(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.[6]
G10L 5/00

(45) 공고일자 1999년04월01일
(11) 등록번호 특0173340
(24) 등록일자 1998년10월29일

(21) 출원번호 특1995-055841
(22) 출원일자 1995년12월23일
(65) 공개번호 특1997-050108
(43) 공개일자 1997년07월29일

(73) 특허권자 한국전자통신연구원 양승택
대전시 유성구 가정동 161번지
(72) 발명자 이정철
대전광역시 중구 태평동 삼부아파트 36-102
이영직
대전광역시 유성구 어은동 한빛아파트 102-105
김상훈
대전광역시 유성구 가정동 236-1
한민수
대전광역시 유성구 신성동 삼성한울아파트 106-1004
(74) 대리인 박해천

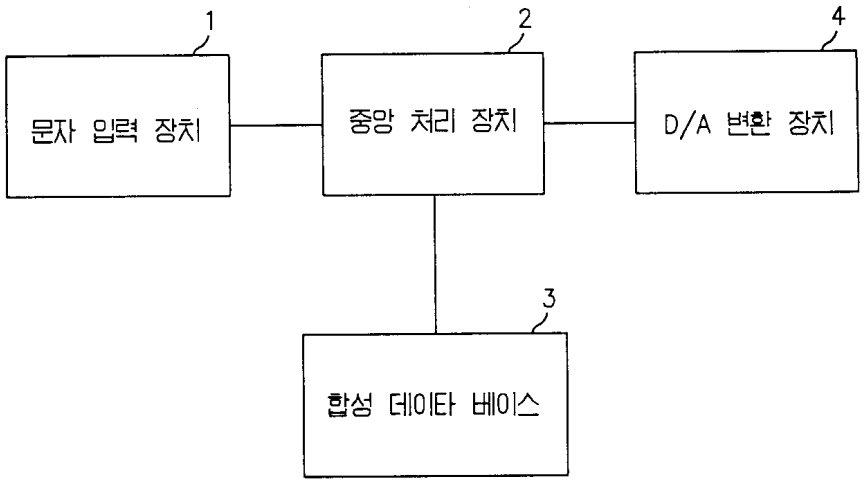
심사관 : 박정학

(54) 텍스트/음성변환기에서 억양패턴 정규화와 신경망 학습을 이용한 억양 생성 방법

요약

본 발명은 텍스트/음성변환기(text-to-speech conversion system)에서 억양패턴 정규화와 신경망 학습을 이용한 억양 생성 방법에 관한 것으로서, 음성합성 장치에 적용되는 텍스트/음성변환장치에 적용되는 억양패턴 정규화와 신경망 학습을 이용한 억양생성 방법에 있어서, 합성 데이터베이스(3)로부터 음성 데이터를 읽어 음절의 피치 패턴을 정규화 및 표준화하고 어절내 각 음절의 평균 피치값에서 어절의 평균 피치값을 뺀 피치값으로 어절피치패턴을 학습하고, 문장내 각 어절의 평균 피치값으로부터 문장의 기준 억양을 추정한 후, 문장의 문맥에 따른 문법 속성열과 그에 해당되는 억양패턴테이블을 작성하는 제1단계(10 내지 15); 한국어 문장과 문법 속성열이 입력되면 문장의 기준억양 생성 과정에서 각 어절에 대해 문장내 위치에 따라 1차 평균 피치값을 할당하고, 비균일 단위의 억양패턴 생성 과정에서 입력된 문법 속성열을 이용하여 왼쪽 우선 검색 방식으로 최장 일치 부분을 억양 패턴 테이블에서 찾아 해당 어절에 2차 평균 피치값을 할당하고, 어절의 피치 패턴 생성 과정에서 신경망을 이용하여 각 음절의 평균 피치값 변화량을 계산하고, 음절의 피치 패턴 생성 과정에서는 각 음절을 구성하는 음소열과 표준 피치 패턴 테이블을 이용하여 음절의 피치 패턴을 계산하고, 각 과정의 결과를 합하여 전체 문장의 억양을 생성하여 출력하는 제2단계(16 내지 21)를 포함하여 실제 음성데이타를 분석하여 억양 제어 규칙을 작성하고 합성음의 억양 제어를 통해 합성음의 자연성과 유창성을 향상시킴으로써 통신 서비스, 사무 자동화, 교육 등의 여러 분야에 응용할 수 있는 효과가 있다.

