인 지 과 학
Korean Journal of Cognitive Science
Vol.6, No. 2(1995)

다중 제어 레벨을 갖는 입모양 중심의 표정 생성

Speech Animation with Multilevel Control

문보희†, 이선우††, 원광연†††
Bohee Moon, Sonou Lee, Kwangyun Wohn

요 약

오래 전부터 컴퓨터 그래픽을 이용한 얼굴의 표정 생성은 여러 분야에서 응용되어 왔고, 요즘에는 가상현실감 분야나 원격 회의 분야 등에서 가상 에이전트의 표정을 생성하는데 사용되고 있다. 그러나 네트워크를 통해 다중 참여자가 상호작용을 하는 상황에서 표정을 생성하는 경우에는 상호작용을 위해 전송되어야 할 정보의 양으로 인해, 실시간에 원하는 표정을 생성하기 어려운 경우가 생긴다. 본 연구에서는 이러한 문제를 해결하기 위해 표정 생성에 Level-of-Detail을 적용하였다. Level-of-Detail은 그래픽스 분야에서 복잡한 물체의 외형을 좀 더 효율적으로 나타내기 위해 오래 전부터 연구되어져 온 기법이지만 아직까지 표정 생성에 적용된 예는 없다. 본 연구에서는 상황을 고려하여 적절하게 상세도를 변경하여 표정을 생성하도록 Level-of-Detail 기법을 적용하는 방법에 대해 연구하였다. 구현된 시스템은 텍스트, 음성, GUI, 사용자의 머리의 움직임 등과 같은 다양한 입력에 대해 입모양과 동기화 되는 표정을 생성한다.

주제어 표정 생성, 다중 참여자, 원격 회의, LOD, 입모양의 동기화

---

† 삼성전자 전략기획실
Samsung Electronics,
Strategic Planning office
e-mail:bhmoon@ dangun.kaist.ac.kr.

†† 한국과학기술원 전산학과
Computer Science Department,
KAIST
e-mail:solee@ dangun.kaist.ac.kr.

††† 한국과학기술원 전산학과 교수
Computer Science Department,
KAIST
e-mail:wohn@ cs.kaist.ac.kr.

※) 본 연구는 '인공지능연구센터 중점연구과제', '삼성멀티미디어 미래기술사업' 의 지원하에 이루어 졌음.

ABSTRACT

Since the early age of computer graphics, facial animation has been applied to various fields, and nowadays it has found several novel applications such as virtual reality(for representing virtual agents), teleconference, and man-machine interface. When we want to apply facial animation to the system with multiple participants connected via network, it is hard to animate facial expression as we desire in real-time because of the size of information to maintain an efficient communication. This paper's major contribution is to adapt 'Level-of-Detail' to the facial animation in order to solve the above problem. Level-of-Detail has been studied in the field of computer graphics to represent the appearance of complicated objects in efficient and adaptive way, but until now no attempt has mode in the field of facial animation. In this paper, we present a systematic scheme which enables this kind of adaptive control using Level-of-Detail. The implemented system can generate speech synchronized facial expressions with various types of user input such as text, voice, GUI, head motion, etc..

Keyword facial expression, multiuser, telecommunication, LOD, speech synchronization

1. 서론

얼굴을 마주하는 상호작용에서의 의사소통은 신체, 목소리, 얼굴 등 여러 개의 의사교환 통로(channel)를 통하여 이루어진다. 얼굴은 의사교환에 있어서 중요하고 복잡한 통로이다. 오래 전부터 이러한 얼굴의 역할을 응용하기 위한 여러 연구들이 있었다. 만화나 영화에서 나오는 주인공의 얼굴을 애니메이션 하거나, 사용자 인터페이스로 애니메이션 되어진 얼굴을 사용하거나, 구축된 얼굴의 이미지와 신상 기록 데이터베이스를 이용하여 범죄자를 조회, 구별, 추론하는데 사용하거나, 얼굴 이미지 전체를 전송하는 대신 사용자의 표정 변화에 따른 파라미터의 값만을 전송하여 통신량을 줄이는 데 사용되기도 했다. 요즘에는 새롭게 등장하는 분야인 가상 현실감(virtual reality)이나 원격 회의 분야에서 가상 에이전트의 얼굴 표정을 생성하는 데 응용되고 있다.

이러한 분야에서는 멀리 떨어져 있는 다중 참여자들이 서로 상호 작용을 하기 위해 실시간으로 피드백을 제공하여야 하므로 일정하면서도 충분히 빠른 프레임 속도(초당 10프레임 이상)가 유지되어야 한다. 이러한 실시간이라는

제약을 지키기 위해 가상 에이전트와의 상호작용은 그 상황에 맞추어 복잡도를 선택할 수 있는 기능이 필요하게 된다.

본 논문에서는 위에서 말한 바와 같이 원거리간에 다중 참여자들의 상호작용을 실시간에 제공할 수 있도록 하기 위해, 상황에 따라서 제어의 복잡도를 조절하는 Level-of-Detail 기법을 가상 에이전트와의 상호 작용의 중요한 통로인 입모양 중심의 표정을 애니메이션 하는데 적용하였다. 그리하여, 가상 세계 내에 있는 가상 에이전트의 표정을 가상 현실 시스템의 필요나 수용능력에 따라, 적절하게 애니메이션의 상세도 레벨을 선택하여 원하는 정도의 상세도로 생성한다. 이와 같이 여러 제어의 레벨을 가지는 것과 동시에 입력에 있어서도 텍스트, 음성, GUI, 사용자의 머리 움직임 등과 같은 표정 생성을 위한 다양한 입력 수단을 지원하며, 텍스트 입력인 경우에는 합성된 음성과 입모양을 동기화 시켜 출력한다.

본 논문의 구성은 다음과 같다. 2장에서는 표정 생성에 관한 간략한 소개와 관련 연구의 내용을 기술하고, 기존 연구에서 보완되어져야 할 점을 지적하여 본 연구의 방향을 제시한다. 3장에서는 제안된 다중 제어 레벨을 갖는 입모양 중심의 표정 생성 시스템의 구조에 대해 알아본다. 4, 5, 6장에서는 실제로 구현된 시스템의 세부 모듈에 대해 기술한 후, 7장에서 결론과 향후 연구 방향을 제시한다.

2. 관련연구

2.1 표정생성

얼굴의 애니메이션에 관한 연구는 여러 연구자들이 각기 다른 관점에서 이루어져 왔다. 처음에는 2차원의 만화와 같은 모습에서 출발하여 점차 3차원의 모델의 연구로, 더 나아가서는 사람의 얼굴의 근육의 움직임을 관찰하여 정밀히 모델링 하게 되었다[Parke74][Ekman77][Waters87]. 또한 더욱 실제감을 주기 위해 실제 사람의 얼굴사진을 모델에 덮어 씌워 사람과 아주 흡사하게 애니메이션 되는 방법도 연구되었다[Yau88]. 얼굴의 애니메이션을 음성과 동기화 시키기 위한 다양한 방법도 여러 연구자들에 의해 연구되었[Lewis87][Hill88][Morishima93][Waters94]. 음성이 아닌 사람의 움직임이나 얼굴의 움직임을 이용하여 애니메이션에 사용하는 예도 있었다[Williams90][Benoit94].

심리학 분야에서는 얼굴의 표정들을 분류하여 놓은 연구가 있다. Ekman과 Friesen의 Facial Action Coding System(FACS)이 바로 그것이다[Ekman77]. 이 두 사람은 심리학적인 관점에서 사람이 얼굴의 움직임을 통해 어떻게 정보를 주고받는지에 관한 연구의 근간으로 얼굴의 표정을 연구한 것이다. FACS에서는 얼굴의 표정을 만드는 많은 근육들을 비슷한 것들끼리 그룹 지어 Action Unit(AU)이라 명명하였다. 그리고 이것은 컴퓨터를 기반으로 한 얼

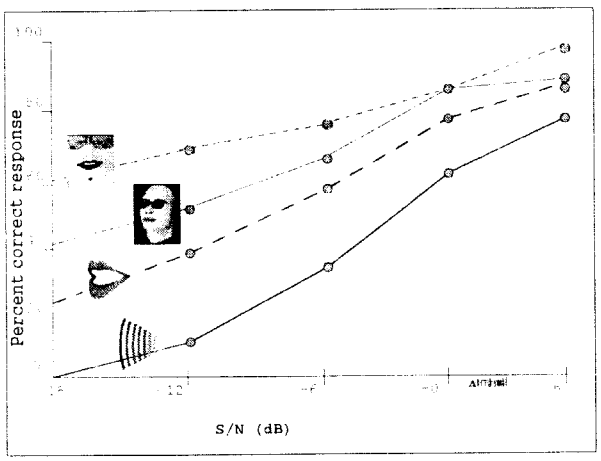
굴의 애니메이션 분야에서 사용하는 표현 방법으로 널리 쓰이고 있다. Ekman과 Friesen은 대략 55,000개의 얼굴의 표정이 있다고 분류했다. Action Unit들은 이들 표정을 생성하는데 사용된다. 대체적으로 이들 표정들은 6가지의 기본적인 형태(노여움, 두려움, 슬픔, 행복, 놀람, 혐오감)로 묶여질 수 있다.

2.2 음성과 동기화 되는 표정 생성

오래 전부터 사람들은 말(speech)을 듣고 이해하는데 있어, 시각과 청각으로부터 얻어진 두 가지의 정보를 이용하는데 익숙해 있다. 일반적으로 청각 장애인만이 독순술(lip-reading)을 사용한다고 생각하지만, 사실은 정상적인 청력을 가진 사람들도 잡음이 있는 경우에는, 말을 듣고 이해하는데 있어 청각과 더불어 시각적 정보를 사용한다. 유창한 연설인 경우에는 얼굴 표정에 의해 말이 강조되거나 잠시 중단되기 때문에, 사람들은 연사의 얼굴 표정으로부터 좀 더 많은 시각적 정보를 얻으려 하게 된다.

이미지가 말을 하는 것처럼 보이도록 하는 시도는 1892년 Demeny에 의해서 Phonoscope라고 부르는 장치를 이용하여 처음 이루어졌다 [Des66]. 이 장치는 원판 위에 얼굴 하부의 이미지들을 올려놓고 회전시킴으로써 잔상을 이용하여 말하는 이미지를 생성하였다. 이 방법의 기본적인 절차는 오늘날 애니메이션에 적용되어 사진을 사용하는 대신 손으로 그린 입술의 이미지를 사용하였다. 그러나 모든 이미지들을 손으로 그려 넣어야 하는 일은 시간과 노력이 많이 드는 일이다.

수작업에 드는 노력을 대신 하고자 컴퓨터를 이용하여 말하는 표정 생성이 시도 되고 있다. 다음절에서는 음성과 동기화 되는 표정이 필요한 이유와 그러한 표정을 만들어 내기 위한 다양한 방법들을 소개한다.



〈그림 1〉 표정이 의사소통에 미치는 영향

2.2.1 음성과 동기화 되는 표정의 필요성

말을 듣고 이해하는데 있어 말하는 사람의 얼굴을 보는 것은 많은 도움을 준다. 말의 시각적인 측면은 특히 말의 청각적인 측면에서 많은 방해가 있을 경우 (잡음, 대역폭 필터링, 청각장애 등)에 특히 유용하다. 그러나 말의 시각적인

측면은 단지 청각적 정보가 감소했을 때에만 영향을 주는 것은 아니다. 시청자들의 이해력에는 시각과 청각 모두가 기여를 하고 있다. 만약 비디오 테이프의 배우가 /da/라고 말하고 있는데 /ba/라는 발음으로 더빙을 했을 경우, 시청자들은 종종 그 배우가 /ə/라고 말하고 는 것으로 인지한다[Massaro90].

얼굴의 표정이 의사 소통에 많은 역할을 담당하고 있다고는 여겨졌지만 얼마나 많은 정보를 얼굴 표정이 전달을 하는지에 대해서는 알 수 없었다. 이에 관련하여 화자의 얼굴 표정이나 입술 모양을 통해 전달되는 정보의 양을 정량적으로 실험한 연구가 있다[Benoit94].

실험의 내용은 잡음정도를 서로 다른 음성을 들려주면서 피실험자들에게 무슨 단어인지를 알아내도록 하는 것이다. 실험은 네 개의 종류로 나뉘는데, 각 실험에서는 음성과 함께 종류가 다른 시각적인 정보를 준다.

