

모바일 SMS용 캐릭터 애니메이션을 위한 감정 기반 제스처 스타일화

변혜원[†], 이정숙^{††}

요 약

새로운 텍스트 입력으로부터 제스처를 생성하는 것은 컴퓨터 게임, 가상현실 등의 응용분야에서 종종 요구되는 중요한 문제이다. 최근에는 유명 아나운서와 같은 특정인의 제스처를 모방하는 제스처 스타일화에 관한 관심이 증가하고 있다. 그러나 기쁨이나 슬픔과 같은 감정을 이용하여 제스처를 스타일화하려는 시도는 많지 않다. 또한 이전의 연구는 대부분 실시간 알고리즘에 초점을 맞추고 있지 않다. 본 논문에서는 SMS 문장을 캐릭터 얼굴 표정과 제스처 애니메이션으로 자동 변환하고 감정 요소를 이용하여 제스처를 스타일화하는 시스템을 제안한다. 이 시스템의 가장 큰 특징은 제스처에 회로애학의 감정을 결합하는 실시간 알고리즘을 제시하는데 있다. 모바일 단말기를 플랫폼으로 하기 때문에 서버와 단말기에 작업을 적절하게 분배하고 초당 20프레임 이상의 실시간 수행을 보장한다. 먼저, 디즈니 영상에서 감정을 표현하는 단어와 이에 해당하는 제스처를 추출하고 통계적으로 모델링 한 후, 감정과 제스처의 결합을 위하여 Laban의 움직임 이론을 도입한다. 제안된 시스템의 타당성과 기존 서비스와의 대체 적정도를 알아보기 위해서 사용자 반응을 조사 분석한다.

Emotion-based Gesture Stylization For Animated SMS

Hae Won Byun[†], Jung Suk Lee^{††}

ABSTRACT

To create gesture from a new text input is an important problem in computer games and virtual reality. Recently, there is increasing interest in gesture stylization to imitate the gestures of celebrities, such as announcer. However, no attempt has been made so far to stylize a gestures using emotion such as happiness and sadness. Previous researches have not focused on real-time algorithm. In this paper, we present a system to automatically make gesture animation from SMS text and stylize the gesture from emotion. A key feature of this system is a real-time algorithm to combine gestures with emotion. Because the system's platform is a mobile phone, we distribute much works on the server and client. Therefore, the system guarantees real-time performance of 15 or more frames per second. At first, we extract words to express feelings and its corresponding gesture from Disney video and model the gesture statistically. And then, we introduce the theory of Laban Movement Analysis to combine gesture and emotion. In order to evaluate our system, we analyze user survey responses.

Key words: Mobile 3D animation(모바일 3D 애니메이션), Emotional gesture generation(감정 기반 제스처 생성), Laban Movement Analysis(라반 움직임 분석)

* 교신저자(Corresponding Author) : 변혜원, 주소 : 서울시 성북구 동선동 3가 249-1 동선동길 54-7(136-742), 전화 : 02)920-7615, FAX : 02)920-7114, E-mail : hyewon@sungshin.ac.kr
접수일 : 2010년 2월 18일, 수정일 : 2010년 4월 23일
완료일 : 2010년 5월 4일

[†] 정회원, 성신여자대학교 IT학부

^{††} 정회원, 성신여자대학교 일반대학원 전산학과
(E-mail : jjungi84@sungshin.ac.kr)

* 이 논문은 2007년 정부(교육인적자원부)의 재원으로
한국학술진흥재단의 지원을 받아 수행된 연구임
(KRF-2007-531-D00026/1)

1. 서 론

캐릭터 애니메이션을 제작하는 하나의 방식으로서 텍스트 입력으로부터 제스처를 생성하는 연구는 많은 주목을 받아 왔다. 특히 최근에는 유명 아나운서와 같은 특정인의 제스처를 모방하는 제스처 스타일화에 관한 관심이 증가하고 있다. 하지만 아직까지 기쁨이나 슬픔과 같은 감정 요소를 이용하여 제스처의 스타일화 문제를 다룬 연구는 많지 않다. 이러한 문제는 휴대폰 단문 서비스(SMS: Short Message-Service)를 제스처 애니메이션으로 재생하는 멀티미디어 서비스에서 매우 유용하게 사용될 수 있다. SMS는 단문을 이용하여 자신의 의사와 특히 감정을 즉각적으로 전달하는 수단으로서 사람의 표정을 기호화한 이모티콘(emoticon) 등을 제공하며 감정 표현에 큰 비중을 두고 있기 때문이다. 이와 같은 휴대폰 서비스를 응용 분야의 대상으로 할 때 대두되는 또 하나의 문제는 실시간 알고리즘을 제시해야 한다는 점이다.

본 논문에서는 SMS 텍스트 입력으로부터 얼굴 표정과 제스처 애니메이션을 자동 생성하는 실시간 시스템을 제시하고 특히 효과적인 감정 전달을 위하여 감정과 제스처를 합성하는 새로운 방법을 제안한다. 감정 상태가 다소 과장되어 있는 디즈니 영상 분석을 통해 제스처를 통계적으로 모델링하고 SMS 텍스트로부터 단어중요도 개념을 정의하며 이를 이용하여 제스처 애니메이션을 생성한다. 이때, SMS 텍스트로부터 추출한 감정 상태를 효과적으로 상대방에게 전달하기 위하여 Laban의 움직임 이론을 도입하여 감정과 제스처 애니메이션을 합성한다. 모바일 단말기 플랫폼에서 많은 계산량을 빠른 속도로 처리하기 위해서 시스템을 클라이언트와 서버 구조로 설계하여 실시간 모바일 콘텐츠 서비스를 가능하게 한다.

제안한 시스템의 효용성을 보이기 위하여 다양한 SMS 텍스트를 입력으로 제안한 방법을 3차원 캐릭터 모델에 적용하고, 안드로이드 모바일 단말기에서 실시간 수행이 가능함을 보인다. 그리고 시스템의 타당성, 사용자 만족도 및 기존 서비스와의 대체 적정도를 평가하기 위해서 사용자 설문조사를 수행한다. 본 논문은 감정과 제스처를 결합하는 실시간 알고리즘을 모바일 단말기에서 동작 가능한 서비스 형태로 적용한 첫 번째 연구결과로서, 스마트폰 어플리케이

션 시장이 확대되는 시점에서 새로운 가능성을 모색한다.

본 논문의 순서는 다음과 같다. 2장에서는 제안하는 시스템과 관련된 기존 연구를 살펴보고 3장에서는 시스템 개요를 소개한다. 4장에서 감정 추출, 5장에서 제스처 모델링, 6장에서 제스처 애니메이션, 그리고 7장에서는 감정과 제스처 합성에 관하여 설명한다. 그리고 8장에서 실험 결과를 분석하고 9장에서 결론을 맺는다.

2. 관련연구

텍스트 입력으로부터 캐릭터의 제스처 애니메이션을 생성하기 위해서 다양한 분야에서 관련 연구가 활발하게 진행되어 왔다[1-6]. Cassell 등의 초기의 연구는 입력 텍스트로부터 제스처를 생성하는 규칙을 정의하는 부분에 주목하고 있다[1]. 이후에 Noma 등은 특정 단어에서 특정 제스처를 수행하도록 수작업으로 표기하거나 규칙을 정의하는 방식을 제안하였다[2]. 이 연구에서는 제스처 애니메이션 생성에 관한 문제보다는 발표자가 청중의 눈을 쳐다봐야 하는 등의 발표자로서의 좋은 자세를 모델링하는 규칙에 초점을 맞추고 있다. Cassel 등은 후속연구에서 언어 구조 분석이나 문맥 구조 분석을 도입하여 보다 복잡한 형태의 제스처를 생성하였다[3]. 이들 대부분의 연구는 텍스트로부터 제스처를 생성하는 규칙을 사용하는 방식으로서 인위적인 규칙에 의해 제스처를 생성하는 반면에 본 논문에서는 디즈니 비디오를 분석하여 주인공의 자연스러운 제스처 특성을 통계적으로 모델링하는 방식을 제안한다.

