A Study for Education Contents Production using Character Agent

Sang-yi Park*, Hea-Jung Lee, Suk-Tae Joung, Dept of Computer Engineering, Won-kwang University

요약
본 논문은 LipSynchro 소프트웨어 개발 카트(SDK)를 이용하여 기존 2차원의 정직된 이미지를 모션생성엔진, 음성분석, 인식엔진과 함께 연동함으로써 캐릭터의 움직임을 자동으로 생성하여 사실성이 높아진 캐릭터 에이전트를 만들어, 이를 멀티미디어 교육용 컨텐츠 저작에 통합하여 보다 나은 교육용 컨텐츠를 생성할 수 있도록 하였다.

1. 서론

교사와 학생의 요구에 따른 교육 컨텐츠를 개발하고, 이를 공유하여 교육정보화의 목표를 달성하는 것이 현 디지털 교육사업의 당연한 과제이다. 정보통신기술은 교사와 학생들이 지식정보사회에서 살아갈 수 있는 힘을 격증하는 교육을 하기 위한 핵심적인 위치를 차지하고 있으며 기존 텍스트만을 이용한 정보 전달에서 영상과 음향이 가미된 멀티미디아기술은 정보공유 및 전달을 위한 대안으로 제시되고 있다. 캐릭터 음성연동 기술은 멀티미디어 교육 및 정보전달을 위한 디지털 컨텐츠 제작에 안정성 및 편리성을 가져올 수 있으며, 또한 강의의 음성만으로 비디오 영상 클립을 생성함으로써 텔레미디어로 영상을 활용해도 향상된 키워드 및 비교하여 보다 나은 교육용 컨텐츠를 생성할 수 있도록 하였다. 바쁜 간편한 교육영상 컨텐츠 제작은 학생들에게 보다 많은 흥미를 유발시키고 유익한 정보를 빠르고 정확하게 전달할 수 있다. 현재는 인간과 컴퓨터간의 상호작용은 단순한 기계적 조작을 벗어나 좀 더 친숙한 형태로 나가고 있다. 그 중에서 극복적인 부분은 방 법으로 간주되고 있는 것이 음성과 시각을 통한 인터페이스이다. 이는 음성과 시각을 통한 면밀한 결과와 대립을 말하며, 이로 인해 음성강각, 음소인식, 캐릭터 에이전트와 연동되는 음성 처리기술 등이 필요하게 되었다. 이와 같은 기술들의 사용과 친숙한 형태를 가진 모델을 구현함으로써 인간과 컴퓨터간의 상호작용을 더욱 친밀하게 하는 인터페이스를 만들 수 있다.[1]

본 논문은 캐릭터 에이전트의 장점을 살려 멀티미디어 교육과 접목시키기 위해 LipSynchro 소프트웨어 개발 카트(SDK)[2]를 이용하여 모션생성엔진, 음성분석, 인식엔진과 함께 연동한다. 모션생성엔진과 연동하여 제작된 3DC캐릭터 에이전트는 마치 살아있는 것 같이 보여지고, 자동으로 3D캐릭터 에이전트의 변화와 음성과 음성의 자동음직임과 자동 생성의 기본적인 음직임을 생성하였다.

본 논문의 2장에서는 관련연구와 LipSynchro SDK 적용기술에 대해서 설명하였다. 3장에서는 시스템설계 및 구현결과를 보이고, 마지막으로 4장에서는 결론을 맺었다.

2. 관련연구와 LipSynchro SDK 적용기술
2.1 관련연구

따라서 본 논문에서는 편리한 인터페이스를 제공하는 SMIL 에디터에 캐릭터 아이얼을 이용한 플레이어를 결합시킴으로써 더 많은 정보를 정확히 인식할 수 있는 교육용 멀티미디어 컨텐츠 저작 시스템을 설계하였다.

2.2 LipSynchro SDK 적용기술
LipSynchro SDK는 입력된 음성에 분석/인식하여 자동으로 3D 캐릭터가 입력된 음성에 따라 움직임을 하게 된다.

2.2.1 음성압축(Speech Compression)
음성 압축은 효과적인 음성 신호의 처리로 14Kbps에서의 양방향 동시전송 방식을 지원한다. ITU-T Vocoder (G.723.1, G.723AB와 5.4Kbps, 4.8Kbps, 2.4Kbps, 1.2Kbps에서 사용되는 Vocoder)를 채택하여 최적의 음성 압축률을 적용한다.

