

## 소 특 집

# 음성합성 기술 개요 및 향후 과제

이 정 철

울산대학교 컴퓨터·정보통신공학부

### I. 서 론

텍스트/음성변환기(Text-to-Speech: TTS)는 문자를 음성으로 변환하는 시스템으로서 컴퓨터가 사용자인 인간에게 다양한 형태의 정보를 음성으로 제공할 수 있는 음성합성 방법이다. 사용자는 TTS를 이용하여 대화상대로부터 제공되는 텍스트 정보 뿐만 아니라 전자우편, 팩스의 내용, 은행잔고 안내, 증권정보 안내, 경기의 점수 안내, 비행시간 안내 등의 표준 데이터베이스와 무제한의 텍스트(의학적, 법률서적, 백과사전) 등을 음성으로 낭독해주는 서비스를 받을 수 있다. 이와 같이 다양한 응용분야에 활용될 수 있는 TTS 기술이 사용자에게 고품질의 음성합성 서비스를 제공하기 위해서는 합성음의 명료도와 자연성이 높아야 하고, 발성속도 및 적절한 강조가 이루어지게 유창해야 하며, H/W, S/W 면에서 쉽게 구현되어야 한다. 이를 위한 연구는 최근 40-50년에 걸쳐 음성언어학, 음향학, 생리학, 심리음향학, 음성신호처리, 컴퓨터공학 등 다양한 분야에서 활발히 진행되고 있으며 괄목할 결과들을 발표하고 있다. 특히 몇몇 연구소/업체들에서 상용화 제품들을 출시하여 일부 분야에서는 일반인을 대상으로 실제 서비스되고 있는 상황이며 앞으로의 파급효과도 매우 높다고 예상할 수 있다.

음성합성 과정은 모자이크에 비유하여 쉽게 설명할 수 있다. 먼저 여러 종류의 그림(음성데이터)들을 수집하여 원하는 크기(합성단위)로 나누고, 나누어진 조각들을 그 색상, 크기, 모양(언어, 음성, 음향적 특성)별로 분리하여 DB(합성단위

음편 DB)로 구성한다. 이제 새로운 밑그림(합성하고자 하는 문장)을 그리고 이를 가상으로 분할한 뒤(합성단위), 각 조각들에 적합한 색상, 크기, 모양(언어, 음성, 운율 추정정보)의 조각들을 앞에서 작성한 DB에서 가져와 짜맞춤으로써 새로운 그림을 완성(음성합성)한다. 따라서 중복되지 않으면서 특성이 다양하고 풍부한 음편들을 수집하는 방법, 이 음편들의 특징정보를 체계적 DB로 구성하는 방법, 밑그림을 분석하여 최적의 음편들을 선정하기 위한 정보들을 추정하는 방법, 실시간으로 후보 음편들에서 최적의 음편을 선정하고 집합시켜 합성음을 생성하는 방법은 합성음의 명료도, 자연성과 매우 밀접한 관련이 있다.

음성합성 연구는 먼저 인간의 간단한 음성신호를 복제할 수 있는 단순한 합성기의 개발에서 출발하였으며 이후 사람이 인지할 수 있는 무제한 음성합성을 위하여 체계적인 음성/언어 분석, 합성단위의 선정, 합성방식연구로 이어졌다. 그리고 실제 음성과 같은 자연성을 합성음에 구현하기 위한 운율처리 연구와 완전 TTS를 구현하기 위한 필수조건으로서 입력 문장으로부터 음소와 운율정보를 추정, 구현하는 연구로 발전하였다. 일반적으로 분석과정에서는 음성합성의 대상인 사람의 언어/음성에 대한 특징을 분석하고 이를 음성학, 음향학, 생리학, 언어학적 지식들을 이용하여 상관관계(규칙)들을 정립한다. 예를 들면 합성하고자 하는 특정 언어에는 어떠한 음소들이 있으며, 이들 음소는 사람의 발화기관 내에서 어떠한 조음점, 조음방법 등에 의해 발화되는지를 조사한다. 그리고 발화된 음소들의 스펙트럼 특성, 운율적 특성에서 나타나는 기본형을 선정하

