데이터베이스 분산을 통한 소용량 문자-음성 합성 단말기 구현

김 영 길, 박 창 현, 양 윤 기
수원대학교 정보통신공학과
전화 : 031-220-2532 / 핸드폰 : 016-9526-0901

Implementation of text to speech terminal system by distributed database

Young-Kil Kim, Chang-Hyun Park, Yoon-Gi Yang
Dept. of Information Telecommunication Engineering, The Univ. Suwon
E-mail : ykcicada@mail.suwon.ac.kr

Abstract

In this research[1], our goal is to realize Korean Distribute TTS system with server/client function in wireless network. The speech databases and some routines of TTS system is stuck with the server which has strong functions and we made Korean speech databases and accomplished research about DB which is suitable for distributed TTS. We designed a terminal has the minimum setting which operate this TTS and designed proper protocol so we will check action of Distributed TTS.

I. 서론

유무선 통신 기술의 발달로 정보통신기기와 인간과의 인터페이스가 보다 긴밀하여질 필요성이 증가하고 있다. 이러한 HCI(Human Computer Interface)의 기본적인 응용이 문자-음성 합성기(text to speech : 이하 TTS) 일 것이다 [1-5]. 전형적인 TTS는 문자의 코드를 분석하여 언어학적인 분석과 이를 다양한 신호처리 기법을 통하여 주어진 음성데이터 베이스를 사용하여 합성하는 방식을 사용한다. 신호처리 기법은 크게 데이터베이스의 크기가 작은, 파라미터를 사용하는 방식과 데이터베이스의 용량은 크지만 음질이 탁월한 시간영역 합성방식이 있다 [1-5]. 어느 경우에나 음성 데이터베이스를 TTS 시스템이 직접가지고 있어야 한다. 그런데, 최근에 무선망의 발달로 SMS(Short Messag- ing Service)와 같은 문자메시지가 많이 사용되고 있으며, 음성으로 전달하는 것보다 저렴한 비용으로 정보를 이용할 수 있다. 그런데, "차량의 이동 중" 과 같은 상황에서, SMS를 음성으로 전환하여줄 필요성이 많이 증가하고 있다. 그 대표적인 예가 최근에 소비자에게 선보인 무선 네트워크 기능이 있는 taxi이다. 현재 이러한 시스템에 TTS를 구현하는 방식은 단말기에 TTS 시스템을 구현하는 수밖에 없다. 이는 고비용, 고성능을 요구하게 되고, 단말기 시장의 특성이 대량생산을 통한 원가절감이 매우 강한 특성이 있으므로, 이러한 문제를 합리적으로 처리하기 위하여 최근에 분산 시스템 기법을 TTS에 도입하려는 움직임이 포착되었다. 분산 TTS는 음성데이터 베이스와 일정량의 프로세싱은 사업자의 강력한 서버에서 담당하고 가입자의 단말기에서는 간단한 프로세싱만 하면 전체적인 TTS가 구현되는 방식이다. 이러한 기술은 선진국에서는 특허출원중이며, 중요기술은 아직 공개되고 있지 않은 상태이다 [1-5].

따라서, 본 연구에서는 무선 데이터 망에서 단말기의 하드웨어의 복잡성을 증가시키지 않고 강력한 서버 기능을 사용한 분산 한국어 문자 - 음성 합성기(Distributed TTS)를 개발하였다.

