

키프레임 얼굴영상을 이용한 시청각음성합성 시스템 구현

김명곤(전남대), 김진영(전남대), 백성준(전남대)

<차 례>

- | | |
|---------------------|-------------------------|
| 1. 서론 | 3.2. 우세함수 |
| 2. 시청각 DB의 구축 | 3.3. 쌍선형 내삽을 이용한 계산량 감소 |
| 2.1. 데이터 수집 | 4. 시청각음성합성시스템 구현 |
| 2.2. 녹화용 viseme set | 4.1. 음성 합성기 |
| 2.3. 특징점 추출 | 4.2. 음성 합성기와 통합 |
| 2.4. 데이터 베이스 구조 | 5. 실험결과 및 평가 |
| 3. 입술 동기화 알고리즘 | 6. 결론 및 향후 계획 |
| 3.1. RBF를 이용한 워핑 | |

<Abstract>

Implementation of Text-to-Audio Visual Speech Synthesis Using Key Frames of Face Images

MyoungGon Kim, JinYoung Kim, SeongJoon Baek

In this paper, for natural facial synthesis, lip-synch algorithm based on key-frame method using RBF(radial bases function) is presented. For lips synthesizing, we make viseme range parameters from phoneme and its duration information that come out from the text-to-speech(TTS) system. And we extract viseme information from AV DB that coincides in each phoneme. We apply dominance function to reflect coarticulation phenomenon, and apply bilinear interpolation to reduce calculation time. At the next time lip-synch is performed by playing the synthesized images obtained by interpolation between each phonemes and the speech sound of TTS.

* 주제어: Lip-Synch, Radial Basis Function, Dominance Function, Bilinear Interpolation, Viseme

1. 서론

음성은 휴먼 인터페이스 중에서 가장 많은 정보를 포함하고 사용이 편리하다는 장점 때문에 여러 분야에서 활용되고 있다[1][2]. 하지만 음성은 주변의 소음이 일정수준 이상이 되면 전달되기 어려운 특성을 가지고 있다. 따라서 음성만 전달하는 것보다 어느 정도 음성 정보를 포함한 입술 영상 정보를 함께 전달할 수 있다면 더욱 효과적일 것이다. 이러한 맥락에서 대두된 것이 입술동기화(lip synch)이다. 입술동기는 음성과 입술영상을 함께 전달함으로써 사람이 듣는 청각정보의 부족함을 보정하고, 컴퓨터를 인간과 친근한 이미지로 구현한다. 입술동기화에 관한 연구는 일찍이 청각 장애인들을 위해 연구되어 왔으며, 최근 비주얼 아바타 분야에서 많은 관심을 불러일으키고 있는 분야이다. 얼굴 애니메이션은 향상된 컴퓨터 인터페이스, 가상친구, 아바타 채팅, 낮은 전송률의 화상회의 시스템 등등에 응용 분야가 넓어지고 있다.

얼굴 이미지 합성 방법은 두 가지 접근 방법을 가진다. 하나는 3차원 메쉬(mesh)를 이용한 기하학적 변형이며, 다른 하나는 이차원 이미지의 평면적 조작이다. 3차원 메쉬 접근 방식은 특정한 얼굴 모델을 바탕으로 해서 이루어지며, 키-프레임 방법 [3], 파라미터를 이용한 내삽 방법[4], 해부적 구조에 기반한 방법[5], 물리적 성질에 기반한 방법[6][7]등이 있다. 이로 인해 아바타의 현실성을 높여준다.

본 논문에서는 캡처된 2차원 주 프레임 비짐 영상(Key-Frame Viseme image)를 이용한 얼굴 애니메이션 방법과 음성합성기와 통합한 시청각음성합성기 구현에 대하여 설명한다. Radial basis 함수를 이용한 변형과 텍스처 블렌딩을 기반으로 하여 구현하고, 동시조음 현상을 구현하기 위해 우세함수(dominance function)을 적용하며, 계산시간의 단축을 위해 쌍선형 내삽 (bilinear Interpolation)을 적용하여, 이러한 방법들로 실시간 애니메이션이 가능하게 하였다. 그리고, 이를 코퍼스 기반의 음성합성기와 통합하였다.

본 논문의 결과는 낮은 전송 비트율로 가능한 화상회의 등 많은 응용분야에 사용할 수 있는 방안을 제시하고 영상과 음성의 동기화와 자연스러운 발화의 문제에 주안점을 두었다.

