



사이버 맨

성균관대학교 홍광석*

1. 서 론

최근 사이버 가수나 사이버 캐릭터 등이 인터넷과 방송 등에 급격히 등장하면서 사람이 하는 일을 사이버 인간이 조금씩 대신하고 있다. 이에 따라 사이버 인간은 어느 정도의 수준을 가질 수 있을까? 사이버 인간은 인간에게 어떠한 존재일까? 하는 궁금증을 갖게 된다.

사이버 맨 또는 버추얼 휴먼이라 불리는 가상 인간은 가상 공간과 현실 세계를 넘나들며 지금도 다양한 일을 하고 있지만 사이버 인간의 능력이 향상되면 앞으로 그 활약이 어디까지 미칠 것인가에 대한 언급보다는 오히려 미치지 않는 것이 무엇인가를 찾는 것이 훨씬 수월해 질 것이다.

이미 국제적으로 몇몇 연구소에서는 제한적이기는 하지만 인간과 가상 인간이 대화를 통하여 의사를 전달하고 할 일을 지시하는 수준에 도달해 있다. 이러한 기술적 발전에 힘입어 2003년 쯤으로 예상되는 마이크로소프트의 차세대 운영체제에서는 가상 인간(Virtual Human)이 표준 인터페이스로 채택될 가능성이 높다[1].

차세대 인터페이스 기술의 핵심은 눈, 귀, 입, 그리고 마인드이다. 이들 기술은 이미 대부분 현실에서 사용하기 시작되었고 결국 이 네 가지 요소가 통합된 새로운 인터페이스를 채택한 운영체제가 완성되면 인간에게는 편리함과 동시에 하나의 커다란 충격으로 다가올지 모른다.

지난 4월 본인의 연구실에서 3차원 얼굴 애

니메이션, 음성합성, 그리고 음성인식 기술을 통합하여 만든 가상 인간을 “사이버 맨”이라는 제목으로 코엑스에서 전시할 기회를 가졌다 [2]. 전시기간중 많은 관람객을 통하여 사이버 맨과 인간과의 관계를 생각해 보고 앞으로의 사이버 맨에 대해 생각해 볼 기회를 갖게 되었다.

본 고에서는 사이버 맨의 여러 가지 부류에 대하여 알아보고, 본인의 연구실에서 만든 사이버 맨의 소개 및 인간과 사이버 맨의 관계에 대하여 살펴 보도록 한다.

2. 사이버 맨

현재 사이버 인간으로 소개되고 있는 사이버 가수, 애버타, 보트 등은 지능과 능력면에서 볼 때 인간에 비해 상당히 하등한 수준을 벗어나지 못하고 있다.

지능을 가진 사이버 인간을 실현하기 위해서는 인공지능을 도입한 지식처리, 추론, 대화처리기술의 발전과 학습, 행동, 인지모델의 모델링 기술의 개발과 통합이 필요하다. 또한 감정을 갖도록 설계하는 것이 필요하다[3]. 이러한 지능을 갖기 위해서는 언어의 인식과 합성, 시각인식 등의 향상된 지능적 인터페이스 능력을 갖도록 하는 것이 선행 되어야 할 것이다.

본 절에서는 현재 활동하고 있는 사이버 맨의 여러 가지 부류에 대해 소개한다.

2.1 사이버 가수

가상 공간과 실제 공간을 넘나들며 활약하고

*통신회원

있는 아담, 류시아, 사이다, 디키, 다테 교코 등의 사이버 가수 이름은 이제 젊은 층을 위주로 한번쯤은 들어 봄직한 이름이 되었다. 이들은 가수로서 뿐만 아니라 광고모델, DJ 등 그 활동 영역을 점차 넓혀가고 있다. 그러나 이들은 잘 생긴 외모를 앞세워 인기를 모으고 있지만 실은 꼭두각시 인형 같은 삶을 산다고 할 수 있다. 이들의 율동은 센서를 부착한 무명의 댄서가 추는 율동을 모션 캡처한 후 랜더링하여 흉내 내고, 목소리 역시 무명 가수의 목소리를 녹음한 후 더빙하여 노래를 부른다.