- 실험1 : 음성만 들려준다.
- 실험2 : 음성과 그 음을 발음하는 3차원 입술 모델을 보여준다.
- 실험3 : 음성과 그 음을 발음하는 3차원 얼굴 모델을 보여준다.
- 실험4 : 음성과 그 음을 발음하는 실제 사람의 얼굴 표정을 보여준다.

실험에 의한 결과는 그림 1에서 볼 수 있듯이 실제 말하는 사람의 얼굴 표정과 음성을 함께 들려 준 실험4가 가장 좋은 인식률을 나타내고 그 다음은 실험3, 실험2, 실험1의 순서로 인식률을 나타내고 있다. 이것은 얼굴의 표정이 의사 소통에 있어 많은 정보를 전달한다는 사실을 정량적으로 검증한다. 잡음이 많이 섞여 있어 그냥 듣기만 해서는 이해하기 힘든 경우에, 음성 정보와 말하는 사람의 얼굴 표정에서 얻어지는 정보를 함께 이용하면 이해하는데 도움을 줄 수 있다.

2.2.2 동기화 방법

말을 하는 모델을 만들기 위해서는 음성과 동기화 되는 입모양을 만드는 것이 상당히 중요하다. 그리고 표정 생성에 있어서 가장 어려운 문제 중에 하나가 바로 음성과 표정을 동기화 시키는 것이다. 애니메이션을 만드는 데 있어서는 전통적으로 이 문제는 두 가지 방법으로 해결되어져 왔다[Lewis91]. 그러나 이러한 수작업은 매우 시간이 오래 걸리므로 컴퓨터를 이용하여 애니메이션을 하려는 시도가 이루어졌다.

자동적으로 동기화 되는 입의 움직임을 얻으려는 연구는 실제 음성을 입력으로 하여 음성과 동기 적인 입의 움직임을 얻으려는 시도와 텍스트와 같은 입력을 분석하여 합성된 음성에 동기적인 입의 움직임을 얻으려는 시도로 분류된다.

실제 음성을 이용하여 실시간에 입모양을 얻어내는 것은 상당히 어려운 문제이다[Lewis87]. 현재의 음성 인식 기술은 다소 잘못된 결과를 찾는 경향(error-pron)이 있기 때문에, 실제 대부분의 시스템은 분절된 단어들을 인식할 수 있을 뿐이다. 이러한 어려운 문제를 실제로는 다 풀지 못하지만 부분적으로 해결하

여 음성과 동기화 되는 입모양을 얻어내는 시스템들이 있다. 가장 간단하게는 사운드의 음량이 입을 벌리는 역할을 하는 턱의 회전각과 비례하도록 지정해 주는 것이다[Lewis91]. 또는 이것의 변형으로 임계값을 지정하여 음량이 이 값보다 클 경우에는 턱의 회전각을 최대로 하고 작을 경우에는 입을 다물고 있게끔 하는 것이다. 이러한 간단한 방법으로는 발음하는 소리와 맞지 않은 입모양을 만들게 될 뿐만 아니라 상당히 어색한 표정이 생성된다.

좀 더 사람이 발음하는 것과 같은 입모양을 생성하기 위한 방법으로 Lewis와 Parke는 선형예측분석(linear prediction analysis)을 사용하여 음성으로부터 포만트 정보를 추출하여 분석한 다음 그 음이 어떤 발음(대체로 모음)을 하는 것인지를 인식하여 그 발음의 입모양을 생성하도록 하였다[Lewis87]. Morishima는 위의 연구에 한 발 더 나아가 텍스트와 음성을 입력으로 받아들여 입모양을 생성하도록 하였다. 음성입력일 경우에는 음성으로부터 포만트를 추출하고 결과를 다시 벡터 양자화와 신경회로망을 사용하여 적절한 입모양을 찾도록 하였다[Morishima91]. 한편으로 합성된 음성을 이용하여 동기화 되는 입의 움직임을 얻고자 하는 시스템들은 다음과 같다. Hill 등은 미리 규칙 형태로 모든 발음에 따라 입모양들의 파라미터 값들을 정의해 놓은 다음, 사용자가 입력하는 문장을 발음 형태로 분석하여 이들을 이용하여 사전 찾는 방식으로 원하는 입모양을 생성하는 연구를 발표하였다[Hill88]. Waters는 텍스트를 입력으로 받아들이고 그 문장을 분석하여 규칙 형태로 정의되어진 입모양과 함께 합성되어진 음성이 동기화 되어 생성되는 시스템을 만들었다[Waters94]. 생성되는 입모양은 좀 더 사람의 움직임을 근사하기 위해 근육의 움직임을 고려한 방법을 사용하였다. 한편, 입모양을 만들 때에 영향을 끼치는 혀의 모델을 만들고 그것의 움직임에 대한 연구가 Pelachaud에 의해서 이루어졌다[Pelachaud94].

2.3 연구방향

본 연구는 가상 현실감 시스템이나 원격 회의와 같은 다수의 사용자가 참여하여 대화를 하는 응용 분야를 주 대상으로 삼고 있다. 이와 같은 응용 분야에서는 각 사용자들이 서로 다른 상호 작용의 도구를 사용할 수 있고 실시간에 상호 작용하는 것이 가능해야 한다.

본 연구는 상호 작용의 중요한 통로인 입모양 중심의 표정을 다양한 입력 방법에 의해 생성한다. 다시 말해서, 텍스트 입력이나 음성 입력 등 다양한 방법에 의한 표정 생성과, 합성된 음성이나 실시간으로 입력되는 음성이 동기화 되어 출력되는 것을 목표로 한다. 음성과 동기화 되는 표정을 생성하는데 있어서 요구되는 것은 입의 애니메이션이다. 그러므로 본 연구에서는 주로 입모양에 관심을 두고 있다.

다수의 사용자가 참여하는 시스템일 경우 실시간으로 피드백을 받는 것은 상당히 중요하다. 그러므로 상세한 표정을 생성하는 것이 가

능하더라도 피드백을 더 빨리 받기 위해 좀 더 단순한 표정을 생성하는 것이 더 합리적인 접근방법이다. 다양한 입력 방법들은 표정을 생성하는데 이용되어지는 정보들을 포함하고 있으나 그 정보의 양은 서로 다르다. 그러므로 입력 데이터가 포함하고 있는 정보의 양에 따라 그에 맞는 표정 생성의 상세도를 결정하는 것이 더 적합하다. 표정을 생성하고자 하는 모델에 따라서도 표정을 생성할 수 있는 상세도에는 차이가 있다. 또한 표정을 생성하는 시스템마다 표정 생성 방법이 다르게 된다. 이러한 경우에는, 한 쪽에 해당되는 상세도로, 또는 그러한 방법으로 다른 쪽에서도 표정을 생성하도록 하는 것은 불합리하다. 모델이 생성할 수 있는 정도의 상세도로 표정을 생성하도록 하거나, 다른 시스템에서도 이용할 수 있도록 표정의 추상적인 개념을 전달하는 것이 적당하다.

이와 같이 상황에 따라 동적으로 표정 제어 레벨이 변경 가능해지면, 전송해야 할 표정 제어 정보의 양과, 실시간의 제약을 고려하는 등의 이유로 다수의 사용자가 참여하는 시스템에 많은 이득을 가져다준다. 그러므로 입모양 중심의 표정을 생성하는데 있어 실시간이라는 제약과 상황을 고려하여 상세도가 다른 표정을 생성할 수 있도록 제어를 하는 것이 적절하다. 이를 위해 Level of Detail 기법을 입모양 중심의 표정 생성에 적용하였으며, 3장에서 자세히 언급하겠다.

3. 다중 제어 레벨을 갖는 입모양 중심의 표정 생성

3.1 Level of Detail

오늘날 고성능 그래픽스 워크스테이션(예:SGI Workstation) 들은 초당 1백만개 이상의 lighted, antialiased, textured triangle mesh들을 렌더링 할 수 있다. 그러나 복잡한 모델들, 예를 들어 세부묘사가 복잡한 건물, 도시, 자연풍경 등은 이보다 더욱 많은 다각형(polygon)들로 이루어지게 되므로 실시간 렌더링이 어렵다. 이러한 복잡한 모델들을 실시간에 렌더링 하기 위해 몇 가지의 개념들이 소개되어 왔다. Hierarchical database, world subdivision, visibility culling, Level-of-Detail 과 같은 여러 방법들 중에서 복잡한 모델에 대한 좀 더 일반적인 해결책으로서 Level-of-Detail방법이 대두되고 있다[Astheimer94].

많은 양의 데이터를 다루는 컴퓨터의 응용분야에서 여러 단계의 상세도 레벨을 갖는 것은, 특히 많은 양의 데이터에 대한 효율적인 관리의 측면에서 중요성을 갖는다. 많은 양의 데이터를 다양한 상세도로 표현함으로써, 데이터를 다루는 데에 효율성을 기할 수 있게 되고, 사용자의 측면에서도 편리한 사용을 기대할 수 있게 된다. 이를 위하여 다중 상세도 데이터 모델(multiresolution data model)과, 이러한 레벨을 효율적으로 다룰 수 있는 방식이 요구되게 된다. 이렇게 여러 상세도의 단계를 두어 데이터를 관리하는 방식을 Level of Detail이라고

하고, 특히 실시간 렌더링에 많이 이용되고 있다. 모의 비행에서와 같이 일정한 프레임 속도(frame-rate)가 필수적인 분야에서는 이러한 이유 때문에 비행기와 같은 물체들은 종종 수작업에 의하여 몇 개의 Level-of-Detail로 만들어진다. 실시간 렌더링에서는 일정수준 이상의 프레임 속도를 만족시켜 주어야 하므로 복잡한 물체를 표현하는데 Level-of-Detail기법을 사용한다. 실시간 렌더링을 위한 Level-of-Detail기법은 복잡한 물체에 대하여 서로 다른 복잡도를 갖는 변형들을 생성한다. 복잡도를 선택하는 기준에 의거하여 변형들 중 적절한 복잡도를 갖는 변형을 선택하여 그 물체 대신 보여주는 것이다.

이제까지 보았던 바와 같이 Level-of-Detail 기법은 주로 복잡한 물체의 형태(geometry)에서 좀 더 효율적인 표현에 관하여 연구되어져 왔다. 본 연구에서는 복잡한 물체의 형태(geometry)가 아닌 효율적인 움직임(behavior) 생성을 위해 Level-of-Detail을 적용한다.

3.2 입모양 중심의 표정 생성에의 LOD

입모양 중심의 표정을 생성하는데 있어 여러 레벨의 상세도를 가지는 이유는 크게 다음 네 가지로 볼 수 있다.

실시간 수행 다중 참여자가 상호 작용하는 응용분야의 경우, 실시간에 상호 작용이 이루어져야 한다는 것은 매우 중요하다. 실시간에 상호작용감을 충분히 느끼기 위해서는 프레임 속도가 일정하게 유지되어져야 한다. 프레임 속도가 낮아서 상호 작용의 결과를 늦게 인지하게 되는 경우에는 좀 더 단순하고, 함축된 입모양을 보여 주는 것이 적절하다. 예를 들어 프레임 속도가 낮을 경우 상대방이 말을 끝냈는데도 사용자가 보는 상대방의 얼굴 모델은 자세하게 표정을 생성하기 위해 아직도 말을 하고 있는 상태라면 문제가 될 것이다. 또한, 참여자와 멀리 있는 상대방의 입모양은 단순하게 생성하고, 가까이 있는 상대방의 입모양은 복잡하게 입모양을 생성하도록 레벨을 조절하는 것도 프레임 속도를 높이는 데 도움을 줄 수 있다.

호환성 컴퓨터를 이용한 얼굴 애니메이션을 위한 시스템들의 대부분은 Ekman 과 Friesen 이 정의한 AU(action unit)의 개념을 이용하여 만들어 졌다. 그러나 개념이 같을 뿐이지 시스템들의 특성에 따라 필요한 AU를 정의하여 사용하는 것이기 때문에 종류와 기능이 다를 수 있다. 또한, AU의 개념을 사용하지 않는 표정 생성 시스템으로 그 AU들의 값을 전송 할 경우에는 각 AU에 대해 상대편의 모습이 어떻게 변화되어야 하는 지에 대한 정의가 따로 이루어 져야 한다. 이 경우 부가적으로 해야 할 일이 많아지면서 경우에 따라서는 시스템 자체의 많은 부분을 수정할 필요가 있게 된다. 이럴 경우에 AU의 개념이 아닌, 좀 더 낮은 레벨에

서 입모양을 대표하는 파라미터가 추출되어져 그것이 전송되어진다면 각 시스템에서는 그러한 입모양과 비슷한 표정을 생성하도록 해주기만 하면 될 것이다.

