

음성기반 회의록 시스템 설계 및 구현

표상호, 이상훈
국방대학교 전산정보학과
e-mail:shpyo00@yahoo.co.kr

Design & Implementation of Audio Minute System

Sang-Ho Pyo, Sang-Hoon Lee
Dept. of Computer Science, KNDU University

요약

인터넷과 컴퓨터 하드웨어 및 통신망의 발달로 멀티미디어 분야는 고도 정보 사회의 핵심적인 영역으로 부상되고 있으며, 멀티미디어 데이터에 대한 사용자의 요구도 날로 증가하고 있다. 문서, 녹취록, 테이프 형태로 관리되는 멀티미디어 데이터의 경우는 파일의 속성(attribute)에 대한 검색만이 아니라 내용(content)에 대한 검색이 필요하고, 따라서 순차검색 위주인 기존방법은 데이터의 활용 측면에서 문제가 있다. 본 논문에서는 멀티미디어 자료 중 음성 데이터를 저장 관리하는 방안으로 음성 데이터를 화면으로 도시하고 음성의 시작점과 끝점구간을 키로 하여 자동으로 데이터베이스에 삽입한 후 여기에 주석을 입력하는 음성기반 회의록 시스템(Audio Minute System)을 구현하고 그에 따른 기대효과를 제시한다.

1. 서론

최근 사무실이나 조직에 있어서 인간의 활동을 컴퓨터에 의해 지원하는 것에 대한 중요성이 인식되어, 컴퓨터를 이용한 인간의 협조작업지원에 관한 연구, CSCW(Computer Supported Cooperative Work)에 대한 관심이 높아지고 있다[1]. 사회 조직 뿐만 아니라 군 조직내의 중요한 지휘결심이나 의사결정, 정보의 전파 혹은 교환은 회의나 세미나를 통해서 이루어진다.

기본적인 회의록의 필요성은, 회의가 시작되어 끝날 때까지 회의에서 처리한 모든 사항을 회의에 참석하지 않은 여러 사람들에게 알리고, 기록으로 남겨 두기 위해서 작성한다. 현재의 회의록 작성 방법은 사람의 말을 정확하고 신속하게 기록하는 것으로 언어속도와 일치하게 문자화하는 속기, 음성 녹음을 서면 증거화 하는 녹취, 비디오 카메라 등을 이용한

촬영 등의 방법이 있다. 하지만, 이 같은 방법은 대화자의 실제 의사를 잘못 이해하고 작성하는 경우나 결론에 이르는 과정과 분위기를 표현하는데 한계가 있고, 원하는 자료만의 검색이 용이하지 않거나 대량의 저장공간을 요구하는 단점이 있다. 이를 해결하기 위한 방안으로 멀티미디어 대용량 데이터를 바이너리 형태로 DB 에 저장할수 있는 멀티미디어 데이터베이스 기술을 이용한 회의록 시스템을 구현하였다.

음성기반 회의록 시스템 AMS(Audio Minute System, 이하 AMS)는 음성데이터를 화면에 출력하고, 이를 이용하여 음성구간을 검출하고 그 구간에 해당하는 내용을 삽입하여 인덱스를 생성함으로써 검색 시 회의록에서 불필요한 부분을 조회할 필요가 없고 키워드에 의한 검색 및 발언자 또는 회의 요약별 검색 등, 고도의 기능을 지닌 회의록 시스템을 구현할 수 있게 되고, 또한 회의록의 이용률도 높일 수 있다.

본 논문의 연구범위와 구성은 다음과 같다. 1장에는 연구목적 및 범위에 대해 기술하고, 2장에는 관련 기술에 대해 소개한다. 3장에는 본 연구에서 개

※ 본 논문은 정보통신부의 정보통신산업 기술개발 지원사업에 의 수행된 연구결과임.

발한 음성기반 회의록 시스템 AMS의 특성 및 설계 내용을 기술한다. 4장에는 구현된 결과를 나타내고, 5장에는 결론 및 향후 연구 과제를 논의한다.