2. 시청각 데이터 베이스 구축

2.1. 데이터 수집

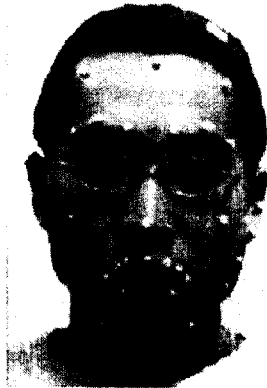
본 논문에서는 표준말을 사용하는 남성화자가 무표정을 포함한 33개의 비짐(viseme) 음절을 보통속도로 발음하는 정면 영상을 디지털 카메라를 사용하여 저

장하였고 이를 다시 컴퓨터상에 30 frame/Sec로 저장하였다.

발화에 따른 화자의 얼굴 움직임 정보를 얻기 위해 얼굴에 마커를 부착하였으며 그 위치정보를 기준으로 분절 영상에서 얼굴움직임 정보를 추측하여 특징점을 뽑아낸다. 얼굴내의 특징점은 얼굴인식, 얼굴모델생성 및 립싱크등 각각의 연구의 목적과 의도에 따라 다른 위치의 특징점들이 사용된다[8-12]. 본 논문에서 사용한 마커는 총 45개를 사용하였으며 양쪽 눈의 가상 마커를 포함해서 총 47개의 특징점을 사용하였고 부착 위치는 그림 1과 같다.

2.2. 녹화용 Viseme Set

언어의 음운학적인 특성을 고려한 영상 매칭을 위해 서는 음성학이나 음운론에 입각한 입술모양, 얼굴 표정이 요구되는데, 음성의 기본단위가 음소라면 그에 상응하는 영상에서의 기본 단위 비짐으로 정의된다.



<그림 1> 표준 얼굴 영상

한국어에 대한 비짐은 표준이 정해진 것이 없으므로 본 논문에서는 음운학적 특성을 고려해 모음과 자음에 대한 비짐을 제시하고 이를 얼굴 애니메이션에 사용할 것이다.

표 1과 같이 모든 한국어 음절에 관하여 33개의 비짐 음절로 표현되며 이에 따르는 자음규칙은 표 2와 같다.

2.3. 특징점 추출

본 논문의 영상기반(image-based) 방식에서 특징점의 위치 정보는 그림 2와 같이 저장된 이미지 열에서 음성레이블링 정보를 바탕으로 음소 이미지를 추출하고 그림 3과 같이 표준 마커 얼굴 영상과 입력 음소 이미지를 비교하면서 입력 음소

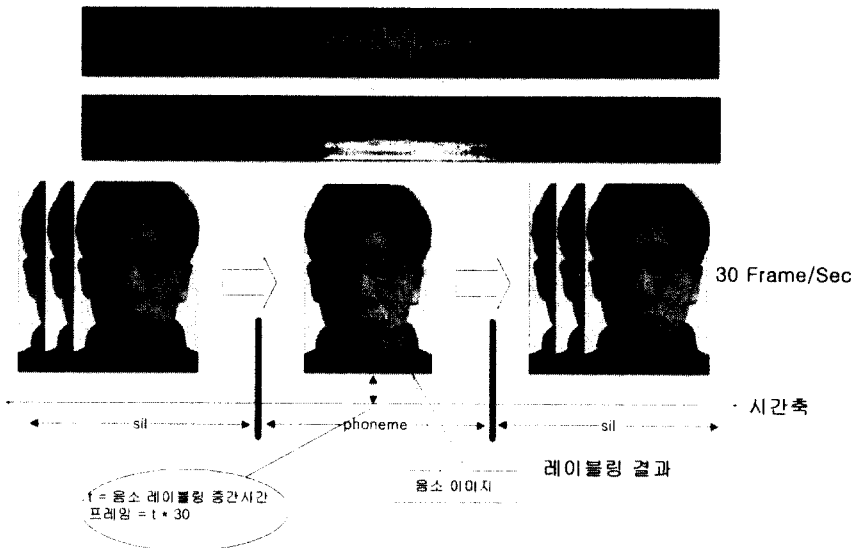
이미지의 특징점을 좌표값으로 표현한다.

