
한글 문자 입력에 따른 얼굴 애니메이션

김태은* · 박유신**

Facial Animation Generation by Korean Text Input

Tae-eun Kim* · You-shin Park**

요약

본 논문에서는 얼굴 애니메이션을 좀 더 실질적으로 표현하기 위하여 한글 생성의 원리와 입 모양 형태의 유사성을 기반으로 기본 음소를 선정하고 이를 이용하는 새로운 알고리즘을 제안한다. 카메라를 통해 얻어진 특징 점 이용이 아닌, 모션 캡처 (Motion Capture) 장비를 사용하여 실제 입모양의 움직임 데이터를 취득하여 취득된 데이터를 지수 증,감 형태로 나타내어 발성에 대한 음소를 표현 하고 연속된 음절을 표현하기 위하여 지배 함수(Dominance Function)와 혼합 함수(Blending Function)를 적용하여 동시 조음에 대한 표현을 해결하였다. 또한 음절 간의 결합 시간을 입 주변의 마커(Marker) 거리 값의 변위를 이용하여 그 기울기 값에 따라 시간 지연을 함으로 현실감 있는 사용자 입력 문자에 대한 입 모양 궤적 데이터를 생성하는 실험 결과를 보여준다.

ABSTRACT

In this paper, we propose a new method which generates the trajectory of the mouth shape for the characters by the user inputs. It is based on the character at a basis syllable and can be suitable to the mouth shape generation. In this paper, we understand the principle of the Korean language creation and find the similarity for the form of the mouth shape and select it as a basic syllable. We also consider the articulation of this phoneme for it and create a new mouth shape trajectory and apply it at face of an 3D avatar.

키워드

Trajectory, Mouth Shape, Face Animation, 3D avatar

1. 서론

최근 컴퓨터 산업의 발전과 함께 컴퓨터 그래픽 하드웨어와 소프트웨어가 크게 발전하였고, 사실과 같이 느낄 수 있도록 3차원 그래픽 렌더링(Rendering) 기술이 부각되고 있다. 이를 기반으로 얼굴 애니메이션은 인간과 컴퓨터의 상호 작용을 할 수 있는 하나의 가

상 현실로 발전하고 있다[1][2][3][4][13].

초기의 얼굴관련 연구는 얼굴 인식과 얼굴 모델을 변형하는 연구로 수행되었다. 얼굴인식은 사람의 얼굴 영역을 알맞게 추출하여 눈, 코, 입 부분의 명암 및 컬러 특징을 이용하여 얼굴 영역을 추출하였다. 그러나 이것은 배경, 조명 상태, 얼굴과 카메라의 거리, 얼굴의 상하좌우 각도, 카메라 시야 내에서의 얼굴의 위

* 남서울대학교 멀티미디어학과
접수일자 : 2009. 04. 01

**중앙대학교 첨단영상학과
심사완료일자 : 2009. 04. 29

치 등을 고려해 시스템의 가정에 부합되도록 조정해야 하는 문제점을 갖고 있다[6]. 얼굴 모델을 변형하는 초기 연구에는 단순히 얼굴의 변형하기 위해 워핑(warping)이나 모핑(morphing)기술을 사용하였다[7]. 이후 인간의 말하는 형태를 언어 학자들과 같이 언어의 형태들을 파악하는 연구가 시작 되었다. 즉, 말하는 경우의 입 모양 형태는 입이 닫히거나 열려 이는 형태를 취하였고 표정에 있어서는 화가 난 경우 눈과 눈썹이 위로 올라가고 슬픔을 표현할 경우는 눈과 눈썹 끝이 아래로 내려가는 형태를 취하였다.Parke로 사 진으로부터 정지된 모양을 찾아 모델을 만들고 얼굴의 특징을 찾아 몇 개의 제어 파라미터(control parameter)를 선정하여 파라미터를 이용해 얼굴을 변형하여 애니메이션을 구현하였다[8][9]. 그 후 Hill과 Wyvill 등은 Parke의 모델을 변형하여 두 이접한 음소들 간의 파라미터 값들을 비선형으로 보간(interpolation)하여 입술 움직임 시스템을 개발 하였으나 동시 조음 현상을 적용하지 않아 자연스럽게 못함을 보였다[10]. 1988년에는 Nahas가 각 음소에 대한 얼굴 형태를 생성하기 위해 B-스플라인 제어 파라미터를 취득 하여 사용하였지만 이 얼굴 영상은 연속된 음소이므로 원하는 음소를 취득하기가 어려웠다[11]. 이후 1993년에 Cohen과 Massoro는 스피치 생성 제스처 이론을 이용하여 동시조음 현상을 처리하였다. 그들은 발음하는 동안의 지배적인 움직임은 지배함수를 소개하고 주어진 스피치 세그먼트에서 지배함수들을 중복시켜 동시 조음을 다루었다[12][13].