대표도



명세서

[발명의 명칭]

텍스트/음성변환기에서 억양패턴 정규화와 신경망 학습을 이용한 억양 생성 방법

[도면의 간단한 설명]

제1도는 본 발명이 적용하는 하드웨어의 구성도.

제2도는 본 발명이 적용되는 한국어 텍스트/음성변환의 흐름도.

제3도는 본 발명에 따른 억양 생성 방법의 일실시예 흐름도.

* 도면의 주요부분에 대한 부호의 설명

1 : 문자 입력 장치 2 : 중앙 처리 장치

3 : 합성 데이터 베이스 4 : D/A 변환 장치

[발명의 상세한 설명]

본 발명은 텍스트/음성변환기(text-to-speech conversion system, 이후 TTS)에서 억양패턴 정규화와 신경망 학습을 이용한 억양 생성 방법에 관한 것이다.

음성 합성기의 기능은 컴퓨터가 사용자인 인간에게 다양한 형태의 정보를 음성으로 제공하는데 있다.

사용자는 음성 합성기를 이용하여 기존의 텍스트 데이터나 대화상대로부터 제공되는 텍스트 정보를 음성으로 출력할 수 있다. 물론 사용자에게 고품질의 음성 합성 서비스를 제공하기 위해서는 합성음은 명료도와 자연성이 높아야 하고, 발성 속도조절 및 적절한 의미적 강조가 이루어지게 유창해야 하며, 하드웨어나 소프트웨어 면에서 쉽게 구현되어야 한다.

그러나, 인간에 가까운 자연스러운 합성음을 만들어 내는데는 아직도 많은 어려움이 있으므로 국내에서 실제 상용화한 제품은 손에 꼽을 정도이다. 특히 무제한 어휘를인간에 가까운 음성으로 변환하기 위한 규칙 합성 방식은 명료도 확보를 위한 합성 단위의 선정과 결합 방법, 자연성 구현을 위한 운율 구현 방법, 그리고 합성음 생성을 위한 합성 방식에 대한 연구가 아직도 미흡한 실정이다.

합성음의 자연성과 밀접한 관련이 있는 운율은 음의 높이, 세기, 길이의 3가지 형태로 나타난다. 이중 음의 높낮이 변화는 억양을, 음의 세기는 의미적 강세를, 음의 길이는조음점, 조음 방법, 조음 결합에 따른 변별적 지속시간 차이와 운율 경계 정보를 내포하고 있다. 이는 실제의 운율 패턴에서 구별가능한 패턴의 종류와 그 의미, 그리고 텍스트 문자장의 형태, 구문 구조, 문맥에 따른 운율 패턴의 관련성을 찾아 운율 구현 규칙을 작성하는 것이 합성음의 명료도 및 자연성 구현에 필수적임을 의미한다.

그러므로, 텍스트 문장의 분석방식, 운율 구현 모델, 운율 패턴 구현 방식은 합성음 명료도 및 자연성을 결정짓는 중요 원천기술이다.

따라서, 본 발명은 텍스트/음성변환기에서 합성음의 자연성과 유창성을 향상시킬 수 있도로 실제 음성데이타를 분석하여 억양 제어 규칙을 작성하고 억양페턴 정규화와 신경막 학습을 이용하여 억양을 생성시키는 방법을 제공하는데 그 목적이 있다.

상기 목적을 달성하기 위한 본 발명은, 텍스트/음성변환장치에 적용되는 억양패턴 정규화와 신경망 학습을 이용한 억양생성 방법에 있어서, 합성 데이터베이스(3)로부터 음성 데이터를 읽어 음절의 피치 패턴을 정규화 및 표준화하고 어절내 각 음절의 평균 피치값에서 어절의 평균 피치값을 뺀 피치값으로 어절피치 패턴을 학습하고, 문장내 각 어절의 평균 피치값으로부터 문장의 기준 억양을 추정한 후, 문장의 문맥에 따른 문법 속성열과 그에 해당되는 억양패턴테이블을 작성하는 제1단계(10 내지 15); 한국어 문장과 문법 속성열이 입력되면 문장의 기준억양 생성 과정에서 각 어절에 대해 문장내 위치에 따라 1차 평균 피치값을 할당하고, 비균일 단위의 억양패턴 생성 과정에서 입력된 문법 속성열을 이용하여 왼쪽 우선 검색 방식으로 최장 일치 부분을 억양 패턴 테이블에서 찾아 해당 어절에 2차 평균 피치값을 할당하고, 어절의 피치 패턴 생성 과정에서 신경망을 이용하여 각 음절의 평균 피치값 변화량을 계산하고, 음절의 피치 패턴 생성 과정에서는 각 음절을 구성하는 음소열과 표준 피치 패턴 테이블을 이용하여 음절의 피치 패턴을 계산하고, 각 과정의 결과를 합하여 전체 문장의 억양을 생성하여 출력하는 제2단계(16 내지 21)를 포함하는 것을 특징으로 한다.