<표 1> 녹화용 비짐 음절 셋

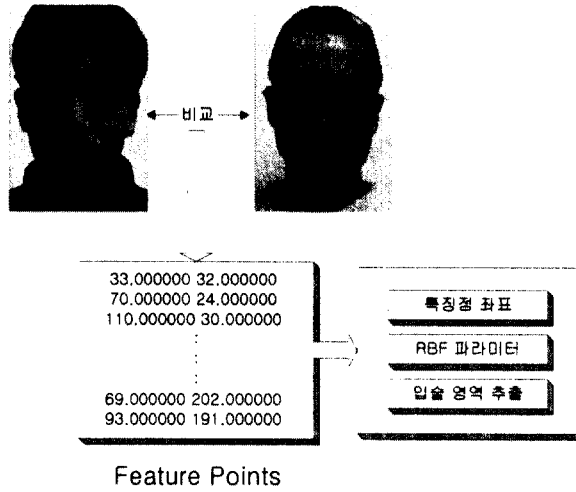
Visemes (Visual Phonemes)	개 수
아, 나, 안, 마, 암, 사	6
어, 너, 언, 머, 엄, 서	6
오, 노, 온, 모, 음, 소	6
우, 무, 움	3
으, 무, 음	3
이, 미, 임	3
에, 네, 엔, 메, 엠, 세	6
계	33

<표 2> 자음 규칙

자 음	Viseme
ㄱ, ㅋ, ㆁ	ㄱ
ㄴ, ㄷ, ㅌ, ㄹ	ㄴ
ㅅ, ㅆ, ㅈ, ㅊ, ㅍ	ㅅ
ㅁ, ㅂ, ㅃ, ㅍ	ㅁ
ㅇ, ㄹ	ㅇ



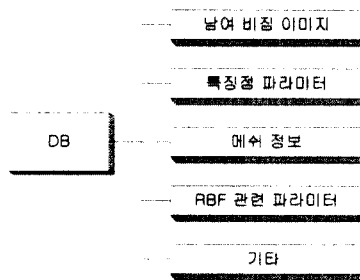
<그림 2> 음소의 대표값 추출



<그림 3> 특징점 추출

2.4. 데이터 베이스 구조

이 논문에서 사용된 립싱크 방식은 특징점의 좌표를 이용한 키프레임 이미지 방식이므로 데이터 베이스는 비침 음소의 이미지와 특징점 파라미터들과 이미지의 메쉬정보를 포함하고 있다. 그림 4는 데이터 베이스의 구조를 보여 주고 있다.



<그림 4> Data Base 구조

3. 입술 동기화 알고리즘

본 논문에서는 캡처된 2D Key-Frame Viseme 이미지를 이용한 얼굴 애니메이션 방법을 사용하였고, radial basis 함수(RBF) 변형과 텍스처 블렌딩을 기반으로 실시간 애니메이션이 가능한 시스템을 구축한다.

본 논문에서 사용한 음성합성기 출력에는 합성음 이외에도 음소정보와 각각 음소에 대한 길이정보가 있는데 음소정보를 바탕으로 DB에서 합성단위를 추출하고, 길이정보를 바탕으로 합성단위의 지속시간이 결정된다. 또한 길이정보는 우세함수(dominance function)가 합성음과 이미지의 동기를 조절하는데 있어서 중요한 역할을 하게 된다. 합성 단위를 불러온 다음 특징점들에 대한 RBF계수를 구하고, 그 계수 값을 바탕으로 이미지 합성시 동시조음 현상을 구현하기 위한 방안으로 우세함수를 적용하였다. 계산시간을 줄이기 위해 쌍선형 내삽을 실시한 뒤, 이 이미지들을 합성음과 함께 화면상에 뿌려주는 방식으로 입술동기화를 구현한다.

3.1. Radial basis function을 이용한 워핑

본 논문에서는 얼굴 애니메이션을 구현하는데 있어 각 음소들이 천이 될 때마다 메쉬의 위치를 비집 영상 안에 산재되어 있는 특징점들의 RBF계수에 의해 구한다[14]. 이러한 키-프레임의 조합으로 입술동기화를 구현하는 방식을 사용한다.

같은 크기의 이미지 안의 어떤 점이 다른 점으로 RBF계수를 바탕으로 이동했다고 가정하고, RBF의 일반식은 다음과 같은 일반식으로 표현된다.

$$(1) \quad F(x) = \sum_{i=0}^N C_i h(\|x - x_i\|)$$

여기서, N은 특징점의 개수이다.

