하기 위해 본 연구팀에서 검토중인 연구방향과 시스템의 구조 및 이를 토대로 현재까지 진행된 결과를 보고한다. 현재 검토중인 시스템은 입력된 문장으로부터 목적치 패턴(Target value pattern)을 생성하고, 이에 근사한 임의 길이 합성단위를 대량의 음성 DB로부터 선택하여 접합시키는 방식을 이용하고자 한다. 본 논문에서는 음성의 왜곡을 최소화 할 수 있는 비정형 합성단위의 추출법에 관한 검토 결과와 본 연구팀에서 성능평가 중인 F0 자동 생성 알고리즘에 대하여 보고한다.

1. 서 론

최근 컴퓨터에 의해 합성된 음성에 대한 평가는 생성된 합성음이 얼마나 자연스러운가를 척도로 하고 있다. 음성 합성기에 있어서 합성단위의 선택은 합성단위의 수, DB의 크기, 합성단위 추출 알고리즘의 복잡도와 밀접한 관련이 있다. 뿐만 아니라 합성단위를 연결할 때 발생하는 불연속성에도 밀접한 관련이 있으므로 합성음의 품질에 큰 영향을 준다.

음성이 가지고 있는 많은 정보 중에서 음성인지(speech perception)에 관련된 많은 정보가 음소와 음소 사이의 천이 부분에 포함되어 있다. 또한 두 음소가 서로 인접할 경우 인접한 음소에 영향을 미쳐 음성의 음향학적 특성을 변형시키게 된다[1][2]. 따라서 합성단위를 선택할 때 이러한 음소의 상호 조음 현상도 충분히 고려되어야 한다.

기존의 음성합성 시스템에 널리 사용되는 합성단위로는 구절(phrase), 단어(word), 음절(syllable), 음소(phoneme)등의 단위와 이들을 음성합성에 알맞은 형태로 변형시킨 반음절(demi-syllable), 다이폰(diphone), CV, VC, VCV, CVC 등 여러 가지가 사용된다. 합성단위가 작을수록 데이터베이스의 크기가 작아 관리하기가 쉬우나, 위에서 언급한 음소간의 조음현상을 구현하기가 어려우므로 합성음의 자연성은 떨어진다. 반면 합성단위가 클수록 음소간의 조음현상(coarticulation effects)을 충분히 반영할 수 장점이 있으나, 데이터 베이스의 크기가 커져 관리하기가 어려워진다[3].

최근, 고품질의 합성음을 생성해 내기 위한 음성합성 기술 중 하나인 코퍼스를 기반으로 한 연결형 합성시스템의 경우, 주어진 문장의 운율정보와 문맥 정보를 이용하여 가장 유사한 음을 코퍼스로부터 추출하여 이를 신호처리 없이 연결함으로써 자연스러운 음을 생성해 낼 수 있다[4]. 또한 합성단위간의 결합점을 최소화하며, 합성단위 내에서의 자연성은 보장할 수 있는 비정형 합성단위를 이용함으로써 합성음의 품질을 높일 수 있다[5].

본 논문에서는 이러한 코퍼스 기반의 연결형 합성시스템을 위하여 우리말에 나타나는 여러 음소들의 결합 특성을 분석하여 음소간의 조음현상(coariticulation effects)을 충분히 반영시키고, 또한 합성단위를 결합할 때 나타나는 불연속을 줄이기 위하여 결합점을 줄일 수 있는 비정형 합성단위를 추출하는 방법에 대하여 검토한 결과를 보고한다.

현재 비정형 합성단위 추출법은 K-ToBI 레이블링된 200문장을 기반으로 설계 및 구현중에 있으며 추후 실험을 통한 객관적인 평가 및 검증이 이루어져야 함을 밝혀둔다.

제 1 장 서론에 이어 제 2 장에서는 비정형 선택기반 합성 시스템에 관하여 기술하였으며, 제 3 장에서는 비정형 합성단위를 추출하는 알고리즘에 대한 검토사항을 서술하였다. 제 4 장에서는 F0 자동 생성 알고리즘을 설명하였으며, 제 5 장에서 본 논문의 결론을 맺었다.

2. 비정형 선택기반 합성 시스템

연결형 합성 시스템에 있어서 고품질의 합성음을 생성해 내기 위해서는 최적의 합성단위 추출이 무엇보다 중요하다. 본 연구에서는 인접 결합간의 불연속을 최소화 할 수 있는 합성단위 경계를 설정한 후 이를 바탕으로 코퍼스로부터 최적의 비정형 합성단위를 추출함으로써 합성음의 품질을 향상시킬 수 있는 연결형 합성시스템을 구축하고자 한다.

