일한 음차 변환을 이용한 음성인식 및 합성기의 구현

이용주, 이현구, 윤재선, 양원렬, 홍광석
성균관대학교 전기전자컴퓨터공학부 휴먼컴퓨터연구실
blueyong@popsmail.com, nymp@chollian.net, sunhci@ece.skku.ac.kr,
idown@ece.skku.ac.kr, kshong@yurim.skku.ac.kr

An Implementation of Speech Recognition and Synthesis System using Japanese-Korean Phonetic Transcription

Yong-Ju Lee, Hyeon-Gu Lee, Jeh-Seon Youn, Won-Ryeol Yang, Kwang-Seok Hong
HCI Lab, Electrical & Computer Engineering, Sungkyunkwan University

요 약

본 논문에서는 일한 음차 변환을 이용한 음성인식 및 합성기를 구현하였다.

음성인식의 경우 CV, VCCV, VCV, VV, VC 단위를 사용하였다. 이와 같이 단위별로 미리 구축된 모델을 결합함으로써 음성인식 시스템을 구축하였다. 따라서 일한 음차 변환을 적용하게 되면 인식 대상이 일어단어일 경우에도 이를 한글 발음으로 변환한 후 그에 해당하는 모델을 생성함으로써 인식이 가능하다. 음성 합성기의 경우 합성에 필요한 한국어 음성 데이터 베이스를 구축하고, 입력되는 텍스트에 따라 이를 연결하여 합성음을 생성한다. 일어가 입력될 경우 일한 음차 변환 규칙을 이용하여 입력된 일어 발음을 한글로 바꾸어 준 후 입력하게 되므로 별도의 일어 합성기 없이도 합성음을 생성할 수 있다.

1. 서 론

사용자와 컴퓨터의 보다 편리한 인터페이스를 위하여 음성인식 및 합성을 이용한 사용자 인터페이스는 최근 널리 연구되고 있다. 음성의 특성상 음성인식 및 음성합성 시스템을 구성할 경우 언어 종속적일 수밖에 없다. 즉 한국어로 구성한 음성인식 및 음성합성 시스템의 경우 다른 나라의 언어에 대해서 적용이 불가능하다. 그러나 외래어 및 일어와 같은 외국어가 자주 쓰이는 실제 상황에서 순수 우리말로만 구성된 시스템만을 가지고는 실제 적용에 어려움이 많다. 그렇다고 우리말 이외에 다른 나라 언어에 대한 시스템을 일일이 구축한다는 것은 효율이지 못하다. 따라서 일어발음을 한글로 변환하는 일한 음차 변환을 이용하여 음성인식 및 음성합성 시스템을 구성할 경우 적은 비용으로 만족할 만한 성능의 효율적인 시스템을 구성할 수 있다. 본 논문의 인식시스템은 CV, VCCV, VC인식단위를 이용하여 훈련 데이터와 인식용 단어가 다른 어휘 독립 환경에서 인식대상 어휘를 사용자가 환경에 따라 자유롭게 가변 하여 사용하는 어휘 독립 시스템을 구성하였다.[1][2] 음성합성의 경우 문서-음성 합성시스템(TTS)을 구성하였다. 합성단위는 CV, VC를 결합하는 반음절 단위를 사용하였으며 합성 알고리즘은 적은 계산량에 우수한 음질을 보이는 TD-PSOLA를 이용하였다.[3]

본 논문에서는 일한 음차 변환을 이용하여 음성인식 시스템에서 일어단어를 인식하도록 구현하였으며, 음성합성 시스템에서 일어로 입력된 텍스트를 출력 가능토록 구현하였다.

2. 일한 음차 변환

일본어에 이용되는 한자어는 그 범위가 넓고 또 예외가 비교적 많기 때문에 일한 변환기에서는 히라가나와 카타카나로 쓰여진 일본어 문장으로만 제한시켰다.

일한 음차 변환 규칙과 음차 변환 시스템 구조는 다음과 같다.