최근에는 제스처 스타일에 관한 연구가 진행되고 있다. Stone 등은 특정인의 제스처 스타일을 표현하기 위해서 미리 캡처한 동작 데이터를 재배열하여 재생하는 데이터 기반 접근 방법을 시도하였다[7]. 이 방법은 미리 캡처한 동작 데이터의 종류와 범위에 의존적이라는 한계를 가지고 있다. Hartmann 등은 시간상의 범위, 공간상의 범위, 활동성, 반복회수 등의 파라메터를 조정하여 제스처 스타일을 표현하고자 하였다[8]. Michael 등은 특정 화자를 촬영한 인터뷰 영상에서 화자의 제스처 스타일을 통계적으로 모델링하고 텁플릿 방식이 아닌 연속적인 제스처 동작으로 애니메이션을 재생하는 방법을 제안하였다[9]. 이

연구에서는 영화 대본에 등장하는 주인공의 스크립트로부터 제스처 애니메이션을 제작하는 응용분야를 대상으로 하므로 실시간성을 보장하지 않는다. 본 논문에서는 이전 연구에서 다루하였다[10]. Fagerberg 등은 LMA를 이용하여 해당 동작이 어떤 감정을 반영하고 있는지를 분석하여 모바일 기기의 사용자 인터페이스에 적용하였다[11]. 그러나 LMA를 이용하여 캐릭터 제스처의 감정을 표현하고 제스처를 스타일화하려는 연구는 진행되지 않고 있다.

모바일 기기에서의 캐릭터 애니메이션 서비스는 대부분 엔터테인먼트 요소로서 제공되고 있다[12-16]. Emura 등은 모바일 이메일 서비스용으로 사용자가 입력한 텍스트와 이모티콘을 캐릭터가 등장하는 애니메이션으로 재구지 않았던 감정 요소 기반의 제스처 스타일화를 시도하며 모바일 단말기에서도 수행 가능한 실시간 알고리즘을 제시한다.

몇몇 연구들은 LMA(Laban Movement Analysis)를 도입하여 제스처에 스타일을 부여하고자 하였다. Chi 등은 LMA의 구성요소인 질적 요소(effort)와 형태적 요소(shape)를 이용하여 사람의 동작을 파라미터 기반으로 제어할 수 있도록 성하는 방법을 제안하였다[12]. 미리 저장된 동작 템플릿을 단지 재생(replay)하는 방법을 사용하고 있으며 이메일 내용 외에 휴대폰에 기록되어 있는 사용자 위치, 수신자 정보 등을 활용하여 배경 세트 등을 포함하는 장면 애니메이션을 생성하는 서비스에 초점을 맞추고 있다. Persson 등은 사용자가 주어진 템플릿들을 합성하는 등의 작업을 통해 아바타의 표정을 편집할 수 있는 SMS(Short Messaging Service) 서비스를 제안하였다[13]. 대부분의 연구들이 표정 생성에 주목하고 있거나 제스처 생성의 경우 미리 제작한 템플릿을 그대로 재생하는 방식을 사용하고 있다. 이와 같

은 템플릿 방식은 템플릿의 종류와 수에 종속적이며 감정을 표현하기 위해서 대량의 템플릿이 필요하기 때문에 휴대폰 용량에 문제가 발생할 수 있다.

3. 시스템 개요

이 시스템은 그림 1과 같이 SMS 텍스트를 캐릭터 표정과 제스처 애니메이션으로 자동 변환시켜 모바일 단말기에서 아바타 애니메이션을 실시간으로 재생한다. SMS 텍스트가 입력되면 이로부터 감정을 추출하여 제스처 애니메이션 부분과 표정 애니메이션 부분으로 전달된다.

제스처 부분은 제스처 모델링, 제스처 애니메이션, 그리고 감정과 제스처 합성의 3단계로 구성된다. 제스처 모델링은 전처리 과정으로서 디즈니 영상을 분석하여 주인공의 제스처 특성을 모델링하는 단계이다. 제스처 애니메이션은 새로운 텍스트를 실시간으로 입력받아서 제스처 모델링 방식대로 음성에 동기화시켜 제스처를 생성한다. 감정과 제스처 합성 부분에서는 LMA(Laban Movement Analysis)를 도입하여 감정과 제스처를 결합함으로써 감정이 과장되게 표현되는 제스처 애니메이션을 생성한다.

표정 애니메이션 부분에서는 TTS(Text To Speech)를 이용하여 입력 텍스트를 음성 신호로 변환하고 사람이 말하는 표정 애니메이션을 생성한다.

4. 감정 추출

감정 추출기는 입력 텍스트로부터 기본적인 7가지 감정(무감정, 행복, 슬픔, 삶음, 놀람, 화남, 두려움)의 종류를 추출한다. 사용자가 모바일 기기에 SMS 텍스트를 입력하면 텍스트를 단어 단위로 구분

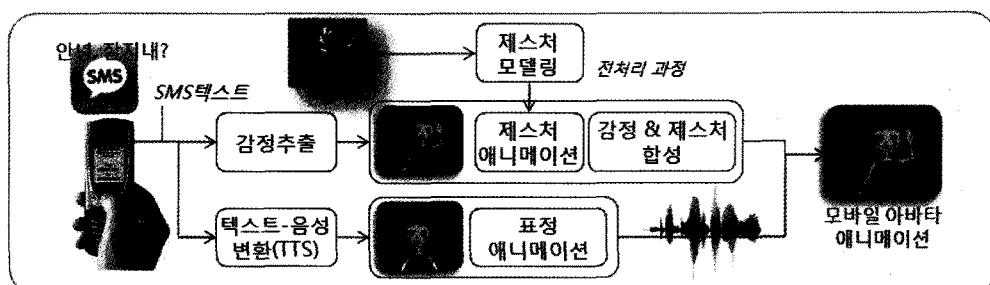


그림 1. 시스템 개요도

하고 각 단어가 어떤 감정을 나타내는지 분석하기 위해서 미리 구축해 놓은 감정 키워드 데이터베이스를 검색한다. 각 단어가 표현하는 감정을 먼저 찾고 난 이후에 전체 문장의 감정을 최종적으로 결정한다.

어떤 상황에서는 한 개의 SMS 텍스트 내에 특정 감정이 드러나는 단어도 존재하지 않고 이모티콘도 사용되지 않을 수 있다. 이러한 경우에는 한 개의 SMS 텍스트에서 특정 감정이 추출되지 않더라도 시간상의 이전 문자와 이후 문자의 감정 히스토리를 사용하여 현재의 감정 상태를 결정한다. 이는 SMS를 사용하는 시나리오를 살펴보면 두 사람이 대화를 주고받는 형식이므로 일정한 감정의 흐름이 보인다는 점에 착안한 것이다. 본 시스템에서는 Java 오픈 소스 'Synesketch' 기반으로 알고리즘을 구현하였다.

5. 제스처 모델링

제스처 모델링 단계는 전처리 과정으로서 감정이다소 과장된 디즈니 비디오 분석을 통해 문장 내 어떤 단어에서 어떤 제스처가 생성되는지를 모델링하고 그 결과를 제스처 프로파일에 저장한다. 제스처 프로파일은 각 단어에 대응되는 제스처를 기록한 것으로서 제스처를 수행하는 시간 정보와 제스처에 연동되는 머리 움직임 정보가 함께 명시되어 있다. 그리고 제스처 프로파일을 이용하여 제스처가 등장하는 빈도수를 기준으로 제스처를 통계적으로 모델링하는데 이때 제스처 그래프를 사용한다.