이하, 첨부된 도면을 참조하여 본 발명의 일실시예를 상세히 설명한다.

제1도는 본 발명이 적용되는 하드웨어 구성도로서, 도면에서 1은 문자입력장치. 2는 중앙 처리 장치, 3은 합성 데이터베이스, 4는 디지털/아날로그(D/A) 변환 장치를 각각 나타낸다.

문자 입력 장치(1)는 KS5601 완성형 및 2 바이트(byte) 조합형으로 표현이 가능한 한국어 문자를 입력받아 중앙 처리 장치(2)로 출력한다.

중앙 처리 장치(2)는 본 발명의 알고리즘의 탑재되어 수행되는 부분이다.

합성 데이터베이스(3)는 합성 알고리즘에 사용되는 CDU 합성 DB로서 기억 장치에 저장되어 있으며 상기 중앙 처리 장치(2)로 필요한 데이터를 전송한다.

디지털/아날로그(D/A)변환장치(4)는 합성이 끝난 디지털 데이터를 아날로그 신호로 변환하여 외부로 출력한다.

제2도는 본 발명이 적용되는 한국어 텍스트/음성변환의 흐름도로서, 크게 언어 처리 단계(5), 운율 처리

단계(6), 합성음 생성 단계(7)로 이루어져 있다.

언어 처리 단계(5)는 먼저 입력된 텍스트 문장을 약어, 문장 기호, 특정 용어에 대해 한국어로 해석하는 전처리 단계를 수행한다. 다음 단계에서는 어휘 사전에 60여개 그룹으로 분류 등록된 한국어의 조사, 활용형 어미, 부사, 접속사등을 이용하여 입력 문장의 각 어절에 문법적 기능을 추정, 할당하는 어절 분석을 한다.

그리고, 한국어 문법을 이용하여 입력 문장의 구문 구조를 추정하는 파싱을 한다.

단어 사전 검색과 한국어 발음 규칙에 따라 입력 문장을 소리나는 데로 바꾼다.

운율 처리 단계(6)에서는 상기 언어 처리 단계(5)의 처리 결과를 받아서 문장 구조에 따라 적합한 운율 규칙을 적용함으로써 합성음의 빠르기, 억양, 뛰어 읽기 등의 자연성 및 유창함과 관계된 정보를 생성한다.

합성음 생성 단계(7)는 상기의 처리 과정을 거쳐 구한 발음 기호와 운율 정보를 이용하여 합성단위들을 조절, 가공 및 결합한 뒤, 합성음을 생성하여 사용자에게 음성으로 출력한다.

제3도는 본 발명에 따른 텍스트/음성변환기에서 억양패턴 정규화와 신경망 학습을 이용한 억양 생성 흐름도를 나타낸다.

먼저, 합성 데이터베이스(3)로부터 음성 데이터를 읽어(10) 한국어에서 가능한 3192개 (19[초성]x[모음]x7[중성])음절의 피치 패턴을 정규화 및 표준화 시킨다(11). 피치 패턴의 정규화에 필요한 4개의 파라미터(p_1.p_2.p_3,Tn)는 음성 신호 데이터베이스에 저장된 단위 음절내 피치 신호에서 음절의 평균 피치값 $\overline{P}_{sj}$를 뺀 값들과, 누적 자승오차를 최소로 하는 값으로 추정한다. 각 음절의 표준화된 피치 패턴 $\overline{P}_{f_k}$ (k=1,.....3192)는 음성 데이터베이스내의 동일 음절들에 대한 각 파라미터값들의 평균으로 구한다.