2.1 일한 음차 변환 규칙

음차 변환 규칙은 국내 특정 일본어 문법책을 기준으로 모두 적용하였다. 음차 변환에서는 일본어 문장을 음차 변환시켜 다시 한국어 문장으로 출력하기 때문에, 일본사람들이 발음하는 대로 정확하게 표현하기엔 다소 무리가 있다. 왜냐하면 일본어 발음에 해당되는 한국어 문자가 없는 경우도 있기 때문이다. 음차 변환의 기준을 일본어 발음의 영어 표기로 정하였다.(か → [ka] → 카)

2.1.1 청음의 변환

청음은 오십음도에 나오는 각 음절의 가나에 탁점을 붙이지 않은 글자로서, 발음할 때 맑은 소리 즉, 성대 진동이 전혀없는 무성음의 발음이다.

2.1.2 탁음의 변환

탁음은 か, さ, た, は행의 문자 오른쪽 윗부분에 탁점(〃)을 붙여 표시하며, 모두 성대의 진동에 의해서 나는 음이며, 우리말 발성음에는 없는 음이다.

2.1.3 반탁음의 변환

반탁음은 は행의 문자 오른쪽에 탁점(°)을 붙여 표시한다.

2.1.4 요음의 변환

요음이란 い 단음글자き, し, ち, に, ひ, み, り, ぎ, じ, ぢ, び, ぴ 에 반모음인ゃ, ゅ, ょ를 작게 표기하여 두글자를 합하여 한 음절로 발음하는 것이다. 이때 작은 글자로 표기하는 ゃ, ゅ, ょ는 한글의 야, 유, 요와 같은 모음 역할을 한다.

2.1.5 ん받침음 변환

ん은 다른 글자 밑에서 받침으로만 쓰이는데, 콧소리 발음이 나며 한글과는 달리 하나의 음절의 길이를 갖고 있다. 로마자 표기로는 n으로 되어 있으나 다음에 오는 글자의 영향을 받아 (m=ㅁ), (n=ㄴ), (ng=ㅇ) 등으로 발음이 변한다.

2.1.6 촉음의 변환

촉음은 다른 글자 밑에서 받침으로만 사용하는데, 발음을 하다가 숨이 막힌 듯이 잠깐 쉬었다가 발음하고 하나의 음절 길이를 갖고 있다. 다음에 오는 글자의 영향을 받아 (k=ㄱ), (s=ㅅ), (t=ㄷ), (p=ㅂ) 받침으로 변한다.

2.1.7 외래어 변환

1) 장음을 나타내는 "-"는 그대로 변환한다
2) [f]의 표기인 "ファ", "フィ", "フ", "フェ", "フォ"는 "후아", "후이", "후", "후에", "후오"로 변환한다
3) [ti], [di]의 표기인 "ティ", "ディ"는 "티","디"로 변환한다.
4) 그외는 일반적인 변환규칙에 따른다

2 . 2 일한 음차 변환 시스템

일한 음차 변환 시스템의 구성은 그림 2.1과 같다.

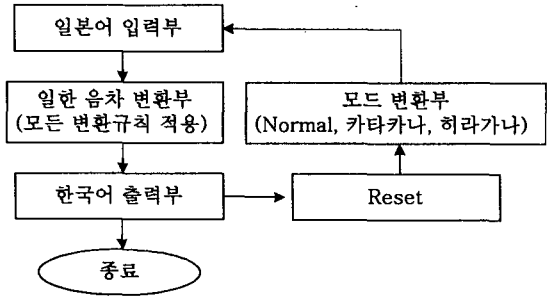

그림 2.1 일한 음차 변환 시스템구성

2.2.1 일본어 입력부

입력부에는 일본어 히라가나, 카타카나 외에도 숫자, 영문, 특수기호 등 모두 하나의 문장안에서 혼용하여 입력할수 있도록 구성하였다. 다만, 변환처리는 히라가나와 카타카나의 경우만 골라서 처리하였다.

2.2.2 일한 음차 변환부

일본어의 표기 특징은 문장의 왼쪽에서 오른쪽으로 순차적으로 진행되며, 하나의 문자를 기준으로 바로 앞과 뒤의 문자만 비교하면 곧바로 발음상의 문자로 정의될 수 있기 때문에 영어에 비해 다소 처리가 간단하다. 2.1에서 설명한 여러 규칙을 토대로 모두 적용하였다.