2. 관련연구

2.1 멀티미디어 데이터 베이스

일반적인 멀티미디어 데이터의 대표적인 특성은 첫째, 다양한 형태의 데이터로 구성되므로 특별한 형식이 없는 비정형 구조를 갖는다. 둘째, 텍스트와는 비교도 되지 않을 만큼 대용량의 저장 공간을 필요로 한다. 셋째, 서로 연관된 데이터가 정해진 시간과 순서에 따라 정확하게, 적시에 표현되어야 한다. 넷째, 멀티미디어 데이터의 검색에 있어서 기존의 순차적인 검색 방법으로는 원활한 검색이 곤란하다. 멀티미디어 데이터의 경우는 파일의 속성(attribute)에 대한 검색만이 아니라 내용(content)에 대한 검색을 사용해야만 한다[2]. 음성 데이터를 검색하는 방법에는 크게 주석을 기반으로 하는 주석기반 검색과 데이터의 내용을 기반으로 하는 내용기반 검색으로 나눌 수 있다. 내용기반 검색은 음성 데이터의 내용만을 기반으로 하며 이에 는 말한 단어를 이해하는 음성인식, 화자인식 등이 있으며 이는 단어 인식률, 앞말과 뒷말의 어학적/의미적 연결 등의 문제점으로 현실 적용에 어려움이 있다. 주석기반 검색은 음성 데이터에 대화 시간, 음성이 저장된 위치, 저장된 음성 데이터의 포맷등의 부가적인 정보를 부여하여 검색에 이용하는 방법으로 연구가 진행되고 있다[3][4]. 그러나 이러한 주석기반의 검색시스템은 입력자의 주관적 판단 등의 문제 그리고 시간적 효율성 문제로 인해 개선되어야 한다.

2.2 음성의 이해

2.2.1 사운드의 제어

윈도우 에서는 MCI(Media control Interface) 서비스를 제공한다. MCI만 조절함으로써 영상과 음성을 조작할수 있다. 하지만 High-Level 오디오 서비스인 MCI 함수 자체는 녹음과 재생만 할 수 있을 뿐 Wave 파일을 수정하거나 특정 부분만을 출력하는 등의 일은 할 수가 없다. Wave에 직접 효과를 주고 편집하고 파형을 도식하는 등의 작업을 하려면 로우레벨 오디오 서비스를 이용해야 한다.

2.2.2 끝점 검출

음성의 끝점검출(End-point detection)이란 언제 말을 시작하고 언제 말을 끝내는가에 대한 정보를 정

확하게 포착하는 기술을 말하며 음성인식을 위한 최초의 기본적인 행위이다[5]. 음성을 화면으로 도식하여 분석 할 때 두 가지 개념을 이해 해야한다.

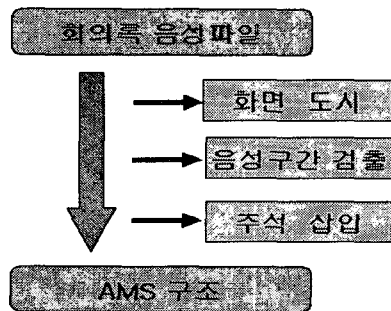
첫째, 절대에너지란 각 샘플의 절대값의 합으로 나타내는 값을 말하며 음성을 8bit 처리하면 256개(0-255)의 데이터로 표현하고 이때 0은 아주 저음을 255는 고음, 중간값 128은 묵음(아무소리도 없음)값이다. 프로그램 상에서 128을 기준으로 +,- 값으로 사용하며 축적된 에너지를 구하기 위해 절대값으로 표현한다.