식 1에 특징점의 이동에 대하여 특징점이 미치는 영향의 범위를 정하기 위하여 RBF함수로 가우시안 함수 $h(r) = e^{-\left(\frac{r}{S}\right)^2}$ 을 적용하면 다음과 같은 식으로 표현된다.

$$(2) \quad \begin{matrix} \rightarrow target \\ x_j \end{matrix} = F\left(\begin{matrix} \rightarrow source \\ x_j \end{matrix} \right) = \sum_{i=0}^N C_i e^{-\frac{(\| \begin{matrix} \rightarrow source \\ x_j \end{matrix} - \begin{matrix} \rightarrow \\ x_i \end{matrix} \|)^2}{S_i}}$$

S_i : stiffness constant

x : 2차원 (Feature Points 차원)

r : Feature Points 간의 기하학적 거리

S_i 를 구하는 식은 다음과 같으며, 이 식을 바탕으로 특징점의 영향 영역이 결정된다.

$$(3) \quad S_i = \min_{j \neq i} \left\| \vec{x}_j^{source} - \vec{x}_i \right\|$$

이동될 점의 좌표를 알 수 있는 계수를 구하는 식은 다음과 같이 표현된다.

$$(4) \quad \vec{C} = H^{-1} \vec{x}^{target}$$

위의 식은 잡음이 없다는 가정 하에 만들어진 식이며 이 식에 잡음이 추가될 경우를 고려한 식은 다음과 같이 전개된다.

$$(5) \quad \vec{C} = (H + \lambda I)^{-1} \vec{x}^{target}$$

여기서 λ 는 잡음의 영향을 제거하기 위한 파라미터이다.

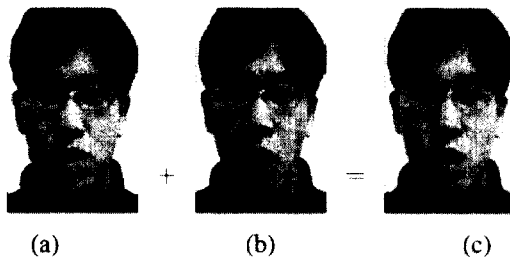
이미지를 워핑하는 과정에서 정방향만으로 진행할 경우 목표 영상에 정확히 부합하는 이미지를 만들어 내는데 어려움이 많기 때문에 정 방향 이미지와 역 방향 이미지를 구한 다음 그 두 이미지의 평균값으로 최종 결과 이미지를 산출해 낸다.

$$(6) \quad I^{new} = (1 - C)I^1(C) + (C)I^2(1 - C)$$

다음 그림 5는 한 방향으로 적용한 결과와 위의 식을 적용해서 평균영상을 취한 그림을 보여준다.

실시간에서 구동되는 얼굴 애니메이션 시스템 구현을 위해 그림 6과 같이 움직임이 많은 입술 부분과 얼굴의 자연스러운 이동과 움직임을 확인 할 수 있는 나머지 부분으로 메쉬를 나누어 알고리즘을 적용한다.

전체 이미지를 5×3개의 메쉬(50×60 Pixels)로 나누고, 입술 부분을 10×12개의 메쉬(5×5 Pixels)로 나눈 다음 Bilinear Interpolation을 적용하여 계산시간을 줄인다.



(a) 순방향 가중치 0.9 (b) 역방향 가중치 0.1 (c) 평균 영상

<그림 5> 순방향과 역방향 보간법



<그림 6> 계산시간을 줄이기 위해 메쉬단위로 나눈 모습

3.2. 우세함수

조음시 음소 각각이 문장중에서 조음 동작이 불가피하게 겹치는 현상이 일어나는데, 이러한 조음 동작의 겹침 현상을 ‘동시조음’이라 한다. 예를 들면 구개음화 현상, 원순음화 현상, 유성음화 현상, 자음약화 현상, 무성음화 현상 등이 있다. 동시조음은 한 음소의 변이음을 결정하는데 가장 중요한 요인이 된다.

동시조음은 앞에 오는 음소에 대한 영향(Backward Coarticulation)과 뒤에 오는 음소의 영향(Forward Coarticulation)을 받는다. 동시조음 현상이 일어날 경우 같은 음소일지라도 다른 입술 모양으로 나타날 수 있으므로 단순히 각 음소를 조합하는 방법[13]으로는 자연스러움이 떨어지게 된다. 이러한 현상을 방지하기 위한 방법으로 연음현상의 모델링에 대한 연구가 시행중이며 크게 Look-ahead model[15], Time locked model[16], 하이브리드 모델 등이 있는데 본 논문에서 적용한 우세함수 본형은 다음과 같다.

$$(7) \quad D = e^{-\theta \tau^C}$$

여기서, $C = \text{power}$

$\theta = \text{비율파라미터}$

이 함수는 시간거리 τ 에 따라 결정되며 음소의 중양으로부터 값의 증감에 따라 지수적 하향 곡선을 그린다.

한국어 음소중 이중 모음은 2개의 단모음으로 구성되어 있다. 이것은 발음시 조음위치가 복합적으로 작용하기 때문에 하나의 이중 모음은 여러개의 단모음의 조합으로 표현이 가능하다.