2.2.3 한국어 출력부

일한 음차 변환부를 통해 넘어온 한국어 문자열을 출력 윈도우에 출력한다. 이때, 일본어에서 외래어에 주로 이용하는 카타카나의 경우 " - "음의 경우와 히라가나의 " ん "음의 위치에는 이후의 이용에 대비하여 출력문장에 " - "를 삽입하였다.

3 . 음성 인식 및 합성 시스템

본 논문에서 사용한 훈련 데이터는 성명데이터 1145개를 남성화자 62명, 단음절 데이터 521개를 남성화자 53명, PBW데이터를 포함한 데이터 1001개를 남성화자 53명이 사무실 환경에서 1회씩 발성한 음성 신호를 16bit, 11.025kHz로 샘플링하여 저장하고, 이로부터 CV단위 383개, VCCV단위 2491개, VC단위 168개의 인식 단위를 분리하고 Reference Model을 구성하였다.[2]

본 논문에서는 HMM을 사용한 가변 어휘 인식 시스템을 구성하였다. 먼저 분할된 데이터로부터 Reference Model을 구성하고, 인식 목록 단어를 CV, VCCV, VC모델로 연결하여 단어 모델을 구성한다. 인식 단계에서는 입력 음성으로부터 Mel cepstrum 특징 파라미터를 추출하고, 입력 받은 데이터로부터 각 단어에 해당하는 확률값들을 비교하여 인식단어를 결정한다. 단어 목록구성 시 일어단어가 일한 음차 변환 시스템을 이용하여 한글로 변환한 후, 인식하도록 구성하였다.

본 논문에서는 입력 문장을 합성음으로 변환하는 문자-음성 변환시스템(TTS)을 음성 데이터 구축하는 부분과 실제 음성합성을 하는 두 부분으로 나누어 시스템을 구성하였고, 실제 음성합성을 하는 부분은 다시 전처리에 해당하는 언어처리 부분과 실제 음성 파형을 합성하는 합성 처리부분으로 나눌 수 있다.

언어처리부의 경우 숫자, 일어, 기호에 대한 처리 와 음운변동처리를 하였다. 숫자의 경우 숫자 뒤에 따라오는 연결어에 따라서 '한','둘'...이나 '일','이'...로 변환하게 된다. 기호의 경우 기호와 한글발음을 일대일로 대응하여 변환한다.

음운 변동의 경우 문자로 표시된 입력 문장을 소리나는 발음대로 바꾸어 주는 과정인데 한국 표준 음운 규칙에 따라 처리하였으며, 이런 처리를 거친 입력 문장은 문자 기호열로 변환되어 합성단계의 입력으로 주어지게 된다.

합성 방식은 TD-PSOLA를 이용하였고, 합성 단위는 반음절 단위로 하였다.[3] 언어처리 부분에서 현재 합성해야할 문장에 해당하는 기호열이 넘어오면 우선 이 기호열을 현재 메모리에 로드된 음성 데이터베이스 중에 적절한 인덱스와 매핑 시키고 외

부입력 파라미터 및 구문 분석에 의해실제 합성 시 적용해야 할 파라미터 값들을 결정한다. 파라미터들이 결정되면 앞에서 결정된 데이터 베이스의 인덱스 열과 각 인덱스 별로 결정된 파라미터에 의해 실제 파형을 합성하게 된다.[4]

4 . 실험 및 결과

일한 음차 변환에서 사용되는 데이터는 일본어 발음의 청음, 탁음, 반탁음, 요음, ん받침음, 촉음과 외래어 변환의 히라가나와 카타카나의 각 20단어씩 임의로 선택하여 테스트한 결과 일한 음차 변환률이 한자를 제외한 것을 빼고 모두 100%로 나타났다.

음성 인식 방법은 먼저 Mel cepstrum 16차 특징 파라미터를 추출한 후, K-means 알고리즘을 이용하여 구성한 VQ 코드북을 통과하여 벡터 인덱스의 sequence열을 얻는 후, HMM 인식 알고리즘을 사용하여 인식한다.