하지만 이들은 캐릭터 산업의 엄청난 부가가치를 향하여 계속 그 기술의 향상과 활동 영역을 확장해 나아가고 있다. 그들은 현재 세계적인 만능 연예인이 된 영국의 라라 크로포드 이상의 꿈을 갖고 있다[4].

2.2 애버타(Avatar)

인터넷 등의 가상의 환경에서 현실의 사용자를 대신하여 가상 공간에서 대화나 게임 등을 하며 사이버 세계의 문화를 형성해 가고 있다. 이들은 사용자를 위하여 명령을 수행해 주고 사이버 공간을 누비며 단순히 사용자의 일방적인 명령만을 받는 것이 아니라 애버타 스스로 사용자에게 많은 정보를 제공하고 사용자에게 보다 친숙한 모습을 가지고 있다.

2.3 보트(Bot)

일반적으로 보트라 불리는 소프트웨어 로봇은 무한한 정보의 바다인 인터넷 가상공간을 돌아다니며 필요한 데이터를 수집하거나 이메일의 필터링 등 사이버 비서역할도 하는 자율적이며 지능적인 프로그램이다.

보트는 외모를 갖기도 하지만 가상공간에서 굳이 인간의 외모를 갖지 않고도 사용자에게 사이버 비서 노릇을 하게 된다.

2.4 기타

최근 일본 전신 전화에서는 말하는 컴퓨터를 개발하여 발표하였다. 여기에서는 자연어처리 기법을 대폭 추가하여 사람이 대화하는 중에 화제를 바꾸어도 이에 대응하여 적절한 대화를 할 수 있도록 음성처리 부분에서 한발 더 나아

간 기술을 적용하였다[5].

현실세계와는 달리 사이버 세계에서는 사용자가 상상하는 대로 원하는 용모와 성을 가질 수 있고 상상하는 일을 마음껏 할 수 있게 된다. 어쩌면 가상 인간은 우리가 상상하는 그 이상의 모습으로 우리에게 다가올 지 모른다.

3. 사이버 맨 기술

사람에 가까운 완전한 사이버 맨을 만들기 위해서는 애니메이션, 음성합성, 음성인식, 시각인식, 촉각, 지식/추론/대화/학습/행동/인지 모델링, 감정/감성 등 다양한 분야의 통합기술을 이용하여야 한다. 그러나 용도에 따라서는 이 모든 것을 굳이 다 이용할 필요는 없고 필요한 기술만을 이용하여 만들게 된다.

인간의 신체에서 가장 많은 정보를 표현하는 것은 얼굴이다. 말을 비롯한 시청각 그리고 감정에 따른 다양한 표정 등 일반적인 개인의 특성이 대부분 얼굴에 있다고 볼 수 있다. 따라서 인간의 신체 중에 얼굴의 애니메이션에 많은 비중을 두고 있다.

컴퓨터 상에서 얼굴 표현은 2차원 얼굴과 3차원 얼굴로 나누어 볼 수 있으며 현재 유통되고 있는 일반적인 컴퓨터의 성능으로 3차원이 무난히 표현되므로 사이버 맨의 얼굴로서 3차원 얼굴 애니메이션을 채택하는 것이 좀더 효율적이라고 생각된다.

본 연구실에서 만들어 소개한 사이버 맨은 음성합성엔진, 3차원 얼굴 애니메이션엔진, 그리고 음성인식엔진을 개발 통합하여 PC환경에서 실시간으로 다양한 응용 프로그램에 이용하도록 한 것이다. 이를 이용하여 차세대 휴먼 인터페이스로 부상할 사이버 맨이 여러 가지 응용 서비스를 제공할 수 있다. 사이버 맨은 방송/영화/광고분야, 인터넷/통신분야, 교육/게임분야, 컴퓨터, 금융서비스 등 다양한 분야에서 많은 활용을 할 수 있다. 전람회에서는 간단한 응용예로서 사이버 맨과 말로써 구구단 게임을 할 수 있도록 통합하여 소개하고, 각각의 엔진별 기능을 실험할 수 있도록 하였다.

음성합성 엔진은 TTS(Text-To-Speech)방식으로 우리말, 영어단어, 숫자 및 기호 등을

합성할 수 있고, 3차원 얼굴 애니메이션은 합성음에 따른 립 싱크, 감정에 따른 표정변화, 얼굴의 자유로운 돌림이 가능하고, 음성인식은 우리말 인식에 효율적인 방법을 적용하여 비교적 인식이 까다로운 연속 숫자음 인식에도 좋은 성능을 보인다[6]. 이 절에서는 이 세가지 기술을 간략히 소개한다.