1) 본 연구는 과학재단 산학 협력 연구(I01 - 2002 - 000 - 00100 - 0)의 지원으로 수행되었음

II. 분산 음성합성 아키텍처

본 연구에서는 무선망에서 서버/클라이언트 기능을 통한 한국어 분산 TTS 시스템을 구현하는 것을 목표로 한다. TTS 시스템의 음성 데이터베이스와 일부 루틴은 강력한 기능의 서버에 장착되어 있고 저가의 DSP(Digital Signal Processor) 와 적은 메모리를 갖는 단말기에 정하여진 계수를 갖는 TTS 파라미터를 무선망에 전송하여 단말기에서 최종적으로 합성음을 재생하는 것을 목표로 한다. 이러한 시스템이 갖는 장점은 다음과 같다. 첫째, 단말기 가격의 저렴화를 들 수 있다. 음성데이터 베이스가 단말기에 존재하는 것이 아니라 서버에 있고, 또한 전처리 과정 등이 서버의 프로세서에서 이루어짐으로, 단말기는 저가의 DSP와 적은 량의 메모리로 구성될 수가 있다. 이는 단말기의 개수가 서버에 비해서 월등히 많으므로 경제적인 효과가 있다. 둘째, 시스템 업그레이드의 용이성을 들 수 있다. 일단 출시된 단말기에서 TTS기능이나 음성 데이터베이스를 업그레이드할 때 서버의 음성 데이터베이스와 알고리즘만 개선하면 일괄처리가 되므로 시스템의 유지 및 보수, 업그레이드가 용이하다.

본 DTTS의 기본적 알고리즘은 크게 Server와 Terminal로 구분해볼 수 있고 다시 Terminal은 송신부와 수신부로 분류되어진다. 송신부는 우리가 합성하고자하는 문자정보를 입력하면 입력을 받아서 전송 protocol에 맞게 변환하여 서버로 전송한다. 서버는 이렇게 입력되어진 값은 구문분석 및 음운 변환시켜 해당되는 DB를 선별하고 이 DB의 위치를 분석해서 Global DB는 전송하여 주고 LocalDB는 DB 정보만을 전송하여 준다. 이렇게 서버에서 분석된 정보가 다시 수신부에 수신되어 DB를 합성하고 PSOLA알고리즘을 거쳐서 더욱 자연스러운 음성을 출력하여 준다. 그림 1은 이러한 알고리즘을 설명하여 주고 있다. 본 연구에서 합성한 음성합성방식 음절 단위 합성 방식을 택하였다. 음절 단위 합성은 자연성과 명령도가 뛰어나다는 특성을 가지고 있다. 본 연구소에서 개발한 음성DB 엔진을 통하여 음절단위로 음성 DB를 생성 분류하였다. 음절 단위 음성합성의 단점은 음의 길이 조절의 어려움 점이 있다. 입력 정보의 분석은 조합형으로 입력된 2 바이트의 한글문자의 초성, 중성, 종성을 분리하여 데이터베이스의 인덱스를 추출한다.

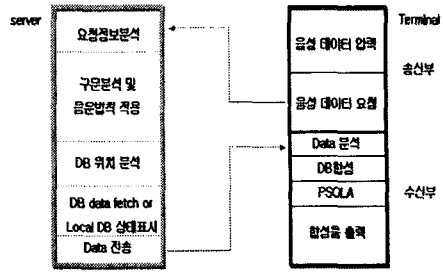

그림 1. 알고리즘 개요도

이 인덱스를 이용하여 구문분석 및 음운법칙으로 본 연구에서는 기초적인 구문분석과 음운법칙을 적용하였다. 적용한 법칙은 '끝소리 규칙', '구개음화' 등의 룰 베이스의 법칙이고 입력문장의 띄어쓰기, 마침표 등의 문장부호를 적절한 형태의 기호로 변환하여 합성신호의 중간에 삽입하여 자연도를 증가시키는 작업을 하였다. 음절연결 알고리즘의 동작에는 PSOLA 방식을 사용하여 시간영역에서 음절을 연결하는 알고리즘을 동작시킨다. 추출된 데이터베이스로부터 음절단위의 데이터를 불러오고 이전의 음절과 시간영역에서 중복하여 음절을 연결하여 이를 출력한다. PSOLA 방식을 사용하여 시간영역에서 음절을 연결하는 알고리즘을 동작시킨다.