따라서 본 논문에서 제안한 입력 문자에 대한 입 모양 궤적 생성 방법은 기존의 방법과는 다른 한글 발성 시 유사한 입 모양으로 분류한 음소를 모션 캡처 장비를 이용하여 3차원 데이터를 취하여 3차원 데이터 정보를 알 수 있으므로 Cohen과 Massaro가 응용한 스피치 생성 제스처 이론이나 기존의 동시 조음 모델보다는 정확한 데이터를 이용하여 사용될 수 있다는 장점을 갖고 있다.

II. 시스템 순서도

본 논문의 구성은 크게 두 부분으로 나눌 수 있다. 전 단계는 새로운 입 모양을 생성하기 위한 사전 작

업으로 한글의 기본 음소를 취득하고 이를 분류하는 단계이고, 후 단계로는 사용자를 통해 입력된 한글을 기본 음소들 간의 조합을 통해 이를 캐릭터에 접목하기 위해 입 모양 궤적을 생성하는 것이다. 그림 1은 시스템의 전체 순서도이고, 두 단계로 나누어 처리하는 전체적인 순서도를 나타내고 있다.

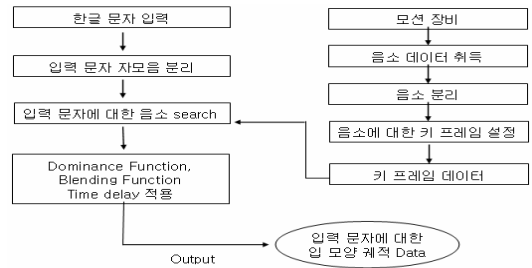


그림 1. 전체적인 순서도

처음 단계로는 모션캡처 장비를 통해 음소 데이터 취득하여 이 연속된 음소들을 각 각의 음소별로 분리하고 각 음소의 키 프레임을 설정하여 키 프레임 데이터를 저장하여 데이터 베이스화 시킨다. 이를 바탕으로 두 번째 단계인 입력 문자에 대한 기본 음소들을 찾는 수행이 가능하게 된다. 이 단계까지는 입 모양의 기본 궤적들을 알고 있기 때문에 사용자가 원하는 문자에 대한 입 모양 생성에 대한 준비가 모두 끝난 것이다. 이 후 각 음소의 입 모양에 대한 동시 조음(articulation)을 고려하기 위하여 우세함수(Dominance Function)을 사용하고 이 우세함수들을 조합하기 위해 혼합함수(Blending Function)를 사용하여 입 모양에 대한 움직임 제어 파라미터 값을 구한다[5]. 이를 통해 최종적으로 입력 문자에 대한 입 모양 궤적 Data를 생성하는 것이다.

III. 한글 음절 데이터 분석

한글은 자음과 모음의 조합으로 되어진 문자이고 ‘초성 + 중성’, ‘초성 + 중성 + 종성’과 같이 2가지 형태로 문자를 표현하고 있다. 자음은 초성과 중성 부분에 들어가고 모음은 중성 부분에 들어간다. 이렇게 하여 한글은 총 11,172자를 표현 하고 있다. 그러나 본 연구에서는 이와 같이 모든 한글을 사용하는 것이 아

나라 입 모양에 대한 한글 유사성 분류에 의한 기본 음소를 선택하여 한글을 표현 가능하였고 그 데이터 양을 줄였다.