설계 중인 시스템[그림2]은 합성문장을 어떤 음성으로 만들어 낼 것인가를 결정할 수 있는 "목적치 패턴 자동생성 모듈"과 목적치 패턴과 가장 유사한 문맥환경과 운율환경을 가지는 합성단위를 대량의 음성 DB에서 선택할 수 있는 "최적의 비정형 합성단위 추출 모듈"로 구성되어 있다.

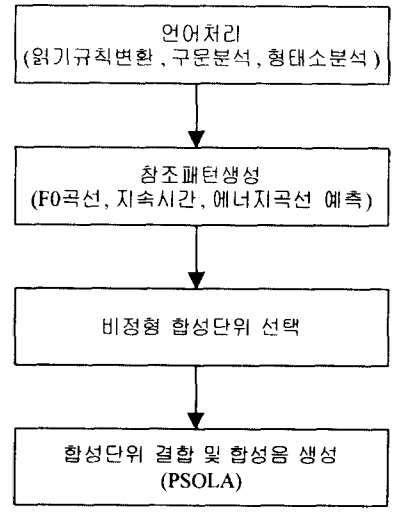

그림 2. 비정형 선택기반 합성시스템

3. 비정형 합성단위 추출 알고리즘

3.1 비정형 합성단위 추출 흐름도

비정형 합성단위 추출을 위한 시스템 흐름도는 그림 3.1과 같다. 전체 시스템은 비정형 합성단위 경계 설정모듈, 합성단위 추출모듈, 합성모듈로 구성되어 있다. 합성단위 추출 시 미리 저장된 3개의 파일들을 이용한다. 이들 중 ①은 현재 실험 계획 중에 있다.

① 음소 결합도

잘 설계된 코퍼스로부터 음소의 피치, 지속시간, 에너지, 두 음소가 인접할 때 나타나는 스펙트럼 변화특성에 따라서 자음(파열음, 파찰음, 마찰음, 비음, 유음), 모음(단모음, 이중모음, 반모음) 사이의 조음현상(coarticulation effects)의 강도에 따라 레벨이 정의되어 있다. 입력문장에 대하여 합성단위 경계를 선정할 때 음소 결합도를 참조하여 두 음소간의 조음현상(coarticulation effects)이 강한 부분은 경계에서 제외되도록 한다.

② 합성단위 사전

현재 시스템 구축에 기반이 되는 K ToBI 레

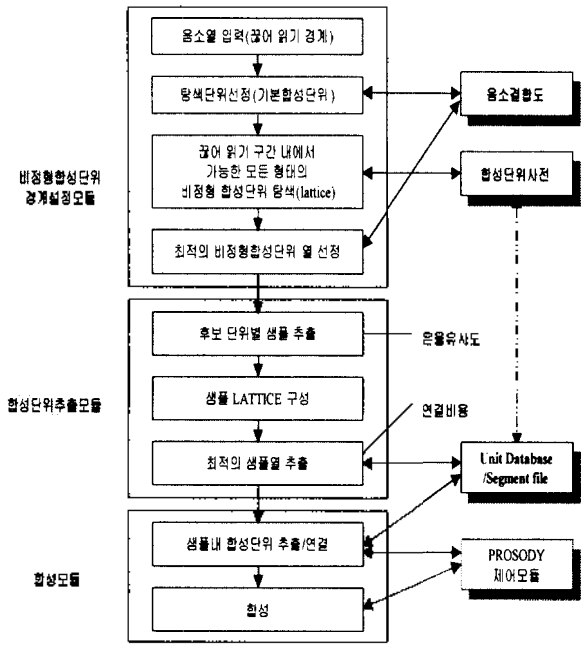

그림 3.1. 비정형 합성단위 추출 흐름도

이블링된 코퍼스에서 출현하는 모든 비정형 음소열을 사전으로 구성한다. 사전에는 비정형 음소열과 음소열의 출현 횟수, 그리고 음소열을 포함하고 있는 파라미터 파일 번호가 등록되어 있다.

그림 3.2는 입력 음소열 /완성에 다라도록/(wansvNedaradorog)에 관련된 합성단위사전의 일부 예를 보여주고 있다.