그런다음, 어절내 각 음절의 평균 피치값 $\overline{P}_{s_j}$에서 어절의 평균 피치값 $\overline{P}_{w_i}$를 뺀 평균피치값 변화량 $\delta\overline{P}_{s_j}$를 계산한다(12). 이들 변화량을 제어하는 규칙은 신경망 학습을 이용하여 작성한다. 신경망의 입력은 대상 음절을 기준으로 동일 어절내의 전후 각 9개의 음절의 평균 피치값, 조음 방법, 조음점에 따른 특성분류를 기준으로 한 직전 음절의 조음 특성 7개, 현재 음절의 조음 특성 11개, 직후 음절의 조음 특성 8개, 어절내 음절의 위치, 어절내 음절수, 어절의 평균 피치값으로 이루어진 49개의 파라미터이다. 신경망의 출력은 $\delta\overline{P}_{s_j}$이며, 이를 목표로 신경망을 학습시켜서 규칙을 구한다.

다음, 문장내 각 어절의 평균 피치값 $\overline{P}_{w_i}$로부터 문장의 기준억양, y=ax+b을 추정한다(13). 각 문장의 기준 억양의 시간축에서 어절수에 따라 정규화된 값들과 누적 자승 오차를 최소로 하는 a,b 값으로 추정된다. 최종적으로 모든 문장에 대해서 구한 a,b 값들의 평균값을 구하여 문장의 표준 기준 억양으로 규칙화한다.

문장의 기준 억양을 추정한 후, 문장의 문맥에 따른 억양 특성분석 및 규칙 테이블을 작성한다(14). 규칙 테이블의 작성은 한국어의 문법을 이용한 구문 구조 분석(파싱)에 의존하지 않고, 단기 1~5개의 연속된 문법속성 열에 대한 운율 패턴을 작성함으로써 문법적 단위가 아닌 비균일(non-uniform)단위의 억양 패턴 처리 규칙을 만들게 된다. 이 방법은 억양 구현이 파서의 정교함과 밀접한 관계가 있는 일반적 방식과 크게 구별되며, 그 구현이 쉬운 장점이 있다. 작성 방법은 다음과 같다.

- 문장의 표준 기준억양을 이용하여 $\overline{P}_{1w_i}$를 구한다.

- 각 어절의 평균 피치 값 $\overline{P}_{w_i}$에서 $\overline{P}_{1w_i}$를 뺀 $\overline{P}_{2w_i}$를 구한다.

- 형태소 분석기를 이용하여 각 어절에 문법적 속성들 {g_k | = | 1,.....,k}중에 하나를 부여한다.

- 음성 데이터베이스에서 n개의 연속된 문법 속성 열을 $G(n)=g_{i1}\cdot g_{i2}\cdots g_{in}$을 만족하는 모든 경우에 대해서 g_{in} 별로 $\overline{P}_{2g_i}$의 평균값 $\overline{P\omega 2_{g_{in}}}$ 을 구한다. 여기서 i는 n개의 연속된 문법 속성열들로 이루어진 집합중의 한 원소를 의미하며, 모든 가능한 i에 대해 억양 패턴을 구한다. 규칙 테이블에 문법 속성열과 그에 해당되는 억양 패턴 즉 피치 평균값열을 저장한다(15).

- n=1,2,.....,5에 대해서 위 과정을 반복한다.

- 두 개의 G(n)조합으로 G(n+1)을 생성할 수 있을 경우 G(n+1)을 규칙 테이블에서 제거한다.

작성된 규칙 테이블이 저장되면, 한국어 문장과 문법 속성열을 입력으로 하여(16) 문장의 기준억양 생성 과정에서 각 어절에 대해 문장내 위치에 따라 1차 평균 피치값 $\overline{P}_{1w_i}$을 할당한다(17). 그런 다음 비균일 단위의 억양패턴 생성 단계에서는 입력된 문법 속성열을 이용하여 왼쪽 우선 검색 방식으로 최장 일치 부분을 규칙테이블에서 찾아 해당 어절의 $\overline{P}_{2w_i}$을 읽어온다(18). 다음 어절의 피치 패턴 생성 단계에서는 신경망을 이용하여 각 음절의 평균 피치값 변화량 $\delta\overline{P}_{s_j}$을 계산한다(19). 다음 음절의 피치 패턴 생성 단계에서는 각 음절을 구성하는 음소열과 표준 피치 패턴 테이블을 이용하여 음절의 피치 패턴 Pf_j를 계산한다(20). 최종적으로 각 단계의 결과를 합하여 전체 문장의 억양을 생성하여 출력한다(21).

