

TTS 음성합성기술

배 명 진

(숭실대학교 정보통신공학과)

■ 차 례 ■

- I. 서론
- II. TTS 음성 합성시스템의 응용예
- III. 음성 합성시스템의 구성요소
- IV. TTS의 고음질 합성방식
- V. TTS시스템의 고음질화에 따르는 문제점
- VI. 결론

I. 서 론

음성합성기의 궁극적인 목적은 무제한 텍스트를 명확하고 자연스러운 목소리로 읽어주는 텍스트의 음성변환(text to speech, TTS)에 있지만, 현재의 기술은 이러한 목표에 아직 도달하지 못하고 있다. 기존의 음성합성기는 텍스트를 발성할 때에 명료성은 우수한 편이나 결정적으로 기계-액센트를 갖는다. 고음질의 음성합성에 대한 목표를 이루도록 하기 위해서는 음향학과 언어학의 두가지 과학적인 원인을 정보통신 알고리즘으로 해결해야 하는 과제가 아직도 남아 있다. 이를 해결하기 위해서는 인간 목소리기관에 의해 음성소리의 생성이 어떻게 이루어지는가의 원리를 이해해야 하고, 어떻게 이들 소리가 인간 청취시스템에 의해 처리되는가를 또한 정확히 이해해야 한다. 즉, 단지 텍스트에 있는 언어적인 구조를 기본으로 하여 목소리기관의 수학적인 모델을 제어하기 위해 필요한 정보를 추출할 수 있어야만 한다.

1971년 초기에 von Kempelen은 짧은 문장을 발성하는 초보적인 음성합성기를 만들었다[1]. 이 장치는 모음을 생성하기 위해 폴무에서 공기흐름으로 공동기를 여기하였고, 자음은 공기 흐름을 협착시킴으로써 발생되었다. Dudley등에 의해 이 기술을 전기회로로 구현한 것이 1939년 세계박람회에서 전시되었다. 이

보코더(VOCODER)는 조음구조를 제어하기 위해 정교한 키보드를 갖추고 있으며, 모음부류와 자음부류를 위해 발진기를 별도로 분리하였다[2]. 현대의 음성합성기는 이들과 동일한 원리에 기본을 두고 있지만, 물론 그 구현은 디지털 컴퓨터나 별도의 DSP하드웨어로 이루어지고, 숙련된 오퍼레이터가 필요없이 합성기는 텍스트를 자동으로 직접 합성해준다.

음성의 소리는 성도를 통해 흐르는 공기의 음향적인 결과이다. 허파에서 압축된 공기는 후두를 통해 흐른다. 성대에 위치한 적절한 텐션은 이들을 공기압력을 주기적으로 변화시켜 발진하게 하고, 실제로 구도 및 비도의 음향적 공명을 여기하여 우리가 잘 알고 있는 모음이나 유사모음의 소리를 생성하게 된다. 조음기관(예를들어 혀, 입술, 치아, 그리고 연구개)들을 움직이기 때문에 공명이 바뀌고 다른 특성의 소리가 생성된다.

성대에서 텐션이 달라진다면 동일한 공기흐름으로 다른 소리가 생성될 수 있다. 공기흐름에 대해 조음기관들을 아주 좁은 관이 이루어지도록 배치하게 되면, 공기의 교란흐름은 예를들어 자음 s, z, sh와 같은 소리를 발생하게 된다. 또한 공기흐름을 모두 모아두었다가 갑자기 방출하면 p, t, k와 같은 소리가 발생된다. 이들 방법들은 여타의 다른 소리를 생성하기 위해 서로 결합될 수도 있다.

이상의 조음동작에 대한 물리음향학은 시변 비균일 단면적인 유행특성 튜브에 의해 이론적으로 잘 모델링되고 있다. 이 튜브는 정상적인 또는 준주기적인 공기흐름에 의해 여기된다. 원리적으로 인간 성도의 기하학적인 측정이 가능하고, 적절한 경계조건에 따라 공기흐름의 편미분 방정식을 푸는 것이 Coker에 의해 연구가 진행되어왔다[3]. 이를 1931년에 Dudley가 전기회로의 형태로 제안하였다. 그의 연구에서는 성도는 시변형 여파기로 모델링되었고, 여파기의 여기는 사육의 경우에 잡음원을 그리고 모음의 경우에는 주기성 펄스열을 사용하였다. 그러한 회로는 여기원/여파기 모델로 알려지게 되었고, 오늘날의 중요한 음성 분석기술의 기본이 되었다. 여기원/여파기 모델에 대한 당대의 구현은 선형예측법에 의해 보통 구현되고 있다[4].

그렇지만 음성의 소리를 생성하는 능력은 말하는 능력을 나타내는 것이 아니다. 음성의 소리를 단순히 연결하게 되면 불쾌하고 불명료한 소리가 발생될 수 있다. 실제의 음성에서 조음기관들은 연속적으로 움직이고, 얻어지는 소리는 독립된 소리들의 연속적인 결합이다. 또한 음성의 목적은 중요한 메시지를 전달하는 것이다. 인간은 의도하는 메시지의 언어적 구조에 따르는 기본소리의 운율구조(위치, 고저, 그리고 지속시간)를 변화시킴으로써 그 메시지를 각각 다르게 이해하게 된다. 텍스트의 음성변환(TTS)시스템은 운율파라미터들은 물론 다른 음성소리를 나타내는 스펙트럼 파라미터들을 연속적으로 변화시키는 계산을 한다. 그리고 이들 파라미터들은 음성신호를 생성하기 위한 여기원/여파기 알고리즘을 제어하는데 쓰인다.

II. TTS 음성 합성시스템의 응용에

TTS는 컴퓨터화된 형태로 정보를 변환, 저장, 유지하는 비용에 기초한 오디오 정보통신 시스템으로 고려된다. 만일 정보가 초기에 문자의 형태로 표현되어 있다면 사람이 그것을 기록하고 다시탈화 하는 것보다는 TTS를 통해 음성으로 변환하는 것이 더 값싸고 빠르다. 그리고 그 문자정보의 용량이 많고 사용가능한 저장공간이 작다면, TTS는 가장 적당한 방법이다. 왜냐하면 자연음성의 평균저장율이 4~6 kbps의 코드화율인데 비해 TTS는 50~60 bps정도이기 때문이다. 그리고 문자정보가 자주 바뀌어서 그것의 일부분이라도 다시 기록하는 비용이 소요된다면, 이때도

TTS의 사용이 더 경제적이 된다.

텍스트 형태로 입력된 문장을 음성으로 변환시켜 주는 TTS시스템은 기존에 사용되고 있는 단순 조합단위의 자동 응답시스템(ARS)에 비해 무제한 문장의 합성이 가능하고 정보의 가공, 변경이 용이하며 저장할 정보량이 대폭 압축될뿐만 아니라 발성 속도 및 음색의 변경이 가능하다는 장점을 지니고 있기 때문에 앞으로 여러 분야에서 활용될 수 있는 기술이다.

구체적인 TTS의 응용분야로는 정보검색시스템, 장애인 복지형 통신, 자동통역전화, 문외예약시스템 등에 사용될 수 있고 그밖에 OA, FA, 의료기기, 교육장비 등의 분야에서 활용될 수 있다. 이것은 TTS가 다음과 같은 장점이 있기 때문이다.