아래의 표 3을 바탕으로 하나의 이중모음은 두 개의 단모음으로 분리되어 우세함수의 적용을 받는다. 이를 통해 이중모음에 대한 입술동기화를 더욱 자연스럽게 구사할 수 있다.

<표 3> 이중모음 조합 규칙

$\uparrow[ja]= \uparrow[i] + \uparrow[a]$	$\uparrow[i?]= \uparrow[i] + \uparrow[?]$
$\downarrow[jo]= \downarrow[i] + \downarrow[o]$	$\downarrow[iu]= \downarrow[i] + \downarrow[u]$
$\uparrow[je]= \uparrow[i] + \uparrow[e]$	$\uparrow[i?]= \uparrow[i] + \uparrow[e]$
$\downarrow[wa]= \downarrow[o] + \downarrow[a]$	$\downarrow[w?]= \downarrow[o] + \downarrow[?]$
$\downarrow[we]= \downarrow[o] + \downarrow[e]$	$\downarrow[we]= \downarrow[o] + \downarrow[e]$
$\downarrow[yi]= \downarrow[o] + \downarrow[i]$	$\downarrow[w?]= \downarrow[o] + \downarrow[e]$
$\downarrow[?i]= \downarrow[?] + \downarrow[i]$	

본 논문에서는 립싱크에 사용할 수 있는 우세함수를 유도해서 사용한다. 그래서 식 7을 다음과 같은 식으로 확장시키면,

$$(8) \quad D_{sp} = a_{sp} e^{-\theta_{sp} |\tau|^\epsilon} \quad \text{if } \tau \geq 0$$

$$(9) \quad D_{sp} = a_{sp} e^{-\theta_{sp} |\tau|^\epsilon} \quad \text{if } \tau \leq 0$$

D_{sp} 는 세그먼트 s의 P제어 파라미터의 우세함수이다. D_{sp} 의 크기는 제어 파라미터의 위치 P에 따라 달라진다. 시간거리 τ 는 우세함수의 꼭지점에서의 거리를 나타내며,

$$(10) \quad \tau = t_{c\ sp} + t_{o\ sp} - t$$

$t_{c\ sp}$: 세그먼트 s의 중심점의 시간.

$t_{o\ sp}$: $t_{c\ sp}$ 로 부터 함수가 최대가 될때까지의 시간.

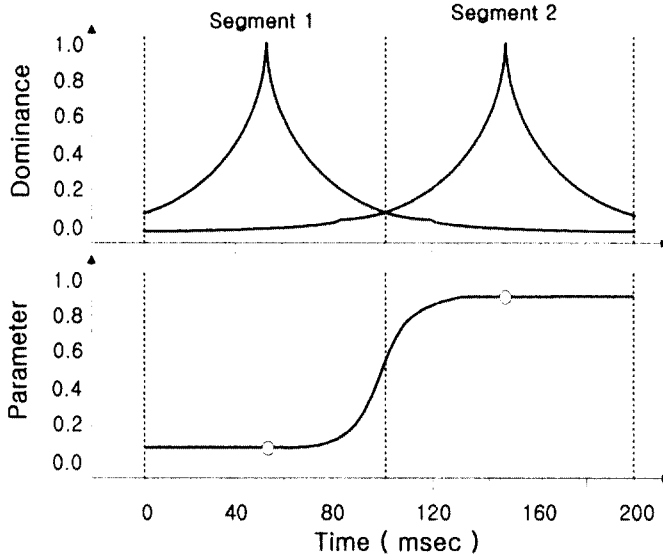
와 같은 식으로 표현된다. 여기서 세그먼트 s의 중심점의 시간을 구하는 식은 다음과 같다.

$$(11) \quad t_{c\ sp} = t_{start\ s} + \frac{duration}{2}$$

$t_{start\ s}$: 세그먼트 시작 시간

$duration$: 세그먼트 지속 시간

시간축상에서 세그먼트된 음소들과 또한 그에 따르는 파라미터의 변화를 그림으로 도시한 것은 그림 7과 같으며, 위의 수식들을 바탕으로 일반적인 블렌딩 함수의 표현은 식 12가 된다.



<그림 7> 립싱크에 적용한 Dominance Function

$$(12) \quad F_p(t) = \frac{\sum_{s=0}^N (D_{sp}(t) \times T_{sp})}{\sum_{s=0}^N D_{sp}(t)}$$

$F_p(t)$: 블렌딩 함수
 s : 음소 세그먼트
 p : *feature points*
 N : 음소의 갯수
 $D_{sp}(t)$: 세그먼트 s 에 도미넌스함수
 T_{sp} : *Target Point*

위와 같은 식을 바탕으로 문장중에서 나오는 비짐 이미지들을 순차적으로 배열한 다음 우세함수를 적용하여 동시조음 현상을 제거할 수 있다.