전송량 아주 정교한 모델의 AU값들을 그대로 전송했을 경우, 정작 AU값이 적용될 모델은 아주 단순하게 만들어졌기 때문에 모든 AU들이 정의되어 있지 않을 수 있다. 이런 경우에는 모든 AU값을 전송시키므로 인해 전송량이 많아질 뿐, 그 효과는 볼 수 없게 된다. 그러므로 단순한 입모양을 나타낼 경우에 AU레벨이 아닌 좀 더 낮은 레벨에서 입모양을 대표하는 보다 적은 개수의 파라미터를 전송한다면 전송량의 측면에서도 상당한 이득을 볼 수 있다.

사용자의 요구 단순한 모델을 사용하여 애니메이션을 하고자 할 경우, 사용자는 모델에 비해 입모양이 너무 복잡하게 생성되는 것보다는 모델에 어울리는 단순한 입모양으로 애니메이션 되는 것을 바랄 수도 있다. 또는 반대로 복잡하고 정교한 모델일 경우, 상세하게 입모양이 생성되기를 기대하게 된다. 그러므로 표정 생성의 상세도를 변경할 수 있다면, 사용자는 자유롭게 원하는 상세도를 선택할 것이다. 또한, 표정 생성을 위한 입력 수단에 있어서도 사용자에게 원하는 것을 선택할 수 있도록 함으로써 여러 수단을 사용하여 적절한 표정을 생성하도록 할 것이다.

이러한 이유로 인해 입모양 중심의 표정을 생성할 때 여러 가지의 상세도 레벨을 갖는 표정을 생성하는 것이 요구된다.

음성과 동기화 되는 입모양 중심의 표정 생성을 위해서는 텍스트나 음성이 주 입력이 된다. 각각의 입력으로부터 입모양을 생성하기 위해서는, 입모양 생성에 이용될 정보를 추출해야 한다. 입력의 종류에 따라서, 아주 단순한 입모양을 생성하는 정보만을 가지고 있는 경우고 있고, 또 아주 복잡한 입모양을 생성할 수 있는 정보를 가지는 경우도 있다.

표정을 생성하는 데 이용되어지는 정보의 함유량에 의해 입력들은 여러 레벨로 나뉘어 지고, 표정을 생성하는데 있어서 얼마나 복잡하게 할 것인지에 따라 표정 제어 레벨이 생기게 된다. 들어온 입력에 따라 입력의 레벨이 정해지고, 입력 레벨과 상황을 고려하여 얼마나 상세한 입모양을 표현할 것인지에 관한 표정 제어 레벨이 결정된다.

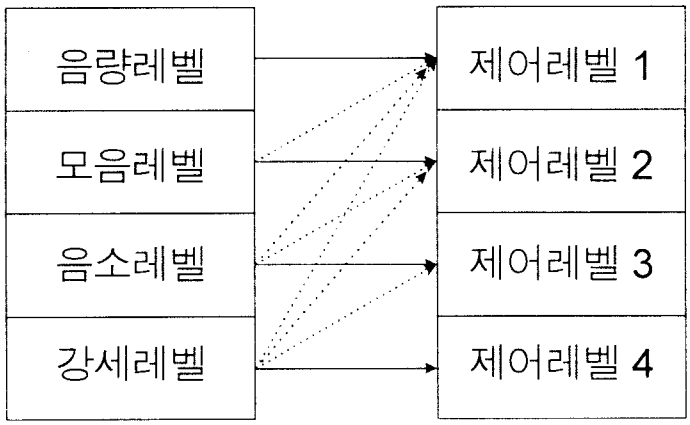
3.2.1 입력 레벨

다양한 입력 수단이 제공되어질 경우, 각각의 입력된 데이터들이 내포하고 있는 정보들은 서로 다르다. 예를 들어 음성을 분석하지 않고 음량만으로 표정을 생성 하고자 한다면 단지 입을 크게 벌릴 것인지 다물 것인지 정도만을 표현할 수 있을 뿐이다.

음성과 동기적인 입모양을 만들기 위해 사용될 수 있는 입력의 종류가 다양하게 제공될 경우에는 그에 포함된 표정 생성 시 사용되는 정

보량도 각기 다르다. 그러므로 이들 입력들은 표정 생성에 이용될 정보량에 따라 〈그림 2〉와 같이 여러 레벨로 분류될 수 있다.

음량의 따른 입벌림 정도의 정보만을 가지게 된다. 음성을 분석하여 모음정보를 가지게 될 때에는 모음 레벨로 분류하고, 말을 할 때의 모음



〈그림 2〉 정의된 입력 레벨과 제어 레벨의 관계

본 연구에서는 입력 레벨을 다음과 같이 네 단계로 나누었다.

- 음량 레벨 (amplitude level)
- 모음 레벨 (vowel level)
- 음소 레벨 (phoneme level)
- 강세 레벨 (accent level)

각 레벨은 입력으로부터 얻는 정보들이 다르며, 이들 정보들은 표정을 생성하는데 이용되어진다. 음성입력을 단순히 디지타이즈 하여 음량의 정보만을 얻게 될 때 음량 레벨로 분류하고, 에 관련된 입모양 정보를 가지게 된다. 텍스트로부터 음소들의 정보를 얻는 경우에는 음소 레벨로 분류하고, 말을 할 때, 음소 각각에 대해 어떤 입모양을 갖는지에 대한 정보를 가지게 된다. 입력으로 텍스트와 함께 GUI를 이용한 제어판으로부터 강세 정보도 함께 주어지는 경우에는 강세 레벨로 분류하고, 음소 각각에 대한 자세한 입모양 정보와 함께 말하는 도중에 강세 변화에 대한 정보도 함께 가지게 된다.

3.2.2 제어 레벨

표정을 생성하는 다양한 방법 중에서 사용하

기 쉽고 계산량이 많지 않아 본 연구에 적합한 방법으로 파라미터 기법을 들 수 있다. 이 방법은 정의된 여러 파라미터들의 값을 변화시켜서 원하는 표정을 생성하는 것이다. 처음 이 방법을 제안한 Parke는 심리학 분야에서 연구된 FACS(Facial Action Coding System)을 기반으로 파라미터를 정의하였다. 그러므로 파라미터란 곧 얼굴의 표정을 생성하는 AU(action unit)에 해당되는 것이다.

얼굴의 표정을 생성하는 것과 마찬가지로 입모양은 입 주위에 정의된 AU들에 의해서 생성된다. 입 주위에 AU들을 충분히 많이 정의하면 사람과 아주 흡사한 입모양을 생성할 수 있게 된다. 그러나 이미 언급했듯이 실시간의 제약을 지키고 호환성과 전송량을 고려하여 여러 상세도 레벨을 갖는 입모양 생성이 필요하게 된다.

제어 레벨은 입모양 중심의 표정을 어느 정도의 상세도로 생성할 것인지에 따라 나뉘어지게 된다. 입모양은 3차원 모델의 입 주위에 정의된 AU에 의해서 변화된다. 그러므로 여러 상세도를 갖는 입모양을 생성하기 위해서는 AU에 의해 움직이는 입술의 모양으로부터 좀 더 상위 레벨에서 전체적인 모습으로 그 입모양을 표현할 수 있어야 한다. 상세한 입모양을 생성하는 제어 레벨일수록 제어해야 하는 파라미터의 수가 많아지게 되고 단순한 입모양을 생성하는 제어 레벨일수록 파라미터의 수는 줄어들게 된다.

본 연구에서는 표정 생성에 있어 제어 레벨을 다음과 같이 네 단계로 나눈다. 제어 레벨은 높아질수록, 좀 더 자세하게 표정을 제어할 수 있도록 정의되었다.

- 제어 레벨 1 : 단지 입이 상하로 벌어지는 표정만을 생성하도록 제어한다.
- 제어 레벨 2 : 상하좌우로 입의 벌어진 정도만을 제어한다.
- 제어 레벨 3 : 사람과 같이 복잡한 입모양을 생성하도록 제어한다.
- 제어 레벨 4 : 복잡한 입모양과 함께 간단한 표정을 생성할 수 있도록 제어한다.

3.2.3 입력 레벨과 제어 레벨의 관계

사용자 인터페이스로부터 들어오는 입력의 성격에 따라 입모양을 생성할 수 있는 정도가 달라진다. 입력에 따라서는 단순한 입모양을 생성하는 것이 더 적합할 경우가 있고, 또는 아주 복잡한 입모양 생성도 가능한 경우가 있다. 입모양을 생성하는데 이용되어지는 정보가 적을수록 단순한 입모양을 생성하고, 반대로 입모양을 생성하는데 이용되어지는 정보가 많을 수록 복잡한 입모양을 생성한다. 입력 레벨은 입력이 가지고 있는 입모양 생성에 이용되어지는 정보의 양에 따라 나뉘어 지게 된다.

표정을 생성하는 측면에서는 입모양을 어느 정도의 상세도로 생성할 것인지에 따라 제어 레벨이 나뉘어지게 된다. 제어 레벨에 따라, 사람이 말하는 것과 같이 상세하게 입의 움직임을 표현하도록 제어할 수도 있고, 또는 만화의

주인공이 말하는 것과 같이 입이 움직이는 정도를 단순하게 제어 할 수도 있다. 제어 레벨은 입모양이 표현할 수 있는 상세함의 정도에 따라 구분되어진다.

입력 레벨이 정해지게 되면 표정 생성의 상세 정도도 다르게 결정되어진다. 다시 말해, 입력 레벨이 높아짐에 따라 높은 제어 레벨의 선택이 가능해진다. 입력 레벨과 제어 레벨은 상관 관계를 갖고 있다. 〈그림 2〉에서 실선으로 정의된 관계와 같이 입력 레벨에 속하는 정보를 사용하여 필요한 제어 파라미터를 추출할 수 있는 경우, 제어 레벨은 입력 레벨에 직접적으로 대응된다고 할 수 있다. 상황에 따라서는 직접적으로 대응되는 제어 레벨보다 좀 더 함축된 제어를 하는 낮은 제어 레벨을 선택해야 하는 경우가 생기게 된다. 이러한 경우에는 그림 2에서 점선으로 나타내는 것과 같이 입력 레벨과 제어 레벨은 간접적 대응 관계가 성립한다.

본 연구에서 정의한 입력 레벨과 제어 레벨간의 관계는 〈그림 2〉와 같다. 먼저 입력 레벨이 음량 레벨인 경우에는 단지 음의 크고 작은 정도의 정보만을 가지고 있다. 이 경우에는 음량이 많으면 입을 벌리고 음량이 적으면 입을 다무는 정도의 입모양 생성을 하는 것이 적합하다. 모음 레벨인 경우에는 말을 하는 도중에 발음되는 모음의 정보만을 가지고 있다. 그러므로 모음의 발음을 하는 데 적당하도록 전체 입 윤곽의 상하좌우의 길이정도만 변화시키는 입모양을 생성하는 것이 적합하다. 음소 레벨일 경우에는 말을 하는 도중에 발음되는 모든 음소의 정보를 가지고 있다. 음소에 들어 있는 모든 발음을 하도록 하기 위해서는 입 주위에 정의되어 있는 모든 AU를 사용하여 입모양을 생성해야 한다. 강세 레벨일 경우에는 모든 AU를 사용하여 말하는 동안의 입모양과 동시에 강조하는 부분에서의 표정도 생성해야 한다.

이와 같이 입력 레벨이 가지고 있는 정보를 사용하여 표정을 생성할 경우에 즉, 입력 레벨과 제어 레벨의 직접적 대응관계는 그림 2에서 실선으로 표시된다. 그러나 상황에 따라서는 입력 레벨에 매핑이 되는 제어 레벨을 선택하지 않고 그 보다 낮은 제어 레벨을 선택하는 경우가 있다. 그와 같은 경우에 입력 레벨과 간접적으로 대응할 수 있는 제어 레벨들과의 관계는 점선으로 나타난다.