- 1) 시각 장애인이나 문맹인에게 글을 읽어 줌으로써 문자정보보다 이용이 용이하다.
- 2) 많은 문서를 읽어 줌으로써 문서 교정에 덜 피로함을 느끼고 다른 일을 하면서도 교정이 가능하다.
- 3) 음성에 의한 정보제공으로 정보에 대한 습득이 용이하다.
- 4) 사용자에게 문자보다 친밀감을 줄 수 있다.

음성 합성시스템의 활용의 예를 구체적으로 몇 가지 들면 다음과 같다.

- 1) 워드프로세서에 응용: 문서교정, 편지의 내용 등을 음성으로 읽어주므로 빠른 시간내에 교정 및 파악.
- 2) 표준시보, 전화 번호, 일기예보, 비행기 및 기차 시간 안내, 증권시세, 교통정보, 은행잔고, 상품 소개 등의 여러 안내들을 자동으로 수행.
- 3) 교육용에 응용: 단일집으로 제한된 수록.
- 4) 전화의 자동통역에 응용: 조회결과를 목소리로 응답.
- 5) 가전기구에 응용: 사용 설명과 기능안내를 음성으로 안내.
- 6) 음성편지: 카드, 전보, 편지 등에 목소리 전송.
- 7) 컴퓨터 사용자 인터페이스: 사용법, 에러처리, 도움말 등을 음성으로.
- 8) 장애인 보조용: 맹인의 생활보조, 발성장애자 목소리 재생, 발음교정.
- 9) 영화제작에 응용: 여러명의 성우를 대신하는 목소리 변조기에.
- 10) 휴대용 통역기에 응용: 외국인과의 자연스러운 대화에.
- 11) 휴대용 수집에 응용: 전화수집, 사진용, 메모알

람 등.

- 12) 비상상태의 방송: 긴급하고 당황하는 실제 상황의 안전 대피유도.
- 13) 말하는 기계: 산업용 로봇 등에 활용.
- 14) 장난감 등의 제조에 응용

TTS시스템을 전화서비스에 적용하는 예는 현재에도 많이 있고, 앞으로도 적극적으로 활용될 예정이다. 지금 전화통신에 연구 및 실용화되고 있는 중인 서비스들로는 호출자 확인서비스, 통합메시지화 서비스, 전화 중계서비스, 자동 가입자 이름/주소 안내서비스, 판매 주문확인 서비스 등이 있다. 이들을 좀더 구체적으로 설명하면 다음과 같다.

1) 호출자 확인(Who's Calling) 서비스: 이 기능은 걸려 온 전화를 전화가입자에게 연결하기 전에 전화를 걸 사람의 이름을 미리 알려주는 서비스이다. 전화 벨이 울린 후, 전화 가입자가 송수화기를 들면 “홍길동씨로부터 전화가 걸려 왔습니다. 전화를 받으시려면 아무번호나 누르십시오”하는 안내가 들려 온다. 전화가입자는 전화를 받기 위해 번호를 누를 수도 있고, 전화를 피하기 위해 그냥 끊을 수도 있다. 전화를 걸 사람은 벨소리만 들을 수 있을 뿐 그 전화가 직격 심사되고 있다는 사실을 알 수가 없다. TTS는 전화기는 사람들의 이름에 대한 데이터베이스가 기대하고 이미 ASCII코드의 형태로 저장되어 있으며 거주지역의 전화가입자의 변화에 따라 매일 바뀌므로 안내에 사용될 수 있다. 이 서비스의 합성 음질은 호출자 이름의 발음이 명료하고 정확하기만 하다면 아주 대중적으로 보급될 수 있다. 그 이유는 전화가입자들이 다른 방법을 통해 받을 수 없었던 유용한 정보와 그들의 전화에 더 많은 통제를 제공해주기 때문이다.

2) 통합 메시지화(Integrated Messaging): 이 기능은 하나의 매체를 통해 전해진 메시지를 다른 매체를 이용해 복구하는 것이다. 예를 들면, 음성 우편시스템에서 사용자는 전화를 통해 전해진 메시지에 대해 ASCII 코드의 형태로 변환한 후에 TTS를 사용해 읽음으로써 E-mail이나 팩시밀리 정보를 음성정보로 복구할 수 있다. 통합 메시지화의 실용 가능성은 가입자확인 서비스의 메시지보다 길고 복잡한 메시지를 번역할 수 있는 능력여하에 달려 있다. 그러함에도 불구하고 간단한 접속에 의해 모든 메시지가 음성으로 복구될 수 있다는 효용성은 사용자들이 경험한 그 무엇보다 더 중요하다.

3) 전화 중계서비스(Telephone Relay Service, TRS): 이 기능은 말하기와 청각기능에 손상을 입은 사람들을 위한 전화-걸기 변환서비스이다. TRS는 표준 음성전화 사용자와 청각장애자를 위한 전화(telecommunication device for the deaf, TDD)사용자 간의 통화를 용이하게 하는 기능을 제공한다. 정상적인 전화 통화자에게 전화를 걸려고 하는 TDD 사용자는 먼저 TDD를 통해 응답할 수 있는 오퍼레이터가 있는 전화 중계센터에 전화를 해야 한다. 그러면 오퍼레이터는 정상 통화자에게 전화를 걸어 정상 통화자에게는 문자-음성 변환기능을 TDD 사용자에게는 음성-문자 변환 기능을 제공한다. TRS는 말하는 것이나 듣는 기능에 손상을 입는 사용자에게는 통신에 큰 이익을 주지만 너무 많은 인력을 필요로 하므로 비용 면에서 비싸질 수밖에 없다. 또한 통화자에게 프라이버시를 침해받을 수도 있다는 부담감을 주게 된다. TRS의 완전 자동화는 오퍼레이터 대신에 자동 음성인식과 TTS의 조합을 사용함으로써 이룩할 수 있다. 오늘날의 ASR기술의 수준은 연속적이고 자연스러운 음성을 다룰 수는 없지만 대화의 반을 TTS를 사용함으로써 부분적인 자동화는 이룰 수 있다.

4) 자동 가입자 이름/주소(Automated Caller Name and Address) 안내서비스: 이 기능은 몇몇 회사에 의해 제공되는 고객 또는 가입자의 번호조회에 의해 자동으로 가입자의 주소와 성명등을 안내해주는 완전히 컴퓨터화된 서비스이다. 사용자는 오퍼레이터에게 전화를 걸어 전화번호를 알려준다. 오퍼레이터는 그 전화번호와 관련된 이름과 주소를 컴퓨터 데이터베이스에서 찾아 전화를 걸 사람에게 알려준다. 이 서비스는 사용자가 전화번호를 누르거나 음성인식기에 전화번호를 불러 주고 컴퓨터 데이터베이스로부터 이름과 주소를 찾아 TTS를 사용해 음성으로 되돌려 줌으로써 완전히 자동화할 수 있다. 이러한 종류의 시스템에 대한 시도는 성씨의 발음과 주소가 정확하기만 하다면 합성음의 음질은 사용자에게 큰 장애요소는 아니라는 사실을 보여준다. 즉, 이 경우에 자연성이 아주 높은 음성정보를 별로 요구하지 않게 된다.