3.3 쌍선형 내삽을 이용한 계산량 감소

주어진 좌표값들로 표현되는 이차원 공간 내부의 어떤 점이 벡터들의 변형으로 인해 새로운 좌표로 이동할 경우 매쉬의 벡터 이동 좌표 값을 알고 있을 때 내부 점들의 새로운 이동 값들은 양방향의 선형 보간법에 의해서 구할 수 있다. R^2 공간에서 다음과 같은 단위 정사각형 위의 네점을 생각해보자.

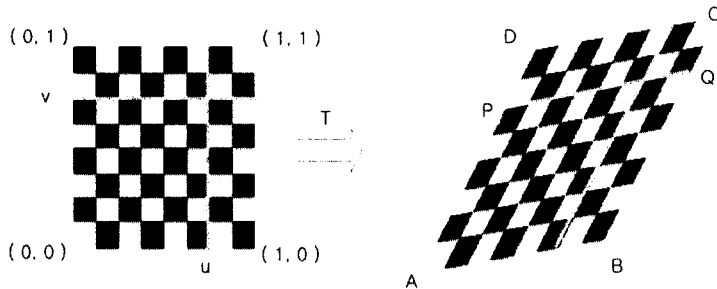
$$(13) \quad \begin{aligned} T(0, 0) &= A, \quad T(1, 0) = B, \\ T(1, 1) &= C, \quad T(0, 1) = D. \end{aligned}$$

위의 4점이 이동할 경우 여기에 해당되는 버텍스(vertex) 안의 변환 점을 찾아야 한다. 변환 식을 아래와 같이 가정한다.

$$(14) \quad T(u, v) = (T_1(u, v), T_2(u, v)), \quad (u, v) \in [0, 1]^2,$$

$$(15) \quad A = (a_1, a_2), \quad B = (b_1, b_2), \quad C = (c_1, c_2), \quad D = (d_1, d_2).$$

쌍선형함수는 4개의 파라미터 즉, a, b, c, d로 정의된다. 그래서 8개의 파라미터 값들이 정의되어야 한다. 쌍선형 변환 T를 수행하기 전에 위 식들을 만족해야 한다. 다음은 쌍선형 변환을 기하학적 해석으로 표현한 것이다.



<그림 8> 쌍선형 보간법

위 그림 8에서 보면 어떤 임의의 점의 변환 점을 알려면 두 번의 1차원 Interpolation을 실시해야 함을 알 수 있다. 그러므로, AD와 BC 두 변 사이에서 보면 다음 두 식과 같이 표현이 된다.

$$(16) \quad P = (1 - v)A + vD$$

$$(17) \quad Q = (1 - v)B + vC$$

다음으로 P와 Q를 파라미터 u에 관한 식으로 나타내면 식 18과 같다.

$$(18) \quad T(u, v) = (1 - u)P + uQ$$

식 16, 식 17, 식 18을 조합한 최종적인 변환 식은 식 19와 같다.

$$(19) \quad T(u, v) = (1 - u, v) \begin{pmatrix} A & D \\ B & C \end{pmatrix} \begin{pmatrix} 1 - v \\ v \end{pmatrix}$$

이런 기하학적 해법은 선형시스템에서 중요한 해법으로 존재한다.

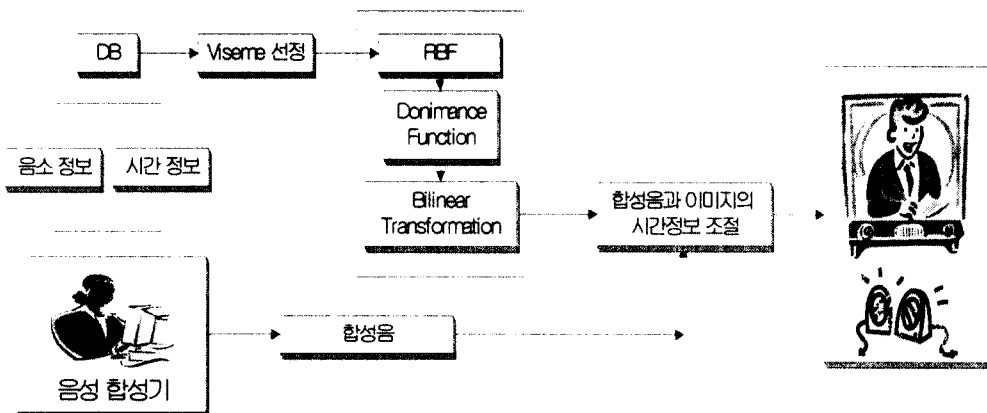
4. 시청각음성합성시스템 구현

본 장에서는 간략하게 본 연구에서 구현된 시청각 음성합성 시스템에 대하여 설명한다.