상기와 같이 동작하는 본 발명은 실제 음성데이타를 분석하여 억양 제어 규칙을 작성하고, 합성음의 억양 제어를 통해 합성음의 자연성과 유창성을 향상시킴으로써 통신 서비스, 사무 자동화, 교육 등의 여러 분

야에 응용할 수 있는 효과가 있다.

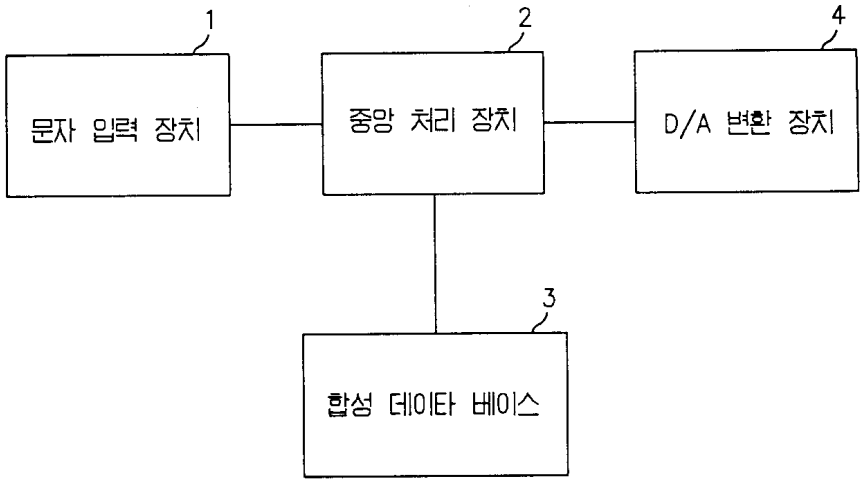
(57) 청구의 범위

청구항 1

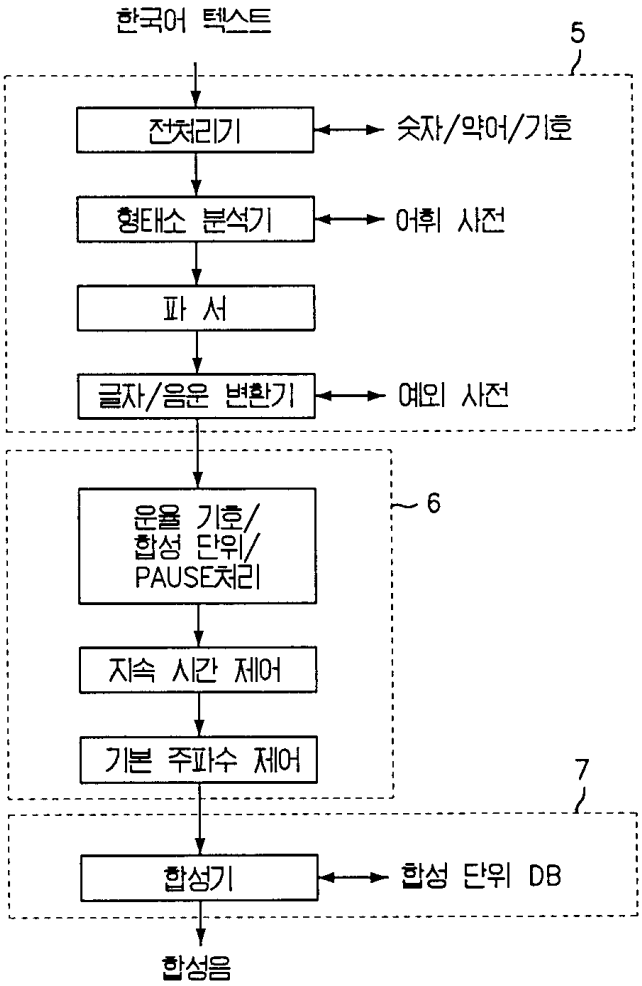
텍스트/음성변환장치에 적용되는 억양패턴 정규화와 신경망 학습을 이용한 억양생성 방법에 있어서, 합성 데이터베이스(3)로부터 음성 데이터를 읽어 음절의 피치 패턴을 정규화 및 표준화하고 어절내 각 음절의 평균 피치값에서 어절의 평균 피치값을 뺀 피치값으로 어절피치패턴을 학습하고, 문장내 각 어절의 평균 피치값으로부터 문장의 기준 억양을 추정한 후, 문장의 문맥에 따른 문법 속성열과 그에 해당되는 억양패턴테이블을 작성하는 제1단계(10 내지 15); 한국어 문장과 문법 속성열이 입력되면 문장의 기준억양 생성 과정에서 각 어절에 대해 문장내 위치에 따라 1차 평균 피치값을 할당하고, 비균일 단위의 억양패턴 생성 과정에서 입력된 문법 속성열을 이용하여 왼쪽 우선 검색 방식으로 최장 일치 부분을 억양 패턴 테이블에서 찾아 해당 어절에 2차 평균 피치값을 할당하고, 어절의 피치 패턴 생성 과정에서 신경망을 이용하여 각 음절의 평균 피치값 변화량을 계산하고, 음절의 피치 패턴 생성 과정에서는 각 음절을 구성하는 음소열과 표준 피치 패턴 테이블을 이용하여 음절의 피치 패턴을 계산하고, 각 과정의 결과를 합하여 전체 문장의 억양을 생성하여 출력하는 제2단계(16 내지 21)를 포함하는 것을 특징으로 하는 억양 생성 방법.