5) 판매 주문확인(Sales Order Verification) 서비스: 이 서비스는 사람들이 주문하려는 물건에 대해 컴퓨터화된 정보를 보고 물건을 구입하게 하는 상업서비스의 일종이다. 이 서비스는 TTS없이 사용할 수 없다. 포장 상품회사에서 이러한 서비스를 적용하

는 경우에 있어 TTS시스템은 소비자가 직접 방문하기 전에 주문상품의 상태를 조사하고 구입할 수 있도록 한다. 이 사업에서 주문의 내용을 바꾸는 것은 가끔 일어나는 문제이고, 크기나 양의 교체, 가격의 변화는 자주 일어난다. 이때 주문을 확인하는 것이 아주 중요하다. 이 회사에서 오퍼레이터들은 주문을 받는 일은 하지만 너무나 바빠서 주문을 항상 확인하는 것이 불가능하다. TTS를 사용한 자동화는 고객 상담 문제를 해결해 준다. 즉, 발음이 정확하다면 그 음성의 질은 중요한 문제가 되지 않는다.

III. 음성 합성시스템의 구성요소

3.1 TTS시스템의 구성요소

음성합성기의 일반적인 구성은 TTS 시스템의 플로우차트로 그림 3-1에 나타내었다. 전체 시스템은 텍스트 전처리, 음성 분할알고리즘의 문법적 분석과 분류, 강세와 억양의 계산, 발성모듈, 억양, 타이밍, 그리고 고저의 윤율변동에 따른 계산, 합성파라미터의 계산, 그리고 파라미터에서 파형의 합성 등이 있다.

이러한 과정을 좀더 잘 알아보기 위해 TTS 합성기를 구성하는 각 단의 기능을 부분별로 설명하기로 한다.

1) 전처리부: 일반 한글 텍스트에서 나올 수 없는 한글 이외의 문자를 한글로 바꾸어 준다. 한글 이외의 문자들이란 영어, 숫자, 약어, 단위 등을 나타내는 기호들을 말한다. 영어는 기본적으로 알파벳 하나하나를 일일이 대응시키거나, 영문을 읽는 법칙에 따르며, 사전처리를 하여 약어나 외래어등을 처리한다. 숫자는 크게 정수, 소수, 분수, 전화번호, 단위 그리고 사전처리 숫자 등을 고려해야하는 복잡한 처리과정을 요한다. 정수는 일반적인 숫자읽는 법칙을 세워 읽게 되고, 소숫점이 있으면 소숫점이 들어온 뒷 부분은 또 다른 법칙으로 처리한다. 마찬가지로 전화번호와 "%", "()", "\$", "#" 등의 특수한 문자도 적절한 법칙과 사전처리를 이용하여 읽게 한다.

2) 형태소 분석: 입력된 한글 텍스트 문장에 대해 품사를 분류한다. 이 정보들이 구문 분석과 발음 변환 등에 이용된다. 형태소 분석을 위하여 적절한 규칙과 체언사전과 용언사전, 어미사전 등을 갖추고 있다.

3) 구문 분석: 출력음성의 윤율 효과를 만들어 내기 위한 정보를 제공한다. 형태소 정보와 적절한 규칙을 이용하여 문장상의 기능정보, 즉, 주어, 목적어 등의 분류와 구, 절 경계, 문장성분 등 한국어 문장에서 어절간의 문법적인 관계 정보를 제공한다. 이 정보를

이 억양처리, 음절길이, 강세처리 등 한국어의 윤율제어에 필요한 정보들이다.

4) 발음변환: 위의 1, 2, 3의 결과를 이용하여 입력된 한글문장을 소리나는대로 바꾸어 준다.

5) 강세처리: 한국어의 한 어절에서 나타나는 볼륨 현상을 처리하여 준다. 이 정보가 음절길이 조절에서도 사용된다.

6) 음절길이 조절: 강세정보, 휴지기여부, 음절수 등의 영향을 고려하여 음절의 지속시간에 대한 정보를 제공한다. 지속시간을 변화시키는 요인으로는 전후음소에 대한 변화, 강세 유무에 의한 영향, 음절수에 의한 영향, 단어의 빈도수에 의한 영향 등이 있다. 한국어의 지속시간에 대한 연구는 2음절로 구성된 부의미단어에 대한 발음을 대상으로 하여 유성자음에 비해 무성자음 앞에 오는 모음의 지속시간이 짧아지고 마찰음에 비해 중지음 앞의 모음길이의 분산이 각 음절길이의 분산의 합에 비해 작어, 음절이 기본단위가 아니고 분절이 기본단위임을 제시하고 있다.

7) 억양 처리: 구문 분석 정보를 이용하여 피치변화에 대한 제어를 수행한다. 합성음의 자연성과 문장의 이해도 증진에 큰 영향을 준다. 강세처리 음절길이

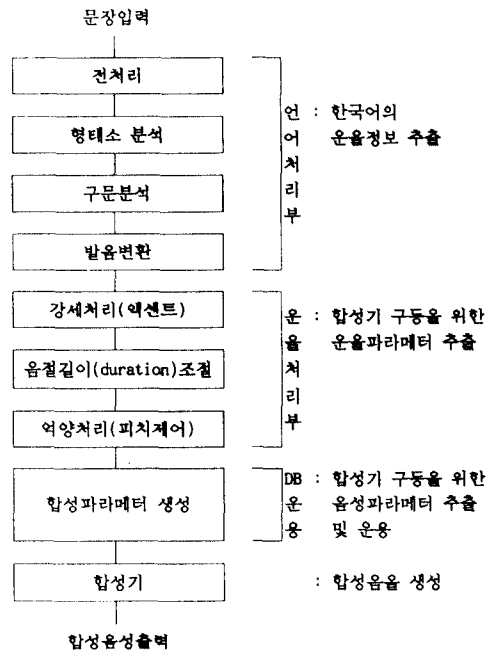


그림 3-1. TTS 시스템의 구성

등과 함께 운율이라고 하는 것이 만들어진다. 이러한 운율은 문장의 이해도와 밀접한 관련이 있으며 정보 전달에도 큰 영향을 미친다.

8) 합성파라미터 생성 : 위의 모든 결과를 이용하여 실제 합성기를 구동하는 파라미터들을 DB로부터 읽어들이어서 합성기에 제공한다. 이때 연결 방식에 따른 합성단위간의 자연스러운 연결을 위한 조절도 해준다. 합성단위는 음소단위, 반음절단위, 음절단위, 다이폰단위 등으로 나누어진다.

9) 합성기 : 합성기를 구동하기 위한 파라미터를 제공받아 실제 합성파형을 만든다. 이 합성기가 어떤 합성방식을 사용하는가에 따라 합성기의 구동 방식이 정해진다고 볼 수 있다. 일반적으로 신호원부호화, 파형부호화, 혼성부호화, 시간영역 부호화 합성기로 구분하고, 다시 이들을 세분하여 LPC합성기, 포먼트합성기, PSOLA합성기 등으로 나눈다.

3.2 데이터베이스에 따른 TTS시스템의 분류

음성합성이란 한마디로 인간이 음성을 발성하는 메카니즘을 전기회로로 실현하여 인공적으로 만들어 내는 것이다. 음성합성의 궁극적인 형태로서, 인간이 어떤 언어로 표현한 정보를 음성으로 출력시키는 과정을 인공적으로 흉내내기 위해서는 그림 3-2와 같이 4가지 처리 단계를 필요로 한다[20-21].

