다이폰을 이용한 한국어 문자-음성 변환 시스템의 설계 및 구현

°정준구, 강명광, 정유현, 이강성, 김순협
광운대학교 컴퓨터공학과

Design and Implementation of Korean Text-to-Speech System

Jeong Jun-Koo, Kang Myeong-Kwang, Jeong You-Hyeon, Lee Kang-Sung, Kim Soon-Hyeob
Kwang-Woon Univ. Dep't of Computer Engineering

Abstract

This paper is a study on the design and implementation of the Korean Text-to-Speech system. In this paper, parameter synthesis method is chosen for speech synthesis method and PARCOR(PARtial autoCORrelation) coeffient, one of the LPC analysis, is used as acoustic parameter. We use a diphone as synthesis unit, it include a basic naturalness of human speech. Diphone DB is consisted of 1228 PCM files. LPC synthesis method has defect that decline clearness of synthesis speech, during synthesizing unvoiced sound. In this paper, we improve clearness of synthesized speech, using residual signal as excitation signal of unvoiced sound. Besides, to improve a naturalness, we control the prosody of synthesized speech through controlling the energy and pitch pattern. Synthesis system is implemented at PC/486 and use a 70Hz-4.5KHz band pass filter for speech input/output, amplifier and TMS320c30 DSP board.

1. 서 론

본 연구에서는 맨-머신 인터페이스(Man-Machine Interface)의 핵심기술인 음성 신호처리 중 무제한 음성합성, 즉 문자-음성 변환 시스템에 대한 연구를 하였다. 최근의 음성합성에 관한 연구는 주로 규칙에 의한 무제한 음성합성으로서 음질의 자연성(naturalness)과 명료도(clearness)의 향상에 초점을 두고 있으며, 기본적인 자연성을 확보하기 위해 보다 복잡하고 세분화된 음성 데이타베이스를 사용하고 있다.[1] 또한 막대한 음성데이타의 크기를 줄이기 위해서 파라메터를 이용한 합성을 하고 있다. 음성합성은 무엇보다도 음성 데이타베이스의 확립과 파라메터의 추출방법 그리고 각 파라메터를 통한 운율제어에 핵심이 있다. 본 논문에 사용된 음성데이타는 한국전자통신연구소 음성응용연구실에서 제작한 1228개의 다이폰 화일로 구성되어 있다.

2. 문자-음성 변환 시스템

문자-음성 변환 시스템이란 텍스트로 입력되는 임의의 문자열을 합성된 음성으로 출력해 주는 장치이다. 본 절에서는 문자-음성 변환 시스템의 대략적인 구조를 소개한다.

문자-음성 변환을 위해서는 그림 1. 과 같은 여러 단계의 처리 과정을 거쳐야 한다.[2] 각 단계에서의 처리 내용을 보면 다음과 같다.

1) 언어 처리부: 입력된 텍스트를 단어단위로 구분하여 구문 분석, 의미 분석, 담화 분석등을 하여 언어 정보를 생성한다. 이 때 생성된 언어 정보들은 효과적인 운율 생성을 위해 음성의 음향적 특징의 조작에 사용된다. 또한 정서법으로 입력된 문장을 소리나는 대로 변환시켜주는 기능도 한다.

2) 음운 및 운율 처리부: 언어처리부로부터 입력된 언어 정보들을 이용하여 적절한 음운 및 운율 기호열을 생성시킨다. 음운 기호열이란 합성에 필요한 합성 단위의 순서를, 운율 기호열은 운율 조절에 필요한 기본주파수 패턴, 음소 길이, 과치 변화 패턴등을 말한다.

3) 합성 및 운율 파라메터 생성: 입력된 음운 기호열로부터 파라메터 사전에 저장되어 있는 해당 파라메터들을 가져온다. 또한 운율 기호열의 정보로부터 실제 운율 파라메터(음소길이, 기본주파수, 유/무성 정보, 에너지 등)을 생성한다.

4) 음성파형 합성부: 입력된 합성 파라메터와 운율 파라메터로써 합성 필터를 구동시켜 합성음을 출력시킨다.