여, 그 변형인 변이음들이 생기는 요인들을 분석하고, 이들이 인지에 어떠한 영향을 미치는 지에 대해 연구한다. 합성단위의 선정에서는 분석과정에서 구한 필요 최소의 변이음들을 이용하여 합성음 생성시 접합시킬 음편의 기본단위를 설계한다. 그리고 실제 발화음을 녹음하여 최적의 음편들을 수집하고, 합성에 필요한 언어/음성/음향학적 정보를 추가한 데이터베이스를 구축한다. 합성방식 연구에서는 음성신호를 신호처리 관점에서 효율적으로 분석/압축/합성하는 방법을 연구한다. 합성음의 자연성은 언어처리 기술을 이용한 구문구조 분석과 운율구 단위의 분리, 그리고 대규모 음성데이터로부터 구한 통계적 운율분석 결과를 기반으로 하고 있다.

본 고에서는 컴퓨터가 사람의 말, 즉 음성언어를 생성하는 음성합성 기술에 대하여 설명하는데, 먼저 II장에서 음성합성 기술 연구동향과 기본적인 시스템 구현 방법들에 대하여 언급하고, III장에서 상용화를 어렵게 만드는 요인들과 향후 극복해야 할 기술적 과제를 간략히 살펴본 후, IV장에서 결론을 맺기로 한다.

## II. 음성합성 기술 개요

### 1. 음성합성 기술 연구동향

음성합성 연구는 먼저 인간의 간단한 음성신호를 복제할 수 있는 단순한 합성기의 개발에서 출발하였으며 이와 같은 의미에서 1950년 Dudley가 선보인 Voder와 1951년 Cooper가 Pattern Playback을 통한 음성신호 복제가 그 출발점으로 볼 수 있다. 이후 음성발성에 대한 음향학적 연구의 결과로 1950년대 말에 source-filter 이론이 정립되었으며 이는 바로 포먼트, 조음기관 합성기의 디자인으로 이어졌다. 그리고 합성단위의 선정 및 결합방식 연구로 확장되었으며 실제 음성과 같은 자연성을 합성음에 구현하기 위한 운율처리 연구와 완전 TTS를 구현하기 위해 필수적인 입력 문장으로부터 음소와 운율정보를 추

정하는 연구로 발전하였다.

최초의 포먼트합성기로는 1953년 Walter Lawrence의 Parametric Artificial Talker (PAT)와 Fant의 Orator Verbis Electricis (OVE I)이지만 사실상 현재와 비슷한 형태를 갖춘 것은 1962년에서야 가능하였다. Holmes에 의해 다듬어진 포먼트합성기는 Klatt에 의해서 1980년에 Klatt 합성기로 완성되었으며 그 이후로도 세부적으로 계속 개선되어지고 있다. 그러나 고품질 포먼트 합성기를 구현하기 위해서는 정교한 규칙과 전문가의 지식이 절대적으로 필요하며 또한 합성기의 구조가 달라지면 이들 정보를 공유하기가 어려운 단점이 있다. 1953년 Stevens가 선보인 M.I.T 모델 조음기관 합성기는, Rosen, Hecker에 의해 구조가 개선되었으며 Flanagan(1975), Liljencrants(1985)에 의해서 현대적 구조를 갖추게 되었다. 조음기관 합성기는 실제 조음기관을 모사하기 때문에 음성합성기의 궁극적인 해결책으로 예상되고 있지만 아직 정교화를 위한 데이터 부족과 계산량 부하로 현재는 경쟁력을 갖추기 못하고 있다. Itakura가 제안한 선형예측 부호화(LPC) 방법과 LSP, MP-LPC 등이 음성신호 분석/합성에 적용되었고, 1976년 Olive에 의해 TTS에 합성기로 도입되어 이러한 단점을 피할 수 있는 방법으로 각광을 받았다. 그러나 단점으로는 원음성의 피치값을 변경하여 합성할 경우 제1포먼트가  $\pm 8\%$  이상 변화하게 되고 더욱 심각한 것은 포먼트 대역폭이 심각하게 변할 수 있다는 것이다. 또한 접합점에서의 smoothing에 의해 합성파라미터가 급격히 변하는 경우 합성음의 자연성도 손상을 받게 된다. 1990년 프랑스의 Moulline에 의해 제안된 TD-PSOLA 방식은 편집합성의 방법을 개량, 발전시킨 방식으로서 시간영역에서 합성단위를 쉽게 접합할 수 있으며 기본주파수의 변경도 가능하기 때문에 많이 이용되는 합성방식이다. 그러나 합성단위 접합방식이므로 접합점에서의 스펙트럼 및 에너지의 불연속성, window 크기에 따른 왜곡이 문제가 되고 있다. 이 단점을 보완하기 위해 FD-PSOLA, MBR-PSOLA 등이 제안되었지만 아

