

MSF 파일형식 작성기 구성에 관한 연구

조철우, 정인화

창원대학교 제어계측공학과

Implementation of MSF encoder

Cheol Woo Jo, In Hwa Chung

Department of Control & Instrumentation Engineering, Changwon National University,

Changwon, Kyeongnam, Korea

cwjo@sarim.changwon.ac.kr

요약

본 논문에서는 기존에 제안되어 있는 MSF(Multimedia Sound File) 멀티미디어 파일을 이용하여 음성화일에 동화상 얼굴정보 및 문자정보를 부가할 수 있는 소프트웨어 및 임의의 소리화일에 동영상 얼굴정보를 부가하는 소프트웨어를 구현하는 과정과 결과를 설명한다.

조를 이용하여 변화하는 영상을 구현함으로써 많은 계산시간이 필요하였으며 취급해야 할 데이터의 양도 많아서 범용 PC시스템이 아닌 빠른 속도를 갖는 그래픽 워크스테이션급의 컴퓨터가 필요하였다.[1][2][3][4] 이를 보완하기 위하여 새로운 멀티미디어 파일형식 MSF가 제안되었다.[5] 본 논문에서는 MSF형식화일을 생성해 주는 부호화 프로그램을 작성하는 과정에 관하여 기술하고 그 결과에 대하여 논의한다.

1. 서론

음성을 출력할 경우 소리만으로 출력하기 보다는 소기와 영상을 같이 출력할 경우 인간의 인지율이 높아진다는 보고가 있다. 컴퓨터에서 음성이 출력될 경우도 영상정보와 같이 출력된다면 효과적인 정보전달이 가능하기 때문에 최근 영상과 소리를 결합시키는 연구가 많이 시도되고 있다. 이와 같은 시도는 영상인식 또는 컴퓨터 그래픽 기술과 음성인식, 합성등 음성처리기술을 결합하여 의사전달과정에 도움을 주고자 하는 시도의 일환이다. 기존의 연구중에도 2차원 또는 3차원 영상을 이용하여 영상정보가 부가된 음성 출력을 구현한 사례가 많이 있다. 그러나 그러한 시도들에서는 얼굴영상을 3차원 메쉬구

2. MSF 멀티미디어 파일형식

영상 및 소리를 포함하는 멀티미디어 파일형식으로는 AVI, MPEG등이 있다. 이러한 파일형식들은 소리와 함께 동영상 데이터를 압축된 형태로 내장하고 있기 때문에 크기가 커지게 된다. 그러나 응용분야에 따라 파일 크기가 크면 전송이나 보관등에 시간이 걸리기 때문에 좋지 않은 경우가 있다. 이와 같은 단점을 보완하기 위하여 고안된 것이 MSF형식이다.

MSF형식은 파일의 내부에 크기가 큰 영상 데이터를 직접 갖고 다니지 않고 정해진 영상데이터의 인덱스만을 소리화일에 부가된 형태로 갖고 있기 때문에 크기가 기존의 동영상화일에 비해서 아주 작아진다는 것이 장

MSF 파일형식의 응용에 관한 연구

점이다. 이러한 장점 때문에 인터넷에서 플러그인으로 동영상 애니메이션을 구현한다면 TTS의 출력을 애니메이션 형태로 내보내던지 할 경우 유용하다.

또 한가지 장점은 소리에 포함된 영상데이터를 선택하는 영상데이터베이스에 따라서 연세를지 바꿀 수 있다는 것이다. 예를 들어 사람의 음성에 따라 변화하는 입모양 영상을 부가할 수도 있고, 또는 관련된 다른 영상을 부가하여 볼 수도 있다.

그림1은 MSF파일형식의 구조이다.

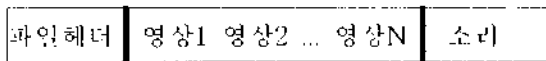


그림1. MSF파일형식

3. 파일작성기의 구성

MSF파일작성기는 음성신호 또는 소리신호를 원하는 부분으로 잘라서 시간축정보를 얻는 부분과 구해진 시간정보를 이용하고 필요한 영상데이터베이스를 선택하여 MSF형식으로 만들어 주는 부분의 두 부분으로 나눌 수 있다.

그림2는 MSF부호화기의 전체 구성을 나타낸 것이다.

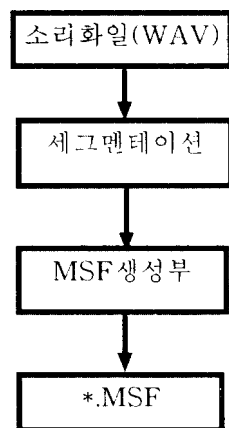


그림2. MSF부호화기의 구성도

세그멘테이션부

여기서는 음성파일을 MSF형식으로 변환하는 과정을 설명한다. 음성파일을 각 음소로 분할하기 위하여 그래픽 음성 편집기를 구성하였다. 편집기는 MATLAB으로 구현하였다. 그림 3은 편집기의 화면을 보여준다.

편집기에는 신호의 시간축 파형, 스펙트로그램, 에너지, ZCR(Zero Crossing Rate) 등이 표시되며 시각적 편집에 의해 각 음소의 시작시간, 지속시간을 기록한다. 이러한 작업의 결과로 시간정보 파일이 만들어진다.

그림4는 시간정보파일의 예이다.

시간(2바이트)	사건코드(2바이트)
10	s
25	a
35	z
54	a

그림4. 시간정보 파일의 형식

그림4에서 시간정보파일은 각 사건에 대한 시간정보와 사건코드를 2바이트씩으로 기록한 것이다. 사건코드는 음성의 경우는 해당 음소기호를 기타 동영상의 경우는 영상의 번호가 부여되게 된다. 영상의 번호는 이미 만들어져 공유하고 있는 영상데이터베이스에 의해 결정된다. 시간정보파일의 확장자는 tdd 로 한다.