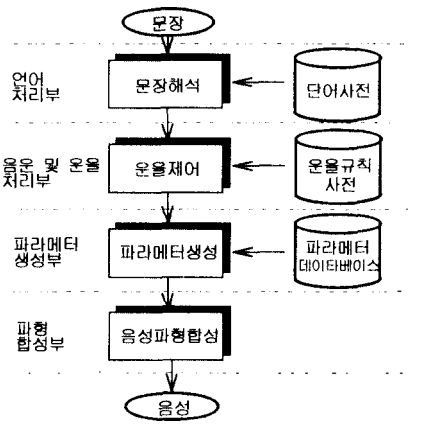

그림 1. 문자-음성 변환 시스템의 일반적 구성

3. 분석 및 합성부의 설계

3.1 유, 무성음 및 묵음구간 결정

PCM 데이타로 저장되어있는 1228개의 다이폰 화일들에 대해 먼저 프레임단위로 유, 무성음 및 묵음 판별을 한다. 본 연구에서는 유성음과 무성음의 합성방식이 다르기 때문에 유, 무성음 정보가 합성음의 음질에 매우 큰 영향을 미친다. 유, 무성음 및 묵음구간에 대한 자동검출 알고리즘은 몇가지 소개된 것이 있지만, 그 성능이 합성 시스템의 구현에 이용하기에는 적합하지 않다. [3] 유, 무성음의 판별에 가장 좋은 방법은 사람의 목측(目測)을 통해 음성 파형을 보면서 판별을 해주는 것으로 이를 위해 그림 2.와 같은 도구를 만들었다.

그림 2.는 다이폰 데이타 21번 /바/ 의 음성 파형으로 47프레임으로 구성되며 각 프레임은 'S'(묵음), 'U'(무성음), 'V'(유성음)중 하나로 결정된다. 이 정보는 저장되어 분석 및 합성시 이용된다.

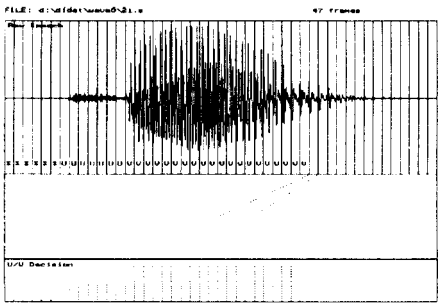

그림 2. 유, 무성음 및 묵음 판별

3.2 합성 파라메터 데이타베이스의 구현

PCM 데이타로 구성된 다이폰 화일에 대해 그림 3.과 같은 방법을 통해 합성파라메터 데이타베이스를 만든다. 유성음 합성을 위한 파라메터로는 PARCOR 계수(10차), 피치, 진폭을 그리고 무성음 합성을 위한 파라메터로 LPC 분석 중 구해지는 잔차신호와 PARCOR 계수(10차)를 구하였다.

잔차신호가 합성필터에 입력되면 분석대상이었던 원래의 음성신호가 그대로 복원된다. LPC계열 합성음의 큰 문제점 중의 하나가 무성음의 합성이 잘 안되는 것인데 잔차신호를 무성음의 여기신호로 사용함에 따라 합성음의 명료도 향상에 큰 효과를 낼 수 있었다. 그러나 파라메터의 저장공간이 커지는 단점이 있다. 본 논문에서는 Levinson-Durbin's 알고리즘을 이용하여 PARCOR 계수를 구하였다. 피치검출에는 3-level center clipping method를 이용하였으며 진폭은 잔차신호의 에너지를 이용하였다. [4] [5]

3.3 Lattice 합성필터의 설계

그림 4.는 합성부의 구조로써 각 합성 파라메터가 어떻게 적용되는가와 Lattice 합성필터의 회로를 나타내고 있다. [6]

PARCOR 음성합성 방식은 입력 여기신호에 순차적으로 correlation을 더해주는 과정으로 평탄한 주파수 스펙트럼을 갖는 임펄스나 잔차신호 입력에 포만트구조를 만들어가는 것이라고 해석된다.

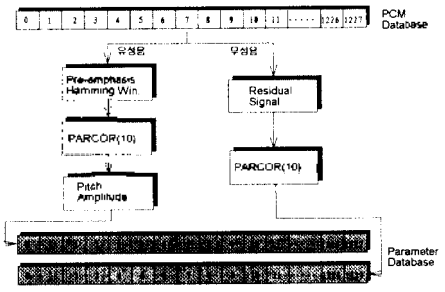

그림 3. 합성 파라메터의 추출과정

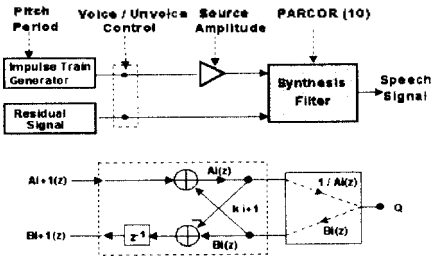

그림 4. 합성부의 구조

그림 5.는 위에서 기술한 분석 및 합성 방법에 의해 합성된 예로써 115번 다이폰 /추/의 원파형 및 합성파형과 각 프레임에서의 에너지와 모음구간의 피치를 나타내고 있다.