III. 분산 알고리즘

음성합성알고리즘을 서버와 클라이언트(터미널)를 분할해서 합성을 한다. 터미널은 음성데이터의 Text를 입력하고 서버에 이 Text정보를 음절단위 패킷 형태로 DTTS 프로토콜을 통하여 서버에 전송하고 서버는 음성합성 연산을 수행하고 수신부에서 얻은 정보를 분석하여 Local DB유무를 판단하여서 Local DB이면 DB를 불러오고 Global DB이면 전송되어진 DB를 사용하여 순차적으로 DB를 합성하여 PSOLA 알고리즘을 거쳐 출력을 해준다. 서버에서는 클라이언트에서 얻은 프로토콜 정보를 통하여 구문분석 및 음운 법칙을 적용하여 DB의 위치를 판별하고 Global DB인 경우는 DB를 전송하고 Local DB인 경우는 그 상태만을 전송한다. 구문분석 및 음운 법칙적용 단계에서 앞 음절 및 뒤 음절의 상태에 정보를 통하여 연산이 이루어지기 때문에 링버퍼를 사용하여 계속적으로 전달되는 패킷정보를 저장한다. 기존의 TTS시스템은 모든 동작을 한 시스템에서 처리하지만 DTTS는 서버와 클라이언트가 합성알고리즘도 분산할 뿐만 아니라 Data Base(DB)분산의 개념을 도입하였다. 음절단위 합성시 요구하는 DB의 양은 초성19개 중성 21개 종성 27개의

조합으로 하면 10773개의 DB가 필요하다. 그러나 음음변환에서 끝소리 규칙을 적용하여 발음되는 종성 7개의 DB만 있으면 되고 잘 사용하지 않는 DB를 비슷한 DB로 대치하면 2000개정도의 DB면 무작위 음성합성을 할 수 있다. 전체 DB 2000개중에서 서버에 저장DB Global DB와 터미널 즉 클라이언트에 저장될 local DB를 분류해서 앞에서 이야기한 DB의 분산을 처리한다. 2000개의 DB중 Global DB와 Local DB의 분류기준은 사용빈도라는 하나의 기준으로 분류하였다. Local DB의 얼마만큼의 DB를 보유할 것인가의 문제는 클라이언트의 메모리와 채널에 걸리는 부하의 상관관계를 가지고 있다. 채널에 걸리는 부하를 최소화하기 위해서는 클라이언트 즉, 터미널에 Local DB의 보유량을 증가하면 될 것이고 시스템의 메모리의 용량을 감소시키기 위해서는 Local DB의 보유량을 줄이면 된다. 이것은 각 시스템에 따라서 유동적으로 조절이 가능하다. 본 연구에서 Local DB의 양 400개로 하도록 하겠다. 1개의 DB의 파일 Size가 Bit rate 8이고 Sampling rate 8000hz인 것을 DB의 시간간격은 0.5초인 DB의 데이터 크기는 4Kbyte 정도이다. 본 연구에서 Local DB를 400개로 정하여 전체 DB의 20%를 Local DB로 정하고 나머지 80%의 DB를 Global DB로 정하였다. 클라이언트와 서버의 통신프로토콜은 그림 2와 같은 구조를 갖는다.

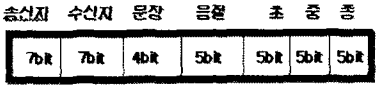

그림 2. 송신프로토콜

두 시스템 즉 서버와 터미널이 통신하기 위해서 DTTS에 맞는 프로토콜 정의가 필요하다. 본 연구를 DTTS 프로토콜을 정의하여서 PC상에서 실험과 ADSP2191보드에서 실험에서 적용하여 서로 통신을 하였다. DTTS 프로토콜은 전송프로토콜과 수신프로토콜로 구분되어진다.