3.1 3D 얼굴 애니메이션

3차원 얼굴 모델링을 위한 방법으로는 Polygon, Nubs 등의 방법이 있으나 비교적 처리가 쉽고 얼굴 모양의 자세한 표현을 하기 위하여 Polygon Mesh를 많이 이용하고 있다. 본 고에서는 Polygon Mesh에 의한 얼굴 구성을 많이 사용한다.

얼굴 구성시 많은 프레임과 Vertex수는 얼굴을 정교하게 표현하는 데는 좋지만 그 수가 너무 많아지면 메모리와 계산량이 많아 진다는 단점이 있다. 본 고에서 소개하는 사이버 맨의 얼굴 구성은 표 1과 같이 하였다.

표 1 얼굴 구성

Frame	내 용	Vertex수
1	속눈썹	186×2
2	치아	519
3	혀	120
4	얼굴 피부	2444
5	입속	176
6	눈썹	120×2
7	입술	176
8	눈흰자	530×2
9	홍채	144×2
10	눈의동공	50×2
11	머리카락	2265
		총 7760

얼굴 모델의 3차원 동작을 위한 방법으로는 Key Frame Animation, Motion Capture, Motion Control Animation 등이 있으나 본 고에서는 사이버 맨의 실시간 동작을 위하여 Motion Control 방법과 Key Frame 방법을 적절히 혼용하여 얼굴의 표정변화 및 립 싱크를 하였다. 여기서 Motion Control 방법은 시뮬레이션 또는 파라메트릭 방법이라고도 불린다. 얼굴 동작에 사용한 Action Unit를 표 2

표 2 3차원 얼굴의 기본동작

Frame	Action Unit
1	입을 벌림
2	입술 양단을 올림
3	입술 양단을 내림
4	입술을 좁힘
5	입술을 서로 벌림
6	입술을 서로 접근
7	혀 후위를 올림
8	혀 앞을 올림
9	좌측 눈썹 올림
10	우측 눈썹 올림
11	좌측 눈썹 끝 내림
12	우측 눈썹 끝 내림
13	좌측 눈꺼풀 내림
14	우측 눈꺼풀 내림
15	눈 좌우 회전
16	눈 상하 회전
17	얼굴 좌우 회전
18	얼굴 상하 회전
19	얼굴 좌우 기울림
20	무표정

에 나타내었다.

안면 근육의 변화에 의해 얼굴에 나타날 수 있는 표현들의 기본요소로 더 많은 Action Unit를 정의[7]해 사용할 수 있지만 표 2의 내용을 조합하여 다양한 얼굴 표정을 나타낼 수 있다.

구체적인 3차원 얼굴 애니메이션 방법은 아무런 표정이 없는 Mesh의 Vertex값들과 특정 표정을 구현한 Mesh의 Vertex값들과의 Vertex Interpolation을 이용하고, 프레임 전체를 회전하거나 움직일 때에는 Frame interpolation을 이용하여 복합적으로 구현하였다[8].

이상에서 언급한 얼굴 모델 구성과 감정에 따른 표정 변화 예를 그림 1에 도시하였다.

3.2 TTS(Text-To-Speech)

합성음의 기본 데이터는 반음절 단위로 하였고 음성합성은 합성음의 명료도나 자연성이 비교적 우수한 TD-PSOLA 방식[9]을 이용하였다. 유니코드를 이용하여 모든 한글과 숫자, 기호 및 영문의 합성이 가능하도록 하였다[10].