4.1. 음성 합성기

본 논문에 사용된 음성합성기는 코퍼스 기반의 음성합성기로서 언어처리부와 운율처리부 그리고 합성부로 구축되어 있다. 코퍼스 기반 음성 합성기는 운율 규칙과 합성단위 선정이 대용량 음성 코퍼스를 기반으로 하여 이루어진다. 무제한 음성합성기로 dll형태로 다루어지며 라이브러리를 추가하여 사용할 수 있다. 본 논문에서는 대용량 코퍼스로서, 훈련받은 대학아나운서가 발음한 3,200 발화문장을 사용하였다.

음성합성기는 입술동기화를 위하여 출력으로 음소 정보와 길이 정보를 입술동기화 모듈 넘겨주게 된다. 입술 동기화 모듈은 넘겨받은 정보를 바탕으로 해서, 얼굴 애니메이션을 구현한다.



<그림 9> 시청각음성합성 시스템 블록도

4.2. 음성 합성기와 통합

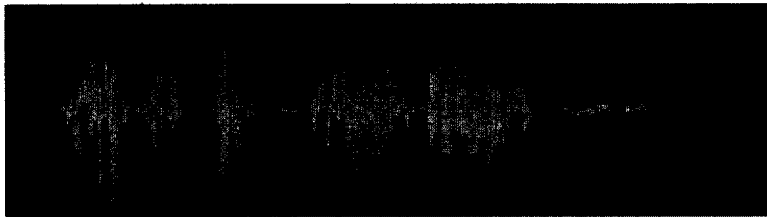
Microsoft Visual C++ 6.0으로 구현되었으며, Form View 형태로 작성되었으며, 칼라 이미지를 화면에 뿌려주는 형태로 재생된다. 그림 9는 구현된 시청각 음성합성 시스템의 블록도를 보여주고 있다.

5. 실험결과 및 평가

프로그램의 성능은 펜티엄 4 1.7Ghz의 컴퓨터를 기준으로 초당 25~30프레임 정도이며, 본 논문에 사용된 립싱크 시스템의 시뮬레이션은 SiliconGraphics 320 모델로 실시하였고, 초당 15Frame의 이미지를 화면에 뿌리게 하였다. 실험결과 그림 11에서와 같이 내삽으로 재생된 이미지 열들은 자연스러운 입술움직임을 보인다. 그림에서 사용된 문장은 다음과 같다.

실험문장 : 아가야 놀자.

단, 그림 10은 합성기로 만들어진 '아가야 놀자'에 대한 합성음의 파형을 그린 것이다.



<그림 10> 음성 출력 파형그림

6. 결론 및 향후 계획

본 논문에서는 캡처된 2차원 주프레임 비집 이미지를 이용하여 얼굴 애니메이션 하는 방법에 대해 기술하였다. RBF 변형과 텍스처 블렌딩을 기반으로 하여 구현 실시간 얼굴 애니메이션 시스템을 개발하였다. 또한 3차원 아바타에 비교하여 낮은 전송비트율과 모델링의 용이함을 바탕으로 2차원 아바타의 가능성과 각종 응용분야에 사용할 수 있는 가능성을 제시하였다.

특징점들의 관계를 계산하고 이차원 입술동기화 시스템의 주축을 이루는 RBF 함수와 입술 애니메이션의 부자연스러움의 원인이 되는 동시조음현상을 구현하기 위해서 사용된 우세함수, 마지막으로 계산시간을 줄이기 위한 방안으로 사용된 쌍선형 변환을 실시간 얼굴 애니메이션 시스템에 적용하였으며 그 결과 효과적인 얼굴 애니메이션을 가능케 할 수 있었다. 이 얼굴 합성기는 음성 합성기의 정보를 바탕으로 구동되며, 단일 음소열에 대한 운율 정보나 움직임의 가중치에 대한 최적화가 필요하다. 향후 계획으로는 얼굴 애니메이션에 표정표현과 자연스런 아바타의 움직임등 더욱 현실감을 높여 줄 수 있는 방안에 대해 연구할 계획이다.