5.1. 비디오 주석 처리

제스처를 모델링하기 위해서 어떤 상황에서 특정 단어에 어떤 제스처가 수반되는지를 관찰할 필요가

있다. SMS 내용에 적합한 제스처 스타일을 생성하기 위해서 감정 표현이 다소 과장되어 있는 디즈니 '인크레더블' 비디오에서 특정 단어에 동반되는 제스처의 종류를 관찰한다. 비디오 내용에서 각 캐릭터들은 자신의 의견을 말하거나 감정을 표현할 때 의미를 더욱 효과적으로 전달하기 위하여 제스처를 사용한다. 주로 특정 단어의 의미 전달에 필요한 제스처를 사용하기도 하고 OK 사인과 같이 문화적 또는 사회적 학습으로 익숙해진 특정 제스처를 습관적으로 사용하기도 한다.

비디오 주석 처리 과정에서는 사용자가 애니메이션을 직접 관찰하며 그림 2와 같이 수작업으로 표시하고 <단어, 제스처> 정보를 수집한다. 각 단어에 대응되는 제스처 종류와 제스처가 수행되는 시간을 기록한다. 예를 들어, 비디오에서 주인공이 'You should do funny thing'이라는 문장을 말할 때, 먼저 문장을 단어 단위로 분할하고 각 단어가 시작되는 시간을 기록하며 'you should do'에서 사람을 지칭하는 제스처를 동반하고, 'funny thing'에서 어깨를 올리는 제스처를 동반한다는 정보를 기록한다. 보다 다양한 종류의 제스처를 수집하기 위해서 구글 비디오를 활용한다. 구글 비디오 검색을 통해서 디즈니 비디오에 상대적으로 등장 빈도가 낮은 OK, Money, Peace, Thumbs up, Handshake 등과 같은 엠블럼 제스처 21개를 추가하여 총 78개의 제스처를 주석 처리한다.

5.2 제스처 프로파일 생성

비디오 주석처리를 통해 얻은 <단어, 제스처> 정보를 이용하여 제스처 프로파일을 생성한다. 제스처 프로파일은 의미론상 분류, 단어, 시간, 제스처, 그리

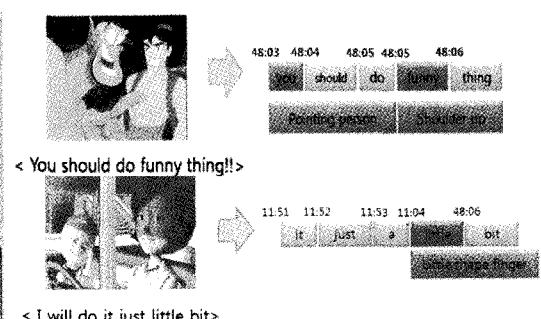
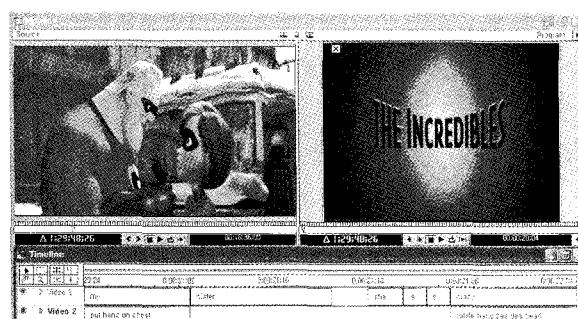


그림 2. 주석처리 과정

고 머리 움직임 및 특징 정보를 표기해 놓은 것으로서 제스처 애니메이션 부분으로 전달된다. 머리 움직임 및 특징의 경우, 손 제스처와 함께 연동되어 자연스러운 제스처 애니메이션을 생성하는데 사용된다. 표 1은 제스처 프로파일의 내용을 보여준다.

표 1에서 ‘의미론상 분류’는 유사한 의미를 가지는 단어들을 통합한 것으로서 제스처 사용 범위를 확대하고 다양성을 증가시키는 기능을 담당한다. 예를 들어 동일한 의미론상 분류에 속하는 ‘yes’, ‘ok’, ‘sure’에 해당하는 ‘Head nodding’ 제스처나 ‘Shoulder down’ 제스처 들은 같은 의미를 상대방에게 전달하기 때문에 어떤 제스처가 사용되어도 무관하다. 이와 같이 의미가 유사한 단어들을 의미론상으로 통합하는 것을 통해 한정된 개수의 제스처로 보다 다양한 단어를 표현함으로써 효용성을 높인다.

5.3 제스처 그래프 생성

제스처 프로파일을 보면 하나의 단어에 대해 여러 가지 제스처가 대응되어 있는 것을 알 수 있다. 예를

들면 단어 ‘yes’에 대해 ‘shoulder down’과 ‘fighting’ 2개의 제스처가 대응되어 있어서 어떤 제스처를 선택할지를 결정할 필요가 있다. 본 논문에서 제안하는 방법은 비디오 분석을 통한 통계적 모델링 기법이다. 예를 들어, 주인공이 ‘yes’라는 단어를 말할 때, ‘shoulder down’ 제스처와 ‘fighting’ 제스처가 등장하는 빈도수를 기준으로 통계적으로 판단한다.

이러한 목적으로 제스처 그래프와 감정 흐름 그래프를 정의한다. 제스처 그래프는 문장에 등장하는 각 단어를 하나의 노드로 표현하고 그 단어를 말할 때 동반하는 제스처를 링크로 연결한다. 그림 3은 제스처 그래프를 보여주고 있다. 감정 흐름 그래프는 감정의 흐름이 정해진 규칙에 의해 전이된다는 이론[8]을 근거로 하여 만들어진 그래프로서 각 감정 간에 전이되는 확률을 나타내고 있다. 감정 흐름 그래프의 각 노드는 행복, 슬픔, 놀람 등의 감정 상태를 나타내고 각 링크는 감정들 간에 전이될 수 있는 확률을 명시한다. 감정 흐름 그래프를 생성하기 위해서 감정 추출기를 이용하여 디즈니 비디오의 대사에서 감정

표 1. 제스처 프로파일

의미론상 분류	단 어	시 간	제 스 쳐	머리 움직임 및 특징
Agreement	‘yes’	00.02.00 00.01.60	Shoulder down Fighting	Head nodding (neutral)
	‘ok’	00.01.30	Ok	
	‘sure’	00.02.22 00.02.80	Cup Shoulder down	Head nodding
Addressee	‘you’	00.00.98 00.01.00 00.01.00	Point with thumb Shooting Pointing you	Forward person
Distance	‘long’, ‘far’	00.02.15	Hand on eyebrows	
Relax	‘relax’ ‘come down’ ‘take it easy’	00.02.00	Both hands move to down	
Negation	‘don’t’ not’ ‘no’	00.01.20	Shake a finger	Shake head
Destruct	‘cancel’ ‘destroy’	00.00.00	Fall down	
Other	‘they’, ‘she’, ‘he’	00.00.99	Pointing out there	
Self	‘I’, ‘my’, ‘me’	00.01.00	Put a hand on the chest	

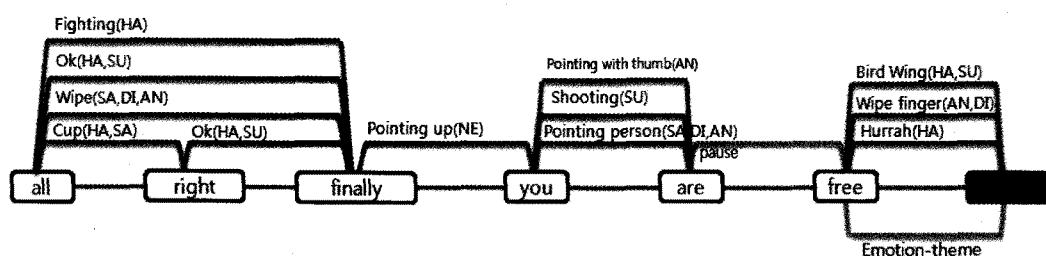


그림 3. 제스처 그래프

을 추출하고 감정이 서로 전이되는 경우의 수를 기준으로 확률을 계산하여 그림 4와 같이 그래프의 해당 링크에 표기한다.