3.3 제어 파라미터의 자동 생성

3.3.1 제어 파라미터

파라미터 기법에 의해 표정을 생성하는 시스템들은 원하는 표정을 생성하기 위해 얼굴의 근육에 해당하는 부분의 움직임을 파라미터로 정의한다. 파라미터들은 일반적으로 통용되고 있는 AU의 개념으로 정의되므로 파라미터의 값을 변화시킴으로써 원하는 근육의 움직임을 모방하게 된다. 다시 말해서, 파라미터들(즉, AU들의 집합)이 시간 축에 따라 변화하는 값을 가지게 되면 그 값들에 의해서 표정이 시간에 따라 변화하게 되는 것이다. 계산량이 적고 원하는 표정을 생성하기 수월하므로 본 연구에서

는 파라미터 기법에 의해 표정을 생성한다.

얼마나 복잡한 상세도로 입모양을 생성할 것인지에 관한 제어 레벨이 정해지게 되면, 입모양을 생성하는 파라미터들도 제어 레벨에 맞게 생성되어져야 한다. 그러나 입력 레벨과 직접적으로 대응되는 제어 레벨보다 낮은 제어 레벨을 선택해야 할 경우, 다시 말해서 좀 더 단순한 입모양을 생성해야 하는 경우가 있다. 이런 경우에는 원래 입력 레벨과 대응되는 제어 레벨의 파라미터로부터, 원래의 상세도를 함축하면서 단순하게 움직임을 나타내도록 하는 낮은 레벨에서의 파라미터를 생성해야 한다.

본 연구에서 정의한 네 단계의 제어 레벨은 표정을 생성하는데 있어서, 제어하기 위해 필요한 파라미터들이 서로 다르다. 각 레벨에 대한 제어 파라미터는 다음과 같다.

- 제어 레벨 1: (θ), θ= 입의 벌어진 정도
- 제어 레벨 2: (ε,γ), ε = 이심률, γ = 장축의 길이
- 제어 레벨 3: 입모양 생성을 위해 정의되어진 AU들
- 제어 레벨 4: 입모양 생성을 포함하여 얼굴의 표정 생성을 위해 정의되어진 AU들

제어 레벨 1에서는 상하로 입이 벌어지는 정도만을 제어하므로 제어 파라미터로는 입의 벌어진 정도를 제어 파라미터로 사용한다. 제어 레벨 2에서는 입의 단순한 움직임만을 제어하기 위해, 전체적인 윤곽을 나타내는 파라미터를 사용한다. AU에 의해 움직이는 입의 전체적인 윤곽은 타원의 형태를 가진다고 가정한다. 타원을 나타내는 파라미터는 이심률과 장축의 길이이므로, 이 두 개의 파라미터를 제어 파라미터로 사용한다. 제어 레벨 3에서는 사람과 흡사하게 말을 하도록 입모양을 제어해야 하므로, 입모양을 생성하는데 사용되어지는 AU들을 제어 파라미터로 사용한다. 제어 레벨 4에서는 사람과 흡사하게 말을 하면서 강세가 있는 부분에서는 표정도 지을 수 있도록 제어해야 하므로, 얼굴 표정을 생성하기 위해 정의되어진 모든 AU들을 제어 파라미터로 사용한다. 사용되어진 AU들에 대해서는 4.4.2에서 자세히 설명한다.

3.3.2 신경회로망을 이용한 제어 파라미터의 자동 생성

3.2절에서 기술하였듯이, 입모양 생성에 LOD(Level-of-Detail)를 적용한 이유는 다음과 같다. 상세한 표정 생성이 가능한 입력이 들어왔더라도 프레임 속도가 너무 느려 피드백이 느릴 경우 상세한 표정을 생성하는 것은 더욱 프레임 속도를 느리게 한다. 또는 모델에 따라서는 입모양을 만들기 위해 사용되는 AU 모두가 필요하지 않은 경우도 있고, AU를 적용하기 위해 다른 시스템으로 전송하는 것이 부적당할 수도 있다. 이와 같은 상황을 고려하여 제어 레벨을 낮추어야 할 경우 각 레벨의 파라미터로부터 낮은 레벨에서 필요한 파라미터를 추출하여야 한다.

하위 레벨에 있는 파라미터들을 사용하여 낮은 레벨의 파라미터를 추출하는 것은 얼마간의 지연시간을 가지게 된다. 실시간을 요구하는 시스템이므로 되도록 이러한 지연시간을 최소화해야 한다. 이러한 지연시간을 줄이기 위하여 계산이 많이 필요한 파라미터 추출단계에서는 신경회로망을 이용하도록 한다.

신경회로망 폰 노이만(von Neumann)방식의 컴퓨터는 자료 관리 및 검색, 수치적 자료의 처리에서 좋은 결과를 보이는 반면, 영상이나 음성 처리와 같은 병렬성이 내재된 실시간 처리나 판단 및 제어 등의 효율적인 수행에 있어서 인간에 크게 못 미친다. 이러한 점을 극복하기 위해, 인간 두뇌의 특이한 구조와 기능을 연구하게 되었고, 그 결과를 응용하여 만들어 낸 모델이 신경회로망(neural network)이다 [Oh91]. 신경회로망은 영상, 음성 인식, 적응 제어 및 최적화 문제 등에 탁월한 기능을 보이고 있다. 인간의 뉴런과 같이 신경회로망은 병렬적으로 작동하는 다수의 단순한 처리기(processing elements)들로 구성되어 있다. 이들 단순 처리기들 간의 연결 강도와 단순 처리기들의 망 구조, 단순 처리기에서의 처리 과정 등에 의해 신경회로망의 기능이 결정된다. 신경회로망의 특징으로는 실시간 응답과 병렬처리 등을 들 수 있다.

신경회로망을 이용하여 파라미터를 추출하도록 훈련시키기 위해서는 입력 노드의 값과 그 입력이 신경회로망을 통과하여 나오기를 희망하는 값을 알려주면서 교사 학습(supervised learning)을 해야 한다. 교사 학습이란 신경회로망을 훈련시키는 하나의 방법이다. 이는 훈련 과정에서는 엄선된 입력 자료와 희망 출력 자료를 주면서 신경회로망을 통과해 나오는 출력 값과 희망 값과의 차이를 이용하여 신경 회로망의 연결 강도를 갱신시키는 방법이다.

교사 학습을 통해 잘 학습된 신경회로망을 사용함으로써 신경회로망의 특징인 실시간 처리에 의해, 파라미터를 추출하는데 드는 지연시간은 되도록 줄일 수 있고, 거의 오차 없는 낮은 레벨의 파라미터 값을 얻을 수 있다.

4. 시스템의 설계 및 구현

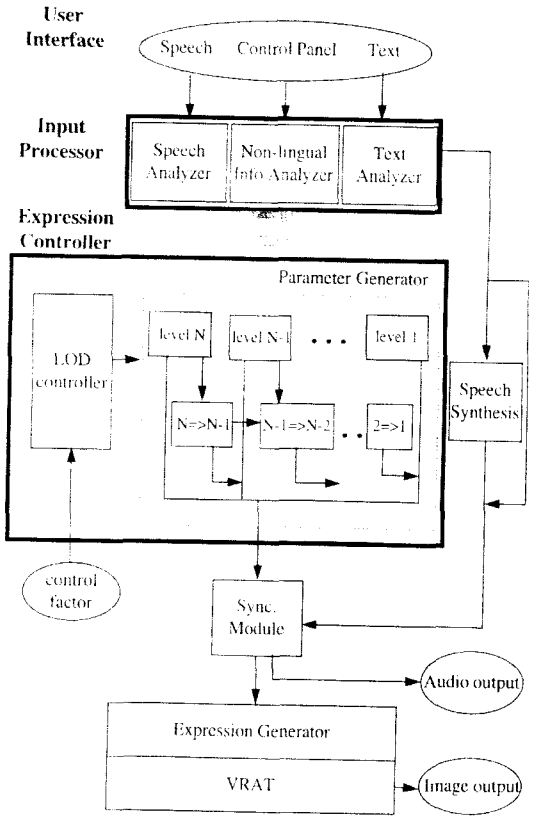
본 장에서는 3장에서 제시한 방법에 따라 구현한 다중 제어 레벨을 갖는 입모양 중심의 표정 생성 시스템에 대해 기술한다. 먼저 전체 시스템의 구조에 대해 살펴보고, 전체 시스템을 구성하고 있는 각 모듈들인 다양한 입력을 받아들이는 입력 처리기와 적절한 표정 생성에 필요한 제어 파라미터를 생성하는 표정 제어기, 그리고 음성 합성 모듈과 동기화 모듈에 대해 알아본다.

본 시스템은 Silicon Graphics사의 *Indigo2*™ Workstation에서 사용자 인터페이스로서 X11/Motif를 사용하였고 모델 독립적 표정 생성 시스템인 MiFegs(Model Independent Facial Expression Generation System)를 기

반으로 구현되었다[Lee95]. 음성 입력을 받아들이기 위해서는 *Indigo2™*에 장착된 마이크를 사용하였고, 사용자의 머리 움직임을 측정하기 위해서는 Ascension Technologies사의 *Bird™*를 이용하였다.

4.1 전체 시스템의 구조

다중 제어 레벨을 갖는 입모양 중심의 표정 생성 시스템의 전반적인 구조는 〈그림 3〉과 같다. 입력 처리기(input processor)는 사용자 인



〈그림 3〉 표정 생성 시스템의 구조

터페이스로부터의 다양한 입력을 처리한다. 표정 제어기(expression controller)는 처리된 입력들에서 얻은 정보들을 이용하여 원하는 복잡도의 제어 레벨을 결정하고 그에 적절한 제어 파라미터를 생성한다. 표정 제어기에서 구해진 적절한 제어 파라미터들은 가상 에이전트를 원하는 복잡도로 제어한다. 그 외에 텍스트 입력인 경우 합성된 음성을 생성해 주는 음성합성 모듈이 있고 합성된 음성과 입모양을 동기화 시켜 주는 모듈이 있다.

4.2 입력 처리기 (Input Processor)

가상 현실감 시스템을 사용하는 경우에, 사용자들은 각기 다른 종류의 입력에 의해 가상 세계에 참여하여 다른 참여자들과 상호 작용을 할 수 있다. 흔히 사용하는 GUI로 원하는 종류의 상호 작용을 선택하거나 상대방에게 하고 싶은 말을 직접 텍스트를 입력하여 전하거나 또는 음성을 이용하기도 한다. 또 참여자의 움직임이나 제스처 등을 입력으로 사용할 수도 있다.

다양한 입력을 표정 생성에 이용하기 위하여 입력 처리기 모듈에서는 각기 다른 입력으로부터 표정 생성에 필요한 정보를 추출한다. 구현된 시스템에서는 입력의 종류로 GUI, 텍스트(한/영), 음성, 사용자의 머리의 움직임을 사용하였다. 사용자의 머리의 움직임을 측정하는 도구로는 3차원의 위치와 방향을 구할 수 있는 Bird를 모자에 부착하여 이용하였다. 구해 낸 사용자의 머리의 움직임을 가상 에이전트의 머리의 움직임으로 매핑하여 좀 더 사실적인 애니메이션을 할 수 있다.

이렇게 입력된 다양한 입력정보들은 표정을 생성하는데 이용되는 정보들을 포함하고 있다. 입력 처리기에서는 사용자 인터페이스로부터 입력된 정보로부터 표정 생성에 이용되는 정보를 추출한다. 입력처리기는 다음과 같은 모듈로 이루어진다.

4.2.1 음성 분석기 (Speech Analyzer)

마이크를 통해 입력받은 사용자의 음성을 분석하여 그에 알맞은 표정을 생성하기 위한 정보를 얻어내는 부분이다. 원하는 표정 제어 레벨에 따라 음성으로부터 추출해 내는 정보들이 달라지게 된다. 사용자의 음성에서 표정 생성에 이용되는 정보를 얻어내는 데에는 어느 정도 한계가 있으므로, 본 시스템에서는 입력된 음성의 음량에 따른 입벌림 정도와 선형 분석을 통해 구한 포만트를 이용하여 모음 정보를 얻어내어 이를 표정 생성에 이용한다.입력된 음성으로부터 음량만을 추출하는 것은 실시간에 수행되는 반면 포만트를 구하기 위해서는 약간의 지연이 생기게 된다. 그러므로 원하는 표정 생성의 상세도 제어 레벨이 낮을 경우에는 음성에서 음량만을 추출하지만 좀 더 자세한 레벨이 가능할 경우에는 포만트와 음량을 함께 구한다.