합성단위	샘플수	파일번호
wa	371	[019]
wan	36	[033]
wans	5	[044]
		[062]
an	420	[178]
ans	26	
ansvNe	1	[116]
s	856	[097]
svNe	6	[122]
svNed	3	[195]
vNe	16	[060]
vNed	5	[089]
		[097]
d	680	[122]
dara	3	[195]
ara	46	[062]
arad	2	[081]
orog	7	

그림 3.2. 합성단위사전

③ Unit Database

코퍼스에 있는 각 문장들에 대하여 문장 내의 음소들에 대한 피치 값, 지속시간, 에너지, 인접 음소의 영향으로 인한 유성음화/무성음화 정보가 저장되어 있다. 두 음소를 비교할 때 각각의 음소에 해당하는 값들을 Unit Database파일로부터 추출한다.

비정형 합성단위를 추출하기 위한 작업순서는 다음과 같다.

(1) 비정형 합성단위 경계 설정

읽기규칙이 적용된 음소열이 입력으로 주어지면 주어진 문장을 자연스럽게 발생시키기 위하여 끊어읽기 구간으로 나눈다. 음소 결합도를 참조하여 끊어 읽기 구간을 기준으로, 가능한 한 인접한 두 음소의 조음현상(coarticulation effects)이 강한 부분은 합성단위 경계로 설정되지 않도록 기본 합성단위를 결정한다.

결정된 기본 합성단위를 기준으로 합성단위 사전내에 있는 모든 형태의 비정형 합성단위를 추출하여 lattice를 구성한다. 음소 결합도를 참조하여 합성단위 경계를 기준으로 두 음소의 결합강도가 낮고, 전체 합성단위의 결합점의 수를 최소화 할 수 있는 최적의 비정형 합성단위열을 선정한다.

그림 3.3는 입력 음소열 /완성에 다라도록/(wansvNedaradorog)에 대한 기본 합성단위 설정 과정을 보여준다.

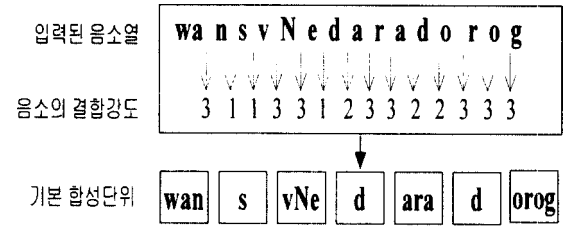

그림 3.3. 음소의 연결강도에 의한 기본단위 결정

(2) 합성단위 추출

(1)단계에서 결정된 비정형 합성단위를 포함하고 있는 모든 샘플들을 Unit Database에서 추출하여 lattice를 구성한다. 주어진 문장과 가장 유

사한 샘플을 선택하기 위하여 각 샘플들에 대해 주어진 문장과의 운율 왜곡 정도와 샘플간의 연결비용을 계산한다[6].

운율 왜곡 정도(Cp)는 주어진 문장(ti)과 샘플(ui)내의 합성단위에 해당하는 부분을 기준으로 앞, 뒤 3음소씩을 포함하여 대응하는 음소의 피치, 지속시간, 에너지, 그리고 음소의 음성학적 분류에 따른 feature들에 대한 distance로 정의한다.

연결비용(Cc)은 서로 다른 문맥에서 추출된 두 샘플(ui,ui+1)간의 자연스러운 연결 정도를 나타낸다. 이를 측정하기 위하여 합성 경계점을 기준으로 두 샘플을 포함하고 있는 코퍼스내의 문장들의 문맥 유사도를 계산한다.

다음 계산식에 의해 운율 왜곡 정도와 연결비용을 최소화시킬 수 있는 합성단위들을 추출한다.

$$\min \sum_{i=0}^{n} (Wp * Cp(ui, ti) + Wc * Cc(ui, ui+1))$$

Wp와 Wc는 각각 운율의 왜곡 측정에 대한, 그리고 연결비용에 대한 weighting value를 나타낸다.

(3) 합성

연결 단위가 결정된 후, 자연스러운 합성을 위하여 prosody 예측 모듈에서 생성된 합성단위의 운율 파라미터 값을 적용시킨다. 최종 결정된 합성단위들을 신호처리 없이 연결한다.