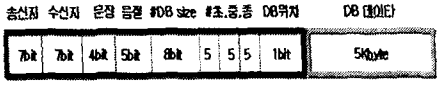

그림 3. 수신프로토콜

● 송신지: 송신프로토콜에서 음성합성을 요구한 클라이언트 주소를 전송한다. 수신프로토콜에서는 수신할 클라이언트의 주소는 표시한다.

● 수신지 : 음성정보를 보내줄 클라이언트를 설정한다.

● 문장/음절 : DB의 위치를 몇 번째 문장 몇 번째 음절에 해당하는지를 표시한다. 이 정보는 클라이언트에서 음성을 합성시 이용한다.

● 초/중/종: DB의 음절의 초성, 중성, 종성의 각 정보를 포함하며 이 정보를 이용하여 각종 음운법칙을 적용한다.

● DB위치: DB가 Local 또는 Global에 위치한 것인가를 표시한다. LocalDB는 DB 데이터를 전송하지 않아도 되고 GlobalDB는 DB위치 뒤에 DB 데이터를 같이 전송한다.

● DB 데이터 : 음절단위 DB는 Bitrate 8bit , Sampling rate 8000hz, DB 지속 시간이 0.5미만이기 때문에 DB 사이즈는 5Kbyte로 하였다.

IV. 단말기의 하드웨어 설계 및 제작

본 연구에서는 ADSP 2191M DSP 프로세서를 이용하여 합성 알고리즘을 처리하게 된다. 그림 4 에 본 연구에서 설계 제작한 DTTS 용 단말기의 블록선도가 제시되어있다.

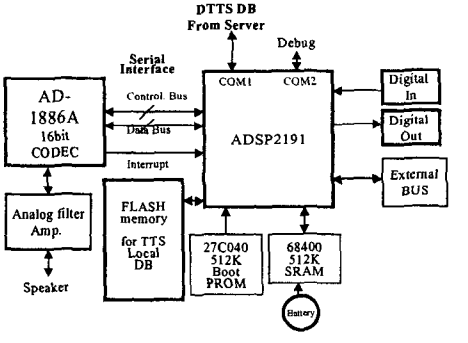

그림 4 DTTS 용 단말기의 블럭선도

본 시스템은 16M words 크기의 Address memory map을 가지고 있어 충분한 양의 외부 메모리의 사용으로 대용량의 데이터베이스의 저장이 가능하고 고성능의 산술 명령어를 이용한 문장에 따른 음원의 변형 및 실시간 처리가 가능하다. 또한 AD1886A 코덱의 장착으로 AC97기반의 고품질로 출력함으로서 음질의 향상을 도모한다. 기존의 PC 기반을 서버로 하여 단말기와 RS232로 연결되어 서버에서 문자 정보를 입력받아 단말기에서 기본적인 TTS 알고리즘으로 처리되어 음성출력을 하게 된다. 단말기의 메모리의 한계가 있으므로 사용 빈도수가 높은 음절단위의 데이터는 단말기에 저장되고 상대적으로 빈도수가 적은 데이터는 서버에서 제공을 받게 된다. 서버에서 문자 정보를 입력받아 문자 정보에 따라 서버의 데이터베이스와 단말기 내의 데이터베이스를 검색하여 단말기 내에 데이터가 없으면 서버로 요청을 하여 데이터베이스를 완성하고

알고리즘에 의해 출력한다. 그림 5에 제작된 DTTS 단말기의 실물사진이 제시되어 있다.

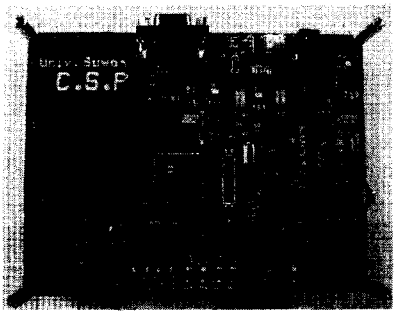

그림 5. 제작한 단말기

V. 시스템 통합실험

본 연구에서는 PC 3대, ROM EMULATOR(3.3 or 5V) 1대, SERIAL CABLE(RS-232) 2개, ADSP2192로 구현한 시스템 보드 2대를 사용하여 실험을 수행하였다. 그림 6에서처럼 서버와 클라이언트를 RS-232 Uart를 통신채널로 가정하고 DTTS 구현을 실험하였다.