3.1 한글의 생성 원리

먼저 한글의 생성은 기본 다섯 자의 제자원칙으로 인해 만들어 졌고, 아(牙)음, 설(舌)음, 순(脣)음, 치(齒)음, 후(喉)음으로 어금니, 혀, 입술, 이, 목구멍의 모양을 본떠서 만든 것이다. 모음의 경우는 기본 세 자의 제자 원리, 즉 천(·), 지(一), 인(丨)의 삼재를 본떠서 만들었으며 기본 삼재를 조합하여 초출자 4개와 재출자 4개를 만들어 사용하였다. 현재 사용하고 있는 한글의 자음과 모음은 다음과 같다.

표 1. 한글 자음과 모음

자음	ㄱ, ㅋ, ㆁ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㆁ, ㅌ, ㅍ, ㅎ
모음	ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ
쌍자음	ㄲ, ㄸ, ㅃ, ㅆ, ㅆ
복자음	ㄱㅅ, ㅋㅌ, ㆁㄷ, ㄹㄹ, ㄹㅌ, ㄹㅅ, ㄹㅆ, ㄹㅈ, ㄹㅊ, ㄹㅋ, ㄹㆁ, ㄹㅌ, ㄹㅍ, ㄹㅎ
복모음	ㅏㅑ, ㅑㅓ, ㅓㅕ, ㅕㅗ, ㅗㅛ, ㅛㅜ, ㅜㅠ, ㅠㅡ, ㅡㅣ

한글의 형태는 2가지로 나눌 수 있는데 초성과 중성만으로 형성되는 경우와 초성과 중성, 중성으로 형성되는 경우가 있다.

- 초성 : 기본자음+ 쌍자음 = 19
- 중성 : 기본모음+ 복모음 = 21
- 중성 : 기본자음+ 쌍자음2 개(ㄲ, ㅆ) + 복자음 = 27

이와 같이 한글 문자로 만들 수 있는 총 가지 수는 ‘초성 + 중성’이므로 $19 \times 21 = 399$ 이고 ‘초성 + 중성 + 중성’이므로 $19 \times 21 \times 27 = 10,773$ 개, 두 경우의 수를 합해서 11,172개의 문자를 생성할 수 있다. 따라서 이와 같이 많은 한글을 데이터를 입 모양 데이터로 갖기는 그 효율성면에서 떨어진다고 볼 수 있다. 그러므로 한글 발음에 대한 입 모양의 유사성을 파악하여 데이터

의 축소가 필요한 것이다.

3.2 한글 유사성 분류

한글은 자음과 모음의 조합으로 구성되어 있고 발음시 입 모양 형태는 자음 보다는 모음의 영향을 많이 받는다. 그러나 자음의 영향을 받는 것은 양 입술이 단혀서 발성되는 순음과 치아 사이에서 발성되는 치음은 발음 시 입 모양에 영향을 준다. 따라서 입 모양 표현에 있어 한글 유사 입 모양의 음소는 다음과 같이 40개로 나누었고 이를 취득 하였다.

실제 ‘가’ 와 ‘아’ 또는 ‘각’에 대해서 그 소리에 대해서는 다르지만 입 모양의 형태에 대해서는 ‘ㅏ’라는 모음에 영향을 받아 동일하다고 볼 수 있다. 따라서 한글의 유사성을 찾을 수가 있다.

표 2. 기본 음절

아	어	오	우	으	이	에	애
바(마)	버	보	부	브	비	베	배
사	서	소	수	스	시	세	새
암	엄	움	움	음	임	엠	엠
안	언	온	운	은	인	엔	앤

이와 같이 각 한글 유사 입 모양으로 음소를 분류한 것은 성대나 혀의 움직임 같은 동작은 소리에 대한 차이는 있어도 발음에 대한 입 모양은 유사하기 때문에 이와 같이 데이터를 분류하였다.

본 연구는 기존 연구[5]와 달리 TTS(text to speech)인 입력 문자에 대한 음성합성이 아닌 립싱크(Lip-Synch) 연구로 음성 부분은 다루지 않고 있다.