둘째, 영교차율(Zero Crossing Rate, ZCR)이란 시간당 영점을 지나는 횟수를 말하며 무성음 검출시 사용한다. 유성음의 특징은 목젓을 울리는 소리이며 큰 에너지를 가진다. 즉 진폭이 길고 복잡하다. 무성음은 에너지가 무음과 큰 차이가 없지만 유성음보다 더 큰값을 가지는 것이 ZCR이다. 무성음은 128에서 위 아래로 더 많이 왔다 갔다 하는 특징이 있다.

3. AMS 의 분석 및 설계

3.1 AMS 의 인덱스구조

AMS 인덱스 구성과정은 <그림 1>에서 보는바와 같이 회의록 음성파일을 화면으로 도식하여 시간적으로 확대/축소 기능을 이용하여 음성의 시작점과 끝점구간을 검출하면 편집작업 시간의 단축과 정확성을 증대시키고 또한 의미 있는 구간을 편집자가 조정할 수 있다. 이 시작과 끝점을 키로 하여 자동으로 데이터베이스에 저장하고 주석을 삽입하는 방식이 음성데이터 관리방안으로 효율적이라 본다.



<그림 1> AMS 인덱스 구성과정

<표 1> 과 <표 2>는 AMS 인덱스를 테이블로 표현한 것으로 <표 1>은 회의록 번호, 회의제목, 발언내용(음성)등 회의록 관련 테이블을 나타낸다. <표 2>는 발언시작점, 발언끝점, 발언내용(주석) 등 발언관련 테이블을 나타낸다.

<표 3> AMS 개발 및 운영환경

회의록 관련 테이블(M_MASTER)		
테이블	내용	TYPE
M_NO	회의록 번호	VARCHAR2(8)
M_TITLE	회의 제목	VARCHAR2(20)
M_TYPE	회의 구분	VARCHAR2(20)
M_LEVEL	회의 구분	VARCHAR2(14)
M_M_NAME	참석자이름	VARCHAR2(100)
M_DATE	회의일자	DATE
M_PLACE	회의장소	VARCHAR2(20)
M_TIME	회의시간	DATE
M_SUMMARY	요약내용	VARCHAR2(200)
M_CONTENT	발언내용(음성)	BLOB

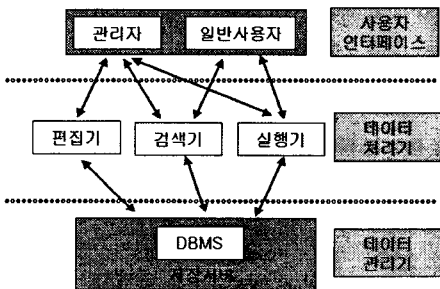
<표 1> 회의록 관련테이블

발언 관련 테이블(S_MASTER)		
테이블	내용	TYPE
S_NO	회의록번호	VARCHAR2(8)
S_START	발언시작점	VARCHAR2(10)
S_END	발언 끝점	VARCHAR2(10)
S_CONTENT	발언내용(주석)	VARCHAR2(200)
S_NAME	발언자	VARCHAR2(30)

<표 2> 발언 관련 테이블

3.2 AMS 구조의 분석 설계

AMS는 <그림 2>에서 보는바와 같이 사용자인터페이스, 데이터 처리기, 데이터 관리기로 구성된다.



<그림 2> AMS 의 구조

사용자 인터페이스는 사용자 권한을 구분하여 일반 사용자용과 관리자용으로 나눈다. 관리자모드는 편집 기능 및 확장검색이 가능하다. 음성데이터 처리기는 편집기, 검색기, 실행기로 구성된다.

데이터 관리기 에서는 음성데이터 파일을 BLOB (Binary Large Object) 필드에 저장하고, 회의관련 정보를 키(회의록번호+음성시작점+끝점)를 중심으로 관리한다. 개발 및 운영환경은 <표 3> 과 같다.