제스처 그래프는 하나의 단어에 대응하는 여러 개의 제스처를 포함하고 있는 경우가 많은데, 이를 제스처 중에서 하나를 선택하기 위해서 감정 흐름 그래프를 사용한다. 문장 내 감정 흐름에 따라 감정 전이 그래프 상에서 감정 전이 확률이 가장 높게 나타나는 제스처를 우선적으로 선택한다.

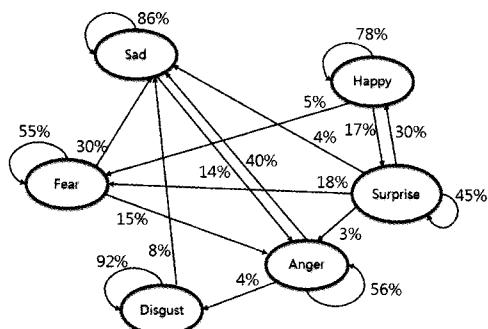


그림 4. 감정 흐름 그래프

6. 제스처 애니메이션

실시간으로 입력되는 SMS 텍스트를 단어 단위로 분할하고 제스처 프로파일을 이용하여 각 단어에 해당하는 제스처를 생성하는 과정을 통해서 제스처 애니메이션을 생성한다. 문장 내 모든 단어에 대해서 제스처 프로파일에 등록되어 있는 제스처를 생성하면 되는데, 이때 인접 단어의 경우 문제가 발생할 수 있다. 예를 들어 'I love you'와 같은 경우 'I', 'love', 'you'의 세 단어 모두가 제스처 프로파일에 등록되어 있다면 모든 단어에 대해서 해당 제스처를 수행하게 된다. 그러나 'I love you'를 말하는 속도보다 제스처 총 수행 시간이 긴 경우 말과 제스처를 연동하기 어려울 뿐만 아니라 비현실적인 애니메이션 결과를 초래한다. 이러한 문제점을 해결하기 위해서 문장의 의미를 결정하는데 각 단어가 기여하는 정도를 단어 중요도 개념으로 정의하고 문장 내 단어 중요도를 결정하는 알고리즘을 설계한다. 총 4가지 경우, 질문/대답 형태의 문장, 감정을 표현하는 문장, 사회적 기호로 통용되는 문장, 그리고 기타 문장에 대해서 중요도를 결정한다.

첫 번째, 질문/대답 형태의 문장에서 같은 문장이라도 질문 내용에 따라서 문장 내 단어 중요도가 달라진다. 예를 들어 다음의 예제에서 동일한 대답 'I like math'이지만 (1)번의 경우 주어를 묻는 질문으로서 "I"가 중요하고 (2)번의 경우는 목적어를 묻는 질문으로 "like math"가 중요하다. 본 시스템에서는 6 하 원칙의 질문(What, Who, When, Why, Where, How)에 대한 대답에서 중요한 의미를 가지는 단어의 형태소를 분석하여 문장 내 단어 중요도를 결정한다. 형태소 분석을 위하여 오픈소스로 제공되는 형태소 분석기 'MorphAdorner'를 사용한다.

- (1) Q: Who likes math? A: I like math.
- (2) Q: What do you like? A: I like math.

두 번째, SMS 핵심 기능이 감정 전달이므로 감정을 표현하는 문장에서는 감정을 나타내는 단어의 문장 내 단어 중요도를 최대한 높여서 우선적으로 제스처를 생성한다. 예를 들어 'I am very tired, so I should go to sleep. T-T' 문장에서 감정을 표현하는 단어 'tired'와 이모티콘 'T-T'의 우선 순위를 'T', 'go', 'sleep' 등의 다른 단어보다 높여서 제스처가 우선적으로 생성되도록 한다.

세 번째, 사람들이 평이한 제스처보다 사회에서 학습된 자신들만의 제스처를 사용하기 좋아한다는 연구결과[17]에 따라서 학습에 의해 사용되는 제스처, 즉 'thumbs up', 'high five', 'hi wave'의 단어 중요도를 다른 단어들보다 증가시킨다. 예를 들어 'Cheer Up! You can do it'의 문장에서 사회적으로 학습된 파이팅 제스처가 'cheer up'에서 선택되어 'you'를 표현하는 제스처보다 우선적으로 제스처가 생성된다.

네 번째, 위와 같은 경우가 아닌 일반 문장에 대하여 보편적으로 적용할 수 있는 규칙을 정의하기 위해서 '인크레더블'의 대사를 기반으로 어떤 품사의 단어에서 우선적으로 제스처를 사용하는지를 관찰한다. 행동을 나타내는 동사(verb), 동사를 꾸며주는 부사(adverb), 문장의 흐름을 결정하는 접속사(conjunction), 사물이나 사람을 나타내는 주어와 목적어(subject and object)의 순서로 제스처를 사용하는 것을 발견하였다. 이와 같은 결과를 반영하여 본 논문에서는 기타 문장의 경우 형태소 분석을 통하여 문장 내 단어의 중요도를 결정한다.

이와 같은 중요도 우선순위 결정 방식을 통하여

문장의 의미를 결정하는 중요한 단어들에서 제스처가 우선적으로 생성되도록 하는 알고리즘을 그림 5와 같이 설계한다.

7. 감정과 제스처 합성

7.1 LMA(Laban Movement Analysis)

Laban은 동작을 의미없는 단순 움직임(movement)과 의미있는 제스처(gesture)로 구분할 수 있다고 주장하고 제스처를 형태적 요소(shape)와 질적 요소(effort)로 정의하고 있다. 형태적 요소(shape)는 제스처의 형태를 나타내는 지표로서 수평성(horizontal), 수직성(vertical), 전후성(sagittal)으로 구성되며 각각에 대한 설명은 다음과 같다.

- 수평성(horizontal) 신체 움직임의 수평적 요소를 나타내는 지표

Spreading: 넓게, Enclosing: 좁게

- 수직성(vertical) 신체 움직임의 수직적 요소를

나타내는 지표

Rising: 위로, Sinking: 아래로

- 전후성(sagittal) 신체 움직임의 전후 방향을 나타내는 지표

Advancing: 전진하는, Retreating: 후진하는

질적인 요소(effort)는 제스처의 성질을 나타내는 지표로서 공간(space), 무게(weight), 시간(time), 흐름(flow)으로 구성되며 각각에 대한 설명은 다음과 같다.