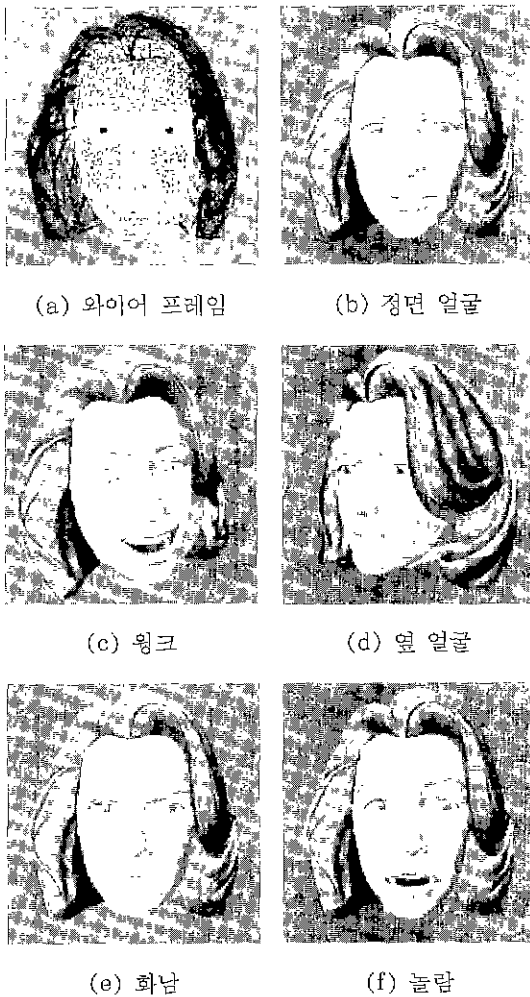


그림 1 얼굴모양

3.3 음성인식

음성인식은 대용량 및 화자 독립 인식에 적합한 HMM(Hidden Markov Model)을 이용하는 것이 효율적이다. HMM 중에서 심볼의 개수가 유한한 DHMM(Discrete HMM)을 이용하여 화자독립 음성인식을 수행한다. 일반적으로 DHMM은 인식단어의 수가 수백 단어 이하일 경우와 훈련용 음성 데이터의 수가 적을 때에는 CHMM(Continuous HMM)보다 인식 성능이 좋은 것으로 알려져 있다[11].

사이버 맨의 음성인식은 우리말의 특성을 살

린 반응절과 반응절+반응절 단위의 HMM모델을 이용하여 효율적인 인식이 되도록 하였다[6].

3.4 립 싱크(Lip Synchronization)

립 싱크는 TTS와 연동하여 이루어져야 한다. 음절이 발생될 때 입모양은 주로 모음이 결정하지만 순음과 같은 자음성분에 의해서도 입모양이 결정된다. 따라서 텍스트에 대응한 입 모양을 변화시키기 위해서는 한글의 자모음 /아, 에, 이, 오, 우, 어, 으, 순음/에 대하여 각각 Action Unit에 weight를 정의하여 사용하였다[8].

사이버 가수의 경우에 립 싱크도 모션 캡처 방식에 의하여 자연스러운 표현을 하지만 음성의 발생시 자음보다 모음 부분이 중요한 역할을 하므로 /아, 에, 이, 오, 우/에 대한 입술의 변화 모델을 만들어 립 싱크를 하기도 한다[1].

3.5 시스템 통합

사이버 맨을 위한 3차원 얼굴 애니메이션, 음성합성, 립 싱크, 음성인식의 통합은 자연스러운 표현과 실시간 통합을 위하여 다음과 같은 원칙을 갖는다.

- (ㄱ) 각각의 엔진이 서로의 프로세스를 침범해서는 안된다.
- (ㄴ) 하나의 프로세스가 실행될 경우에도 다른 프로세스가 실행 권한을 잃지 않아야 한다.
- (ㄷ) 각 엔진들은 이벤트 방식으로 서로간에 통신을 한다.
- (ㄹ) 동적 메모리 할당과 루프를 가능한 짧게 만들어야 한다.

이러한 원칙하에 구성되는 통합시스템의 기본구조는 그림 2와 같다. 그림 2에서 TTS와 음성인식, 렌더링은 각각 독립된 엔진으로 구성하였고, 초기화에서 3D Mesh DB와 TTS DB를 로드하여 각종 응용 프로그램을 설계할 수 있다. 특히 idle 상태에서는 3차원 얼굴과 눈동자의 미세한 흔들림과 눈깜빡임을 주어 좀더 자연스럽게 보이도록 하였다.