직 미흡한 상황이다. 1996년 A.J. Hunt, A.W. Black, N. Campbell은 음질의 열화를 야기시키는 신호처리 과정을 완전히 배제한 새로운 합성방식을 제안하였다. 이 방법은 다양한 음운환경 뿐만 아니라 다양한 운율변화까지 포함된 대용량의 음성데이터를 사용하여 합성음의 품질을 한단계 향상시킴으로써 현재 세계적으로 많이 채택하여 사용되고 있다.

합성단위의 접합방식에서 중요하게 연구되는 항목 중에 합성단위의 선정 및 제작방법이 있다. 문장이나 단어는 그 수가 무한하므로 TTS에서는 고려의 대상이 되질 못한다. 음절의 경우는 그 수가 유한하지만 인접 음운환경에 따른 변이음수가 많아지게 된다. 음소의 경우는 그 수가 작지만 실제 발성음에서 음소간의 조음결합 현상을 올바르게 구현하기 어렵다. 조음결합 현상을 고려하여 1958년 Peterson은 이음소(diphone)를 제안하였고, 1968년 Dixon에 의해서 포만트 합성기에 구현되었으며 1977년 Olive에 의해 LPC 합성기에 접목되었다. 그러나 접합점에서의 smoothing, 이음소 지속시간의 변경, 기본주파수의 변경으로 인하여 합성음의 자연성은 기대만큼의 성과를 얻지 못했다. 1978년 Fujimura가 제안한 반음절(demisyllable)은 조음결합이 강한 음절내 자음군을 단위로 묶는 장점이 있지만 음절경계에서의 조음결합 처리가 미흡한 단점이 있다. 이외에도 1961년 Sivertsen이 제안한 VCV 단위인 syllable dyad와, 1980, 90년대 사용되기 시작한 3음소(triphone), 4음소(tetraphone) 및 1988년 Sagisaka의 비정형 길이의 합성단위, Nakajima의 자동회귀 방법을 채택한 문맥지향 군집화(Context Oriented Clustering) 3음소 작성방법도 시도되었다. 1996년 Donovan은 음성인식기를 이용하여 합성단위 DB의 제작을 통계기반으로 자동화하였다. 그러나 이 방법은 여러가지 장점에도 불구하고 음성인식기의 성능에 의존적이며, 음편내 천이구간의 연결이 자연스럽지 못한 단점이 있다. 이와 같이 합성단위와 관련된 연구는 초기의 합성음 명료도와 자연성에서 한계가 분명한 소량의 음소, 이음소, 반음

절 제작에서 시작하여 다음소열과 운율특성을 고려한 합성단위로 확장되고 있으며 그 결과 합성음의 명료도 향상을 가능하게 하고 있다.