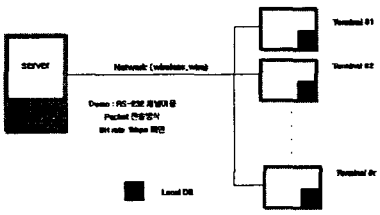

그림 6. 시스템 통합 실험 개요도

클라이언트는 크게 2종류로 하여 PC와 ADSP2191보드를 가지고 실험을 하였다. PC상에서는 서버와 클라이언트는 비주얼 C++ MFC를 이용하여 프로그램 하였다. 그림 6은 클라이언트 화면이다. Text 입력창에 "안녕하세요 반갑습니다"라로 입력하고 Send Data를 클릭하면 각 음절별로 분석하여서 DTTS 프로토콜에 음절별로 정보를 실어서 Uart를 통하여 서버에 송신한다. 서버는 이 정보를 수신지 정보를 이용하여 수신지로 정보를 보낸다. 수신정보를 받은 클라이언트는 앞에서 설명한 분산과정을 거쳐 음성출력을 얻을 수 있다. 음성출력은 입력문장의 음절의 수신프로토콜이 모두 수신되어지면 음성DB를 결합해서 PSOLA 알고리즘을 거쳐 합성음이 출력되어진다. 그림 7에 서버측의 메뉴창이 제시되어 있다.

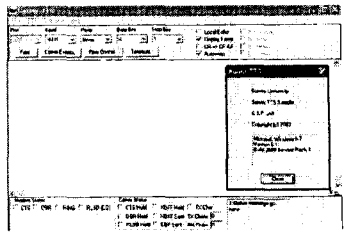

그림 7. 서버 DTTS 화면

VI. 결론

본 연구에서는 서버와 단말클라이언트에 TTS 데이터베이스를 분산한 분산음성 합성기의 하드웨어를 제작하여 실험하는 일련의 작업을 완성하였다. 단말기는 저가의 DSP를 가지고 음성합성을 수행할 수 있었으며 이는 추후에 무선망에서도 활용 가능한 매우 상업적인 가치를 지닌 것이라 할 수 있다. 음성데이터 베이스를 작성하고 한국어의 통계적인 특성에 따라 과학적으로 데이터베이스를 분산하였는데 분산하고자 하는 데이터베이스의 양은 현재 단말기에서 보유 가능한 메모리의 양에 전적으로 의존한다. 따라서 플래시 메모리 등의 가격하락에 따라 유연하게 분산할 수 있는 방식을 고려하여볼 수 있다.

참 고 문 헌

[1] Thierry Dutoit, "An Introduction to Text-to-Speech Synthesis", Kluwer Academic Publisher (Dordrecht), 1997, ISBN 0-7923-4498-7, 312 pages. Volume 3 in the series on Text, Speech and Language Technology.

[2] Douglas O'Shaughnessy, **Speech Communication: Human and Machine,** Addison Wesley series in Electrical Engineering: Digital Signal Processing, 1987.

[3] T.V. Raman, **Auditory User Interfaces - Toward The Speaking Computer,** Kluwer Academic Publishers, Boston, ISBN 0-7923-9984-6, August 1997, 168 pp.

[4] D. H. Klatt, "Review of Text-To-Speech Conversion for English", Jnl. of the Acoustic Society of America (JASA), Vol 82, pp 737-793.

[5] Eds, G. Bailly & C. Benoit, "Talking Machines, Theories, Models and Designs" (Elsevier: North Holland)

[6] I. H. Witten. **Principles of Computer Speech,** London: Academic Press, Inc., 1982.