- 공간(space) 신체가 공간상에서 어떤 방향, 궤도로 움직이는지 나타내는 지표

Indirect: 간접적으로, 유연한, Direct: 직접적으로, 한 곳을 향하는

- 무게(weight) 신체 동작의 균육 수축 강도를 나타내는 지표

Light: 약하게, Strong: 강하게

- 시간(time) 신체 동작의 속도를 나타내는 지표

Sustained: 느리게, 여유있게, Sudden: 빠르게,

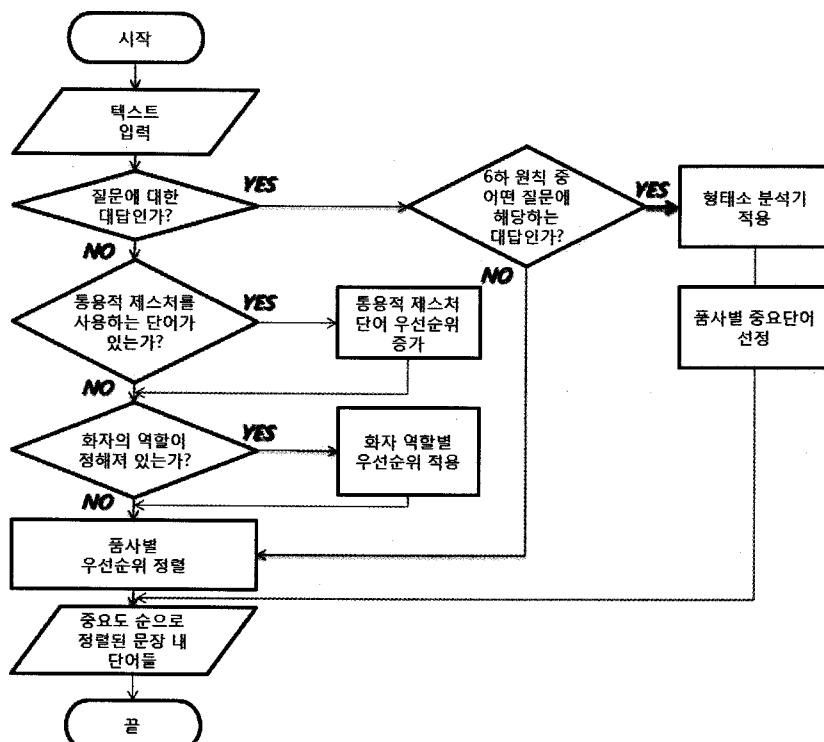


그림 5. 문장 내 단어 중요도 결정 알고리즘

긴급한

- 흐름(flow) 신체 동작의 연속성을 나타내는 지표
Free: 자유로운, Bound: 통제받는

7.2 LMA 기반 감정과 제스처 간의 관계 정의

LMA에서 기술하는 형태적, 질적 요소들을 관찰해 보면, ‘강하게’ 또는 ‘약하게’와 같은 속성들이 감정과 연관이 있음을 알 수 있다[8]. 관찰을 통해, LMA의 질적 요소와 형태적 요소들을 조합하여 행복, 슬픔, 두려움, 삶음, 놀람, 화남의 6가지 감정을 표 2와 같이 경험적으로 정의하였다. 예를 들어 ‘행복’의 경우 형태적 요소에서는 극단적으로 넓게, 위로, 전진하는 특성을 포함하고, 질적인 요소에서는 강하게, 간접적으로, 빠르게, 자유로운 특성을 포함한다.

7.3 LMA 형태적 요소에서 제스처로의 변환

감정과 제스처를 결합하기 위해서 6가지 감정을 표 2와 같이 LMA의 형태적 요소로 나타내는 1단계와 이를 제스처로 변환하는 2단계를 거친다. 1단계에서는 각 감정을 LMA의 형태적 요소로 표현하기 위해서 순서쌍을 이용한다. 예를 들어 행복을 나타내는 형태적 요소는 ‘극단적으로’, ‘넓게’, ‘위로’, ‘전진하는’의 속성을 가지므로 이를 순서쌍으로 표현하면 다음과 같다.

```

emotion = (hor, ver, sag)
happy = (1,1,1)
sad = (-1,-1,-1)
fear = (-1,-1,-1)
disgust = (-1,0.5,-1)
surprise = (0.5,0,1)
anger = (0.5,1,1)

```

감정을 나타내는 파라메터 hor, ver, sag을 이용하여 형태적 요소를 제스처로 변환하는 방법은 Chi의 수식을 다음과 같이 도입한다.

$$\begin{aligned}x' &= x + hdx + sdx \\y' &= y + hdy + vdy \\z' &= z + vdz + sdz\end{aligned}$$

여기에서 x, y, z 는 현재의 관절 데이터이고, x', y', z' 은 변형된 관절 데이터를 의미한다. (hdy, hdx)는 수평성의 변화량이고, (vdy, vdz)는 수직성의 변화량이며 (sdx, sdz)는 전후성의 변화량이다. (vdy, vdz)는 다음의 수식으로 구할 수 있다.

표 2. 감정과 LMA 질적 요소와 형태적 요소 간의 관계 정의

감정	파라메터	속성
행복	형태적 요소	극단적으로 넓게, 위로, 전진하는
	질적 요소	강하게, 간접적으로, 빠르게, 자유로운
슬픔	형태적 요소	좁게, 아래로, 후퇴하는
	질적 요소	약하게, 직접적으로, 느리게, 자유로운
두려움	형태적 요소	좁게, 아래로, 후퇴하는
	질적 요소	약하게, 직접적으로, 느리게, 자유로운
삶음	형태적 요소	좁게, 약간 위로, 후퇴하는
	질적 요소	약하게, 직접적으로, 느리게, 통제받는
놀람	형태적 요소	약간 넓게, 중립, 전진하는
	질적 요소	약하게, 직접적으로, 느리게, 통제받는
화남	형태적 요소	약간 넓게, 위로, 전진하는
	질적 요소	강하게, 간접적으로, 빠르게, 자유롭게

$$\begin{aligned}vdz &= -(a^* \cos(\text{rot}(\theta))) - y \\vdy &= (a^* 1 / \text{abratio} * \sin(\text{rot}(\theta))) - y\end{aligned}$$

$$\text{rot}(\theta) = \begin{cases} 0 & \text{ver} = 0 \\ \min(\theta - \text{ver}^* \max d\theta, 2\pi) & \text{ver} < 0, \theta \geq \pi \\ \max(\theta + \text{ver}^* \max d\theta, 0) & \text{ver} < 0, \theta < \pi \\ \max(\theta - \text{ver}^* \max d\theta, \pi) & \text{ver} > 0, \theta \geq \pi \\ \min(\theta + \text{ver}^* \max d\theta, \pi) & \text{ver} > 0, \theta < \pi \end{cases}$$

이때 a 는 팔을 움직이는 궤적을 나타내는 YZ 평면상의 타원의 장축의 길이이고, b 는 단축의 길이이며 abratio 는 장축과 단축의 길이의 비를 나타낸다. ver 는 수평성을 나타내는 파라메터이다. 이와 유사한 방법으로 (hdy, hdx)와 (sdx, sdz)을 계산한다.

7.4 LMA 질적 요소에서 제스처로의 변환

질적 요소(effort)를 제스처로 변환하기 위해서 Chi 등의 연구[7]를 확장하여 모바일에서 제어 가능하며 계산 성능을 높일 수 있는 방법을 제안한다. 첫 번째, 공간(space, 직접성/간접성)에 대해서는 손목의 운동 궤적을 이용하여 직접성(direct)인 감정(슬픔, 삶음, 놀람)에서는 손목의 움직임을 최소화하여

물건을 가리키는 동작 등으로 움직임의 정확성을 높였고, 간접성(Indirect)인 감정(행복, 두려움, 화남)에서는 손목의 기본 움직임에 상하 움직임을 적용하여 움직임의 자유도를 높였다. 두 번째, 무게(weight, 약함/강함)는 프레임 비율과 동작 간의 시간 차이의 두 가지 요소로 제어하였다. 프레임 비율(frame rate)은 동작의 세기를 표현하고 마지막 프레임에서 반복되는 동작에 한하여 동작과 동작간의 시간 차이를 조절한다. 세 번째, 시간(time, 느긋함/급함)에 대해서는 애니메이션의 프레임 비율(frame rate)을 조절하여 적용한다. 네 번째, 흐름(flow)에 대해서는 통제받음(bound)을 가지는 감정(싫음, 놀람)에 대해서는 마지막 프레임에서 제스처를 정지시키고 반대로 자유로움(free)인 나머지 감정에 대해서는 반복하여 제스처를 재생한다.

8. 실험결과

다양한 영문 SMS 텍스트를 입력으로 감정과 합성된 제스처와 표정 애니메이션을 실시간으로 재생하는 시스템을 구현하고 ‘단어 중요도’ 실험과 ‘감정과 제스처간의 합성’ 실험을 주로 수행하였으며 결과에 대해 사용자 평가를 진행하였다. 플랫폼으로써 안드로이드 모바일 단말기를 사용하였으며 시스템 구현 당시 한글 사용에 다소 불편함이 있어서 영문 SMS 부분을 먼저 구현하였다.