MSF생성부

영상정보를 주어진 음성 파일에 부가하는 과정은 다음과 같다.

- (1) 시간정보 파일을 읽어 들인다.
- (2) 음성(또는 소리) 파일을 읽어 들인다.
- (3) MSF헤더부를 기록하여 새로운 파일로

지장한다.

이 과정에서는 tdd 파일과 wav형태로 기록된 소리화일을 읽어들이 각 시간구분에 따른 사건코드를 추출하고, 이에 따른 영상데이터 베이스를 선택한 뒤 tdd정보를 wav 파일의 앞에 부가하여 MSF형식으로 기록한다.

4. 실험 및 검토

이와 같은 방법에 의해 구성된 MSF화일은 적은 용량을 가지면서도 동영상 효과적으로 표시해 줄 수 있다. 본 실험에서는 음성 에 따른 입술영상데이터베이스와 음악소리에 따른 동영상애니메이션의 두가지에 대하여 구현해 보았는데 효과적인 동영상 구현이 가능하였다. 보다 자연스러운 영상의 구현은 세그멘테이션과정에서의 섬세함과 그래픽구현의 자연성을 통해 얻어질 수 있다.

앞으로 세그멘테이션과정에서 신호의 특성을 측정하여 자동으로 음소분 추출한다던지 음의 종류에 따라 감성적인 방법으로 적합한 영상을 찾아내어 부가하는 형태로 영상을 부가하는 지능형 부호화기를 개발할 예정이다.

5. 결 론

본 연구에서는 MSF형식을 쉽게 만들도록 도와주는 MSF 부호화기의 구현과정에 대하여 서술하였다. 현재는 세그멘테이션부와 MSF생성부가 별도로 구성되어 있으나 차후 통합된 형태로 구현된다면 MSF형식을 활용하는 데 유용하게 사용될 수 있을 것으로 생각된다. 차후에는 MSF형식을 인터넷을 통한 데이터표현등에 효과적으로 사용할 수 있도록 플러그인형태로도 구현할 예정이다.

감사의 글

이 연구는 1997년도 한국과학재단 핵심전문 연구과제 "멀티미디어 환경을 위한 정서음성

의 모델링 및 합성에 관한 연구"의 연구결과 의 일부입니다. 연구비를 지원해 주신 한국 과학재단에 감사드립니다.

참고문헌

1. Shigeo Morishima, Hiroshi Harashima, "A Media Conversion from Speech to Facial Image for Intelligent Man-Machine Interface", IEEE Journal on Selected Areas in Communications, pp.594-600, Vol.9, No.4, May 1991
2. W.Goldenthal, K.Waters, J-M Van Thong, O.Glickman, "Driving Synthetic Mouth Gestures: Phonetic Recognition for FaceMe!", Proc. Eurospeech'97, Rhodes, Vol.4, pp.1995-1998, 1997
3. F.Lavagetto, P.Lavagetto, "A New Algorithm for Visual Synthesis of Speech", Proc. Eurospeech'95, Madrid, pp.303-306, 1995
4. Jonas Beskow, "Rule-based Visual Speech Synthesis", Proc. Eurospeech'95, Madrid, pp.299-302, 1995
5. 조철우, 성인화, "애니메이션 영상을 이용한 멀티미디어 음성출력기의 구현", 창원대학교 정보통신연구소 논문집, 제2집, pp.147-150, 1998

MSF 파일형식의 응용에 관한 연구

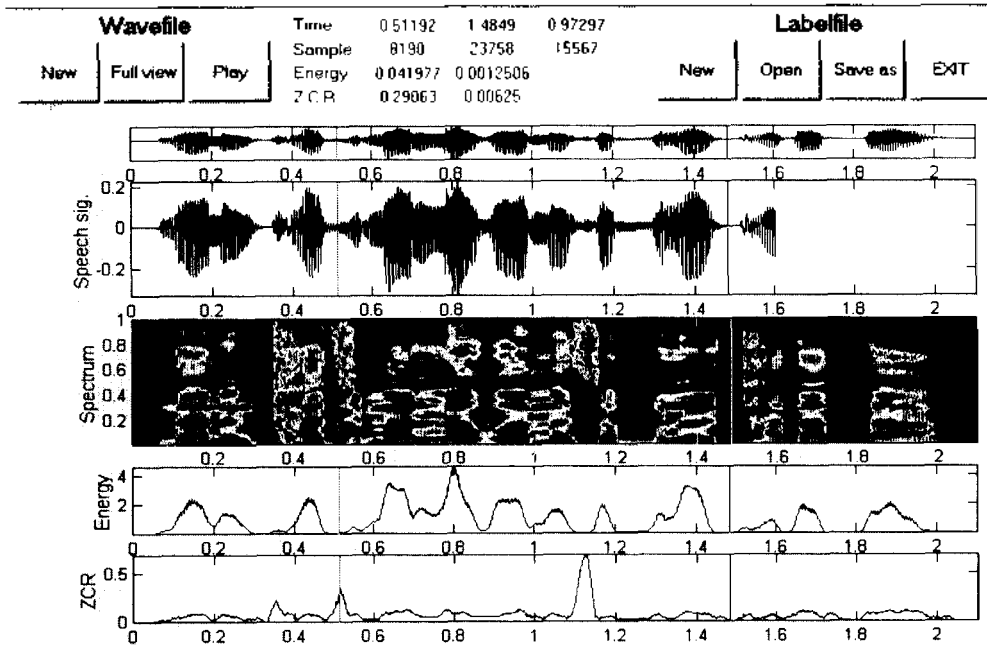


그림3 세그멘테이션 도구