2.2.2 음소일치(Phoneme Recognition)
음소일치는 인공지능 네트워크 기술(ART - Artificial Neural Networks Technology)에 중점을 두었다. 음소일치는 신경네트워크와 음소 데이터베이스를 이용하여 음성의 특성을 분석하고 음소를 분류하여 개발하였다. 음소일치 연산은 음성에서 음소

2.3.4 캐릭터 연동 음성 처리
사실과 같은 액션의 음직임은 얼굴 모델링 (Face Modeling)을 기반으로 생성되며, 이는 음성의 음직임이 3D모델의 프레임 시퀀스를 이용하여 보여진다. 3D 캐릭터의 입 모양을 실시간으로 연동시키는 과정은 음성 신호를 분석해서 음성 정보를 이용하여 받은 음소에 맞는 음직임을 보여 주게 된다. 캐릭터 입방울을 포함한 아이얼은 음성으로 자동 생성하게 된다.

3. 시스템 설계 및 구현
3.1 시스템 설계
기존에는 사용자가 SMIL을 기반으로 교육용 멀티미디어 컨텐츠 제작을 위해서는 메모장이나 워드프로세서를 이용해 SMIL의 태그를 수작해야하는 어려움과 이를 구현하는데 많은 시간과 노력을 요구하는 등의 문제점을 가지고 있었다. 또한 청각장애자들은 수업강의내용을 들을 수 없어 정보를 얻는데 많은 어려움을 가지고 있었다.

본 논문에서는 이러한 문제점을 해결하고 사용자가 좀 더 쉽게 SMIL을 이용하여 교육용 멀티미디어 컨텐츠를 제작할 수 있도록 SMIL과 캐릭터 아이얼을 기반으로 하는 교육용 컨텐츠 저작 시스템을 설계하였으며, 이 시스템에 의해 만들어진 컨텐츠는 오디오정보와 비주얼정보를 상호보완적으로 사용함으로써 교육의 효과를 향상 시킬 수 있다. [그림 1]는 캐릭터 아이얼을 이용한 교육용 컨텐츠 저작 시스템의 전체적인 시스템 구성도를 나타낸다.

본 연구실에서 개발한 멀티미디어 교육용 컨텐츠

[그림 1] 전체적인 시스템 구성도
저작 몰인 SMIL에디터와 캐릭터 에이전트 플레이어를 결합한 전체적인 에디터의 초기화면은 [그림 2]와 같다. 이 그림에서 보는 바와 같이 SMIL에디터의 사용자 인터페이스와 캐릭터 에이전트 테스트 플레이어, Wave 레코드가 동시에 작동된다.

[그림 2] 전체적인 에디터 초기화면

3.1 사용자 인터페이스

기존에는 사용자가 SMIL 기반의 컨텐츠를 제작하는데 있어 단순한 텍스트 편집기를 이용하여 직접 코딩해야 하는 불편함과 텍스트 편집 위주의 연속적인 단조로운 작업으로 인해 많은 시간과 노력을 요구하는 문제점이 있었던 것으로서, 이러한 문제점을 해결하기 위해 사용자 인터페이스 부분에서는 WYSIWYG 방식으로 밀리미터 단위로 공간적 동기화 정보를 설정하는 방식과 시간적 동기화 정보를 설정하는 시간편집기능으로 분리하였다.

사용자 인터페이스[그림 2] 참고는 그림의 사용자 인터페이스를 사용하여 SMIL 문법의 기본적 지식만을 가진 사용자가 쉽게 SMIL 문서를 작성할 수 있도록 설계하였으며, 공간적 동기화 정보를 설정하는 ①레이아웃 튜브와 시간적 동기화 정보를 설정하는 ②타임라인 튜브, 그리고 ③트리 튜브, ④속성 튜브, [그림 3]의 소스 튜브로 구성되어 있다.

[그림 3] SMIL에디터 소스뷰

3.1.2 캐릭터 에이전트 테스트 플레이어

캐릭터 에이전트 플레이어 모듈에서는 먼저 8KHz, 16bit의 wav음성녹음부분에서 mono의 녹음 방식을 사용하여 음성을 분석하여 Ls_record.dll을 통해 음성암축 및 음소인식을 하게 된다. 또한 3D미디어로 추출된 캐릭터 정보를 Ls_Player.dll을 통해 암축 표준을 인식하게 된다. 이렇게 어조분석 및 발음정보가 생성되고 음량 표준인식 정보를 함하여 캐릭터 연동 음성 처리(3D Phonem Visualization)를 통해 발음은 음소에 맞는 음직임을 보여준다. 마지막으로 웹용 멀티미디어 자료와 캐릭터 에이전트 플레이어가 결합된 기능을 수행하게 된다.