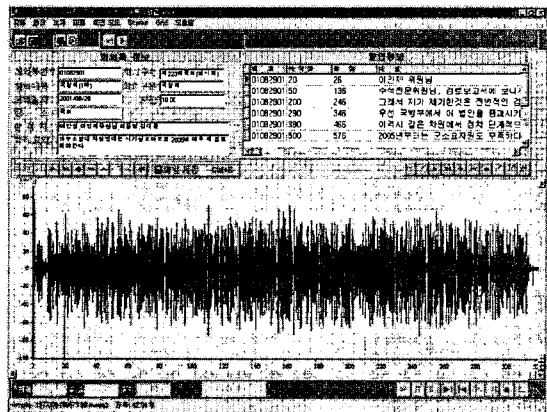
구분	OS	DB	개발도구
클라이언트	WIN95이상		DELPHI5.0
서버	WINNT4.0 이상	ORACLE8.0	

4. AMS의 구현

4장은 3장에서 설계한 AMS의 구현을 나타낸다.

4.1 편집기

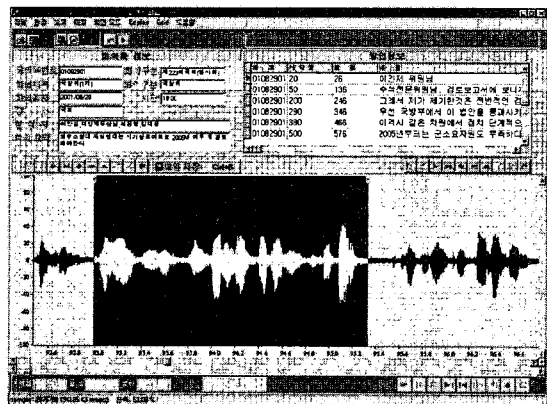
편집기를 각각의 기능중심으로 설명하면, 먼저 편집자는 편집할 회의록의 회의제목, 회의종류, 참석자, 회의장소 등의 정보를 화면 입력창에 입력한다. 다음 관련 음성파일을 파일 열기에서 지정하여 버퍼에 불러온다. 그러면 <그림 3>에서 보는 바와 같이 음성 파형이 화면 하단에 도시 된다.



<그림 3>편집 화면

이때 확대/축소 아이콘을 이용해 편집이 용이하도록 조절한다.

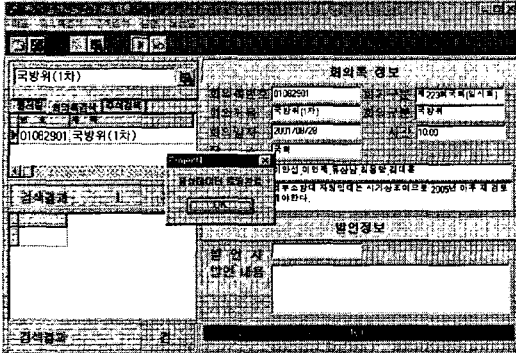
편집자는 적용범위를 마우스로 드래그 하여 선택영역을 설정하면 자동으로 선택구간 값이 DB 에 저장되고 편집자는 이 음성구간을 반복 청취 하면서 주석을 입력한다.



<그림 4>구간 확대 화면

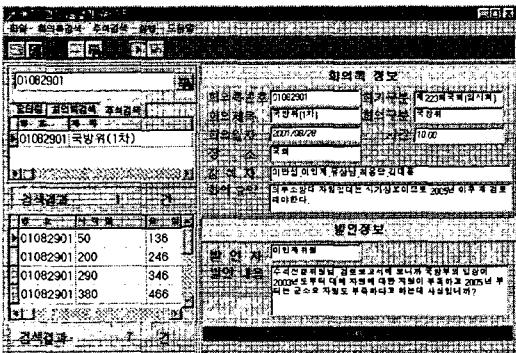
4.2 검색기 및 실행기

검색 기능은 검색창을 통해 편집창에서 편집자가 작성한 음성구간에 대한 주석 정보를 이용하여 일반사용자가 원하는 정보를 찾는 검색이다.