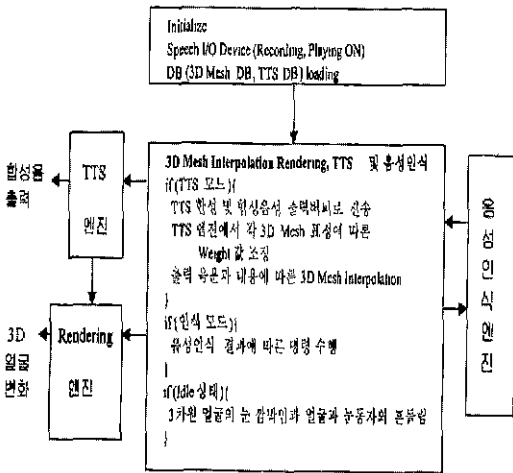


그림 2 통합시스템의 기본구조

4. 인간과 사이버 맨

고대로부터 우리 인간은 인간의 신격화 또는 신의 인격화에 많은 정성을 기울여 왔다. 이를 위하여 신상을 인간과 비슷한 형상으로 만들어 놓고 메시지를 전하기 위해 말을 하도록 하기 위한 노력 등을 기울여 왔다[12]. 또한 주술사나 예언자들은 실제의 세계가 아닌 또다른 세계를 상상하며 현실과 관계를 맺는데 노력하였다.

현재 사이버 맨은 가상 공간에서 실제 공간으로, 인간은 실제 공간에서 가상 공간으로의 여행을 자유로이 할 수 있다. 그렇다면 한편으로는 기술의 발전에 의해 가상 공간과 실제 공간의 영역 구분이 없이 혼재하는 가상-실제 공간 통합 세계가 도래하고 있다고 보아야 할 것이다. 그러나 이러한 가상공간과 실제공간의 혼재는 인간에게 많은 영향을 미칠 수 있다. 몇가지 예를 살펴보면 다음과 같다.

가상 공간에서는 현실의 여러 가지 구속으로부터 벗어날 수 있다. 시공간을 초월한다든지 유명처럼 서로 교차하여 지나가는 것이 가능하여 자칫 허상을 추구하는 인간의 경우 현실에 대한 적응력 문제가 대두될 수 있다.

가상 세계에서의 수많은 사이버 맨(일종의 신중 생명체)의 출현과 통제 불능의 사이버 맨

의 출현 등에 의한 혼돈과 무질서 상태의 문제가 발생 할 수 있다[4].

더우기 기술의 발전에 힘입어 인간과 같은 외모, 감각 등을 갖춘 사이버 맨들이 등장할 것이 예측된다. 진짜 인간과 같은 정도의 아니면 그 이상의 능력을 갖춘 사이버 맨의 출현시 사이버 맨과 인간은 무엇이 다른가로부터 시작 되는 인간의 정체성 문제가 야기될 수 있다.

또한 개인의 행동을 감시할 수 있는 인터페이스의 등장으로 인해 생길지 모르는 사생활 침해의 문제도 생각해 볼 문제이다[1].

그러나 사이버 맨은 지체 장애인에게 다양한 도우미 역할을 할 수 있다. 또한 사이버 공간에서 자신의 분신을 통해 장애가 없는 사람과 대등한 입장에서 게임을 한다든가 공동 작업을 하며 자신감과 만족감을 갖게 될 것이다.

지난 봄 코엑스에서 5일간 사이버 맨을 전시했을 때 어른보다는 아이들이 더 많은 관심을 보였다. 아이들은 단순히 캐릭터가 말하고 웃는 것에 대해 무척 신기해 하고 단순히 텍스트로 자기이름을 타이프하면 사이버 맨이 발성하는 것을 보고 듣는 것만으로 컴퓨터에서 눈을 떼지 못하였다.

관람자 중 뇌성마비 장애자와 그 아버지가 방문하였을 때 그분은 아들이 몸을 잘 움직일 수 없어 책 등을 일일이 읽어주거나 손으로 짚어주면서 보여주어야 하는데 사이버 맨을 보시고는 매우 기뻐하였다. 그분은 마지막에 앞 못 보는 분들을 위해서도 좋은 일이니 합성음을 지금보다 더 듣기 좋게 해 달라는 부탁도 잊지 않으셨다. 또한 유관한 전문직에 종사하시는 분들은 캐릭터가 만화같다, 인식이 잘 안된다, 움직임이 부자연스럽다, 합성음이 이상하다 등의 따끔한 일침과 함께 격려도 해주셨다.