그러나 이러한 분절음에서의 음질향상으로 인해 많은 음성합성기에 사용되고 있는 초보적인 운율제어 규칙이 부적절함을 부각시키는 결과를 낳게 되었다. 그 결과 운율모형을 개선하기 위한 연구는 다량의 데이터베이스를 이용한 연구로 방향을 정립하게 되었다. 그리고 억양커브의 모델링은 macroprosodic 성분과 microprosodic 성분을 분리하는 방법이 도입되어 피치값 계산의 단순, 편리성, 실제 데이터와의 근접성, 관측데이터로부터 파라미터 추정의 용이성, 음성 발생모델과의 호환성 등의 측도에 준하여 개발되고 있다. 음운학적 표기와 억양과의 상관관계에 대한 연구는 1968년 Chomsky에 의해 시도되었지만 1975년 Liberman의 단어강세, 1976년 Goldsmith의 어휘론 이론이 나오면서부터 억양의 여러 양상을 다룰 수 있게 된다. 이들 이론은 바로 운율음운학, 자동분절 음운학으로 발전되면서 현재까지 그 연구의 흐름을 이어오고 있다. 그러나 아직 대다수의 음운학적 연구가 연속된 음성의 억양을 다루기보다는 어휘로 분리된 특성에 대한 연구에 머물러 있어서 억양과 관련해서는 현재 만족할 만한 결과를 보이지 못하고 있다. 또한 이들 연구는 구문 구조, 문법적 단위, 문법적 성분의 경계 등에 의해 수반되는 음성학적 현상의 분석/구현에 있어서 정교한 형태소 분석 및 파서를 전제로 하고 있으며, 현재 그 결과도 여전히 만족스럽지 못한 상황이다.

입력 문장에 대한 운율현상 분석과 운율규칙 작성 및 적용을 위해서 구문구조 분석이 필요하다. 구문 구조 분석 기술에는 Chart parsing, LR parsing, Link parsing, Example based parsing, 통계 기반의 은닉 마르코프 모델(hidden Markov model; HMM)을 이용한 parsing 등이 있다. 이 중 통계 기반의 HMM을 이용한 parsing이 비교적 짧은 기간에 만족할 수준의 성능을 얻을 수 있고 무제한 어휘 및 문형을 대상으로 확장도 용이하다.

이상의 연구를 토대로 1970년대 말에 TTS 시 제품이 등장하기 시작하였고 일부는 실용화되었으며, 일부 기술선진국에서는 자국어는 물론이고 다국어 합성기 개발을 진행하여 이를 상용화하기 위한 연구가 계속되고 있다. 그러나 기존의 음성 합성기의 기술적 수준은 문장단위의 낭독체 스타일의 합성음을 명료하게 생성하는데 머물러 있다. 이는 실용화하는데 있어서 사용자의 요구를 만족시키기에는 자연성, 개인성, 대화체, 감정구현 등의 측면에서 미흡한 수준이다. 현재의 음성 합성 개발연구는 이러한 조건을 만족시키는 방향으로 진행되고 있다. 자연성 개선을 위해서 문맥, 의미 이해, 담화구조분석 및 모델링 연구가 실험적으로 진행되고 있고, 개인성 확보를 위해 개인성 분석, 음색변환, 운율특성 부여 방법에 관한 연구가 수행되어 진전을 이루고 있다.

2. 한국어 텍스트/음성 변환기의 구현방법

텍스트/음성변환기는 일반적으로 <그림 1>에서와 같이 언어 처리기, 운율 처리기, 음편 선택기, 합성음 생성기, 합성단위 음편 데이터베이스, 음성신호 출력기로 구성되어 있다.

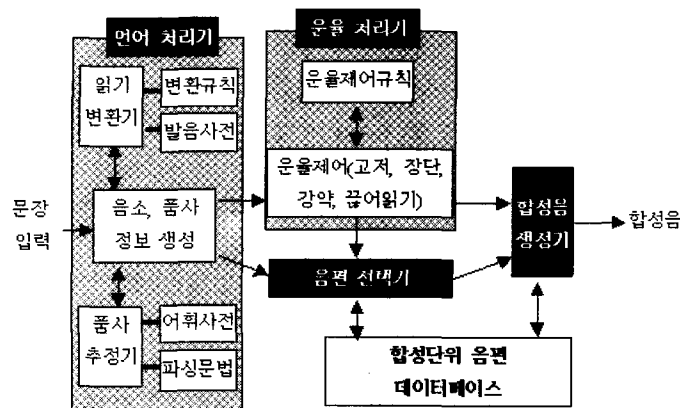
언어 처리기는 먼저 입력된 텍스트에 포함된 숫자, 심볼, 영어문자, 한자를 한글로 변환한 뒤 품사 추정기를 이용하여 각 형태소의 품사를 추정한다. 품사 추정기는 형태소 사전과 파싱 문법을 이용하여 입력 문장을 형태소 단위로 분석하