4.2.2 텍스트 분석기 (Text Analyzer)

키보드를 통해 입력받은 텍스트를 분석하여,

표정을 생성하는데 필요한 정보를 추출해 내는 부분이다. 사용자가 의도하는 바를 한글과 영어가 포함된 텍스트의 형태로 입력하면 입력된 텍스트들을 분석하여 발음 형태로 바꾸고 그 발음에 따른 입모양의 변화 정보를 출력해 준다.

· **한/영 분리 모듈** : 한글이나 영어를 포함한 문장이 들어올 경우 한글이면 음절 분석 모듈로, 영어면 단어 분석 모듈로 보낸다.

· **음절 분석 모듈** : 한글 음절은 한글 코드가 정의되어져 있는 테이블을 참조(look up)하여, 초성, 중성, 종성으로 분리한 다음 입모양 표기 모듈로 보낸다.

· **단어 분석 모듈** : 영어 문장이 들어올 경우에는 공백을 이용한 단어 구분을 하여, 약 11만 어휘를 가진 발음기호 데이터베이스[1]를 참조하여 발음 표기로 나타낸다. 구해 낸 한 단어의 발음 표기들을 입모양 표기 모듈로 보낸다.

· **입모양 표기 모듈** : 발음 정보가 입력되면 이들에 대한 정보를 가지고 있는 데이터베이스를 참조하여, 입모양을 변화시키는 입모양 표기로 변환한다. 예를 들어 /b/와 /p/를 발음하는 경우에, 발음하는 입모양은 같게 되므로 입모양 표기는 같게 되며, 영어 발음 표기인 "AO"의 경우에는 "A"와 "O", 두 개의 입모양 표기로서 나타내어진다.

4.3 비언어적 표정정보 분석기 (Non-lingual Information Analyzer)

사용자는GUI 제어판을 조절함으로써 말을 하는 도중에 강세를 주고 싶은 부분을 선택할 수 있다. GUI 제어판을 이용하여 연출되어진 표정의 예는 〈그림 4〉와 같다. 이와 같이 비언어적 표정정보 분석기는 제어판으로부터 입력되어진 그래프로부터 시간 경과에 따라, 감정이나 눈 깜박임 등과 같은 직접적인 언어 전달과는 관계없으나, 말하는 것을 강조하는 역할을 하는 비언어적 표정정보를 추출해 낸다.

〈그림 4〉 GUI제어판을 이용하여 말하는 동안에 강세를 준 표정

주1)The Carnegie Mellon Pronouncing Dictionary [cmudict.0.3]

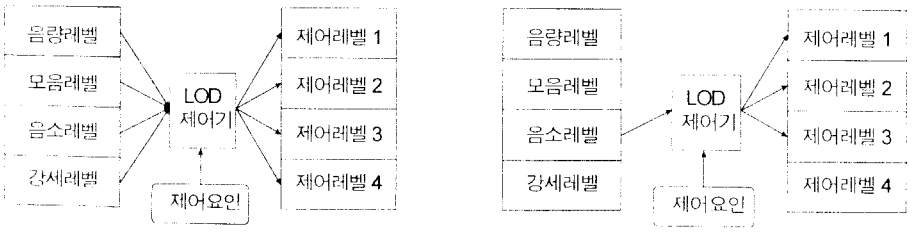
4.4 표정 제어기 (Expression Controller)

표정 제어기(expression controller)는 LOD 제어기와 표정 제어 파라미터 생성 모듈로 구성되어 있다. LOD 제어기는 입력 처리기로부터 들어오는 입력으로부터 입력 레벨을 결정하고, 입력 레벨과 상세도 제어 요인을 고려하여 적당한 표정 제어 레벨을 결정하는 모듈이다. LOD 제어기에 의해서 제어 레벨이 결정되어지면 그에 해당하는 적당한 제어 파라미터가 생성되어져야 한다. 표정 제어 파라미터 생성 모듈에서는 제어 레벨에 따른 제어 파라미터를 생성한다. 결정된 제어 레벨이, 입력이 가지고 있는 표정 생성 정보들을 사용하는 레벨일 경우에는 입력으로부터 제어 파라미터를 생성한다. 제어 레벨이 그보다 낮은 레벨인 경우에는 입력으로부터 얻어진 제어 파라미터로부터 그 보다 낮은 레벨의 제어 파라미터를 구해야 한다.

4.4.1 LOD 제어기

LOD 제어기(Level-of-Detail controller)는 적절한 상세도의 표정 생성을 위해 가능한 입력 레벨과 상세도 제어 요인을 고려하여 원하는 상세도에 해당하는 표정 제어 레벨을 결정하는 모듈이다. 각 입력이 들어오게 되면 표정을 생성하는 데 사용되는 정보의 양이 많을수록 입력 레벨이 높게 정해지게 된다. 따라서 입력 레벨이 높을 수록 표정 생성의 상세도에 관련된 제어 레벨도 높아지게 된다.

그러나 입력 레벨이 높을 경우라도 여러 가지 요인(예를 들어 프레임 속도, 모델의 복잡도, 호환성)으로 인하여, 입력 레벨이 표현할 수 있는 만큼의 제어 레벨로 표정을 생성할 수 없거나 그럴 필요가 없는 경우가 있다. 이 경우에는 적당하게 제어 레벨을 낮추어야 한다. 상황에 따라 충분히 상세한 표정을 생성할 수 있는 경우에는 가능한 가장 상세한 제어 레벨을 선택하도록 한다. 이와 같이 사용자의 요구와 상황에 따



〈그림 5〉 (좌) LOD제어기와 (우)결정되어지는 제어 레벨의 예

〈표 1〉 사용된 AU(action unit)들의 종류

AU	표정 생성	AU	표정 생성
AU1	눈까풀 상하운동(좌)	AU2	눈까풀 상하운동(우)
AU3	눈동자 상하운동(좌)	AU4	눈동자 좌우운동(좌)
AU5	눈동자 상하운동(우)	AU6	눈동자 좌우운동(우)
AU7	입의 벌림	AU8	입술을 당김(좌)
AU9	입술을 당김(우)	AU10	입술을 당김(좌하)
AU11	입술을 당김(우하)	AU12	눈썹 상하운동
AU13	코를 찡그림(좌)	AU14	코를 찡그림(우)
AU15	눈썹을 모음	AU16	입술을 오므림
AU17	눈을 크게 뜸(좌)	AU18	눈을 크게 뜸(우)
AU19	입술을 돌출(상)	AU20	입술을 돌출(하)

라 동적으로 표정 제어 레벨이 변경 가능해지면, 3.2 절에서 살펴보았듯이 전송해야 할 표정 제어 정보의 양과, 실시간의 제약을 고려하는 등의 이유로 가상 현실감 시스템에 많은 이득을 가져다준다.

입력 처리기로부터 구해진 입력들은 LOD 제어기로 전해진다. LOD 제어기는 들어온 입력들로부터 입력 레벨을 결정하고, 〈그림 5〉 (좌)와 같이 입력 레벨과 제어 요인들을 고려하여 적절한 제어 레벨을 선택하도록 하는 역할을 한다. 입력으로부터 얻어진 표정 생성 정보들을 충분히 사용할 수 있는 경우에는 입력 레벨과 직접적으로 대응되는 제어 레벨을 선택하도록 한다. 그러나 상황에 따라서는 입력 레벨과 대응되는 제어 레벨을 선택하는 대신, 그보다 낮은 제어 레벨을 선택하는 경우가 있다.

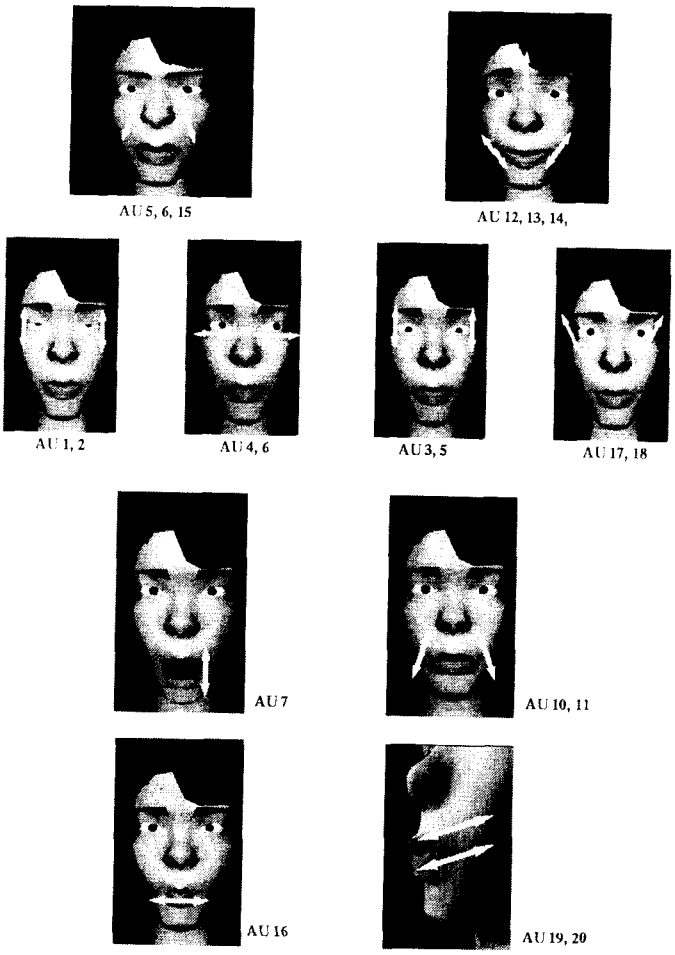
〈그림 5〉 (우)와 같이 제어 레벨은 입력 레벨과 직접적으로 대응되는 레벨이거나 또는 그보다 낮은 레벨로 결정되어진다. 입력이 가지고 있는 표정 생성 정보로부터 제어 파라미터를 구할 수 있는, 즉, 입력 레벨과 직접적으로 대응되는 제어 레벨과의 관계는 실선으로 표시된다. 제어 요인에 의해 그 보다 낮은 레벨로 매핑되는 경우는 입력 레벨과 제어 레벨은 간접적으로 대응되고, 점선으로 표시된다. 제어 요인으로서는 실시간에 수행되고 있는지를 검사할 수 있는 프레임 속도, 네트워크의 속도와 모델의 복잡도, 모델과 관찰자와의 거리나 모델의 가시성, 또는 사용자의 요구 등이 있다. 구현상에서는 제어 요인으로서 사용자의 요구만을 고려하였다.

4. 4. 2 표정 제어 파라미터 생성 모듈

LOD 제어기로부터 제어 레벨이 결정되면,

표정을 생성하는데 사용되는 제어 파라미터를 구해야 한다. 제어 레벨에 해당되는 표정을 생성하기 위해 사용되어지는 제어 파라미터는 3.3.1절에서 보았듯이 다음과 같이 정의된다.

- 제어 레벨 1: = 입의 벌린 정도
- 제어 레벨 2: = 이심률, = 장축의 길이
- 제어 레벨 3: (AU_7,AU_{10},AU_{11},AU_{16}, AU_{19},AU_{20}) 입모양 생성을 위해 정의되어진 AU들
- 제어 레벨 4: (AU_1,AU_2,AU_3,AU_4, AU_5,....,AU_{19},AU_{20}) 얼굴의 표정 생성을 위해 정의되어진 AU들



〈그림 6〉 표정과 입모양에 관계된 AU에 의한 움직임

제어 레벨 1의 파라미터인 입의 벌린 정도는 입이 상하로 벌어진 정도를 0과 1사이의 값으로 정규화 하여 사용하므로 가장 많이 벌렸을 때의 값이 1을 갖게 되고 입을 다물고 있는 경우 0 값을 가지게 된다.

제어 레벨 2의 파라미터인 이심률과 장축의 길이는 AU값에 의해 변화된 모델의 입을 구성하는 3차원 점들과 가장 근접한 타원을 나타내는 것이다. 일반적으로 사용되는 정의와는 달리 본 구현에서는 입의 좌우방향을 장축의 방향으로 하고, 상하방향을 단축의 방향으로 정의하여 타원의 방향을 고정함으로써, 일반적인 정의를 사용할 때에 생길 수 있는 모호성(ambiguity)을 배제하였다.