기본적인 접속에 의한 음성의 규칙합성을 위해서는 합성의 기본이 되는 단위를 가능한 적은 양으로 구현하는 것이 효율적이다. 규칙에 의한 음성합성에서는 임의 어휘의 출력이 가능해야 하므로 단어보다

작은 단일 음운(음소), 또는 2-3개의 음운 연쇄를 합성의 기본 단위로 한다. 이들 단위를 이용할 경우 합성에 필요한 총수는 약 수십개에서 수천개에 이르며, 단위가 길고 복잡해짐에 따라 소요되는 dB의 규모는 증가한다[22]. 이와 같이 규칙에 의한 음성합성에 있어서는 다음 조건을 모두 만족하는 단위를 필요로 한다.

- 임의 어휘의 합성이 가능할 것.
- 음운의 명료성, 자연성에는 손상을 주지 않고 고품질의 합성음을 얻을 수 있을 것
- 합성에 필요한 모든 음성 단위를 간단히 작성할 수 있을 것.

일반적으로 이들 단위로는 음소, 반음절, 음절, 다이폰 등이 사용되고 음절, 다이폰에서 실현되지 않는 조음 결합에 의한 음운 변동을 흡수하기 위해 CVC나 VCV와 같은 더 큰 단위도 이용되며 최근에는 고정된 길이의 합성단위 보다 합성하고자 하는 음소열의 다량의 음성 데이터베이스를 이용하여 가장 적합한 환경중에서 임의 길이를 선택적으로 사용하는 방법도 시도되고 있다.

1) 단어단위 합성 : 단어단위의 경우 문장단위에 비해 데이터베이스의 규모가 줄어들 것이나 한 언어체계에서 사용하는 단어가 보통 10만 단어 이상이 되므로 무제한 어휘합성에는 적합하지 않고 또한 단어 사이의 천이구간에 대한 변화나 운율에 대한 법칙을 만들어 주어야 한다. 따라서 무제한 어휘 합성보다는 일기예보나 전화번호 안내와 같은 제한적인 상황에서 사용된다[16][23-24].

2) 음절단위 합성 : 언어체계에 따라 음절수가 4,000-

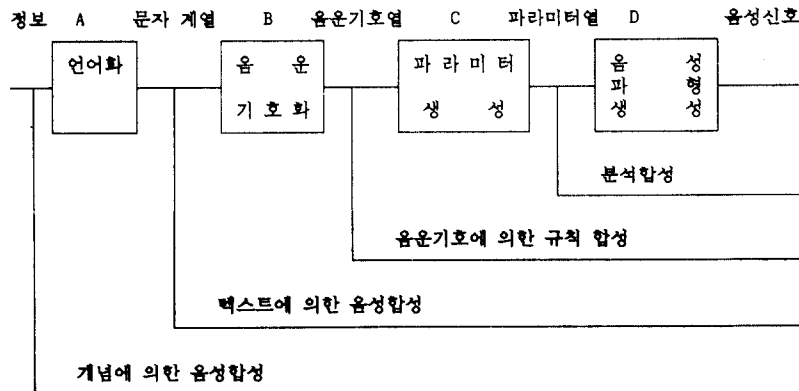


그림 3-2. 음성합성 시스템 구분을 위한 처리 4단계

10,000개 정도가 되나 한국어의 경우 대표음으로 조성 19개, 중성 8개(반침없는 경우도 포함)가 있어서 이론적으로 총 3,192개의 음절이 존재한다. 그러나 실제 모니터나 프린터에서 사용하는 한글의 경우 대표음으로 표시할 수 있는 음절이 약 2,000개가 되며 이중 사용하지 않는 음절을 고려한다면 대략 2,000개 정도의 음절 데이터베이스를 이용하면 부제한 어휘 합성이 가능하다. 각 음절의 평균길이를 250msec 정도로 하고 프레임 길이 256, 이동길이 128, 윈도우차수 10으로 하는 선형예측방식으로 데이터베이스를 작성하여 저장하기 위해서는 총 14 Mbyte 정도의 메모리 용량이 필요하고 자연스러운 소리를 생성하기 위해서는 운율에 대한 법칙도 따로 필요하다[23][25-27]. 더 작은 단위의 합성보다는 합성음의 명료도를 어느 정도 확보할 수 있으나 이 단위 역시 음소변화를 완전히 실현시킬 수는 없으므로 여러가지 음운변형규칙이 필요하고 변이음 처리 및 타이핑 제어 등에 문제가 있다.

3) 반음절단위 합성: 음절은 초성자음과 중성모음 그리고 중성자음으로 구성되는데 중성모음의 경우 상당 구간동안 일정한 특성을 유지하는 부분이 지속되므로 이 부분을 중심으로 나누면 합성할 때 용이하게 연결이 가능하고 음절에 비해 데이터베이스의 규모를 대폭 줄일 수 있다는 장점이 있다. 즉, 음절해를 중심으로 해서 그 전부분과 음절끝의 자음부로 된 suffix라고 부르는 단위를 이용한다. 언어체계에 따라 1,000-2,000개 정도의 반음절이 존재한다. 한국어의 경우 초성자음 19개와 중성모음 21개의 조합에 의한 반음절과 중성 21개와 중성 8개의 조합에 의한 반음절을 합해 총 567개의 반음절이 만들어지 음절에 비하면 상당한 절감효과가 있음을 알 수 있다[28-29].

4) 다이폰단위 합성: 반음절에서처럼 한 음절을 2부분으로 나누는 것이 아니고 연속발음의 경계를 고려할 때 각 음소 중 안정된 부분을 찾아 그 점을 중심으로 양분하여 전이구간에 대한 변화규칙을 가능하면 줄여 줌으로써 변화규칙의 오류로 인한 음질저하를 방지하고자 하는 목적에서 다이폰을 단위로 한다. 다이폰의 보다 정확한 정의는 '다이폰 원소는 하나의 단음의 정적 중심에서 시작해서 다음 단음의 정적 중심으로 끝난다'라는 설명으로 대신할 수 있다. 이 방법은 음절이 가진 장점을 살리고 합성단위의 갯수를 줄이는 방법이 된다. 다이폰은 자음-모음(CV), 모음-자음(VC), 모음-모음(VV)과 같이 2개의 음소쌍으로 된 단위이므로 음운 과도부를 단위속에 포함하고 있

으므로 음소단위 합성에 비해 음운 명료성을 확보하기 쉽다. 길이조절이나 강세조절, 피치조절은 연속하는 두 다이폰의 정상상태의 값을 목표값으로 인터폴레이션하는 방법등으로 구현한다. 한국어의 경우 음질구성이 초, 중, 중성으로 구성되어 영어와 같이 자음이 계속되는 경우가 없어 다이폰 부분이 용이하며 이론적으로 1,598개의 다이폰이 가능하나 실제로는 1,100개 정도의 다이폰을 이용하여 부제한 어휘 합성이 가능한 것으로 발표된 바 있다.