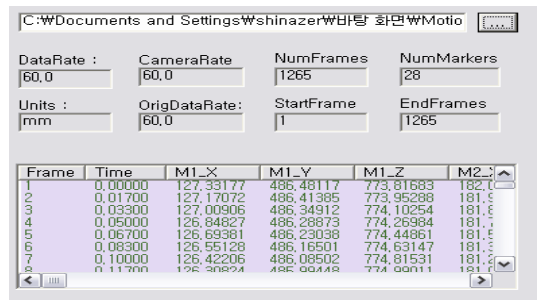


그림 2. 기본 음절에 대한 데이터

3.3 한글 조합형 코드 변환

한글은 영어와 달리 2 바이트에 해당한다. 과거 기준이 없었지만 지금은 유니 코드에 의해 한글이 표현 가능하게 되었다. 입력에 의한 코드는 완성형이다. 따라서 입력된 완성형 코드를 조합형으로 변환하여 각 초성, 중성, 종성에 해당하는 값들을 찾아 사용자가 입력한 문자에 해당하는 음소 데이터를 찾기 위한 코드 변환을 하는 것이다. 한글의 비트 표현은 다음과 같다.

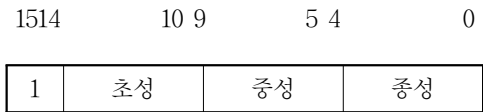


그림 3. 한글 문자에 대한 Bit 표현

한글은 최 상위 비트를 제외하고 5 비트씩 초성, 중성, 종성에 해당하는 값들을 갖고 있다. 최 상위 비트는 한글과 한자를 표시하는 비트이다.

다음은 입력된 완성형 한글을 조합형으로 분리하는 예를 보여 주고 있다.

표 3. 코드값에 대한 예 (정의된 테이블)

번호	00000000000111111111122222222
	0123456789012345678901234567
초성	ㄱ ㅋ ㆁ ㄷ ㅌ ㄹ ㄴ ㄷ ㄹ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㆁ ㄷ ㅌ ㄹ ㅅ
중성	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅝ ㅞ ㅟ ㅡ ㅣ ㅥ ㅧ ㅩ ㅫ ㅭ ㅯ ㅱ ㅳ ㅵ ㅷ ㅹ ㅻ ㅽ ㅿ ㅿ ㅿ
종성	ㄱ ㅋ ㆁ ㄷ ㅌ ㄹ ㄴ ㄷ ㄹ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㆁ ㄷ ㅌ ㄹ ㅅ

$$\begin{aligned}
 & "AC00h + (a*21+b)*28 + c" \quad (1) \\
 & c := code - \$AC00; \\
 & a := c \text{ div } (21*28); \\
 & c := c \text{ mod } (21*28); \\
 & b := c \text{ div } 28; \\
 & c := c \text{ mod } 28;
 \end{aligned}$$

글자	코드값 (hex)	계산식
가	AC00h	AC00h+(0*21+0)*28+0
각	AC01h	AC00h+(0*21+0)*28+1
개	AC1Ch	AC00h+(0*21+1)*28+0
까	AE4Ch	AC00h+(1*21+0)*28+0
럼	B7FCh	AC00h+(5*21+4)*28+16

문자 코드에서 "가"를 처음으로 놓기 위한 값으로 정의 하고 만약 \$AC00에 "가"라는 입력 문자가 들어 오면 식 (1)의 a, b, c 값이 0이 되어 테이블에 명시된 "가"에 대한 각각의 "ㄱ"과 "ㄴ"의 16진수 조합형 코드를 알 수가 있다.

3.4 음절의 우세함수와 혼합함수

입 모양 제어 파라미터들은 Lofqvist의 스피치 생성 제스처 이론을 이용하여 입술 움직임 함수들로 정의하였다[5]. 따라서 다음 식 (2)은 제스처 이론을 기반한 시간거리 τ 에 따른 exponential의 일반적 형태이다.

$$D = e^{-\theta \tau} \quad (2)$$

식 (2)의 형태를 립싱크(lip-synch.)에 사용할 수 있는 우세함수로 확장시킨 것이 다음의 식 (3)이다.