고 품사를 추정함으로써 각 어절이 문장 내에서 담당하고 있는 기능을 알려준다. 그리고 한국어 문장을 읽기형태로 변환한 뒤 한국어 음소열을 생성한다.

숫자음 및 심볼 처리는 아라비아 숫자, 영어단어, 혹은 %, \$ 등의 심볼이 입력되는 경우 이를 발음기호로 변환해 주는 기능을 갖는다. KS-5601 코드로 작성된 한자는 한자코드에 대응하는 한글로 변환하는 것을 기본으로 한다. 예를 들면 金(김)의 코드값은 0xD1D1이고 金(금)의 코드값은 0xD0DD으로 분리되어 있다. 그러나 나인(內人)과 같은 경우는 앞에서 언급한 1:1의 변환이 불가능하므로 예외사건으로 처리한다.

읽기 형태로의 변환은 한국어 음운변동 규칙과 예외 발음사전을 이용하여 한국어 표준 맞춤법에 준해 작성된 텍스트 문장을 소리나는 대로 변환한다. 음운변동 규칙은 음소 체계의 제약성, 발음의 편의를 위한 자연적인 현상이 대상이다. 한국어 발음규칙 적용이 불가능한 예외 단어들은 예외 발음사전에 등록하여 예외처리 한다.

운율 처리기는 언어 처리기로부터 음소열과 어절별 품사정보를 전달 받아서 입력 문장에 적합한 운율을 갖도록 끊어읽기, 음소별 지속시간, 피치값 및 에너지값을 추정한 뒤 음편 선정기와 합성음 생성기로 전달한다. 끊어읽기는 품사열 정보 등을 이용하여 각 어절 경계에서의 끊어읽기 유형을 추정한다. 음소별 지속시간과 피치값은



<그림 1> 텍스트/음성변환기의 구조

품사열, 끊어읽기, 음소열, 문장내 위치, 어절을 구성하는 음절 수 등의 정보를 이용하여 추정한다.

일반적으로 텍스트에서 띄어쓰기, 구두점 등을 이용하여 독자에 정확히 의미를 전달하고자 하는 것과 마찬가지로 발화시에도 운율을 이용하는데 운율의 표현수단은 끊어읽기, 음소 지속시간의 장단, 소리의 고저(억양), 소리의 세기(에너지 컨투어)의 4가지 형태이다. 발성자가 자신의 의사를 표현하고자 할 때 먼저 의미구조를 형성하고, 이를 전달하기에 적합한 단어들을 선택하며, 단어들의 연결이 자연스러운 구문 구조를 구상한 뒤, 조음기관을 이용하여 음절단위로 발성하며 이 때 조음기관의 물리적 제약을 받게 된다. 물론 이 과정이 반드시 순차적으로 일어나는 것은 아니다. 즉 언어학습을 통해 숙지된 의사표현과 발성자의 표현 습관 등은 반사적으로 흔히 나타나며, 특히 대화체와 같은 경우는 의미구조를 계속 생성하며 발화하는 것이 일반적이다. 그러나 이 경우라 할지라도 그 발화가 자연스럽다면 이미 상기 과정이 충분히 사전 학습된 결과라 볼 수 있다. 효과적인 운율제어를 위해서는 한국어 문장에 대한 다양한 구문구조 특징과 의미구조의 특징을 도출할 수 있는 고성능의 문장분석 시스템이 요구된다. 그러나 여기서 얻어진 구문론적 또는 의미론적 정보만 가지고는 실제 운율정보를 생성할 수 없다. 즉, 구문론적 의미론적 특징들과 운율정보와의 관련성에 대한 연구도 동시에 진행되어야 텍스트 정보로부터 자연스러운 운율을 생성함으로써 자연스런 합성음을 얻을 수 있을 것이다. 이를 위해서 다양한 환경에서 녹음된 대량의 음성데이터에 적절한 운율정보를 표기한 DB가 활용된다.