5) 변이음단위 합성: 음절을 구성하는 자음, 모음의 경우 그 전후 배열에 따라 음향음성학의 관점에서 보면 같은 음소라 하더라도 각기 다른 소리값을 갖는다. 이러한 소리값이 각기 다른 음소를 변이음(allophone)이라 하며 보다 정확한 변이음의 정의는 '특정 음소를 구성하는 조음적으로나 음향학적으로 다른 두가지 이상의 음형태의 하나이다. 각 음소에 대한 변이음 분류는 연구자에 따른 분류 기준이 서로 달라 조금씩 다르나 일반적으로 초, 중성 자음의 경우 약 50개 정도의 변이음으로 분류할 수 있고 모음의 경우 따로 변이음을 분류하지 않고 운율법칙의 조절을 통해 변이음을 구현하면 총 80개 정도의 변이음으로 부제한 어휘 합성이 가능하다. 이같은 데이터 베이스의 규모는 음소단위를 제외한 다른 단위의 규모에 비하면 극히 작아 실시간 하드웨어 구현을 위해 바람직한 합성단위로 판단된다. 그러나 합성단위가 간단할수록 그에 따르는 법칙이 복잡해지므로 잘못 만들어진 법칙이 전체 시스템 성능에 끼치는 영향은 치명적이다. 따라서 지속적인 법칙연구 및 수정보완이 필요하다.

6) 음소단위 합성: 여러 단위중 가장 원리적인 것이 단일 음소를 이용한 것이다. 말을 표현하는 최소단위 음소는 언어에 따라 그 수는 다르나 보통 30-50개 정도이고, 가장 적은 수로 합성할 수 있다. 그러나 음소는 소리 값을 나타내는 것이 아니라 의미를 나타내는 것이므로 하나의 음소에 대응하는 실제 음성 파형은 여러가지 음소 환경, 발성지, 발성속도 등에 의해 크게 변화하고 따라서 이를 하나의 음소데이터로 실현하기는 매우 어렵다. 즉, 데이터의 양은 적으나 인접한 음소사이의 전이를 실현하기 곤란한 점이 있다.

7) 발성환경단위 합성: 기대한 음성 데이터베이스의 통계적 분석을 토대로 발성 환경단위를 선택한다. 일본어의 경우에 가장 빈번하게 나타나는 1-4 음소 계열의 20%가 임의의 문장에서 나타나는 음소열의 80%를 차지한다. 이와 같은 관측으로부터 길이를 변화시키는 단위음성을 사용하여 고음질의 음성을 생

성할 수 있음이 보고되었다. 그러나 아직 한국어에 있어서는 충분한 통계적 자료가 마련되어 있지 않음으로서 인하여 사용되지는 않는다.

3.3 표준 데이터 베이스

최근 들어 국내에서도 음성합성 및 인식 분야의 연구가 본격적으로 시작되고 있어 개발된 여러 인식 알고리즘들의 성능을 객관적으로 평가할 필요가 생기고 있으며 이러한 경우 적절히 구성된 공통 음성 데이터 베이스가 필요하다. 더구나 무제한 음성 합성시스템의 연구 개발에 있어서는 성별, 나이별, 지역별로 다양한 개성을 포함하는 다수의 화자에 의한 대용량의 음성 데이터 베이스가 요구된다. 이러한 방대한 양의 음성 데이터 베이스 구축에는 많은 인력과 경비 및 시간이 소요될 뿐만 아니라 국내외의 한국어 음성처리 기술 연구 개발에 공동으로 사용될 수 있도록 구축되어야 하기 때문에 국가적인 지원이 요구된다. 미국, 일본 및 유럽에서도 국가적인 지원하에 자국의 음성 데이터 베이스 구축 연구가 활발히 진행되고 있고 국내에서도 일부 연구가 진행중이나[25] 제한된 범위의 음성 데이터 베이스 구축에 머무르고 있는 실정이다.

표준 음성 데이터 베이스 구축을 위해서는 다음과 같은 사항을 고려한 연구가 필요하다.

1) 언어학적 통계성 고려: 한국어 유문 구조에 대한 통계학적인 연구를 거쳐 음운의 조합된 대표적인 단어군 및 문장들을 구성하여야 한다. 이는 합성 시스템이 음소, 음절, 단어 등의 어떠한 단위로도 구현될 수 있도록 합과 동시에 될 수 있는 한 최소한의 데이터 베이스를 이용하여 최대의 효과를 얻기 위함이다.

2) 데이터 채집환경 고려: 앞에서 선정된 단어군 및 문장 등에 대해 자연스러운 음성을 녹음 채취하여 디지털화하는 과정이 요구된다. 앞으로의 한국어 음성처리 기술 연구에 공동으로 사용할 수 있도록 하기 위하여는 성별, 나이별, 지역별로 화자들을 선정하여 음성을 녹음토록 하여야 하는데 이 경우 많은 시간과 노력이 필요하다.

3) 음성분할시의 고려: 녹음되어 디지털화된 음성 신호는 적절히 분할(labeling)하는 과정을 거쳐야 한다. 음성분할의 정확도를 유지하기 위하여 직접 음성 파형이나 스펙트럼을 관찰하면서 음성을 분할하여야 하는데 이 경우 음성 분할에 방대한 인력과 시간이 소요되기 때문에 자동 음성분할에 대한 연구개발이 또한 필요하다.

4) 데이터베이스 구조의 고려: 분할된 음성 데이터를 저장 이용할 수 있는 효율적인 데이터 베이스 구조의 연구 개발이다. 앞서서도 언급하였듯이 음성 데이터 베이스와 달리 음성파형 그 자체가 저장되는 것으로 그 양이 방대한 뿐만 아니라 음성분할 정보와 함께 효율적인 액세스 방안도 고려되어야 한다. 또한 데이터베이스들간의 상호연관, 수정, 보완 등이 가능하도록 구성되어야 하기 때문에 데이터 베이스 구조를 결정하는데 많은 노력이 필요하다.

5) 도구개발의 고려: 음성 데이터베이스 구축을 위한 도구로서 음성 신호처리를 위한 워크스테이션의 개발을 들 수 있다. 방대한 양의 음성신호를 녹음하여 디지털화 한 후에 음성분할 등을 수행하기 위해서는 고속이고 큰 용량의 저장공간, 정밀급 고속 A/D 및 D/A 변환기, 그래픽 소프트웨어, 시뮬레이션용 소프트웨어 등의 기능이 용이한 신호처리 속도가 충분히 빠른 컴퓨터 시스템이 요구된다.

IV. TTS의 고음질 합성방식

고음질의 한국어 TTS 시스템을 개발하기 위해서는 먼저 음성의 특징을 추출하는 해석 방법과 합성음을 내기 위하여 추출된 특징을 합성하는 방식에 대한 연구가 선행되어야 한다. 이러한 음성 해석 및 합성 방식은 데이터베이스의 메모리용량 및 합성음의 음질 등에 큰 영향을 미치므로 신중히 고려해야 할 사항이다. 기존의 음성 응답 시스템은 PCM 또는 ADPCM 등의 파형부호화 합성방식을 채택하여 데이터베이스를 구성하고 있는데, 이러한 방식은 방대한 양의 데이터 베이스를 구성하기에 부적절할 뿐만 아니라, 단어를 연결하여 문장을 구성할 때 좋은 음질의 합성음을 기대할 수가 없다. 따라서 위에서 서술한 문제점을 극복하기 위하여 다양한 합성방식에 관한 연구가 필요하다.

음성신호의 합성방식으로는 파형부호화, 신호원부호화, 혼성부호화법이 있으며, 메모리 절약을 위해서는 신호원부호화법을, 음질을 높이기 위해서는 파형부호화법이 주로 사용되고 있다.