참 고 문 헌

- [1] Sumbly, W. H. and I. Pollack (1954), Visual contribution to speech intelligibility in noise, *Journal of the Acoustical Society of America* Vol.26, pp.212~215.
- [2] Summerfield, Q., A. MacLeod, M. McGrath and M. Brooke (1989), Lips, teeth, and the benefits of lipreading, *Handbook of Research on Face Processing*. A. W. Young and H.D.Ellis Editors, Elsevier Science Publishers, pp.223~233.
- [3] Ezzat, Tony and Tomaso Poggio (1999), Visual Speech Synthesis by Morphing Visemes, MIT AI Memo No 1658/CBCL Memo No 173, May.
- [4] Parke, F. (1982), Parameterized models for facial animation, *IEEE Computer Graphics and Applications*. November, pp.61~68.
- [5] Waters, Keith (1987), A Muscle models for Animating 3D Facial Expression, *Proc. SIGGRAPH 87*, In computer Graphics, Vol.21, No.4, July, pp.17~24.
- [6] Lee, Yuencheng, Demetri Terzopoulos, and Keith Waters (1995), Realistic Modeling for Facial Animation, *Proc. SIGGRAPH 95*. In Computer Graphics 99. pp.55~62,
- [7] Terzopoulos, Demetri and Keith Waters (1990), Physically-based Facial Modeling, Analysis and Animation, In *The Journal of Visualization and Computer Animation*. Vol. 1, pp. 73~80.
- [8] Yuille, A. L., D. Cohen, P. Hallinan (1989), Feature Extraction Form Face Using Deformable Templates, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.104~109.
- [9] Reisfeld, D., H.Wolfson and Y. Yeshurun (1990), Detection of Interest Points Using Symmetry, *Proceedings of the 3rd International Conference on Computer Vision*, Dec., pp.62~65.
- [10] Brunelli, R. and T. Poggio (1993), Face Recognition: Features Versus Templates, *IEEE Trans. PAMI*, Vol.15. No.10, Oct., pp.1042~1052.
- [11] Silva, Litanage C. Ee, Kyoharu Aizawa and Mitsutoshi Hatori (1995), Detection and Tracking of Facial Features by using edge pixel counting and deformable circular template matching, *IEICE Trans. Information and System*, Vol.E78-D. No.9, Sep., pp.1195~1207.

- [12] Huang, Chung-Lin and Ching-Wen Chen (1992), Human Facial Feature Extraction for Face Interpretation and Recognition, *Pattern Recognition*, Vol.25. No.12, pp.1435 ~ 1444.
- [13] Noh, Jun-yong and Ulrich Neumann (2000), Talking Faces, *IEEE International Conference on Multimedia and Expo II*, pp.627 ~ 630.
- [14] Elson, M. (1990), Displacement facial animation techniques, *ACM SIGGRSPH Course Note #26*.
- [15] Kozhevnikov, V. A. and L. A. Chistovich (1965), *Rech: artikulyatsiya I Vospriyatiye* (Moscow-Leningrad, 1965). *TransArticulation and perception*, Washington, D.C : Joint Publication Research Service Vol.30, pp 543.
- [16] Bell-Berti, F. and K. S. Harris (1982), Temporal patterns of coarticulation: Lip rounding, *Journal of the Acoustical Society of America* 71, pp.449 ~ 459.
- [17] Young, Steve, Julian Odell, Dave Ollason, Valtcho Valtchev and Phil Woodland, *The HTK Book*.
- [18] Cohen, M. and D. Massaro (1993), Modeling coarticulation in synthetic visual speech, Michael In D. Thalmann N. Magnenat-Thalmann, editor, *Computer Animation '93.*, Springer-Verlag, pp.139 ~ 156.

접수일자: 2002년 5월 6일

게재결정: 2002년 5월 24일

▶ 김명곤(MyoungGon Kim)

주소: 광주광역시 북구 용봉동 300번지 전남대학교 전자공학과

소속: 전남대학교 전자공학과 신호처리 실험실

전화: 062)530-0472

fax: 062)530-1759

E-mail: ledmaster76@empal.com

▶ 김진영(JinYoung Kim)

주소: 광주광역시 북구 용봉동 300번지 전남대학교 전자공학과

소속: 전남대학교 전자공학과 신호처리 실험실

전화: 062)530-1757

fax: 062)530-1759

E-mail: kimjin@dsp.chonnam.ac.kr

▶ 백성준(SeongJoon Baek)

주소: 광주광역시 북구 용봉동 300번지 전남대학교 전자공학과

소속: 전남대학교 전자공학과 정보신호처리 실험실

전화: 062)530-1795

fax: 062)530-1759

E-mail: tozero@chonnam.ac.kr