8.1. 캐릭터 모델

실험에 사용된 캐릭터는 그림 6에서 볼 수 있듯이 MAYA에서 제작된 OBJ 형식의 파일로서 텍스처 맵 평원 3차원 모델이다. 캐릭터 애니메이션 엔진에서 SMS 텍스트로부터 추출된 감정에 따라 그림 7과 같이 미리 제작해 놓은 행복, 슬픔, 화남 등의 표정 이미지를 캐릭터 얼굴에 텍스처 맵 평함으로써 표정 애니메이션을 생성한다. 제스처 생성을 위해서 SMS 텍스트를 분석하여 기본 제스처를 생성하고 감정에 따라 이를 변형한다.

8.2. 문장 내 단어 중요도 실험

첫 번째 실험은 단어 중요도 알고리즘을 검증하는 실험으로서 동일한 문장이지만 그 내용이 질문에 대한 대답인 경우와 감정을 포함하는 문장인 경우를

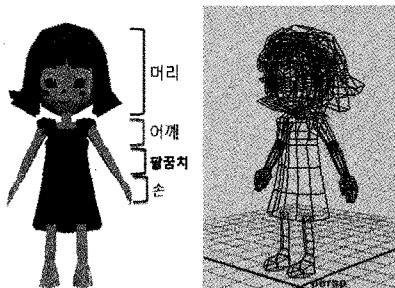


그림 6. 캐릭터 관절 및 모델링

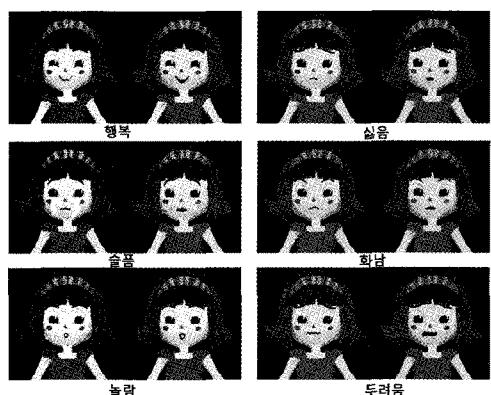


그림 7. 감정별 캐릭터 표정 텍스터

비교한다. 그림 8 (a)의 경우, 문장 ‘I like music’은 감정을 표현하는 단어의 중요도가 다른 단어보다 가장 높으므로 감정을 나타내는 ‘like’에서 해당 제스처가 우선적으로 생성된다. 두 번째 예제 (b)에서 ‘Who likes music?’이라는 질문에 대한 대답의 형태로 문장이 주어지는 경우에는 주어를 묻는 질문의 대답으로 ‘T’를 나타내는 제스처가 우선적으로 생성된다. 세 번째 예제 (c)에서 문장 ‘I like my teddy bear a lot.’의 경우에 단어 중요도를 판단하는 알고리즘이 감정을 표현하는 단어의 중요도가 가장 높다는 결정을 하므로 ‘like’에 해당하는 제스처가 우선적으로 생성된다. 네 번째 예제 (c)에서 문장 ‘How much do you like your teddy bear?’에서 ‘how’라는 의문사에 대한 대답으로 단어 ‘a lot’이 가장 높은 중요도를 가지고므로 ‘a lot’에 해당하는 제스처가 우선적으로 생성된다.

그림 9은 제스처 프로파일에 등록되어 있는 단어들이 다수 존재하고 특정 감정을 반영하지 않는 단어와 감정을 표현하는 단어들이 혼재해 있는 문장을 입력하였을 때 생성된 제스처 애니메이션을 보여준다. 감정을 반영하는 단어가 단어 중요도가 높아서

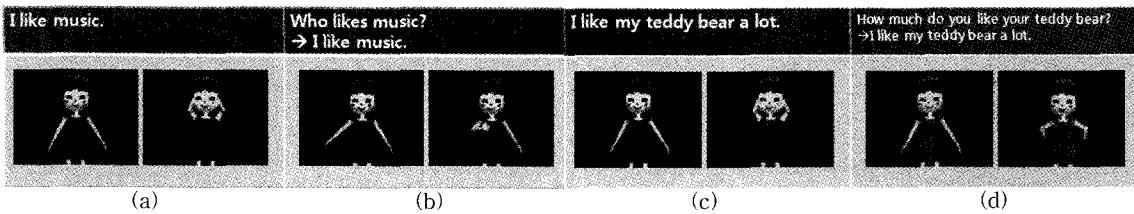


그림 8. 질문에 대한 대답일 경우와 아닐 경우 제스처 결과 비교

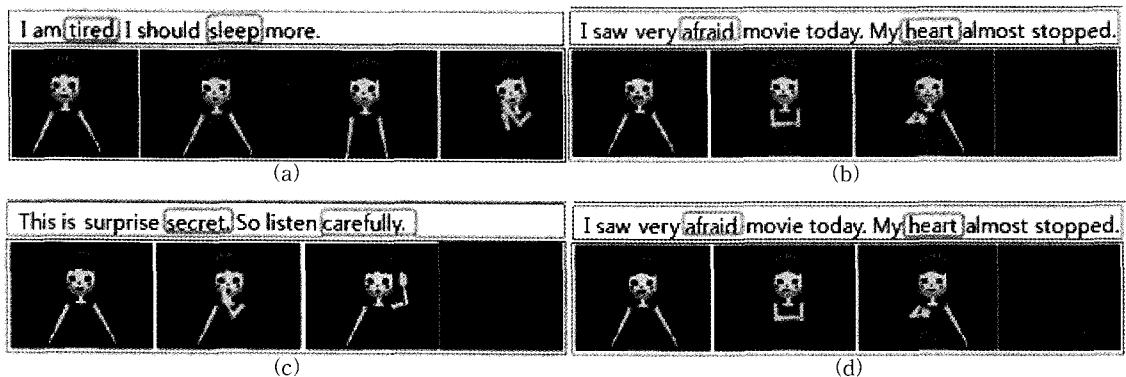


그림 9. 문장별 생성된 제스처 애니메이션

대부분 제스처로 생성되는 것을 확인할 수 있다. 예를 들어, 그림 9의 (a)에서 문장 'I am tired. I should sleep more.'의 경우에 슬픔을 반영하는 'tired'와 'sleep' 단어에서 제스처가 생성되는 것을 보여준다.

그림 10, 11은 SMS 텍스트 문장이 입력될 때 단어 중요도 결정 알고리즘으로 생성한 다양한 제스처 애니메이션 결과를 보여준다.

8.3. 감정 파라미터 기반 제스처 애니메이션 생성 실험

다음 실험은 SMS 텍스트로부터 생성한 제스처 애니메이션과 감정을 합성한 결과를 보여 준다. 행

복, 슬픔, 놀람, 두려움의 감정을 표현하기 위해서 도입한 LMA의 형태적 요소와 질적 요소를 적용한 결과이다. 그림 12의 두 번째 행에서 볼 수 있듯이, 동일한 제스처 'Shake hand horizontal way'에 대해서도 감정에 따라서 제스처의 느낌이나 분위기가 다르게 표현되는 것을 확인할 수 있다.