도면

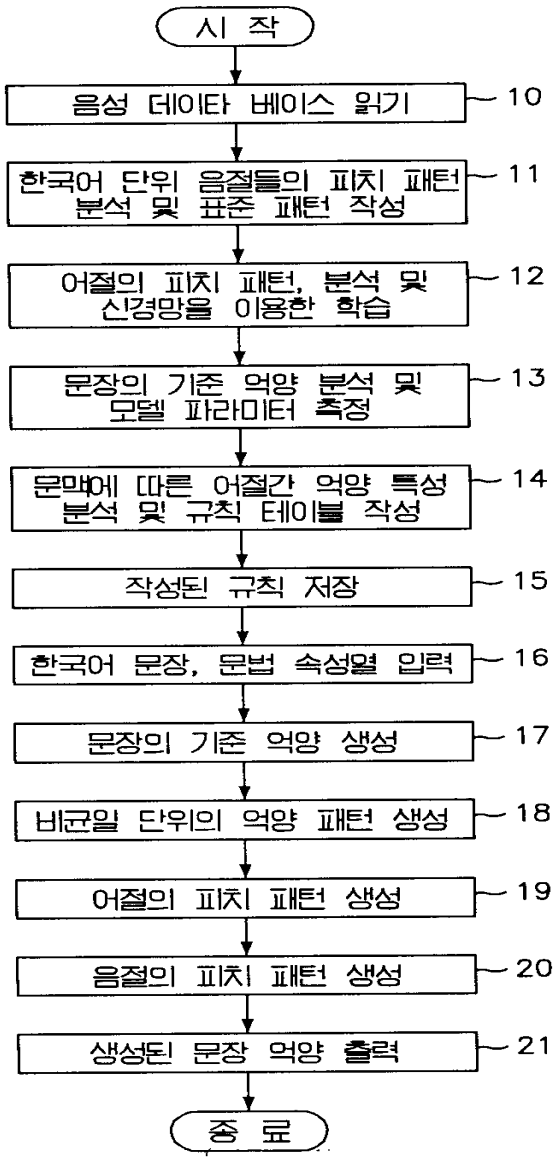
도면1



도면2


한국어 텍스트
5
전처리기
숫자/약어/기호
형태소 분석기
어휘 사전
파 서
글자/음운 변환기
예외 사전
6
운율 기호/
합성 단위/
PAUSE처리
지속 시간 제어
기본 주파수 제어
7
합성기
합성 단위 DB
합성음

도면3


시 작
음성 데이타 베이스 읽기
10
한국어 단위 음절들의 피치 패턴 분석 및 표준 패턴 작성
11
어절의 피치 패턴, 분석 및 신경망을 이용한 학습
12
문장의 기준 억양 분석 및 모델 파라미터 측정
13
문맥에 따른 어절간 억양 특성 분석 및 규칙 테이블 작성
14
작성된 규칙 저장
15
한국어 문장, 문법 속성열 입력
16
문장의 기준 억양 생성
17
비균일 단위의 억양 패턴 생성
18
어절의 피치 패턴 생성
19
음절의 피치 패턴 생성
20
생성된 문장 억양 출력
21
종 료