<그림 5> 회의록검색 및 데이터로딩

검색의 시나리오는 우선 회의록검색을 선택하여 회의록번호를 입력후 검색 버튼을 누르면 해당 자료가 화면왼쪽 상단에 디스플레이 되고 리스트를 선택한후 다운로드 아이콘을 누르면 BLOB 필드의 데이터가 클라이언트에 회의록번호.WAV 파일로 저장된다. 이때 이 파일을 지정한 후 로딩아이콘을 누르면 데이터로딩이 완료 되었다는 메시지와 함께 버퍼에 저장된다.



<그림 6> 주석검색 화면

다음, 주석검색 탭을 클릭한후 회의록번호를 입력하고 검색 아이콘을 클릭하면, 다운로드 받은 음성 데이터에 연결된 주석정보가 왼쪽 하단과 발언정보 영역에 디스플레이 된다. 이때 조회화면은 회의록에 대한 모든정보를 보여주며, 일반사용자가 편집할 권한은 없다. 왼쪽 하단 리스트를 클릭하고 실행 아이콘을 클릭 하면 단독 실행되고, 연속실행버튼을 누르면 선택위치이후부터 연속실행 된다.

4.3 분석

AMS의 장점은 음성데이터의 관리가 용이하고, 편집자가 직접 음성 파형을 보며 편집할 수 있다. 또한 주석 및 회의 정보를 통한 텍스트 검색을 제공하기 때문에 검색 속도가 빠르고, 리스트 형태의 검색을 통해 보다 손쉬운 검색환경을 제공한다. 현재 개발된 AMS에는 몇가지 개선점이 있다. 첫째, 동영상상에서 오디오 부분을 추출하는 기능이 추가되어야 한다. 둘째, 업무의 효율을 높이기 위해 자동 끝점검출 기능과 병행 해야 된다. 셋째, 편집작업이 용이하도록 사용자 인터페이스가 단순해야 한다. 즉 주석 입력은 키보드작업에만 의존하므로 모든 편집작업이 키보드만으로 작업가능하도록 해야한다. 넷째, 음성뿐 아니라 영상부분을 포함해서 내용기반 검색 기능이 추가적으로 필요하다. AMS의 구현에서 멀티미디어 제공 컴포넌트 6개와 응용 컴포넌트 3개를 이용하였다. 현재, 개발된 AMS는 초기 버전으로 로컬 환경에서만 테스트하였고 실제운영은 C/S(Client/Server) 환경에서 사용해야 한다.

5. 결론 및 향후 연구과제

지난 십여년간 멀티미디어 정보검색에 주석을 기반으로 한 연구가 많이 발표되었다. 그러나 음성데이터를 화면으로 도시하고 음성구간을 검출하여 주석을 입력하는 방법에 관한 연구는 드물다. 본 연구에서는 음성을 기반으로 데이터를 분석하여 저장 관리하는 회의록 시스템을 개발하였다. 분석 결과 미흡한 내용들을 개선해 나간다면 군 뿐만 아니라 여러 분야에서의 회의나, 세미나등의 귀중한 멀티미디어 자료들을 활용할 수 있는 좋은 시스템이 될 것이다.

참고문헌

- [1] Sang-Hoon Lee, "Dynamic Functions to Support Integrated Meeting Systems", IPSJ International Symposium on Information Systems and Technologies for Network Society, Fukuoka, Japan, 1997
- [2] 김명호,이윤준 "멀티미디어 개념 및 응용" 홍릉과학출판사, 1996
- [3] 허진용, 김인홍, 배종민, 강현석 "MPEG-2 압축 동영상 관리 시스템에 대한 연구" 정보과학회 '97 가을 학술발표논문집", pp 159-162, Oct. 1997
- [4] 김기욱, 김형주 "비디오 주석 시스템 설계 및 구현" 정보과학회논문지, Vol 24, No 6 pp.588-597, June 1997.
- [5] 김화수 「음성 인식 기술」, 국방대학원, 1997.