3.2 시스템 구현

본 논문에서는 SMIL 기반 교육 저작 시스템의 사용자 인터페이스와 LipSynchron 소프트웨어 개발 키트(SDK)를 활용한 캐릭터 에이전트 플레이어에 중점을 두고 설계하였다.

사용되는 시스템 환경은 패키징 64MB이상과 OS는 윈도우98이상의 버전을 사용하며, 먼저 Directx SDK 8.X 버전 설치를 하고 LipSynchron 라이브러리 파일(ls_player.dll, ls_recorder.dll) 두개를 레이스러브에 등록한다.

3.2.1. 음소(Phoneme)와 캐릭터

본 논문에서 사용되는 음절표정은 총 16가지로, 감정을 표현하는 음절표정으로는 기본표정, 두운 감은표정, 흔들음 감은표정, 오른쪽 눈 감은표정, 웃는 표정, 놀란 표정, 화내는 표정, 위쪽눈썹 음란표정, 오른쪽눈썹 음란표정의 총 9가지가 사용되며, 음소에 따른 음절표정은 a(아), o(오), e(에), f(프), s(쓰), m(멍, 응), n(우, 우) other(기타) 발음으로 총 7가지가 사용된다.

이와 같은 음절표정은 몽블래틀방식으로 디자인되었으며 이렇게 만들어진 각각의 음절표정에 Mesh name을 지정해주면서 프로그램에서 동작하게 된다. [그림 4]의 첫 번째 줄은 필요한 음소를 나타내고 있으며, 두 번째 줄은 감정에 따른 캐릭터의 음절표정을 나타내는 것으로 ①은 기본표정, ②는 웃음, ③은 화나, ④는 오른쪽 눈 감음, ⑤는 놀람, ⑥은 늘람을 표현하고 있다.

[그림 4] 음소(Phoneme)와 감정에 따른 캐릭터표정
3.2.2 캐릭터 에이전트 플레이어 기능 및 동작


[그림 5] 캐릭터 테스트 플레이어

먼저 웨이브 파일을 가져와 테스트하는 기능에 대해 살펴보면, [그림 6]과 같이 디지털 카테고리적 음악 이미지를 3D MAX를 이용하여 캐릭터의 기본모습을 나타내도록 하였다. (1) 이 캐릭터는 움직임을 음직이게 해주는 기능 (2)를 넘어갈 수 있으며 원하는 캐릭터를 선택 (3) 하여 바꾸어 줄 수 있다. 캐릭터 에이전트를 동작시키기 위해서는 음원을 가져와야하는 데 파일을 불러오는 버튼 (4) 가 필요하여 가져온 웨이브 파일과 캐릭터 에이전트와 병합시켜주는 기능 버튼 (5) 이 있다. 음원을 불러들이면 실행버튼 (6) 이 활성화되어 실행될때 캐릭터 에이전트가 소리에 따라 동작하게 되며, 중단할 경우 Stop버튼 (7) 실행하고 동작 중 일시정지와 해제는 (8) 버튼을 사용하면 된다.

[그림 6] 실행화면


4. 결론

본 논문에서는 인관에게 가까운 측면에서의 컴퓨터 사용을 통한 3D 캐릭터를 이용하여 캐릭터 에이전트 플레이어를 구현하였다. 이 캐릭터 에이전트 플레이어는 음성만으로 음소인식을 통해 정확한 발음에 따른 모양을 보이며 캐릭터 연동 음성 처리를 통해 발음에 따른 머리, 눈동자, 눈썹, 녹음백그라운드의 다양한 음작용을 자동으로 생성하여 살아있는 캐릭터 에이전트를 멀티미디어 교육용 컨텐츠 저작물과 결합하여 보다 나은 교육용 컨텐츠를 생성할 수 있도록 하였다.

참고 문헌
  http://www.seestorm.com/lipsynchrosdk.jsp
[6] Tagfree, Tagfree 2000 SMIL Editor v1.0,
[8] 씨마일미디어, EZer SMIL 1.0,
[9] RealNetworks, RealPlayer,
  http://www.real.com/?PV=6
[11] Oratrix, GRiNS Pro Editor for SMIL 2.0,