파형부호화법은 음성의 정보를 발생모델에 따라 분리하지 않고 파형자체의 잉여성분을 제거한 후에 부호화하는 방법으로 PCM, ADPCM, ADM법 등이 제안되어져 있다. 최근에는 디지털 신호처리 전용칩의 제조기술과 파형부호화의 분석 및 합성방식법이 잘 개발되어 32k bps 전송율을 갖는 ADPCM의 표준화가 실현되어 있다[29]. 그렇지만 파형부호화법은 인간

의 개성과 감정을 나타내는 여기정보와 언어의 의사 전달을 나타내는 성도의 여파기정보를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다.

신호원부호화법은 분석시에 여기정보와 여파기정보를 분리시켜서 독립적으로 부호화하는 방법으로써 LPC, PARCOR, LSP 등의 알고리즘이 제안되어져 있다. 이들 알고리즘은 1k bps 이내로 전송율을 낮출 수 있기 때문에 메모리 효율적인 부호화법이 된다. 또한 분석시에 추출된 여기정보나 필터정보를 합성시에 인위적으로 변경시킬 수 있기 때문에 음절단위나 음소단위의 합성기법으로 적용하기에 용이하다. 그렇지만, 분석시에 성분을 분리하고 다시 그 정보를 이용해서 합성하기 때문에 분석시의 오차와 합성시의 오차가 합해져서 합성음의 음질은 자연성이나 명료성이 크게 떨어진다.

신호원부호화의 메모리 효율성과 파형부호화의 명료성과 자연성을 적당히 유지하기 위해 이 두가지 부호화법을 결합시킨 혼성부호화법으로는 MLPC, RELP, VELP, CELP법 등이 제안되어져 있다. 그렇지만 혼성부호화법에서는 성도의 여파기 정보를 부호화하는데는 신호원부호화법을 적용하고, 여기정보의 부호화에는 파형부호화법을 주로 적용하고 있다. 이 때문에 여기정보를 변경시켜야 하는 음절단위나 음소단위의 합성알고리즘으로 적용하기에는 문제점이 있다.

최근 메모리 제조업체에서는 칩당 16M bit를 집적화하여 사판하고 있고, 이것을 바이트단위로 쓰기 위해 8개를 사용하면 32k bps의 ADPCM 파형부호화법으로 합성하여도 $(16 \times 10^6) \times 8 / (32 \times 10^3) = 4000$ 초의 음성 데이터를 수록할 수 있는 많은 양이 된다. 따라서 메모리 용량의 절약을 위해 합성시스템에 신호원부호화 합성법을 채택하는 것은 현실적이지 못하며 음질을 보장받기 위해서도 상용화될 음성합성 부호화법으로는 파형부호화이나 혼성부호화법이 바람직하게 된다. 그렇지만, 파형부호화법이나 혼성부호화법은 분석후 합성을 하는 분장단위의 합성법으로는 오랫동안 적용되었으나 음원의 변경이 용이하지 못하기 때문에 단어나 음절 및 음소단위의 합성기법으로는 사용되지 못하는 실정이다. 따라서 규칙합성이 가능한 파형부호화 합성방식이 필요하게 된다.

파형 및 혼성부호화법을 고음질 합성방식에 사하려면 음성 생성모델에서 성분 및 성도의 특성을 그대로 보존하면서 여기원의 피치를 변경시킬 수 있어야 한다. 고음질 합성에 사용하려면 무엇보다도 음성

신호의 포먼트정보와 위상정보를 훼손시키지 않은 피치변경법이 필요하다. 피치변경에 대한 연구가 최근 10년동안 다음과 같이 제안되어지고 있다.

일반적인 선형 예측시스템에서는 여기임펄스들의 시간간격을 변경함으로써 유성음의 피치가 간단히 변경되고, 지속시간(duration)은 예측계수의 갱신을 변경시킴으로 바뀌게 된다. Caspers와 Atal은 이러한 선형예측기법의 관점에서, 멀티 펄스여기원의 피치주기를 변경하기 위해 잉-추가나 제거 방식을 적용하였다. 또한 지속시간은 피치주기를 추가 또는 제거함으로써 변경하였다[30].

그렇지만 혼성부호화법인 멀티펄스-LPC법에서는 임펄스들 사이의 간격조절에 의해 피치만 간단히 변경될 수 없다. 왜냐하면 멀티펄스-LPC법에서는 각 임펄스들이 이전에 찾아진 모든 임펄스들의 효과를 고려하여 최소의 오류가 되도록 계산되기 때문이다. 임펄스위치의 변경은 합성파형의 왜곡을 초래할 뿐이다.

Stella와 Charpentier는 멀티펄스-여기원을 성도모델과 여기원의 한 결합으로 보는 관점에서, 이러한 문제를 해결하였다. 그들은 우선 멀티펄스분석법으로 다이폰단위의 데이터압축을 하였고, 합성시에는 원래 형태로 다이폰을 합성한 다음에, 위상(phase)코더를 적용하여 배율을 변경하였다[31].

Varga와 Fallsides는 파형 편집부분과 선형예측에 의한 합성부분으로 구성된 혼성방식을 적용하여 파형부호화의 피치변경을 시도하였다[32]. 피치주기를 인상하기 위해 음성표본의 일부분과 선형예측 여파기계수를 사용하여 일반적인 합성여파기에 의해 합성하고 연결부분을 평활화하였다. 합성여파기의 초기조건으로 이들을 사용하면, 합성여파기는 별도의 새로운 여기원없이 필요한 음성의 나머지부분을 간단히 발생할 수 있지만, 피치주기를 줄이는 경우에는 단순히 한 피치 구간과형의 일부분을 제거하고 평활하는 것으로 피치변경을 시도하였기 때문에 피치가 짧은 여성이나 어린 화자의 경우에는 스펙트럼왜곡이 많아진다.

Quatieri와 McAulay는 그림 4-1과 같이 주파수영역에서 위상을 보존하는 피치변경법을 제안하였는데 [39] 이것은 입력된 음성에 대해 진폭 및 위상스펙트럼을 추출하여 별도로 처리하는 방법이다. 진폭스펙트럼에 대해서는 두드러진 스펙트럼 봉우리들을 추출한 다음에 이것을 피치변경율(ρ) 만큼 인터폴레이션하여 진폭스펙트럼의 피치를 변경시킨다. 위상스펙트럼에 대해서는 시간영역에서 구한 피치 개시시

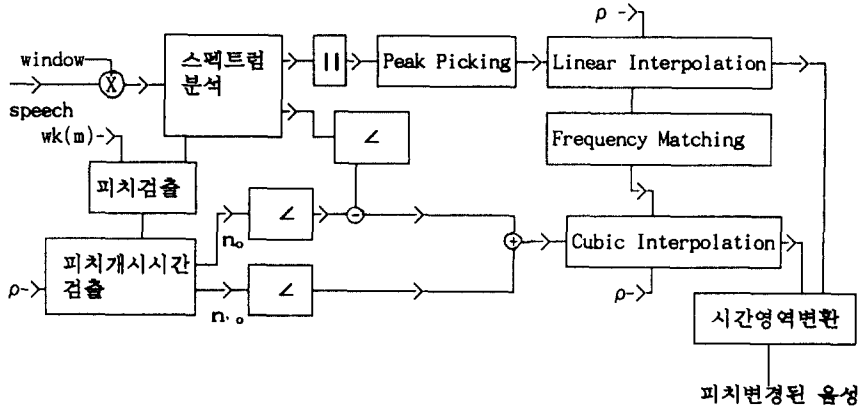


그림 4-1. 피치변경에 대한 분석 및 합성

간(pitch onset time)에 해당하는 위상을 제거하고 나서 피치가 변경되었을 때의 새로운 피치 개시시간의 위상을 더해줌으로써 새로운 위상을 구성하게 된다.