8.4 사용자 설문조사 및 결과

본 논문에서 제안하는 시스템에 대한 사용자 선호도, 기존 서비스와의 비교 분석, 감정과 제스처 합성을 위한 LMA의 형태적 요소와 질적 요소 파라미터

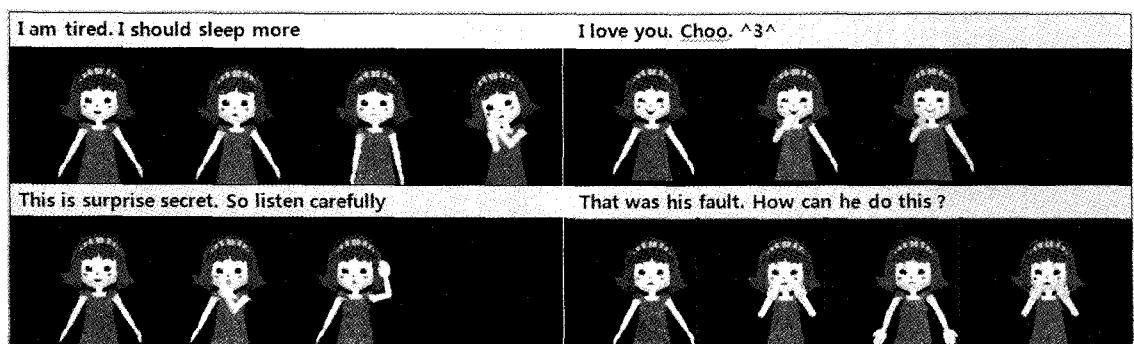


그림 10. 단어중요도를 적용한 제스처 애니메이션 결과

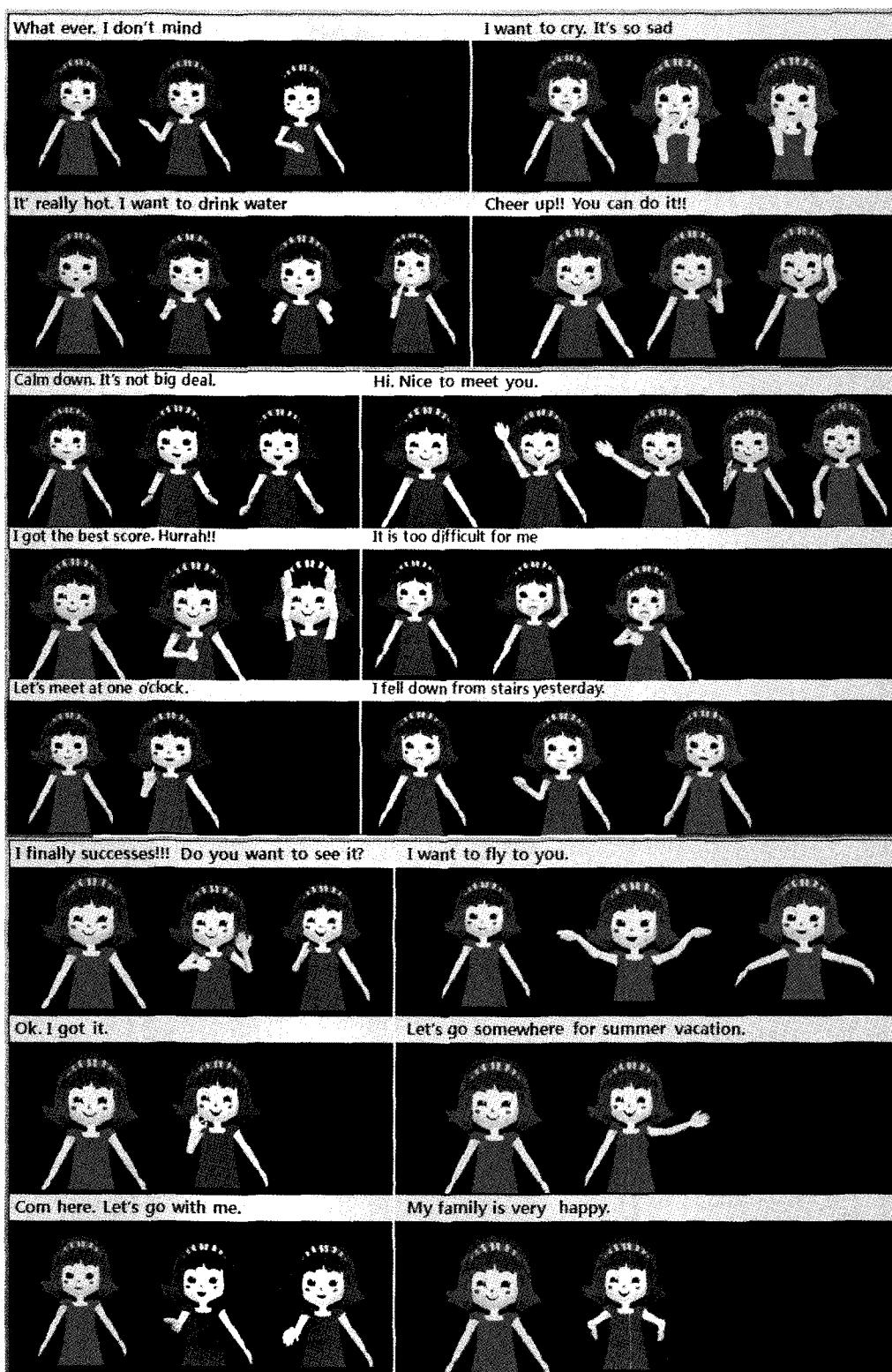


그림 11. 단어중요도를 적용한 제스처 애니메이션 결과

	행복	슬픔	놀람	두려움		행복	슬픔	놀람	두려움
OK					There with a hand				
Shake hand horizontal way					Sleep				
					You				
Fall down					Point with hand				
					Round shape				
Put down both hands					Clam down				
					come				
Hand -shake					Fly				
One on the chest the other cup									

그림 12. 감정과 제스처 애니메이션 합성 결과

의 타당성, SMS 내용 및 감정 전달의 정확성 등을 알아보기 위해서 다음의 3가지 실험을 진행하였다.

첫 번째 실험에서는 LMA의 형태적 요소와 질적 요소 파라미터의 타당성에 대하여 조사한다. 제스처와 행복, 슬픔, 두려움, 삶음, 놀람, 화남의 6가지 감정을 각각 합성한 제스처 애니메이션 결과에 임의로 번호를 매긴 후, 총 50명의 피실험자에게 보여주고 제스처가 표현하는 감정을 유추하도록 하였다. 그 결과 그림 13과 같이 전반적으로 비교적 높은 적중률을 보였다.

두 번째 실험에서는 이모티콘과 텍스트만으로 이루어진 기존의 SMS 서비스와 본 논문에서 제안하는 감정 기반 제스처 애니메이션을 결합한 SMS 서비스를 비교한다. 50명의 사용자에게 총 30번의 시스템 사용기회를 준 후, 만족도 비교를 위한 설문조사를 수행하였다. 그림 14의 결과에서 알 수 있듯이, 기존 서비스와 비교했을 때 본 논문에서 제안하는 서비스의 가장 큰 장점을 '재미있다'라고 대답한 사용자가 29%로 가장 많았다. 그리고 대다수의 사용자가 '내용전달에 효과적이다', '내용에 공감이 잘 간다', '감

정 전달에 효과적이다'라고 대답하였다. 그러나 기존 서비스보다 이용하기 번거롭다고 응답한 사용자가 8%로써 모바일에서 애니메이션이 재생되기까지의 자연시간과 음성 데이터를 이용해야 한다는 점을 번거롭다고 느끼는 사용자도 있었다.

9. 결론 및 향후 연구

본 논문에서는 SMS 텍스트 내용을 캐릭터 표정과 제스처 애니메이션으로 수작업 없이 실시간으로 자동 변환하고 감정 제어가 가능한 모바일 SMS용 캐릭터 애니메이션 시스템을 제안하였다. 다소 제스처가 과장되어 있는 디즈니 비디오로부터 제스처를 통계적으로 모델링함으로써 결과 제스처의 인위적인 측면을 어느 정도 제거할 수 있었다. 또한, 연속적으로 등장하는 단어 중에서 어느 단어에서 제스처를 생성해야 하는지에 관한 문제에서 '단어 중요도' 개념을 도입하여 결과적으로 타당한 제스처를 생성하였다. 질문과 대답 형태의 문장, 감정을 반영하는 문장, 앰블럼을 나타내는 문장, 그리고 품사를 단어 중요도의 기준으로 선정함으로써 다양한 범위의 문장에 대해서 최적의 제스처를 찾을 수 있었다.

