

# 차량항법 시스템을 위한 소형 음성합성 엔진

## Speech synthesis engine for car navigation systems

°김 경 하\*, 서 흥 석\*, 박 찬 식\*\*, 성 태 경\*\*\*, 이 상 정\*\*\*\*

\* 충남대학교 전자공학과 (Tel:042-825-3991; Fax:042-823-4494; E-mail:kim9689@hanmail.net)  
\*\* 충북대학교 전기전자공학부 (Tel:043-261-3259; Fax:043-268-2386; E-mail:chansp@cbucc.chungbuk.ac.kr)  
\*\*\* 충남대학교 정보통신공학부 (Tel:042-821-5660; Fax: 042-823-4494; E-mail:tkssaint@elpwer1.chungnam.ac.kr)  
\*\*\*\* 충남대학교 정보통신공학부 (Tel:042-825-3991; Fax:042-823-4494; E-mail:sjlee@csllab.chungnam.ac.kr)

**Abstract :** This paper proposes a modified TD-PSOLA algorithm for Korean speech synthesis. A WSS(Weighted score search) algorithm is proposed for pitch detection and speech synthesis engine is designed using 46 phones database.

**Keywords :** Speech synthesis, WSS, Pitch detection, TD-PSOLA.

### 1. 서론

음성합성(Speech synthesis)과 TTS(Text to speech)분야의 기술 수준은 최근에 PC 또는 전용하드웨어에서 사람의 목소리와 매우 유사한 음성을 구현할 수 있는 수준에 와 있으며, 이러한 기술은 여러 응용분야에서 요구되고 사용되고 있다. Mobile-PC, 장애자를 위한 Text reader, 자동 응답 전화(ARS), 인터넷 응용 프로그램, Multimedia-Car 등을 대표적인 응용분야로 들 수 있으며, 특히 Handheld PC와 차량항법장치 등에서는 음성을 이용한 사용자 인터페이스가 가장 경제적이고 효율적인 방법이라 할 수 있다.

이러한 기술은 이미 선진국을 중심으로 70년대부터 연구가 시작되어 80년 대부터 상용화가 시작되었으며, 컴퓨터와 인간과의 interface 측면에서 음성인식과 함께 많은 연구가 진행되고 있다.

음성신호의 합성방식은 크게 음성신호 발생구조의 특징을 근간으로 하는 생성원 처리 합성방식(Source coding)과 음성파형을 직접 부호화하는 파형 처리 합성방식(Waveform coding)으로 나눌 수 있다.

선형 예측 부호화(LPC:Linear predictive coding)방식은 생성원 처리에 의한 음성 합성 방식의 일종으로 음성 파형은 일정한 구간 동안 상관도가 높다는 점을 이용하여 과거의 음성신호에 의해 미래의 신호예측이 가능하다는 성질을 이용한 방법이다[3]. 이 방법은 음성처리 분야에 중요한 비중을 차지 하고 있다. 이밖에 포먼트 주파수를 이용하는 합성방법, 발생기관을 직접 모델링하여 합성하는 방법등이 생성원 처리방식에 속한다. 생성원 처리 합성방식은 운율조절이나 음의 장단 조절이 비교적 쉬워 파형 처리 합성방식에 비해 합성음의 자연성이 우수하나 기계적인 합성음이 출력되어 명료성이 떨어진다[4].

파형 처리 합성 방식은 음소, 음절, 반음절 또는 단어를 미리 녹음하고, 필요시 연결하여 합성하는 방법으로 생성원 처리 합성방식에 비해, 음질과 명료성이 우수하나 음의 길이 높낮이 조절이 어려워 합성음의 자연성 확보가 어렵다. 그러나, 파형처리 합성방식에서 어려웠던 음성신호의 운율제어를 간단한 방법으로 해결 함으로서 명료성이 보장된 상태에서 자연성을 높인 TD-PSOLA(Time-Domain Pitch Synchronous Overlap and Add)방법이 개발되어 최근 각광 받고 있다.

최근의 연구 동향은 시스템의 고성능화와 대용량화에 따라

TD-PSOLA 알고리즘을 기반으로 하며, 대용량 음성 데이터 베이스(코퍼스 : 100- 500Mbyte)를 사용하여 실제 인간의 발성음과 유사한 수준의 합성기 개발이 주류를 이루고 있다. 이러한 합성기는 인터넷, 자동 응답기, 자동 예약 시스템등의 응용분야를 위한 것이고, 대형 시스템을 요구한다.

본 논문에서는 차량 항법 장치와 소형 이동환경 등에 적합한 음소 또는 반음절 을 합성단위로 사용할 수 있는 음성합성엔진 개발과 소용량 시스템에 적합하며 계산량을 줄이고 더욱 안정적인 pitch detection알고리즘을 갖는 TD-PSOLA를 제안하였다.

### 2. 음성 합성 엔진 개발

#### 2.1. 피치검출과 TD-PSOLA 알고리즘

Pitch는 유성음 신호에서 가장크기가 큰 신호에서 다음 크기가 큰 신호 또는 유사한 모양이 반복되는 주기로 정의 할 수 있으며 pitch의 역수가 유성음 신호의 기본주파수라 할 수 있다. 근본적으로 폐에서 방출되는 공기가 성문의 떨림에 의해 나타나게 된다[1]. 음성신호의 특징으로 운율, 소리의 장단, 크기 등이 있는데 이 중 사람의 귀가 가장 민감하게 느끼는 것은 소리의 운율이다. 운율은 음성신호의 무성음과 유성음 중 유성음의 pitch의 변화로 나타난다.

생성원 처리 합성방식은 기본 pitch를 가지는 impulse train 신호를 성도 또는 음성신호를 모델링한 디지털 필터에 통과시켜 출력음을 얻는 방식이므로 입력 impulse train의 주기를 조절함으로써 쉽게 출력음의 운율을 조절할 수 있다. 그러나 파형처리 합성방식은 미리 저장된 음성신호를 그대로 이용하는 방법이므로 유성음의 짧은 구간의 주파수 특징을 유지하면서 pitch를 변경하기가 쉽지 않았다. 최근 개발된 TD-PSOLA방법은 유성음의 pitch의 위치를 계산한 후 pitch단위로 음성을 분해한 후 분해된 pitch의 위치를 원래의 순서에 따르되 재배치 함으로써 유성음의 pitch를 조절하여 운율 효과를 낼 수 있다[4].

Pitch를 찾는 방법으로는 크게 시간 영역 계산법, 주파수 영역 계산법, 보조신호를 이용하는 방법으로 나눌 수 있다[2].

시간 영역 계산법의 경우 신호의 크기와 threshold값을 이용하는 방법으로 비교적 간단히 피치의 위치를 찾을 수 있다.