제어 레벨 3과 4에서 정의된 제어 파라미터인 AU들은, 표정 생성 시 가장 하위 레벨에서의 움직임을 변화시키는 단위이다. 얼굴의 표정을 모두 생성하기 위해 FACS에서 정의한 44개의 AU들을 구현하는 것이 이상적이나, 본 논문에서는 주로 말하는 동안의 입모양 생성에 관심이 있으므로, 〈그림 6〉과 같이 인간의 표정에 기본이 되는 6가지의 표정과 자연스러운 입모양에 대해서만 파라미터를 정의하여 사용하였다[Lee95]. 사용되어진 AU들의 종류는 표 1과 같다. 이들 AU들은 각각 0에서 1사이의 값을 가지게 된다. 예를 들어 AU_7의 경우, 값이 0이면 입을 다물고, 1로 가까이 갈수록 입을 크게 벌리게 된다. 이들 AU중에서 입모양을 생성하는데 영향을 끼치는 AU는 AU_7, AU_{10}, AU_{11}, AU_{16}, AU_{19}, AU_{20}이다.

표정 제어 파라미터의 생성은 두 가지 경우로 나누어 볼 수 있다. 입력 레벨과 직접적으로 대응되는 제어 레벨을 위해 파라미터 생성을 해야 하는 경우와 간접적으로 대응되는 제어 레벨을 위해 파라미터를 생성해야 하는 경우이다. 제어 레벨이 입력 레벨과 직접적으로 대응되는 경우에는, 입력 데이터에 포함되어 있는 표정 생성 정보로부터, 제어 레벨에서 필요로 하는 제어 파라미터를 추출해야 한다. 제어 레벨이 입력 레벨과 간접적으로 대응되는 경우에는, 제어 파라미터는 직접적으로 대응되는 제어 레벨의 파라미터로부터 추출하여야 한다. 그러므로, 표정 제어 파라미터의 모듈은 직접 제어 레벨의 제어 파라미터를 구하는 모듈들과, 간접 제어 레벨로 제어 파라미터를 자동 생성해 주는 모듈들로 이루어져 있다.

직접 제어 레벨의 제어 파라미터를 생성해 주는 모듈

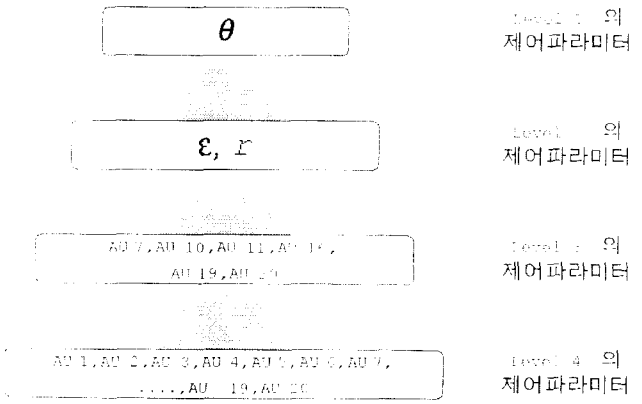
이 모듈은 결정되어진 제어 레벨이 입력 레벨과 직접적으로 대응되는 제어 레벨인 경우에, 제어 파라미터를 생성해 준다. 입력 레벨은 각각의 입력 데이터들이 가지고 있는 정보에 의해 제어 레벨과 대응된다. 입력 레벨에 해당하는 입력 데이터에서 얻을 수 있는 정보를 이용하여, 테이블 참조(table-lookup)나 정해진 매핑에 의해 제어 파라미터의 값을 구한다. 각 제어 레벨에서의 제어 파라미터는 다음과 같이 구한다.

· 제어 레벨 4 : GUI를 사용한 제어판을 통해 들어온 입력은 비언어적 표정 분석기에 의해 각각의 비언어적 표정 변화의 정도로 변환되어 진다. 입력된 텍스트는 텍스트 분석기에 의해 입모양 표기로 변환되어 진다. 비언어적 표정 정보로는 행복과 노여움의 정도와 눈의 깜박임 정보 등이 있다. 이러한 비언어적 표정과 입모양 표기들은 각기 어떠한 AU 값을 가져야 한다는 것이 정의되어져 있는 데이터베이스를 가지고 있다. 예를 들어 행복정도 1, 눈의 깜박임 정도 0(눈을 감고 있는 상태)일 경우에 AU_8, AU_9, AU_{12}의 값이 0.5이고, AU_1, AU_2의 값이 0이다. 입력 처리기를 통하여 얻어진 비언어적 표정 변화의 정도와 입모양 표기들은 데이터베이스를 참조하여 AU값을 구해 낸다. 이렇게 구해진 모든 AU의 값이 제어 파라미터가 된다.

· 제어 레벨 3 : 입력된 텍스트는 입모양 표기로 변환되어 진다. 입모양 표기에 대해 각기 어떠한 AU 값을 가져야 한다는 것이 정의되어져 있는 데이터베이스를 참조하여 각 입모양을 생성하는 데 필요한 AU값을 구해 낸. 구해진 입모양 생성에 관련된 AU들이 제어 파라미터가 된다.

· 제어 레벨 2 : 음성이 입력으로 들어오면 음성 분석기에서 말하는 도중에 모음의 변화를 구할 수 있다. 3차원 모델이 각 모음의 입모양을 할 경우, 입술을 구성하는 점들을 근사하는 타원의 이심률과 장축의 길이를 구해 놓은 테이블을 참조하여, 이심률과 장축의 길이를 구해 낸다.

· 제어 레벨 1 : 음성 분석기에서는 입력으로 들어온 음성으로부터 단지 음량정보만을 구해



〈그림 7〉 제어 파라미터의 자동 생성

낸다. 음량정보는 입의 벌어진 정도로 매핑될 수 있고, 매핑되어진 입의 벌어진 정도는 제어 파라미터가 된다.

간접 제어 레벨로 제어 파라미터의 자동 생성 모듈

이 모듈에서는 결정되어진 제어 레벨이 입력 레벨과 간접적으로 대응되는 제어 레벨인 경우에, 제어 파라미터를 생성해 준다. 제어 레벨은 구현상에서 네 단계로 정의되어져 있다. 이들 하위 제어 레벨에서 낮은 제어 레벨로의 파라미터 자동 생성 과정은 〈그림 7〉과 같다. 제어 파라미터의 자동 생성 모듈은 N 레벨에서 N-1 레벨로 파라미터를 생성하도록 이루어졌다. 그러므로 두 레벨이상 낮은 레벨의 제어 파라미터를 생성하려면 N레벨에서 N-1레벨로 파라미터를 생성하고, 생성되어진 파라미터로 다시 N-2레벨의 파라미터를 생성해야 한다. 각 제어 레벨 N에서 제어 레벨 N-1로의 제어 파라미터의 생성은 다음과 같다.

· 제어 레벨 4 ⇒ 제어 레벨 3

얼굴 표정을 생성하기 위해 정의되어진 전체 AU 중에서, 제어 레벨 3의 제어 파라미터인 입모양 생성에만 관여하고 있는 AU를 추출한다.

· 제어 레벨 3 ⇒ 제어 레벨 2

입모양 생성에 사용되는 AU들에서, 입모양의 윤곽을 나타내 주는 타원의 이심률과 장축의 길이를 구해 내야 한다. AU의 변화에 의해 입을 구성하는 3차원 점들이 어떠한 타원의 형태를 갖는지를 계산하는 것은 많은 반복을 거쳐 근사하는 타원의 식을 찾는 것이므로 실시간을 요구하는 본 시스템에는 부적당하다. 이러한 지연시간을 줄이면서 복잡한 매핑에 대해 좋은 결과를 얻을 수 있도록 신경회로망을 사용하여 제어 레벨 2의 파라미터인 이심률과 장축의 길이를 구한다.

· 제어 레벨 2 ⇒ 제어 레벨 1

이심률과 장축의 길이로부터 입의 벌린 정도를 나타내는 파라미터를 추출하여야 한다. 본 구현에서는 타원의 이심률과 장축의 길이로부터 입의 전체적인 윤곽을 알 수 있어야 하므로 일반적인 타원의 정의 (단축의 길이 / 장축의 길이)를 사용하지 않고 입의 좌우방향을 장축의 방향으로 정의하였다. 그러므로, 이심률과 장축의 길이로부터 타원의 상하방향의 길이를 구해 낸다. 이는 입의 벌린 정도를 나타내므로 이를 제어 레벨 1의 파라미터로 사용한다.

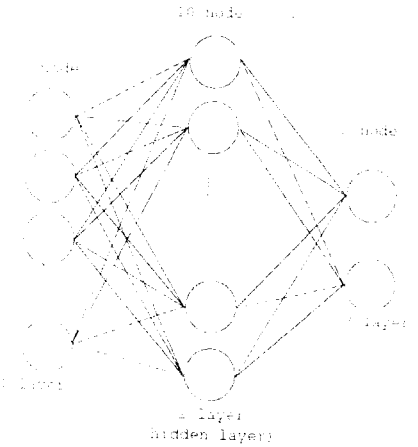
제어 레벨 2의 제어 파라미터를 생성하기 위해 사용된 신경회로망은 실시간 응답과 병렬처리의 특성을 가지고 있다. 신경회로망의 한 유형인 다층망(multilayer network)은 일반적으로 가장 많이 쓰이고 있는 신경회로망으로서, 가장 중요한 특성은 어떠한 복잡도의 매핑이라도 배울 수 있다는 점이다[Zurada92]. 다층망

을 사용하려면 우선, 교사 학습방법(supervised learning)을 사용하여 망을 훈련시켜야 한다. 교사 학습이란 엄선된 입력 자료와 희망하는 결과를 함께 사용하여 신경망을 훈련시키는 방법이다. 입력 자료가 망을 통과하여 얻어지는 출력 결과와, 희망하는 결과와의 차이를 이용하여 망의 연결 강도가 갱신된다.

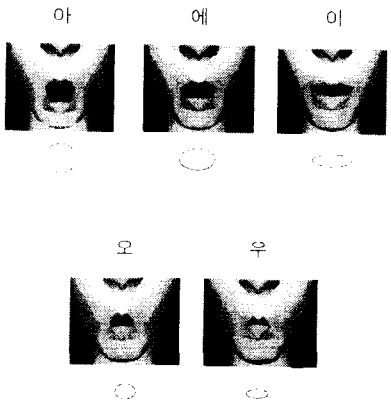
본 시스템에서는 제어 파라미터를 실시간에 구하기 위해 3층으로 된 다층망을 사용하였다. 엄선되어진 훈련 자료를 사용하여 교사 학습을 시켰고, Error-back-propagation 알고리즘을 사용하여 다층망의 훈련이 좀 더 빨리 이루어질 수 있도록 하였다. 사용한 다층망의 구조는 〈그림 8〉과 같이 입력 노드의 개수는 6개이고, 은닉층의 노드수는 10개이며 출력 노드의 개수는 2개이다. AU들의 값을 변화시킴에 따른 입모양의 이심률과 장축의 길이를 구하고자 하므로, 입력 노드에는 입모양을 생성하는 6개의 AU값이 주어지게 되고 출력 노드에서는 이심률과 장축의 길이를 얻는다.

훈련에 사용되어진 자료는, 40여 개의 입력(6개의 AU값들)과 각각의 입력에 대해 원하는 출력(이심률과 장축의 길이)의 쌍으로 이루어진 자료로써 수 십만 번의 훈련으로 원하는 값과 실제 구해 낸 값 사이에 오차가 거의 없는 출력을 얻게 되었다.