인식에 사용되는 데이터는 임의로 선택하여 한글 및 일어 단어 각 20개를 인식 목록으로 작성하였으며, 목록은 다음과같다.

あさい, ちいいわたし, ありさま, ひちのひと, えもの, くすりや, こうこく, おじいさん, がいぎん, ごしんぶ,とっぱん, えんきょり, かきゅう, ぎょうしゃ, ようじゅつおんな, こうねんき, だんぼう, たんぼぼ, がっさん, すっぽん, せっだい, ちっこう, ジョーク, ターミナル, ネーチュア

과일, 결과, 나루터, 날개, 담배, 다리, 로마, 마음, 벌레, 바위, 사진, 사이다, 여름, 아이, 정신, 차표, 카드, 토시, 펀치, 헌신

인식 방법은 한글 단어 및 일어 단어를 화자 독립 3명과 화자 종속 2명으로 5명이 발성하여 인식 성능을 평가하였으며, 인식 결과는 표 4.1과 같다.

표 4.1 임의의 데이터 인식결과

화자		한글	일어
화자종속	화자1	19/20	19/20
	화자2	18/20	17/20
화자독립	화자3	19/20	18/20
	화자4	17/20	17/20
	화자5	18/20	18/20
	평균	91%	89%

한글과 일어의 성능 비교에서 거의 차이가 없음을 알 수 있었으며, 인식의 Reject률은 한글 1%, 일어 4%가 있었다. 일어의 단어 발음이 한글 발음과 매우 유사하여 성능에 거의 차이가 없음을 알 수 있었다.

합성의 경우 여성화자 1인이 사무실환경에서 1회씩 발성한 음절 단위의 데이터를 가공하여 데이터 베이스를 구축하였고, 이를 이용하여 반음절 단위 음성합성시스템을 구성하였다.

실험에 사용되는 데이터는 인식 실험에서 사용했던 데이터와 동일한 각 20개의 데이터를 사용하였고 합성음의 명료성과 자연성을 MOS방법(1-5범위)으로 청취자 5인에게 평가하였으며 그 결과는 표4.2와 같다.

표 4.2 합성음의 명료성과 자연성
평가(한글/일어)

청취자	명료성	자연성
청취자1	4.25 / 4.2	4.0 / 3.9
청취자2	3.35 / 3.4	3.5 / 3.25
청취자3	3.25 / 3.25	3.4 / 3.2
청취자4	3.35 / 3.1	3.05 / 3.0
청취자5	2.95 / 3.35	3.1 / 3.25
평균	3.43 / 3.46	3.41 / 3.32

명료성의 경우 비교적 좋은 결과를 나타내었으나 억양 및 강세처리가 우리말과 일어가 조금 다른 이유로 자연성의 경우 비교적 성능이 떨어짐을 알 수 있었다.

5 . 결론

본 논문에서는 일한 음차 변환을 이용하여 음성인식 및 음성합성 시스템을 구현하였다. 일한 음차 변환을 이용하게 된다면 별도의 음성인식 및 음성 합성 시스템 없이도 일어 인식 및 합성에 대해 만족할 만한 성능을 보임을 알 수 있었다.

6 . 참고 문헌

[1] 김태환, 박순철, "문맥종속 반음소 단위 모델을 이용한 자동 음소분할 및 레이블링 시스템의 구현," 한국음향학회, 제 17권 2호, 1998.

[2] 윤재선, 홍광석, "무제한 어휘 독립 단어 인식 시스템의 구현", 제17회 음성 통신 및 신호처리 학술대회(KSCSP2000 17권 1호), pp.145-148, 2000

[3] Jon R. W. Yi, "Time-Domain PSOLA Concatenative Speech Synthesis Using Diphones"

[4] 양원렬, 윤재선, 홍광석, "영한 음차 변환을 이용한 무제한 음성인식 및 합성기의 구현", 한국신호처리·시스템학회 하계 종합 학술대회 논문집 1권 1호, pp181-184, 2000.