합성단위 음편 데이터베이스는 무제한 텍스트를 합성할 수 있도록 다양한 음운환경을 수용하고 있으며, 운율의 다양성도 포함하도록 각 음운환경에 복수의 후보가 등록되어 있다. 합성음 생성에 있어서 언어처리나 운율처리 뿐만 아니라 합성단위와 변이음 갯수 선정, 합성단위 데이터베이스의 작성은 명료도 및 자연성과 아주 밀접한 관련이 있다. 합성단위의 선정은 조음결합 현

상과 관련된 변이음 갯수를 결정짓고, 변이음 선정은 바로 합성단위 데이터베이스의 크기를 결정짓게 되며, 합성단위 데이터베이스는 음편선정과 저장형태에 따라 합성음의 명료도 및 자연성을 결정짓기 때문이다. 따라서 합성단위의 선정과 합성단위 데이터베이스의 제작은 가능한 모든 조음결합 현상을 수용할 수 있어야 하며, 접합점에서 스펙트럼의 불연속이 적고, 연결이 용이해야 하며, 가능한 한 그 수가 적도록 설계되어야 한다.

합성단위 설계에서 최소 음편 집합을 작성하여도 동일한 음운환경의 중복은 피할 수 없다. 연속된 어절 혹은 음절 발화의 경우 단순히 낱음절의 단독 발성이 아니라 좌우 음운환경, 문장내의 구문구조에 따른 운율 파라미터값, 발성 속도, 의미적 중요도, 끊어읽기와 상대적 위치 등에 의해 음성의 스펙트럼 패턴, 음원특성 등이 변화하게 되는데 이를 굳이 단순화, 단일화 하기도 불가능하다. 따라서 자연스러운 합성음 생성을 위해서는 이들 조건을 사용, 합성단위 DB에 기록, 저장한 뒤, 복수후보 중에서 입력 조건을 만족시키는 최적 후보를 선정하여 접합하는 것이 바람직하다. 합성 단위 데이터베이스는 연결구간에서의 스펙트럼 연속성 및 target 운율에 가장 가까운 합성단위를 선택하기 위해 음운환경, 운율 정보, 스펙트럼 정보 등을 가지고 있다.

음편 선택기는 운율 처리기로부터 음소열과 끊어읽기, 음소별 지속시간, 피치값, 에너지값 정보를 전달 받아서 합성단위 음편 데이터베이스에 등록된 음편들을 검색한다. 그리고 음운환경, 스펙트럼의 연속, 운율 특성의 연속성을 고려하여 복수후보들 중에서 최적의 음편들을 선택하며 합성음 생성 모듈에 이 정보를 전달한다.

합성음 생성기는 음편 선정기로부터 전달받은 음편 번호열을 이용하여 합성단위 데이터베이스에서 음편들을 읽어오고, 운율 처리기에서 추정된 운율 파라미터 값을 이용하여 음편들을 가공하기도 한다. 그리고 끊어읽기 정보를 이용하여 음편들을 접합하고 해당 위치에 묵음구간을 삽입함으로써 합성음을 생성한다. 생성된 합성음은 음성신호 출력기로 전송된다.

### Ⅲ. 상용화를 어렵게 만드는 요인들, 기술적 과제

최근 음성합성 시스템 구현에 필요한 각 모듈별 요소기술과 컴퓨터 성능이 급속히 발전함에 따라 다양한 응용분야에서 상용 음성합성기가 활용되고 있지만, 아직도 일반 사용자가 편리하게 이용하기 위해서는 극복해야 할 많은 문제점들이 남아 있고, 이것이 상용화 시스템의 확산에 걸림돌이 되고 있다. 이러한 제약 요인 및 기술적 과제들을 요약하면 다음과 같다.