이러한 방법은 과형의 꼴을 그대로 유지하기 때문에 프레임단위로 분석처리하는 통상의 처리법에서 인접 프레임간의 연결이 아주 용이해 진다는 장점이 있다. 그렇지만, 피치변경시에 피치주기와는 별도로 피치의 개시시간을 공급해 주어야 하고, 또한 진폭 스펙트럼상에서 두드러진 봉우리 위주로 고조파의 인터폴레이션을 수행하기 때문에 스펙트럼의 왜곡이 높아진다는 단점이 있다.

V. TTS시스템의 고음질화에 따르는 문제점

음성 합성기술의 현황은 상기의 실제적인 응용기술을 가능하게 하지만, 자연성의 음질이 요구되는 여러 다른 응용에서는 아직도 어려움이 많다. 한국에 있어서 음성 합성분야에 대한 문제 해결의 답은 단일한 양상이 아니다. 현재의 연구는 다음의 몇가지 주제로 집약된다[16-19].

첫째로, 출력 음성의 자연성을 높이기 위한 음운 및 운율처리에 대한 기술이 확보되어야 한다. 보다 자연적인 억양과 지속시간을 생성하려면 아주 많은 운율 규칙들이 필요하게 된다. 자연음에서 바로 추출된 자연 운율이 다른쪽 합성발성에 부과되었을 때 자연성에서 중요하게 개선되어야할 사항들이 얻어진다.

두번째로는 합성기가 자연성을 유지하려면 현재 합성중인 내용이 무엇을 말하고 있는가를 이해하는 것이 필요하다. 현재 의미모델은 아주 빈약하기 때문

에, 운율구조적인 측면에서 의미하는 엔코딩 알고리즘들과 의미를 내포하는 아주 정확한 표현이 필요하게 된다.

세번째로는 고품질의 합성음질을 좌우하는 것으로 합성단위를 적절히 선정하는 방법이 필요하다. 단위 음성의 자동작성이 독일어의 다이폰, 일본어의 CV음질의 자동작성법[26], 한국어의 계통적 작성법에 의해 고품질의 음성이 합성되었음이 보고되었다.

네번째로, 자연스러운 합성단위간의 연결방식에 대한 기술을 고려하여야 한다. 즉, 스펙트럼의 불연속성이 최소화될 수 있도록 음성단위를 접속하는 기술이 필요하다. 독일어의 경우 반음절 접속법, 한국어의 경우 자음의 변이음을 고려한 접속법과 반음절 접속법이 연구되었다.

다섯번째로, TTS를 구현하기 위한 핵심기술은 일단 합성방식이 정하여지면 그 합성방식에 따르는 파라미터를 정확히 추출하는 기술이 필요하다. 그렇지만 현재 음성 합성방식으로 적용되고 있는 음성 생성 모델인 여기원/여파기 모델은 50년전의 것이기 때문에 몇가지 결함이 알려져 있다. 여기원과 여파기는 서로 강한 상관관계를 갖기 때문에 이들을 분리하여 모델링하는 것은 바람직하지 못하다. 또한 음향 여파기의 현재 모델들은 성도내에서 음파의 진행이 평면과의 전파를 가정하지만 실제의 소리기관은 그렇지 않다. 반면, 음향과는 선형과동방정식이나 또는 현재 많이 쓰이는 그의 선형 예측등가식 보다는 비선형 부분 미분방정식들의 아주 복잡한 집합에 의해 지배를 받는 공기흐름의 결과이다.

여섯번째로, 추출한 합성 파라미터들을 효율적으로

로 저장하고 이용할 수 있도록 데이터베이스화 하는 기술을 생각하여야 한다.

끝으로, 이들 합성기를 실시간 합성기로 구현하기 위한 DSP 칩 등을 다루는 하드웨어 기술을 가지고 있어야 한다.

VI. 결 론

정보화 사회에서는 정보통신은 사람과 사람사이의 정보교환에 컴퓨터등 편리한 기기들이 중간에 개입되는 새로운 정보통신의 형태가 상당한 비중을 차지하게 되었다. 따라서 사람과 컴퓨터 사이의 정보전달을 원활하게 하는 인간-기계 접속기술의 중요성이 대두되었다. 인간-기계 접속기술에는 음성인식기술(음성인력가능)과 음성합성기술(음성출력가능) 등을 들 수가 있다. 결과적으로 기계가 사람처럼 그 처리 결과를 자연스럽게 음성으로 들려주는 음성 합성기술은 인간이 현대의 정보통신 관련 서비스를 효율적으로 이용하기 위해서 반드시 필요한 기술이다.

우리는 지금까지 TTS 합성법의 응용, 구성원리, 요소기술 등에 대해 간단히 알아보았다. TTS 시스템이 수많은 응용분야에 대해 유용하게 만들어졌다 할지라도 명료성은 우수하나 자연성이 부족하기 때문에, 지금으로서는 제한된 응용분야에만 사용되고 있다. 인간-기계접속이 필요한 모든 분야에서 TTS시스템을 응용하려면 지금까지의 기술을 발판으로 하여 자연성이 유지되는 고품질의 음성 합성기술이 이루어져야 한다. 고품질의 음성 합성시스템을 구현하기 위해서는 TTS 구성의 요소기술에 대한 전반적인 연구개발이 필요하다. 현재의 기술로 특히 부족한 것은 언어처리 및 구문분석 기술의 알고리즘화, 언어 인지도 모델의 알고리즘화, 음성 생성모델링의 개선, 데이터베이스기술, 고품질 음성 합성 방식기술 등이다.

국내에서 음성 합성분야에 관한 연구 및 실용화 개발 여건은 그 정도가 매우 빈약한 실정이며 특히 부존자원이 빈약한 국내의 여건으로 볼 때 고부가가치를 창출할 수 있는 이에 대한 연구개발은 시급히 서둘러야 할 과제이다. 또한 한국이 음성 합성기술은 사용 언어에 따른 우리나라 고유의 문체이므로 이를 해결하기 위한 국내 기술의 독자적인 연구개발이 절실히 필요하다.

참 고 문 헌

1. W. von Kempelen, "Le mecanisme de la parole suivi de la description de un machine parlante," J. V. Degen, Vienna, 1971.
2. H. Dudley, "Remaking Speech," J. Acoust. Soc. Am, Vol. 11, pp. 169-77, 1939.
3. C. H. Coker, "A model of articulatory dynamics and control," Proc. IEEE, Vol. 64, pp. 452-460, 1976.
4. D. Markel, and A. H. Gray, Linear Prediction of Speech, Springer-Verlag, Berlin, 1976.
5. J. Bachenko, E. Fitzpatrick, and J. Lacy, "Implementing Prosodic Phrasing for an Experimental Text-to-Speech System," J. Acoust. Soc. Amer., Vol.81, No.51, pp. S79, 1987.
6. D. Hindle, "Deterministic Parsing of Syntactic Non-fluencies," Proc. 21st Meeting of the ACL, pp. 123-28, June 1983.
7. J. Hirschberg, and J. Pierrehumbert, "The Intonational Structure of Discourse," ACL-86 NY, pp. 136-44, 1986.
8. C. H. Coker, "A Dictionary-Intensive Letter to Sound Program," J. Acoust. Soc. Amer., Vol.78, No.51, pp. S7, 1985.
9. R. W. Sproat, "Stress Assignment in Complex Nominals for English Text-to-Speech," Proc. Tutorial and Research Workshop on Speech Synthesis, Autrans, France, pp. 129-132, 1990.
10. K. Church, "Stress Assignment in Letter to Sound Rules for Speech Synthesis," ICASSP Proceedings pp. 2423-2426, 1986.
11. A. L. Buchsbaum, "Drcomp: A compiler for Phonemic Duration Rules," TM 11227-860606-860606-08, June 6, 1986.
12. J. D. Anderson, J. B. Pierrehumbert, and M.Y. Liberman, "Synthesis by Rule of English Intonation Patterns," ICASSP Proceedings pp. 2.8.1-4, 1984.
13. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. America, Vol. 50, pp. 637-660, 1971.
14. J. P. Olive and N. Spickenagel, "Speech Resynthesis from Phonemic-Related Parameters," J. Acoust. Soc. Am., Vol.59, No.4, pp. 993-996, April 1976.