신경회로망을 이용하여 제어 레벨 3에서 제어 레벨 2로 파라미터를 추출하여 표정을 생성한 예는 〈그림 9〉와 같다. 이들은 각각 텍스트로 "아 에 이 오 우"가 입력되었을 때 제어 레벨 2로 파라미터를 추출하여 입모양을 생성한 것이다. 이들의 입모양은 신경회로망으로부터 구해진 이심률과 장축의 길이를 제어 파라미터로 사용하여 생성된다. 제어 레벨 3에서의 입모양

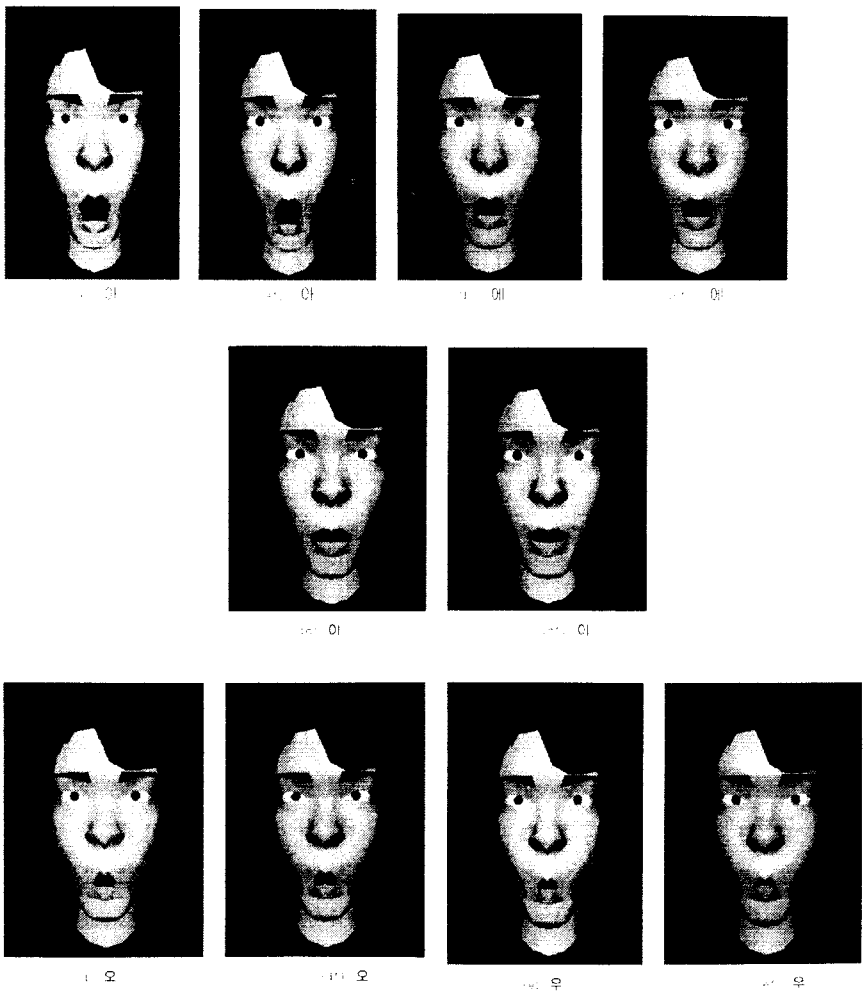


〈그림 8〉 사용한 신경회로망의 구조



〈그림 9〉 신경회로망을 통해 추출한 제어 파라미터로 생성한 입모양

과 파라미터 추출을 통해 구해진 제어 레벨 2에서의 입모양의 비교는 〈그림 10〉과 같다. 구해진 각 제어 레벨에 해당되는 제어 파라미터들은 동기화 모듈로 전달된다.



〈그림 10〉 제어 레벨3 (좌)에서의 입모양과 파라미터 추출을 이용한 제어 레벨2 (우)에서의 입모양 비교

5. 음성 합성 모듈

한글이나 영어 문장이 입력으로 들어 올 경우에는 입모양을 만드는 동시에 텍스트를 분석하여 그에 해당하는 음성을 합성해 주는 부분이 필요하다. 입력 처리기내에 있는 텍스트 분석기에서는 한글이나 영어 문장을 분석하여 발음기호를 음성 합성 모듈로 보내준다. 음성 합성 모듈에서는 각 발음별로 녹음되어진 음성 데이터들을 찾아서 합성하여 출력한다.

5.1한글 텍스트로부터 음성 합성

한글일 경우에는 한 음절별로 발음이 이루어지므로 모든 음절에 대하여 녹음되어져 있는 데이터가 있을 경우에는 어떤 글자인지만으로 그에 해당하는 완전한 음성 데이터를 구할 수 있다. 한글의 음절은 자음-모음-자음의 음절 형태를 이루며 한글이 가질 수 있는 모든 음절의 개수는 동음이어를 동일시하더라도 약 2800여개의 음성 데이터가 필요하게 된다. 이렇게 많은 음성 데이터를 사용하는 것은 메모리를 굉장히 많이 차지하게 될 뿐 아니라 각 음절에 대하여 녹음된 음성 데이터를 얻는 것도 쉽지 않은 일이다.

이러한 어려움으로 인해 구현에서는 다음과 같이 온음절과 반음절을 함께 사용했다. 온음절이라하면 일반적으로 생각하는 한 글자를 말하는 것으로서, 상시 많이 사용되어지는 글자일 경우에는 온음절로 녹음 데이터를 만든다. 한 글자를 두 부분으로 나누어 놓은 것을 반음절이라 한다. "문"이란 글자인 경우 "무" + "운"으로 분리되고, 이 각각이 반음절이다. 한글에 있는 모든 음절들이 같은 빈도수로 사용되지 않으므로 잘 사용하지 않는 음절일 경우에 반음절로 만들어 놓으면 필요한 녹음 데이터가 상당히 줄어들게 된다. 구현에서는 자주 사용되어지는 83개의 온음절과 359개의 반음절을 사용하고 있다.

텍스트 분석기로부터 어떤 글자인지에 대한 정보가 들어오게 되면 그 글자에 해당하는 온음절로 녹음된 데이터가 존재하는지 검사하여 온음절 데이터가 존재하면 그 데이터와 데이터의 지속시간을 동기화 모듈에 보낸다. 온음절 데이터가 없을 경우에는 반음절로 녹음된 데이터들을 합성하여 온음절 데이터를 만들고 만들어진 데이터와 그것의 지속시간을 동기화 모듈에 보낸다.

5.2 영어 텍스트로부터 음성 합성

영어인 경우에는 한 단어 별로 발음이 이루어지므로 발음기호 별로 녹음되어져 있는 데이터를 이용하여 한 단어를 합성하여 사용한다.

구현에서는 영어 단어를 이루고 있는 발음기호들을 표 2와 같이 정의하여 사용하였다. 영어 사전을 참조하여 단어의 발음 기호를 표 2와 같은 39개의 발음 표기로 나타낸다. 한 단어의 발음 표기가 구해지면 발음 표기 별로 녹음되어 있는 음성 데이터들을 합성하여, 그 단어에 해

당하는 음성을 만들어 낸다. 예를 들어 "Hello"라는 단어가 들어오게 되면 텍스트 분석기에서 먼저 표 2에서 정의되어진 표기로 바꾼다. 그러면 "HH AH L OW"라는 발음 표기가 얻어 지게 된다. 각 발음 표기에 대한 음성 데이터는 이미 저장되어 있으므로 이들을 합성하여 한 단어에 대한 음성을 만들어 낸다. 각 단어에 대한 합성되어진 음성이 만들어지면 그 음성의 지속 시간과 합성되어진 음성을 동기화 모듈로 보낸다.

〈표 2〉 사용된 영어 발음 표기

AA	AE	AH	AO	AW	AY	B	CH	D
DH	EH	ER	EY	F	G	HH	IH	IY
JH	K	L	M	N	NG	OW	OY	P
R	S	SH	T	TH	UH	UW	V	W
Y	Z	ZH						

5.3 동기화 모듈

표정 제어기에서 만들어 낸 표정 제어 파라미터와 음성 합성 모듈에서 만들어 낸 합성된 음성이 동기화 되어 출력되기 위해서 동기화 모듈로 들어오게 된다. 실제 음성이 들어오는 경우에는 실시간에 음성에 대한 입모양이 생성되므로 동기화가 필요 없으나 텍스트가 입력으로 들어올 경우에는 합성되어진 음성과 입모양의 동기화가 필요하게 된다.

동기화는 전역적인 시간 축을 근거로 하여 이루어진다. 애니메이션의 시작이 알려지게 되는 시간을, 초기 시간(initial time) t_0라하고, 맨 처음 발음하게 될 합성 음성의 지속 시간, d가 알려지면, 시간 축이 $t_0 + d$를 지나지 않을 때까지는 맨 처음의 발음을 위한 표정 제어 파라미터를 표정 생성기로 보내주는 동시에 맨 처음 발음을 위한 합성 음성을 오디오로 출력한다. $t_0 + d$시간이 지나게 되면 다음 발음을 위해 합성되어진 음성의 지속 시간을 이용하여 다음 발음을 위한 표정 제어 파라미터를 표정 생성기로 보내준다. 이러한 방식으로 텍스트가 끝날 때까지 반복된다.

이 때 표정 제어 파라미터는 어떤 시간의 간격을 두고 입력되어진다. 이렇게 시간에 따라 서로 떨어져 있는 표정 제어 파라미터를 그대로 표정 생성기에 보내게 되면 로보트와같이 연결되지 않은 입모양을 생성하게 된다. 이러한 점을 보완하기 위해 현재 사용되어지는 표정 제어 파라미터의 값과 다음에 오게 될 표정 제어 파라미터의 값 사이를 보간하여(interpolate)

입모양을 좀 더 사람과 흡사하게 만들어 주어야 한다.

하나의 입모양 중심의 표정을 나타내는 표정 제어 파라미터는 그 입모양을 나타내기 위하여 얼마간의 지속시간을 갖고 있다. 동기화 모듈에서는, 자연스러운 움직임을 생성하기 위해, 각 발음에 대한 표정 제어 파라미터들 사이를 보간하여, 시간 축에 따라 비선형적으로 변화하는 표정 제어 파라미터를 만들어 낸다. 구현상에서는 합성되어진 음성의 지속시간을 d라고 했을 때, 현재 제어 파라미터가 생성되어져야 하는 시간(t_{cur})과 처음 그 음성이 출력되기 시작한 시간(t_{start})과의 차이를 Δt라 하고, Δt가 d를 넘지 않을 동안, 파라미터를 코사인 함수에 의해 변형시킨다. 선형 함수 대신 코사인 함수를 이용하여 인터폴레이션 하는 것은 효과적이면서 만족할 만한 결과를 가져다준다.

구해진 표정 제어 파라미터들은 표정 생성기로 전해져 실제 표정을 생성하게 된다. 표정 생성기는 6절에서 소개할 모델 독립적 표정 생성기인 MiFegs를 기반으로 하였다.

6. 표정 생성기

표정 생성을 위해 사용된 MiFegs(Model Independent Facial Expression Generating System)는 한국과학기술원 전산학과에서 개발한 모델 독립적 표정 생성 시스템이다[Lee95]. 이 시스템은 다양한 모델들의 표정 생성을 쉽게 하기 위하여 각 모델에 대한 그래픽 정보(꼭지점과 모서리의 집합)와 설정 정보(표정 생성을 위한 자료)를 사용자가 제공하도록 하여 시스템이 모델 독립적으로 원하는 표정을 생성하도록 한다. 생성된 표정은 렌더링 모듈을 통해 사용자에게 제공된다.

6.1 MiFegs에서의 표정 생성

MiFegs에서는 표정 생성을 위해 감정 레벨에 의한 상위 레벨의 제어, AU의 변화량에 의한 근육의 움직임을 통한 하위 레벨의 제어 단계로 나누어 정의 하였다.

6.1.1 표정 생성을 위한 상위 레벨의 제어

MiFegs에서는 얼굴의 구성요소나 근육의 변화에 대한 지식이 없는 사용자를 위해 개념적인 단계에서의 표정변화를 위한 기능을 제공한다. 즉, 사용자는 "슬프다", "기쁘다" 등과 같이 감정과 관련된 개념적인 지식만으로 원하는 표정을 생성할 수 있도록 하였다. 또한, 단일 감정에 대해서도 미세한 변화에 따른 표정을 생성하기 위해 감정의 정도를 몇단계로 구분하고 이들 단계에 따른 표정을 생성한다. 이를 위해 각각의 감정에 대한 최고 단계에서의 AU 집합들의 변화량을 정의하고 감정의 단계에 따라 이를 보간해서 사용하였다. 감정 단계에 따른 표정을 생성하기 위해서는 각 감정의 최대 표현시 AU의 값이 설정되어 있다고 할 때 임의의 단계 E에 대한 AU의 레벨값은

$$AU[i] = (1 - E)(AU_{netural}[i]) - E(AU_{emotional}[i])$$
$$\text{for } i = 1,2,\ldots,N$$

where

N: Number of AUs

E: Level of Emotion

$AU_{netural}$: AU value at the netural status

$AU_{emotional}$: AU value at some emotional status

와 같이 정의된다.

6.1.2 표정 생성을 위한 하위 레벨의 제어

표정 생성기의 하위 레벨은 AU의 변화에 의한 근육의 움직임, 구성 요소인 눈과 입의 움직임으로 구성된다.