$$\begin{aligned}
 D_{sp} = \alpha_{sp} e^{-\theta_{c-sp} \tau^c}, & \text{ if } \tau \geq 0 \\
 \alpha_{sp} e^{-\theta_{c-sp} \tau^c}, & \text{ if } \tau < 0
 \end{aligned} \quad (3)$$

$$\tau = t_{csp} + t_{osp} - t \quad (4)$$

$$t_{csp} = t_{start\ s} + \frac{duration}{2} \quad (5)$$

여기서 D_{sp} 는 세그먼트 s의 p 제어 파라미터의 우세함수이고, α_{sp} 는 세그먼트 s의 p 제어 파라미터의 크기이다. 또 t_{csp} 는 세그먼트 s의 중심점 시간이고, t_{osp} 는 t_{csp} 로부터 함수가 최대가 될 때까지의 시간이다. $t_{start\ s}$ 는 세그먼트 시작 시간이며, duration은 세그먼트 지속 시간이다.

$$B_p(t) = \frac{\sum_{s=1}^N (D_{sp}(t) * T_{sp})}{\sum_{s=1}^N D_{sp}(t)} \quad (6)$$

$B_p(t)$ 는 혼합 함수이고 T_{sp} 는 세그먼트 s의 target point이다.

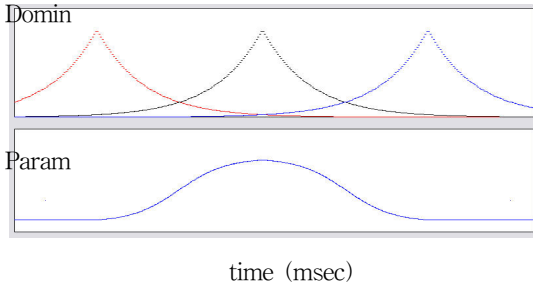


그림 4. Dominance function 와 Blending function

그림 4는 우세함수와 혼합함수의 그래프이다.

IV. 실험 결과

본 실험에서는 VC++ 6.0을 사용하였고 기본 음절 데이터를 취득하기 위해 25개의 마커를 사용하였다. 각 음절마다 키 프레임(Key frame)을 선정하여 선형 보간(linear interpolation)을 통해 우세함수에 적용하였다. 또한 프레임율은 초당 20프레임으로 하였다.

다음 실험 결과는 “안녕하세요”에 대한 문장을 실험을 통해 나타낸 것이다.

그림 5는 사용자가 “안녕하세요”라는 문자를 입력 하였을 때 초, 중, 종성으로 분리한 실험 결과이다.