prosody 예측 모듈은 현재 본 연구실에서 개발중인 F0 생성모듈을 적용하여 합성음의 품질을 개선시키고자 한다

4. F0 모델링[7][8]

비정형 합성단위 검출 및 접합기술을 기반으로 한 합성 시스템에서 가장 중요한 요인은 실제 합성하고자 하는 음과 가장 유사한 특성을 가지는 단위음을 음성DB에서 검출해 내는 것이다. 최적의 비정형 합성단위를 음성DB에서 검출하기 위해서는 먼저 합성하고자 하는 문장에 대한 참조패턴-외국의 문헌에서는 이러한 참조패턴의 생성기능을 Target value generation이라 한다.-을 생성할 수 있어야 한다. 참조 패턴이 생성되면 참조패턴을 기반으로 음성DB에서 다양한 합성단위를 검출하고 각각을 참조패턴과 비교하여 유사도 또는 불연속성을 계산하게 된다. 그러므로, 본 연구에서는 대상 문장을 입력으로 받아 최적의 참조패턴을 생성할 수 있는 자동 참조 패턴 생성 기능을 비정형 선택 및 접합 기반 합성 시스템의 핵심 기능으로 간주한다.

이에 따라, 최적의 참조 패턴을 생성 또는 예측할 수 있는 자동 참조 패턴 생성 모듈을 그림 4.1과 같이 분류하여 연구를 진행하고 있다.

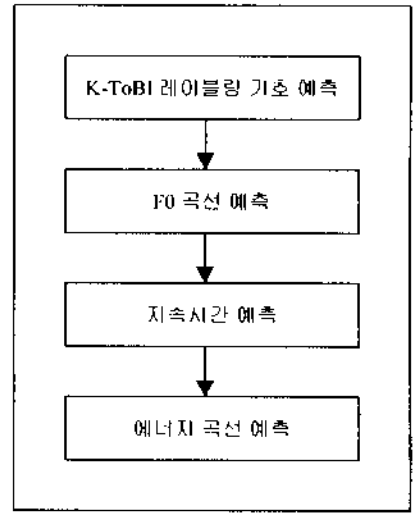

그림 4.1. 자동 참조 패턴 생성 모듈

현재까지 진행된 결과는 K-ToBI 레이블링 기호 예측 모듈과 이를 이용한 f0곡선 생성 알고리즘을 개발하였으며, 이에 대한 성능평가를 진행하고 있으며 문제점 검토가 이루어졌다. 또한 문제점으로 검토된 바를 해결하기 위해 K-ToBI 레이블링 데이터의 추가 확보 계획을 진행 중에 있으며, 예측 모델의 성능향상을 위한 연구를 병행으로 진행하고 있다.

이후 연구에서는 F0생성 시스템의 성능개선을 위한 포우스 검출 알고리즘과 지속시간 모델링 및 에너지 곡선 모델링에 관한 연구를 진행시킬 예정에 있다. 다음은 현재까지 진행된 F0곡선 참조 패턴 자동 생성 모듈에 대해서 간략하게 기술한다.

4.1 F0 곡선 참조 패턴 자동 생성 모듈

본 연구에서는 F0곡선 참조 패턴을 자동 생성하기 위해서 K-ToBI 레이블링 기호를 이용한 F0 자동 생성 알고리즘을 개발하였으며, 현재 이에 대한 성능평가를 진행하고 있다. 현재까지 개발된 F0곡선 생성 시스템의 프로시져는 그림 4.2와 같다.

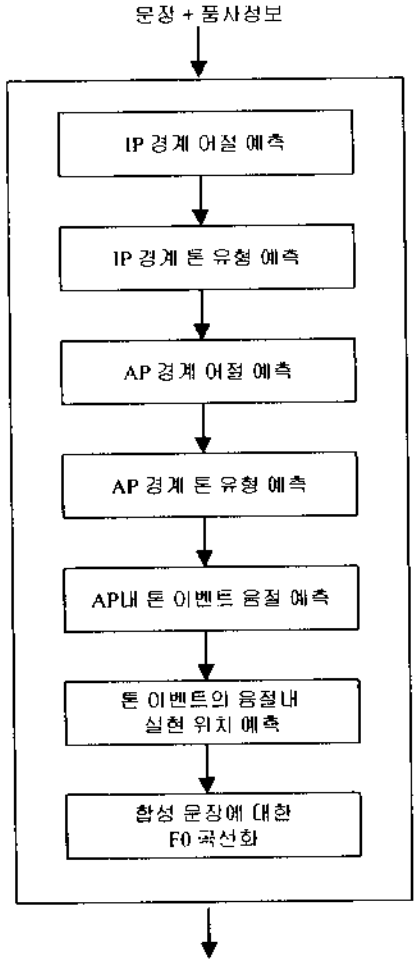

그림 4.2

F0곡선 생성 시스템에서는 어절단위 IP 경계 위치 및 유형 예측과 AP 경계 위치 및 유형 예측을 위한 4개의 Classification Tree와 AP내 음절단위 Optional/Low/High 피크를 할당하는 규칙, 그리고 각 K-ToBI 레이블링 기호별 F0곡선을 예측할 수 있는 14개의 Regression Tree로 구성되어져 있으며, 모델은 현재까지 K-ToBI 레이블링된 남녀 각 1인이 발성한 400문장을 이용해 훈련되었다.