표정 생성을 위한 근육의 움직임

얼굴의 내부적 구성요소로는 얼굴의 형태를 이루는 두개골(뼈)과 움직임의 근간이 되는 근육, 근육과 피부를 이어주는 근막, 그리고 피부로 이루어져 있다. 이들 중 두개골은 턱뼈의 움직임으로 인한 입의 벌림을 제공하며, 근육과 피부의 움직임은 표정의 변화를 생성한다. MiFegs에서는 근육모델에 기반을 둔 표정생성 방식을 채택하고 있으며, 근막과 피부의 움직임에 의한 표정 변화는 고려하지 않는다.

MiFegs에서는 얼굴에 위치한 근육을 그 움직임 특성에 따라 Linear, Flat, Sphincter 세 종류로 나누어 정의하는 방법을 사용하였다 [Waters87]. Flat 유형은 이마와 같은 부분에 위치한 근육과 같이 넓은 영역에 걸쳐 평행하게 움직이는 근육을 정의할 때 사용된다. Linear 유형은 코 옆이나 뺨을 움직이는 것과 같이 좁은 영역에 대해 움직임을 갖는 근육을 정의할 때 사용된다. Sphincter 유형은 입주위나 눈주위같이 원형적인 움직임을 갖는 근육을 정의할 때 사용된다. 이렇게 정의된 근육의 종류에 대해 AU의 레벨 값을 근육으로 정의된 영역을 AU 포인트상으로 끌어 당기는 힘의 정도로, AU 포인트는 근육의 변화에 있어서 정의된 영역이 모이는 곳에 위치한다고 정의하면, AU의 레벨값이 높아질 수록 주위의 영역은 AU 포인트의 위치로 몰려 들게 된다. 이러한 움직임은 파라미터를 이용한 간단한 형태의 선형 보간법을 이용하거나 복잡한 형태의 근육 방정식을 통해 구현할 수 있다[Parke74][Waters87]. 파라미터를 이용한 선형 보간법은 실시간에 처리될 수 있는 간단한 형태를 가지며 제어를 위한 파라미터가 직관적인 반면, 움직임이 제한되고 사실적이지 못한 단점이 있으며, 근육 방정식을 통한 방법은 근육 방정식의 복잡도에 따라 더욱더 사실적인 근육의 움직임 생성이 가능하지만 제어를 위한 파라미터가 직관적이지 못하고 실시간에 처리되기 힘든 문제를 가지고 있다. MiFegs는 이러한 두가지 요소를 적당히 결합해서 아래와 같은 두가지 종류의 근육 방정식을 사용하였다.

$$Type1: \quad P_{x,y,z} = (1-A)S_{x,y,z} + AE_{x,y,z}$$

$$Type2: \quad d = \sqrt{(E_x - S_x)^2 + (E_y - S_y)^2 + (E_z - S_z)^2}$$

$$d_{x,y,z} = \sqrt{(E_{x,y,z} - S_{x,y,z})^2}$$

$$P_{x,y,z} = \frac{d_{x,y,z}}{d} F_{x,y,z} A + S_{x,y,z}$$

where
A: AU level value
F: Factor value
S: vertex at start position of muscle
E: vertex at end position of muscle
P: vertex at new position of skin

유형 1의 경우 A값의 변화에 따라 단순한 선형 보간을 통한 움직임을 나타내며, 유형 2의 경우, F값의 설정에 따라 다양한 형태를 지니는 근육 방정식이 될 수 있다.

얼굴 구성 요소의 움직임

얼굴의 외부적 구성 요소는 크게 눈, 코, 입, 귀 가 있으며 이들 중 표정 변화에 영향을 주는 요소로는 눈, 입이 있다. 눈의 움직임은 깜박임과 눈동자의 움직임으로 구분되며, 입의 움직임은 입의 벌림과 다뭄 그리고 입술의 움직임으로 구분된다. 눈의 깜박임은 무의식 중에 일어나는 행동으로 심리상태에 따라 그 주기가 변화한다. 눈의 깜빡임은 눈을 뜬 상태와 감은 상태의 일련의 주기로 이루어지며 이러한 주기를 사용자가 조절할 수 있게 해 줌으로써 보다 자연스러운 깜박임을 생성할 수 있다. 눈동자의 움직임은 상하 회전과 좌우 회전이 있다. 이는 눈동자로 정의된 그래픽 모델을 회전시킴으로서 생성할 수 있다. 입의 일반적인 움직임인 벌림과 다뭄은 턱으로 정의된 영역을 두개골 영역에 대해 회전시킴으로서 생성할 수 있다.

6.1.3 LOD에 의한 표정 생성

동기화 모듈로부터 전해지게 되는 표정 제어 파라미터에 의해, 표정 생성기는 3차원 모델의 얼굴 표정을 생성하게 된다. 제어 레벨 3과 4의 표정 제어 파라미터가 들어오게 되면, 제어 파라미터를 그대로 AU값으로 사용한다. 제어 레벨 2의 제어 파라미터인 (이심률, 장축의 길이) 나, 또는 제어 레벨 1의 (입의 벌어진 정도)와 같은 제어 파라미터가 들어오게 되면, 이들 파라미터가 나타내고 있는 효과를 낼 수 있도록, 입 주위에 정의되어진 AU와 매핑하여 표정을 생성한다.

구현상에서는 제어 레벨 2의 파라미터들이 표정 생성기에 전해졌을 경우에, 이심률과, 장축의 길이를 이용하여 입의 상하좌우의 길이를 구해 내어, 입모양을 상하로 움직이는데 주로 기여하는 AU와 좌우로 움직이는데 주로 기여하는 AU의 값에 매핑을 시키고, 이들 값에 비례적으로 다른 AU값들도 정해진다. 제어 레벨 1의 파라미터인 '입의 벌어진 정도' 가 들어오게 되는 경우, 이 파라미터는 입의 상하움직임을 주로 제어하는 AU값에 매핑시킨다. AU값들은 MiFegs의 입력으로 들어가 표정을 생성한다. 생성되어진 표정의 출력은 렌더링 되어져서 사용자에게 보이게 된다.

7. 결론 및 향후 연구방향

본 연구는 표정 생성을 위해 다양한 입력수단을 제공하고, 상황에 따라 적절한 상세도로 표정을 생성할 수 있도록 다중 제어 레벨을 갖는 입모양 중심의 표정 생성 시스템을 설계하고 구현하였다.

기존의 표정 생성 시스템들에서는 다중 참여자들이 상호작용을 하는데 있어서, 효율적으로 수행하기 위한 고려는 하지 않았다. 그러나 실시간에 이루어져야 하는 상호작용을 위해, 되도록 전송량을 줄이면서, 또는 서로 다른 표정 생성 시스템 사이에서도 효율적으로 호환될 수 있도록 하기 위해서는 적절한 대응책이 필요하다.

본 연구에서는 물체의 복잡한 외형(geometry)에 주로 적용되었던 Level-of-Detail개념을 표정 생성에 적용하였다. 그리하여 상황(실시간의 제약, 호환성, 전송량)등을 고려하여 대화하는데 필요한 입모양과 간단한 표정의 애니메이션의 상세도 레벨을 적절하게 선택하여, 원하는 정도의 상세도로 표정을 생성할 수 있도록 한다.

다중 제어 레벨을 갖도록 하기 위해서, 입력 데이터가 포함하고 있는 표정 생성에 이용될 수 있는 정보의 양에 따라 입력 레벨을 정의하였고, 표정을 생성하는데 있어서의 상세도에 따라 제어 레벨을 정의하였다. 그리고, 제어 레벨을 동적으로 변경시키기 위해서는 신경회로망을 사용하여 필요한 제어 파라미터를 실시간에 효율적으로 구하도록 하였다.

이와 같이 상황에 따라 동적으로 표정 제어 레벨이 변경 가능해지면, 전송해야 할 표정 제어 정보의 양과, 실시간의 제약, 호환성 등을 고려하므로 다수의 사용자가 참여하는 시스템에 유용하게 사용될 수 있다.

본 연구와 관련되어 앞으로 연구해야 할 과제는 다음과 같다.

· 입모양의 전체적인 윤곽을 표현하는 방법에 대한 연구

본 연구에서는 입모양의 전체적인 윤곽을 나타내기 위해 타원을 사용하였으나, 더 복잡하고 미묘한 입모양에 대한 생성을 위해 입모양에 굽어진 정도와 상하좌우로의 길이변화 등을 표현할 수 있는 방법에 대한 연구가 필요하다.

· 제어 요인에 대한 동적인 제어 레벨 변경 방법에 대한 연구

본 연구에서는 사용자의 요구만을 제어 요인으로 삼아 구현하였으나, 이를 확장하여 다중 참여자가 사용하는 시스템에서, 모든 제어 요인들(프레임 속도나 네트워크 상태, 모델의 상세도 등)을 고려하여 동적으로 제어 레벨을 변경하는 방법에 대한 연구가 필요하다.

참고 문헌

[Oh91] 오 영환, *패턴 인식론*, 정익사, 1991.

[Lee95] 이 선우, "모델 독립적 표정 생성 시스템의 설계 및 구현에 관한 연구," *석사 학위 논문, KAIST 전산학과*, 1995.

[Astheimer94] P. Astheimer and P. Maria-Luise, "Level-of-Detail Generation and its Application in Virtual Reality." *Proceedings of VRST'94*, pp. 299-309, 1994.

[Benoit94] C. Benoit, "Real-Time Analysis and Intelligibility of Talking Faces," *2nd International Conference on Speech Synthesis*, 1994.

[Des66] J. Destandes, "Histoire comparee du cinema," 1, 1966.

[Ekman77] P. Ekman and W.V. Friesen, *Manual for the Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, 1977.

[Hill88] D. R. Hill, A. Pearce, and B. Wyvill, "Animating Speech: An Automated Approach Using Speech Synthesis by Rules," *Visual Computer*, Vol.3, pp.277-289, 1988.

[Johnson93] S. K. Johnson, "Neural Modeling of Face Animation for Telecommuting in Virtual Reality," *IEEE VRAIS 93*, pp.478-485, 1993.

[Lewis87] J. P. Lewis and F. I. Parke, "Automated Lip-Synch and Speech Synthesis for Character Animation," *Proceedings of Human Factors in Computing Systems and Graphic Interface*, pp.143-147, Apr. 1987.

[Lewis91] J. P. Lewis, "Automated Lip-Synch: Background and Techniques," *The Journal of Visualization and Computer Animation*, pp.118-122, 1990.

[Massaro90] W. Massaro and M. Cohen, "Perception of Synthesized Audible and Visible Speech," *Psychological Science*, pp.55-63, 1990.

[Morishima93] S. Morishima and H. Harashima, "Facial Expression Synthesis Based on Natural Voice for Virtual Face to Face Communication with Machine," *IEEE Virtual Reality Annual International Symposium 93*, pp.486-491, 1993.

[Nagao94] K. Nagao, "Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation," *Proceedings of the 32nd Annual Meeting of the Association for Computational*

Linguistics, pp.102-109, 1994.

[Ohya93] J. Ohya and Y. Kitanura, "Realtime Production of 3D Human Images in Virtual Space Teleconferencing," *IEEE Virtual Reality Annual International Symposium 93*, pp.408-414, 1993.

[Parke74] F. I. Parke, "A Parametric Model for Human Faces," *Ph. D. Thesis, TR UTEC-CSc-75-047*, University of Utah, 1974.

[Pelachaud94] C. Pelachaud and C. Seah, "Modeling and Animating the Human Tongue During Speech," *Proceedings of Computer Animation*, 1994.

[Waters87] K. Waters, "A Muscle Model for Animating Three-dimensional Facial Animation," *Proceedings of SIGGRAPH*, pp. 17-24, July 1987.

[Waters91] K. Waters and D. Terzopoulos, "Modeling and Animating Faces Using Scanned Data," *Journal of Visualization and Computer Animation*, Vol.2, pp.123-128, 1991.

[Waters94] K. Waters and T. Levergood, "An Automatic Lip-Synchronization Algorithm for Synthetic Faces," *Proceeding of ACM Multimedia '94*, 1994

[Williams90] L. Williams, "Performance Driven Facial Animation," *Proceedings of SIGGRAPH*, Aug. 1990.

[Yau88] D. Yau, "A Texture Mapping Approach to 3D Facial Image Synthesis," *Computer Graphics Forum*, pp.120 -134, 1988.

[Yoshida95] M. Yoshida, "A Virtual Space Teleconferencing System that Supports Intuitive Interaction for Creative and Cooperative Work," *Proceedings 1995 Symposium on Interactive 3D Graphics*, 1995.

[Zurada92] M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, 1992.