제스처의 감정 제어 부분에서 LMA를 도입하여 질적 요소와 형태적 요소를 파라미터로 설정하고 감정과 제스처 간의 관계를 정의함으로써 동일한 문장이라도 감정에 따라 다르게 나타나는 제스처를 표현할 수 있음을 보였다. 이에 따라 캐릭터 애니메이션으로 표현한 SMS를 받아보는 수신자가 발신자의 감정 상태를 비교적 정확하게 유추할 수 있다는 것을 실험 결과를 통해 입증하고 있다. Text-To-Speech, 감정 추출, 형태소 분석기를 도입하여 각각 구현하였으며 그 결과 생성된 제스처 애니메이션은 모바일 캐릭터 애니메이션 엔진을 통해서 스피치 애니메이션과 합성 및 혼합 과정을 거쳐 최종 애니메이션을 생성하고 모바일 단말기 상에서 실시간으로 재생된다. 본 시스템은 사용자가 전달하고자 하는 감정을 보다 정확하게 표현하고 현실감 있는 대화 애니메이션을 제공하여 흥미도와 만족도를 높이는데 주력하고 있다.

제안된 시스템의 타당성과 기존 서비스와의 대체적 정도를 알아보기 위해서 사용자 반응을 조사 분석하였다. 사용자 선호도, 기존 서비스와의 비교 분석,

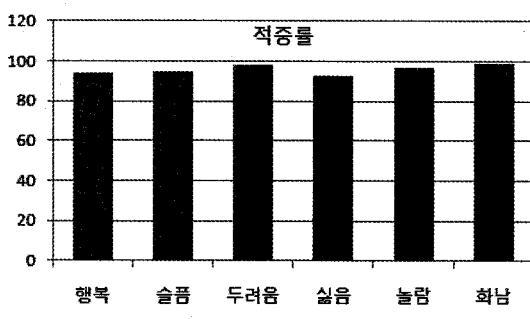


그림 13. 사용자 감정 유추 적중률

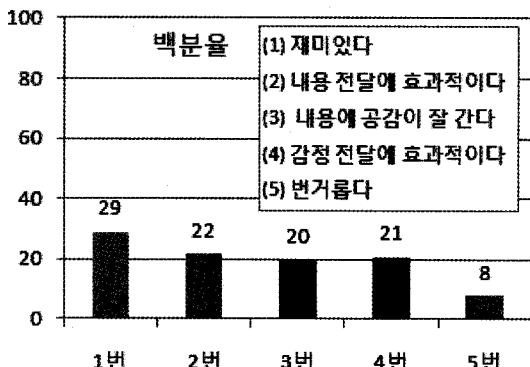


그림 14. 기존 서비스와의 비교 설문조사

SMS 내용 전달의 정확성 등을 조사한 결과 사람들은 이 서비스를 대체적으로 흥미롭게 생각하고 기존 SMS 서비스와 비교하여 자신의 의도와 감정을 보다 효과적으로 전달하며 이 서비스가 상용화될 경우 가입하겠다는 사람들은 79%로 나타났다.

향후에는 한글 SMS 텍스트도 사용할 수 있도록 지원하며 본 논문에서 제안하는 감정 요소 기반 제스처 스타일화 방법을 확장하여 발신자의 개성과 특성을 파라메터함으로써 제스처 스타일화를 시도하고자 한다.

참 고 문 헌

- [1] Cassel, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M., "Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents," *ACM SIGGRAPH Proceedings*, 413-420, 1994.
- [2] Noma, T., Zhao, L., and Badler, N., "Design of a Virtual Human Presenter," *IEEE Computer Graphics and Applications*, Vol.4, pp. 79-85, 2000.
- [3] Cassell H., Vilhjalmsson T., and Bickmore, "BEAT: the Behavior Expression Animation Toolkit," *ACM SIGGRAPH Proceedings*, 2001.
- [4] Sergey Levine, Christian Theobalt, and Vladlen Koltun, "Real-time Prosody-driven Synthesis of Body Language," *ACM SIGGRAPH ASIA Proceedings*, 2009.
- [5] Catherine Pelachaud, "Studies on Gesture Expressivity for a Virtual Agent," *Speech Communication*, Vol.51, Issue7, pp.630-639, 2009.
- [6] Juan Abadia, Alun Evans, Eduard Gonzales, Sergi Gonzales, Daniel Soto, Santi Fort, Marco Romeo, and Josep Blat, "Assisted animated production creation and programme generation," *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, pp. 207-214, 2009.
- [7] Stone, M., Decarlo, D., Oh, I., Rodriguez, C., Stere, A., Lees, A., and Bregler, C., "Speaking with Hands: Creating Animated Conversational Characters from Recordings of Human Performance," *ACM SIGGRAPH proceedings*, pp.506-513, 2004.
- [8] Hartmann, B., Mancini, M., and Pelachaud, C., "Implementing Expressive Gesture Synthesis for Embodied Conversational Agents," *Lecture Notes Artificial Intelligence*, Vol.3881, pp. 45-55, 2006.
- [9] Michael N., Michael K., Irene A., and Hans-peter S., "Gesture Modeling and Animation Based on a Probabilistic Re-Creation of Speaker Style," *ACM Transactions on Graphics*, Vol.27, 2008.
- [10] Chi D., Costa M., Zhao L., and Badler N., "The EMOTE Model for Effort and Shape," *ACM SIGGRAPH Proceedings*, pp.173-182, 2000.
- [11] Fagerberg P., Stahl A., and Hook K., "Designing Gestures for Affective Input : An Analysis of Shape, Effort and Valence," *ACM 1-58113- 826-1/01/12*, 2003.
- [12] Emura K., Yasugi M., Tanaka T., Miyazaki S., and Motoike S., "Personal Media Producer: A System for Creating 3D CG Animation from Mobile Phone E-mail," *Journal of Multimedia*, Vol.1, No.2, 2006.
- [13] Per Persson, "ExMS: an Animated and Avatar-based Messaging System for Expressive peer Communication," *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, pp.31-39, 2003.
- [14] 최승혁, 김재경, 임순범, 최윤철, "모바일 환경에서 스크립트를 이용한 3차원 아바타 동작 제어," *한국정보과학회 한국컴퓨터종합학술대회 논문집(B)*, pp.574-576, 2005.
- [15] 윤재홍, 송용규, 김은석, 허기택, "SMS 메시지에 따른 모바일 3D 아바타의 입술 모양과 음성

- 동기화,” 한국콘텐츠학회 추계종합학술대회 논문집, 제4권 제2호, pp.682-686, 2006.
- [16] 이정윤, 이경진, 송왕철, “휴대폰에서 Avata컨텐츠 기반의 E-mail Client / Server 모델 설계 및 구현,” 한국콘텐츠학회 춘계종합학술대회 논문집, 제2권 제1호, pp.385-391, 2004.
- [17] McNeil, D., “Hand and Mind: What Gestures Reveal about Thought.” University of Chicago, 1992.



변 혜 원

1990년 연세대학교 전산과학과
이학사
1992년 KAIST 전산학과 공학 석사
2004년 KAIST 전산학과 공학 박사
2004년 KBS 방송기술연구소 선임연구원
2006년 ~현재 성신여자대학교 IT 학부 교수.

관심분야: Computer Graphics, Virtual Reality, HCI, Game, User Perception



이 정 숙

2007년 성신여자대학교 미디어정보학부 학사
2009년 성신여자대학교 미디어정보학부 석사
관심분야: Computer Graphics, Virtual Reality, Game, User Interface