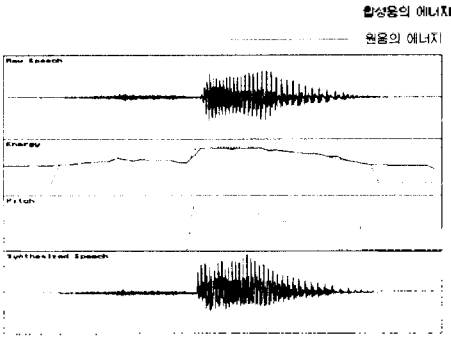

그림 5 115번 다이폰 /추/의 합성 예

위의 그림을 통해 원음과 합성음의 에너지가 거의 비슷하다는 것을 알 수 있었고 청취결과 원음과 거의 동일한 합성음을 얻을 수 있었다. 이를 통해 앞에서 기술한 합성부가 문자-음성 변환 시스템에 적용 가능하다는 것을 알 수 있다.

4. 합성 시스템의 구현

그림 6. 에서는 본 논문에서 구현된 문자-음성 변환 시스템의 구성을 보여준다. 각 모듈의 기능을 보면 다음과 같다.

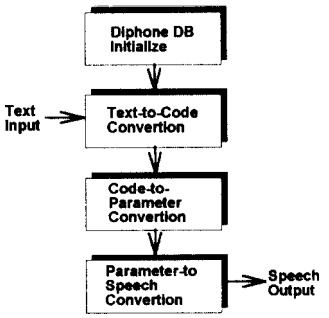

그림 6. 구현된 문자-음성 변환 시스템의 구성도

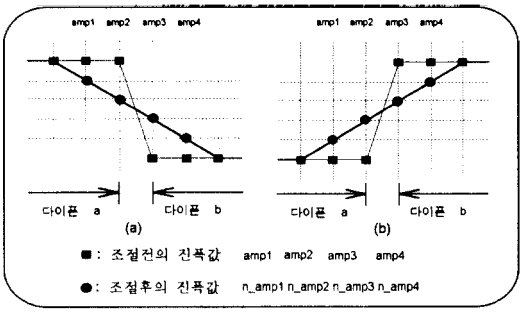

그림 7. 인접 다이폰간의 에너지 조절

4.1 Text-to-Code 변환부

입력된 2바이트 조합형 한글을 음소단위로 분리한다. 이때 각음소의 자음, 단모음, 이중모음여부와 초성, 중성, 종성여부가 결정된다. 이 정보에 따라 음소열로 부터 인접된 음소간의 결합규칙을 찾아 각 규칙에 따른 해당 다이폰코드(조합형 한글)열을 생성해 낸다. 생성된 다이폰코드들에 해당하는 다이폰 번호열과 세그멘테이션 정보열이 출력된다.

4.2 Code-to-Parameter 변환부

TTC에서 생성된 다이폰 열과 세그멘테이션 정보열이 입력으로 들어온다. 음소길이에 맞게 해당 다이폰 파라메터 화일로부터 파라메터를 가지고 와 메모리에 저장한다. 이 부분에서 합성음의 운율조절을 위한 에너지 평활화 작업과 피치제어가 이루어진다. 운율조절에 관한 부분은 5장에서 설명한다.

4.3 Parameter-to-Speech 변환부

합성필터 부분으로 Lattice 필터를 이용하였다. 운율제어가 끝난 메모리내의 합성 파라메터 열이 입력으로 들어와 합성파형이 생성된다. 합성음은 PCM데이타로 저장되고 A/D 변환되어 스피커를 통해 출력된다.

5. 운율처리부의 구현

자연성 향상을 위한 합성음의 운율조절은 문자-변환 시스템에 있어서 필수적이다. 합성음의 운율조절을 위해 사용되는 파라메터로는 에너지 패턴, 기본주파수의 변화 패턴등이 있다.