F0곡선 생성 시스템에 의해 생성된 F0곡선과 원음의 F0곡선은 그림 4.3과 같다.

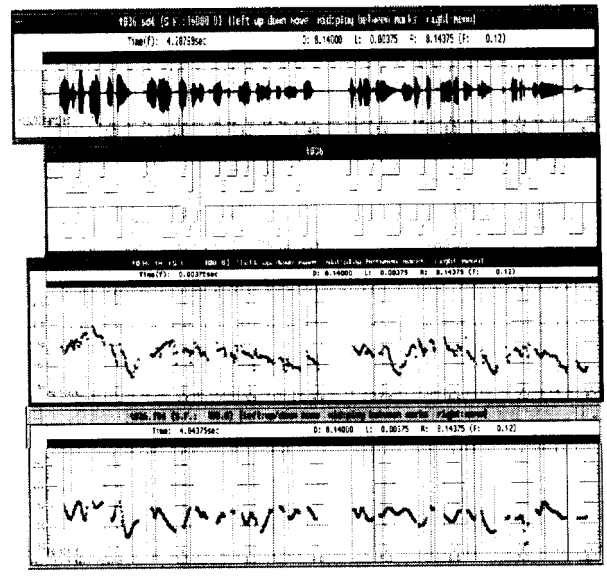

그림 4.3 . 합성된 f0곡선과 원음의 f0곡선과 비교: (a) 원음파형, (b)첫줄:예측된 K-ToBI 레이블링기호, 두번째줄:레이블러가 레이블링한 K-ToBI 레이블링 기호, (c)원음의 f0곡선, (d) 합성된 f0곡선

5. 결 론

본 논문에서는 연결형 합성시스템의 전반적인 시스템 설계와, 그에 따른 비정형 합성단위 추출법에 관한 검토 사항, 그리고 F0 자동 생성 알고리즘을 기술하였다.

합성단위를 결합할 때 발생하는 음성의 불연속을 줄이기 위해서는 먼저 우리말에 나타나는 각 음소간의 조음현상(coarticulation effects)을 분석하여 이러한 특징들을 합성단위에 충분히 반영시키기 위한 실험이 선행되어져야 한다.

현재 음소의 피치, 지속시간, 에너지, 두 음소의 결합 시 발생하는 스펙트럼 천이 정보를 이용하여 두 음소의 결합강도 분석에 관한 실험을 계획 중이다.

본 시스템은 아직 설계 및 구현중이며 추후 실험을 통한 객관적인 평가 및 검증이 이루어져야 함을 밝혀둔다.

참 고 문 헌

[1] S. Furui, "On the role of spectral transition for speech perception, J. Acoust. Soc. Amer., vol.80, pp.1016-1025, Oct. 1979.

[2] K. N. Stevens, "Acoustic correlates of some phonetic categories," J. Acoust. Soc. Amer., vol.68, pp.836-842, Sep. 1980.

[3] Yunkeun Lee, Seungkwon Ahn, Trend in Speech Synthesis, KITE Review Vol. 20, No.5, pp.523-532, 1993.

[4] Andrew J. Hunt and Alan W. Black, "Unit Selection in a Concatenative speech Synthesis System using a Large Speech Database," ICASSP, pp.373-376, 1996.

[5] K. Takeda, K. Abe and Y. Sagisaka, "On the basic scheme and algorithms in non-uniform unit speech synthesis," in *Talking machines*. pp.93-105, 1992.

[6] Nick Campbell and Alan W. Black, "Prosody and the Selection of Source Units for Concatenative Synthesis", *Progress in Speech Synthesis*, pp.279-282, Springer Verlag, 1996.

[7] Jong-Jin Kim and et al., "An analysis of some prosodic aspects of Korean utterances using K-ToBI labelling system," ICSP'97, pp.87-91, Seoul, Korea, Aug. 1997.

[8] 이용주, K-ToBI 기호에 준한 F0 contours 생성 알고리듬 연구, 위탁과제 최종 연구보고서, 한국전자통신연구원, 1997.