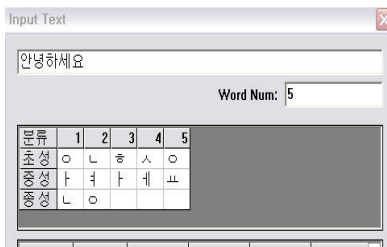


그림 5. 입력된 한글의 초성, 중성, 종성 분리

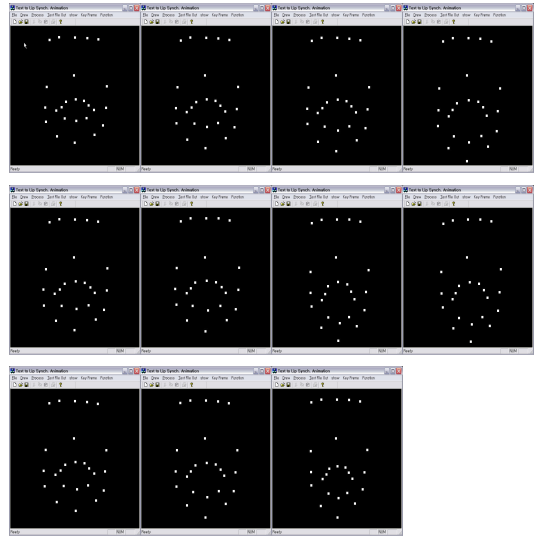


그림 6. “안녕하세요” 발음시 입 모양

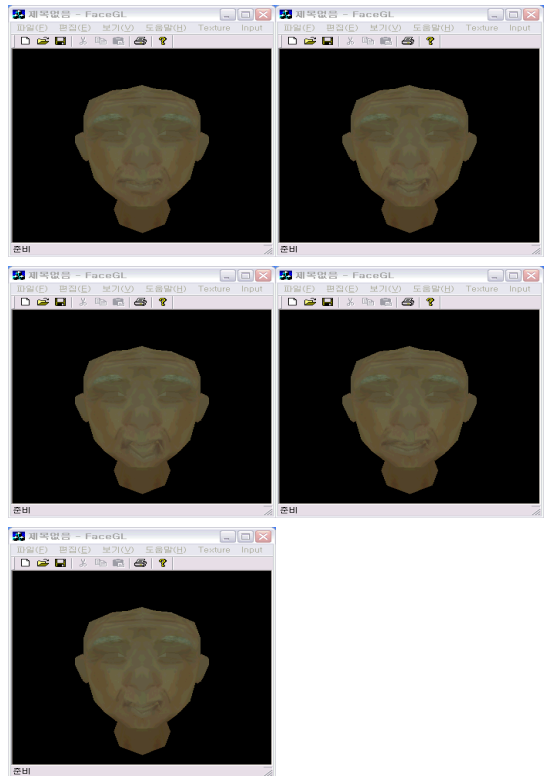


그림 7. 3D 모델에 적용한 결과

V. 결론

본 논문에서는 가상의 얼굴 모델에 적용할 수 있는 입 모양의 궤적을 음소들간의 조합을 이용하여 얼굴 애니메이션 입 모양 궤적 시스템을 제안하였다. 기존의 방법과는 다른 발성 시 유사한 입모양으로 분류한 음소를 모션 캡처 장비를 이용하여 3차원 데이터를 취하였다. 이 취득 데이터를 지수 증가, 감소의 형태로 표현을 하고 각 음소를 3개의 영역으로 분리를 하여 지배함수와 혼합함수를 이용하였으며 4개의 마커의 거리의 변위 값을 지배함수에 적용하여 정확한 입 모양의 궤적을 생성함을 실험 결과를 통해서 우수함을 확인할 수 있었다.

제안하는 입력 문자에 대한 입 모양 궤적 생성 방법은 3차원 궤적 데이터로 가상의 얼굴 모델에 합성할 수 있기 때문에 방송, 영화, 게임 등의 콘텐츠 분야에서 널리 응용될 것으로 기대 된다.

참고 문헌

- [1] R. Chellappa, C. H. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces : A Survey," Proc. of the IEEE, Vol. 83, No. 5, pp. 705-740, May 1995.
- [2] 한영환, 홍승홍, "연속 영상에서의 얼굴표정 및 제스처 인식," 의공학회지, Vol. 20, No. 4, pp.419-425, 1999.
- [3] Jun-yong Noh, Ulrich Neumann " Expression Cloning" SIGGRAPH 2001.
- [4] 이인서, 박운기, 전병우, "MPEG-4 FAP기반 얼굴근육모델을 이용한 Facial Animation", 춘천멀티미디어 학술회의, pp.147-151, Feb. 2000.
- [5] Cohen, M.M. & Massoro, D.M. , " Modeling Coarticulation in Synthetic Visual Speech." In Thalmann N.M. & Thalmann D. (Eds) Models and Techniques in Computer Animation , Tokyo : Springer-Verlag.
- [6] G. Yang and T.S. Huang, " Human face detection in a complex background," Pattern Recognition, Vol.27, No. 1, pp 55-63, 1994.
- [7] Sims, Karl. Particle Animation and Rendering Using Data Parallel Computation. Computer Graphics(SIGGRAPH'90 Proceedings), 24, pp. 405-413, 1990.