5.1 다이폰 인접구간의 에너지 평활화

에너지는 음성의 단어나 문장내에서 강조하려고 하는 부분에서 크게 나타나는 등 문장내의 운율성분과 관계가 있으나 합성음의 명료도와 자연성에는 그리 큰 영향을 주지는 않는다. [7] 그러나 다이폰간의 연결 부분에서의 에너지 차이는 합성음에 큰 영향을 미친다. 이러한 음질저하를 줄이기 위하여 모음구간에 대하여 그림 7.과 같이 합성필터의 입력 진폭값을 조절하여 다이폰간의 에너지 평활화를 하였다.

(a)는 전다이폰의 에너지가 후 다이폰의 에너지 보다 큰 경우이고 (b)는 후 다이폰의 에너지가 전 다이폰의 에너지보다 큰 경우이다.

5.2 유성음 구간의 에너지 평활화

본 연구에서는 합성음의 자연성 향상을 위하여 합성음의 전체 유성음 구간에 대한 에너지 평활화를 하였다. 본 논문에서 제안한 에너지 평활화는 이동평균(Moving Average) 개념을 이용한 것으로 인접한 유성음 다섯 프레임의 평균 진폭값이 중간 프레임의 진폭값으로 갱신된다. 그림 8.은 이러한 과정을 보여준다.

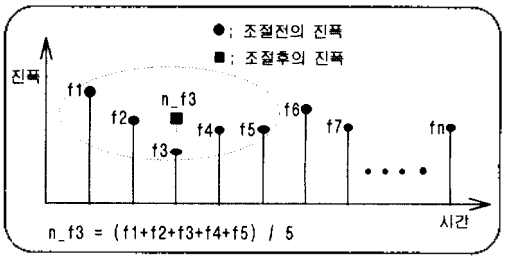

그림 8. 유성음 구간의 에너지 평활화

유성음 전 구간에 대한 에너지 평활화 결과 전체적으로 에너지가 상향 평활화 되었다.

5.3 기본주파수의 제어

기본주파수는 합성음의 억양조절에 관한 파라메터이다. 억양은 말의 뜻을 구별시키는 변별적 기능이 있기 때문에 억양의 조절은 음성합성에 있어 매우 중요하다. [8] 그러나 우리말에 대한 기본주파수 제어규칙은 연구의 어려움으로 아직은 정립되지 않은 상태이다. 본 논문에서는 다음과 같은 간단한 규칙을 사용하였다.

1. 평서문에서는 하강패턴을 생성한다.
2. 의문문에서는 하강-상승패턴을 생성한다.
3. 원음의 기본주파수 값에서 일정 비율만큼의 Base 피치값을 설정하여 그이하의 피치값은 생성하지 않는다.

1, 2번은 일반적으로 조사되어 있는 규칙이고 3번은 본 논문에서 제안한 사항으로 기본주파수 값이 원음에 비해 너무 낮아지면 떨리는 소리가 합성되기 때문에 이를 막기 위하여 실험적으로 제안한 것이다. 실험에 의하면 원음의 피치보다 약 90% 이하의 피치값이 적용되었을 때 음질이 떨어졌고 85% 이하의 피치값이 적용되었을 때 떨리는 소리가 합성되었다.

6. 합성 결과

앞에서 설명된 각 규칙들을 이용하여 한국어 문자-음성 변환 시스템을 컴퓨터상에서 소프트웨어로 구현하였다. 문자-음성 변환 시스템을 통해 합성한 음성 /안녕하세요/의 예를 그림 9에서 보여주고 있다. 문장의 시작과 끝 부분에 5 프레임 씩 묵음구간을 삽입하였다. (a)는 다이폰 PCM 데이타를 파라메터 합성이 아닌 파형 합성법으로 접속한 파형이고 (b)는 PARCOR 합성법에 의해 합성된 파형이다. (c)는 에너지 윤곽선으로 앞에서 설명한 다이폰 인접구간의 에너지 평활화 규칙을 적용한 결과이다. (d)는 유성음 구간의 피치 패턴을 보여주고 있다. 피치의 경우 역시 하강 패턴을 형성하도록 조절하여 주었다. (e)는 유성음 구간의 에너지 평활화 후의 합성 파형이다. 그림 10. 역시 입력문장 /광운대학교/의 음성합성 예이다.