15. B. Caspers and B. Athal, "Role of Multi-Pulse Excitation for Synthesis of Natural-Sounding Voiced Speech," ICASSP Proceedings pp.34, 14, 1987.
16. J. Allen, S. Hunnicutt, and D. Klatt, From text-to-speech : the MITalk System, Cambridge University Press, Cambridge, U.K., 1987.
17. J. L. Flanagan, "Speech Analysis, Synthesis and Perception, 2nd ed., Springer-Verlag, Berlin, 1972.
18. J. N. Holmes, Speech Synthesis and Recognition, Van Nostrand Reinhold, 1988.
19. R. Linggard, Electronic Synthesis of Speech, Cambridge University Press, Cambridge, U.K., 1985.
20. R. J. Perdue and E. L. Rissanen, "CONVERSANT 1 voice system : architecture and applications," AT&T Tech. J., pp. 34-47, Sep/Oct. 1986.
21. A. J. Uhrbach and M. R. Diltz, "Concatenative synthesis : problems and prospects," Proc. of Speech Tech '88, pp. 37-39, 1988.
22. D. H. Klatt, "Review of text-to-speech conversion for English," J. Acoust. Soc. Amer., Vol. 82, No. 3, Sep. 1987.
23. P. Wallich, "Putting speech recognizers to work," IEEE Spectrum, pp. 55-57, Apr. 1987.
24. J. Mariani, "Speech technologies in western Europe-a review," Speech Technology, pp. 48-57, Aug/Sep. 1985.
25. 이용주, 김형태 외, "음성 데이터 베이스의 구축에 관하여," 한국음향학회지, Vol.7, No.5, 1988.
26. S. Chiba, "Recent speech technology topics in Japan," Speech Technology, pp.38-42, Aug/Sep. 1985.
27. G. Bristow, *Electronic Speech Synthesis*, McGraw-Hill, 1984.
28. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Cliffs, NJ : Prentice-Hall, 1984.
29. J. B. Reimer, M. L. McMahan and M. M. Arjmand, "32k bps ADPCM with the TMS 32010," Appl. Report, Texas Instruments, Dallas TX, 1985.
30. B. E. Caspers and B. S. Atal, "Changing pitch and duration in LPC synthesized speech using multipulse excitation," J. Acoust. Soc. Amer., suppl., Vol.73, No.1, pp.55, Spring, 1983.
31. M. G. Stella and F. J. Charpentier, "Diphone synthesis using multipulse coding and a phase vocoder," in Proc. IEEE ICASSP '85, pp.740-744, 1985.
32. A. Varga and F. Fallside, "A technique for using Multipulse Linear Predictive Speech Synthesis in Text-to-Speech Type Systems," IEEE Trans. Acoust., Speech, Signal processing, Vol.ASSP-35, No.4, pp. 586-587, April 1987.
33. 강동규, 김을제, 배명진, 안수길, "음성 파형의 halving 기법에 의한 파형코딩의 피치변경에 관한 연구," 한국음향학회 추계발표회(국제음향학회) 논문집, pp.107-111, Nov.10, 1990.
34. 배명진, 이미숙, 이해균, 안수길, "켈스트럼 분석에 의한 음성 파형코딩의 피치변경에 관한 연구," 제4회 신호처리합동학술대회 논문집, Vol.4, No.1, pp.304-309, September 1991.
35. M. Bae, H. Yoon, S. Ann, "On Altering the Pitch of Speech Signals in Waveform Coding- Alteration Method by the LPC and the Pitch Halving-," J., Acoustics of Korea, Vol.10, No.5, pp.11-19, October 1991.
36. 민경중, 배명진, 윤희상, 안수길, "음성 파형 코딩의 음원피치 변경에 관한 연구," 한국음향학회 학술발표회 논문집, Vol.10, No.1(s), pp.45-49, Nov.9, 1991.
37. 박찬수, 김형태, 배명진, 안수길, "음성파형 분절의 스프딩 기법에 관한 연구," 한국음향학회 학술발표회 논문집, Vol.10, No.1(s), pp.7-10, Nov.9, 1991.
38. 이미숙, 배명진, 안수길, "전, 후방향 LPC에 의한 음성 파형 분절의 연결부분 스프딩," 한국음향학회 학술발표회 논문집, Vol.10, No.1(s), pp.15-20, Nov.9, 1991.
39. T.F. Quatieri, R.J. McAulay, "Shape Invariant Time-Scale and Pitch Modification of Speech," IEEE Trans., Signal Processing, Vol.40, No.3, pp. 497-510, March 1992.
40. M. Bae, M. Lee, "On a Pitch Change of the Waveform Coding by the Cepstrum Analysis for Speech Waveforms," J., Acoustics of Korea, Vol.11, No.4, pp.14-21, August 1992.
41. U. Kim, J. Lee, M. Bae, S. Ann, "On a Pitch Change of the Waveform Coding by the Cepstrum Analysis for Speech Waveforms-on the Phase Compensation for the Pitch Change-," J., Acoustics of Korea, Vol. 11, No.5, pp.51-58, October 1992.
42. 유건수, 이성우, 배명진, 안수길, "고조파 스케일

링에 의한 주파수영역 피치 변경에 관한 연구,"
대한전자공학회 하계종합학술대회 논문집, pp.
637-640, June 1992.

43. 이해군, 이미숙, 한기남, 배명진, "스펙트럼 평탄
화에 의한 음성신호의 피치변경법," 대한전자공
학회 추계종합학술대회 논문집, pp.627-630, No-
vember 1992.

44. S. E. Levinson, J. P. Olive, J. S. Tschirgi, "Speech
Synthesis in Telecommunications," IEEE Commu.
Magazine, Vol.31, No.11, pp.46-53, November 1993.



배 명 진

- 1956년 5월 20일생
- 1981년 2월 : 숭실대학교 공대 전자공학과(공학사)
- 1983년 2월 : 서울대학교 대학원 전자공학과 석사
과정수료(공학석사)
- 1987년 8월 : 서울대학교 대학원 전자공학과 박사
과정수료(공학박사)
- 1986년 3월 ~ 1992년 8월 : 호서대학교 전자공학과
조교수
- 1992년 9월 ~ 현재 : 숭실대학교 공대 정보통신공학
과 부교수
- 주관심분야 : 음성신호처리, 음성통신