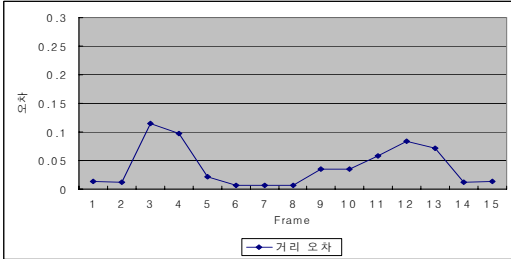
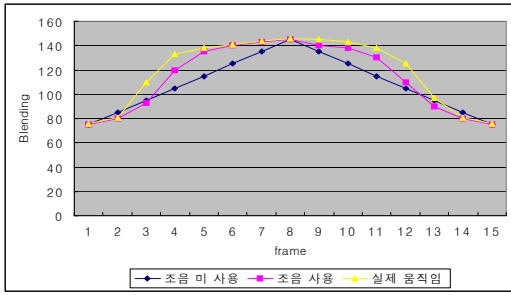


그림 8. 음절/어의 혼합함수와 거리 오차

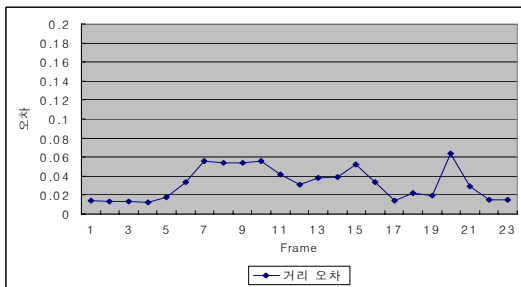
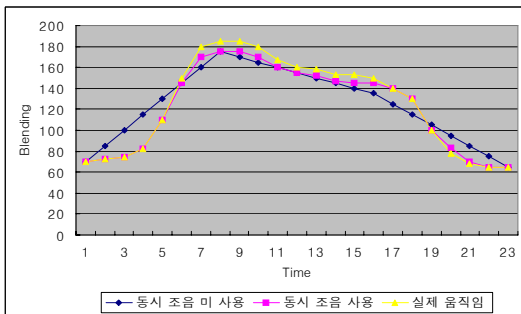


그림 9. /아이/의 혼합함수와 거리 오차

그림 8, 9는 동시조음을 적용한 방법과 적용하지 않은 방법의 혼합 함수를 비교한 것이다. 또한 키 프 로입을 중심으로 선형 보간한 경우에 대해서 거리 오차를 구하여 표현한 것이다.

[8] Parke, F. I., "Parametrized models of facial animation." IEEE Computer Graphics, 2(9), pp. 61-68, 1982.

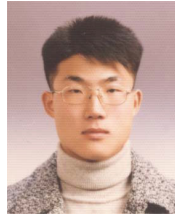
[9] Fu Jie Huang, and Tsuhan Chen. "Real-time lip-synch face animation driven by human voice". In IEEE Multimedia Signal Processing Workshop, Los Angeles, California, 1998.

[10] Hill, D.R., Pearce., & Wyvill, B.m, "Animating Speech: And automated Approach using speech synthesized by Rules, " The Visual Computer, 3: 277- 289, 1988.

[11] Nahas, M., Huitric, H., & Saintourens, M.(1988) Animation of a B-spline figure. The Visual computer, 3, 272-276.

[12] Cohen, M.M. & Massaro, D.W., " Modeling Coarticulation in Synthetic Visual Speech." In Thalmann N.M. & Thalmann D. (Eds) Models and Techniques in Computer Animation , Tokyo : Springer-Verlag, 1993.

[13] Kwang-Sik Kong and Chang-Hun Kim, "MODELING COARTICULATION IN KOREAN VISUAL SPEECH" Proceedings of the first CGIM conference, pp.207-210, . 03 Jun. 1998.



박유신(You-shin Park)

2004년 2월 : 중앙대학교 첨단영상 대학원 졸업

2009년 2월 : 중앙대학교 첨단영상 대학원 박사과정 재학중

※ 관심분야 : 3D영상처리, 웹공학, 증강현실 등

저자 소개



김태은(Tae-eun Kim)

1987년 8월 : 중앙대학교 공대 전 기공학과 졸업

1992년 2월 : 중앙대학교 대학원 전자공학과 졸업(석사)

1997년 2월 : 중앙대학교 대학원 전자공학과 졸업(박사)

1995년 삼성전자 휴먼테크논문대상은상수상

1997년 3월 ~ 현재 : 남서울대학교 멀티미디어학과 교수.

※ 관심분야 : 3D영상처리, 웹공학, 증강현실 등