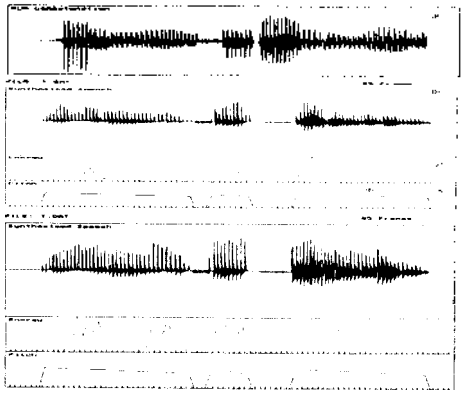

그림 10 합성음성 /안녕하세요/

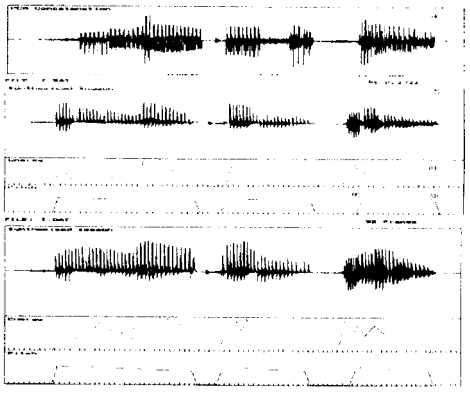

그림 10 합성음성 /광운대학교/

7. 결 론

본 논문에서는 한국어 문자-음성 변환 시스템을 구현하였다. 합성음의 명료도 향상을 위하여 모든 다이폰 데이타에 대해 유, 무성, 묵음 판정을 목측(目測)으로 하였으며 유, 무성 여부에 따라 분석 및 합성 방식을 다르게 하였다. 또한 간단한 운율제어 규칙을 적용하여 자연성을 어느정도 향상시켰다. 특히 다이폰 인접 구간의 에너지 불균형 문제와 유성음 구간의 에너지 평활화로 합성음의 명료도와 자연성을 향상시킬 수 있었다. 무성음의 음질은 잔차신호를 필터의 여기신호로 사용하므로써 어느정도 향상시킬 수 있었으나 운율조절에 있어서는 연구에 많은 어려움이 있었다. 위의 합성실험 결과 명료도 및 이해성 측면에서는 비교적 좋은 합성음을 얻을 수 있었으나 자연성의 측면에서는 부족한 부분이 많았다. 자연성을 향상시키기 위해서는 우리말에 일반적으로 적용할 수 있는 운율제어 규칙들을 찾아내야 한다. 무엇보다도 먼저 음성합성을 위해 선행되어야 할 연구는 우리말의 강세와 억양에 대한 것이다. 현재의 음성합성에서는 단일 파라메터로써 강세와 억양을 조절하는데 강세는 합성파라메터에서 에너지와 기본주파수 그리고 지속시간등 모든 요소의 복합된 형태로 나타난다. 억양은 기본주파수로 보통 제어하는데 보다 고품질의 합성음을 위해 위해서는 에너지도 고려해야할 것이다.

음색조절을 통한 개인성의 확보도 앞으로 연구해야 할 분야이다. 개인성에 관한 연구는 음성합성 뿐만 아니라 화자인식(Speaker Recognition) 분야에도 영향을 미칠 수 있을 것이다.

참고 문헌

[1] J.Allen, "Synthesis of Speech from Unrestricted Text", Proc. IEEE, vol.64, 1976

[2] 이용주, 정유현, "사동통역전화 시스템구성을 위한 검토", 대한전자공학회 통신,교환 연구회 합동발표논문집, 1987.11

[3] 김명환, 김순협, "유, 무성음 및 묵음 식별에 관한 연구", 광운대학교 전자계산기공학과 석사학위 논문, 1985

[4] S. Saito, K Nagata, Fundamentals of Speech Signal Processing Academic Press, 1985

[5] L. R. Rabiner, R. W. Schafer, "Digital Processing of Speech Signal", Presentice-Hall, 1978

[6] S. Furui, "Digital Signal Processing, Synthesis and Recognition", MARCEL DEKKER, pp. 168 - 169, 1992

[7] 이 용주, "반음절 단위, LSP 방식에 의한 한국어 음성의 규칙합성에 관한 연구", 고려대 박사학위 논문, 1992

[8] 양질의 음성합성을 위한 최적의 합성단위 추출에 관한 연구, 한국전자통신연구소, 1993

[9] D. Klatt, "Review of Text-to-Speech Conversion for English", JASA. 82(3), pp. 737 - 793, 1987

[10] 이기문외, 국어음운론, 학연사, 1991