#### 1. 일부 불명료한 합성음

- 숫자, 심볼, 영어단어를 정확하게 한글로 변환
- 음운변동규칙/예외발음처리의 정교화
- Micro-segmental 피치특성 고려
- 조음 음향학적 특성 고려

#### 2. 어색한 운율

- 적절한 음소별 지속시간의 구현
- 리듬의 자연성 구현
- 문법적 단위와 운율단위의 체계 정립
- 복합단어의 적절한 경계 분리 및 문법적 결합 관계 정립
- 강약처리 기능을 통한 beat 발생 억제

#### 3. 고정적 패턴의 반복

- 문맥에 따른 자연스러운 속도의 변화, 억양의 변화를 제어
- 문단, 담화 처리 기능
- 중심어, 의미어, 기능어의 분류적 처리

#### 4. 사용자가 원하는 음색, 운율패턴 등을 구현 불가능

- 음색변환기술
- 운율 학습 기능
- 대화체 구현

### Ⅳ. 결 론

본 고에서는 음성합성 기술의 개요, 당면 문제점 및 향후 과제를 개괄적으로 살펴 보았다. 지난 30여년간 관련분야 기술의 발전으로 상용화 시스템 및 서비스가 확산되고 있지만, 아직 극복해야 될 수많은 과제를 안고 있다. 관련기술분야에 대한 음성합성기술의 파급효과를 고려하면 상용 시스템 개발과 병행하여 핵심 요소기술에 대한 투자 및 연구가 지속적으로 이루어져야 할 것이다.

#### 참 고 문 헌

- [1] D.H. Klatt, "Review of text-to-speech conversion for English," *J. Acoust. Soc. Am.* 82(3), pp.737-792, 1987. 9.
- [2] R.E. Donovan, Trainable Speech Synthesis, Ph.D dissertation, University of Edinburgh, 1996.
- [3] A.J. Hunt, A.W. Black, "Unit selection in a concatenative speech synthesis system using a large speech database," in *Proc. ICASSP'96*, pp.373-376, 1996.
- [4] W.N. Campbell, A.W. Black, "CHATR : a multilingual speech re-sequencing synthesis system," *SP96-7 Tech Rept. IEICE*, pp.45-52, 1996.
- [5] M.S.Hahn, S.H.Kim, J.C.Lee, Y.J.Lee, "Construction of a multi-level speech database for spontaneous speech processing," in *Proc. ICSLP'96*, pp.1930-1933, 1996.
- [6] 이정철, 한민수, 성평모, "신경망 학습을 이용한 한국어 음성 합성기용 에너지 컨트롤러 제어," *한국음향학회지*, Vol. 17, No. 4, pp.76-81, 1998.
- [7] 이정철, 김상훈, 성평모, "한국어 음성합성기용 끊어읽기 추정기," *한국음향학회지*, Vol.

17, No. 5, pp.51-56, 1998.

- [8] Mark Huckvale, "Speech synthesis, speech simulation and speech science," in *Proc. ICSLP-2002*, Vol. 2, pp.1261-1264, 2002.

## 저자 소개



李政哲

1984년 2월 서울대학교 학사,  
 1988년 2월 서울대학교 석사,  
 1998년 8월 서울대학교 박사,  
 1985년 9월~2000년 3월 : 한국  
 전자통신연구원 책임연구원,  
 2000년 3월~2001년 1월 : L&

H Korea 전문위원, 2001년 1월~2002년 2월 : (주)  
 보이스텍 전문위원, 2002년 3월~2002년 8월 : (주)  
 코난테크놀로지 책임연구원, 2002년 9월~현재 : 울  
 산대학교 컴퓨터·정보통신공학부 조교수, <주관심  
 분야 : 디지털신호처리, 음성신호처리, 음성합성>