이와 같이 사이버 맨은 여러 관점에서 인간에게 어떠한 모습으로 다가올 지는 정확히 예측하기 어렵다. 한 가지 분명한 일은 막대한 부가가치를 지닌 인간에 못지 않은 능력을 가진 가상 인간을 만들기 위해 세계적으로 많은 연구자들이 노력하고 있고 우리도 그러한 경쟁에서 뒤지지 않기 위해 더욱 노력을 기울여야 한다는 점이다.

5. 결 론

지금까지 가상 인간의 부류와 가상 인간의 기술 및 인간과 가상 인간의 관계에 대하여 간략하게 살펴보았다. 다양한 형태의 가상 인간은 기술의 발전에 힘입어 인간의 모습과 능력을 갖게 될 것이며 그 결과 가상 세계와 실제 세계에서 많은 긍정적 부정적인 역할을 하게 될 것이다.

가까운 미래에 우리는 실제 공간보다는 가상 공간에서, 실제 인물보다는 가상 인물과 더 많은 접촉과 정보를 얻게 될 것이다. 더 나아가 우리의 자손들은 태어나면서부터 실제 세계보다 가상 세계의 생활에 더 익숙하게 될지도 모른다. 이러한 시대가 도래하면 우리가 가지고 있는 현재의 가치관과 윤리가 한낱 허상에 불과할지도 이미 과거의 유물로 박물관에나 가서 느낄 수 있을지도 모른다.

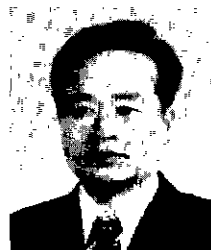
그러나 실제 인간이 가상 인간과 가상 공간을 창조하였기에 차후에 우리가 만든 피조물에 대한 적절한 통제와 이용법을 만들어 인간성의 훼손과 현실의 적응에 대한 우려를 불식해 나아가갈 것이라 생각된다.

참고문헌

1. <http://www.howpc.com/>
2. "99 신기술 창업 박람회 official directory", 과학기술부, p. 131, 1999.
3. "뉴로·퍼지·인공지능 핸드북", 일본 제측자동제어학회편, 1994.
4. <http://www2.donga.com/>
5. <http://www.mbcnewsdesk.com/>
6. Jeh-Seon Youn, Kwang-Woo Chung, Doo-Sung Lee, Kwang-Seok Hong, "Korean Speech Recognition using Semisyllable HMM", ICSPAT 98, pp. 1339-1343, 1998.

7. P. Ekman and W. V. Friensen, "Facial Action Coding System", Consulting Psychologist Press, 1977.
8. 최광표, 윤재선, 이현구, 홍광석, "사이버 에이전트를 위한 3D 얼굴 애니메이션", 제 13회 산학연 멀티미디어 산업기술 학술대회, Vol 13, pp. 204-207, 1999.
9. T. Dutoit and H. Leich, "Improving the TD-PSOLA Text-to-Speech Synthesis with a Specially Designed MBE Re-Synthesis of the Segments Database", EUSIPCO 92, pp. 343-347, 1992.
10. 김종우, 최광표, 이현구, 홍광석, "지능적 휴먼-컴퓨터 인터페이스를 위한 무제한 음성합 성시스템 구현", 대한전자공학회 멀티미디어연구회 학술발표, pp. 209-212, 1999.6.
11. Chin-Hui Lee and Frank K. Soong, "Automatic Speech and Speaker Recognition", Kluwer Academic Publishers, 1996.
12. David B. Roe and Jay G. Wilpon, "Voice Communication between Humans and Machines", National Academy Press, 1994.

홍 광 석



1985 성균관대학교 전자공학과 학사
 1988 성균관대학교 전자공학과 석사
 1992 성균관대학교 전자공학과 박사
 1990~1993 서울보건전문대학 전산정보처리과 전임강사
 1993~1995 제주대학교 정보공학과 전임강사
 1995~현재 성균관대학교 전기전자 및 컴퓨터공학부 부교수

관심분야 : 휴먼-컴퓨터 인터페이스, 음성인식 및 합성
 E-mail : kshong@yurim.skku.ac.kr