한국음향학회지 제21권 제1호 pp. 19~26 (2002)

한국어 모음 입술독해를 위한 시공간적 특징에 관한 연구

A Study on Spatio-temporal Features for Korean Vowel Lipreading

오 현 화*, 김 인 철*, 김 동 수*, 진 성 일*
(Hyun Hwa Oh, In Cheol Kim, Dong Su Kim, Sung Il Chien)

*경북대학교 전자전기컴퓨터학부

(접수일자: 2001년 9월 17일; 채택일자: 2001년 10월 18일)

본 논문에서는 한국어 입술독해를 위한 기반 연구로서 음성학에 기반하여 음성의 시각적 기본 단위인 viseme을 정의하고 입술의 움직임을 적절히 표현할 수 있는 특징들을 추출하여 그 성능을 분석하였다. 먼저, 다수의 화자로부터 한국어 모음에 해당하는 입술의 동영상 데이터베이스를 획득하고 각 모음별 시각적 특성을 분석하여 7개의 한국어 모음 viseme을 정의하였으며 입술 윤곽선상의 특징점과 시공간적 특징 벡터들을 추출하여 은닉 마르코프 모델에 적용함으로써 효과적인 입술독해를 위한 각 특징 벡터별 성능을 비교하였다. 7개의 한국어 각 viseme에 대한 인식 실험 결과에서 입술의 안팎 윤곽선의 정보가 모두 반영된 특징 벡터가 입술독해에 효과적으로 적용될 수 있으며 윤곽선 상의 특징점들의 시간적 움직임 크기와 방향이 입술독해를 위하여 매우 중요한 요소임을 확인할 수 있었다.

핵심용어: 한국어 입술독해, 한국어 viseme, 시공간적 특징
투고분야: 음성처리 분야 (2.5)

This paper defines the visual basic speech units, visemes and investigates various visual features of a lip for the effective Korean lipreading. First, we analyzed the visual characteristics of the Korean vowels from the database of the lip image sequences obtained from the multi-speakers, thereby giving a definition of seven Korean vowel visemes. Various spatio-temporal features of a lip are extracted from the feature points located on both inner and outer lip contours of image sequences and their classification performances are evaluated by using a hidden Markov model based classifier for effective lipreading. The experimental results for recognizing the Korean visemes have demonstrated that the feature vector containing the information of inner and outer lip contours can be effectively applied to lipreading and also the direction and magnitude of the movement of a lip feature point over time is quite useful for Korean lipreading.

Keywords: *Korean lipreading, Korean viseme, Spatio-temporal features*
ASK subject classification: *Speech signal processing (2.5)*

I. 서 론

인간은 음성인식에 있어서 음향 정보와 더불어 입술의 모양과 움직임, 제스처, 화자의 얼굴 표정 등과 같은 시각적인 정보를 함께 이용함으로써 음성을 좀더 정확히 인식하게 된다. 특히 청각 장애자나 잡음이 강한 환경 하에서의 대화에는 입술독해 (lipreading)가 의사소통의 중요한 보조 수단이 됨은 잘 알려진 사실이다[1,2]. 현재 음향 정보만을 이용한 음성인식 시스템은 잡음이나 화자간의 음성 간섭 (cocktail party effect)이 존재하는 환경에서는 인식 성능이 현저히 저하되어 그 적용에 어려움을 가지고

책임저자: 오현화 (ohh@palgong.knu.ac.kr)
702-710 대구시 북구 산격동 1370번지
경북대학교 전자전기컴퓨터학부 대학원
(전화: 053-940-8645; 팩스: 053-950-5505)

인식 성능이 현저히 저하되어 그 적용에 어려움을 가지고 있다. 최근 들어 이러한 문제를 해결하기 위하여 음성의 시각정보 기반의 입술독해 기술과 음향정보 기반의 음성처리 기술이 결합된 바이모달 (bimodal) 시스템을 구현하여 음성인식의 성능을 향상시키고자 하는 연구가 진행되고 있다[3-6].

입술독해에 대한 기존의 연구에서는 주로 고립 단어를 인식 단위로 하여 화자 종속 및 화자 독립 기반의 시스템 개발에 중점을 두고 있으나 음소와 문장을 인식단위로 한 연구도 점차 진행되고 있다. 입술독해를 위한 특징 추출 방법에는 입술 영상의 명암 정보로부터 필터링 (filtering)과 차원감소 (dimension reduction) 등의 전처리 과정을 거쳐 입술 영역의 특징을 추출하는 영상기반 방법 (image-based method)[4]과 입술의 기하학적 형태에 근거하여 입술 모델을 사전에 구성하고 그 모델의 파라메터를 추정하는 모델기반 방법 (model-based method)[5] 등이 있다. 또한 입술의 시각 정보 해석 및 인식을 시간 지연 신경망 (time delayed neural network: TDNN)[3] 기반의 인식기와 은닉 마르코프 모델 (hidden Markov model: HMM) 기반의 인식기[4,5] 등이 주로 사용되고 있다. 국내에서도 입술독해에 관한 연구가 수행된 바가 있으나[6] 현재까지 한국어 입술독해의 성능을 객관적으로 평가하기 위한 표준화된 데이터베이스가 구축되어 있지 않으며, 입술독해에 유용한 특징 벡터 및 이에 기반한 시스템 구현에 관한 연구가 아직 미진한 상태이다.

본 논문에서는 다수의 화자로부터 한국어 모음에 해당하는 입술의 동영상 데이터베이스를 획득하고 각 모음별 시각적 특성을 음성학에 기반하여 분석함으로써 한국어 입술독해에 적합한 음성의 시각적 기본 단위인 viseme을 정의하였다. 또한 입술 윤곽선 상의 특징점을 검출하고 동영상에서의 이들 특징점으로부터 입술의 안팎 높이 및 폭, 그리고 입술의 움직임 궤적, 크기, 및 방향 등의 정보를 포함하는 시공간적 특징 벡터들을 추출하여 각 viseme에 대한 입술독해 실험을 수행하였다. 실험에서는 신호의 시간 및 공간적인 변환을 잘 처리할 수 있어 음성인식, 온라인 필기체 인식 등의 연구분야에 많이 적용되고 있는 이산 HMM을 주 인식기로 사용하였으며 각 특징 벡터별 인식 성능을 비교 분석함으로써 각 viseme에 해당하는 입술의 고유한 모양과 움직임 정보가 적절히 반영된 특징 벡터를 획득하고자 하였다.

II. 한국어 viseme 정의

Viseme은 음성의 기본 단위인 음소에 상응하는 음성의 시각적 기본 단위를 의미하며 국외의 일부 연구그룹에서는 영어에 대한 viseme을 정의하여 입술독해에 이용하고 있으나[7,8] 현재까지 그 표준은 정해지지 않았다. 국내에서는 입술독해를 위한 한국어 viseme에 대한 연구가 현재 전무한 상태이므로, 본 논문에서는 먼저 한국어의 음성학적 특성에 기반하여 입술독해에 적합한 viseme을 정의한다.

Viseme을 정의하기 위해서는 후두, 성대, 성도 등 시각적으로 관찰할 수 없는 조음 기관의 구조적 작용보다는 시각적인 관찰이 가능한 혀와 이 그리고 특히 입술 모양에 대한 분석이 중요하다. 한국어에서 음성 및 음운학적 단위인 음절은 기본적으로 모음과 자음의 조합으로 구성되며 발성과정에서의 화자의 입술 모양은 자음보다 모음의 종류에 의하여 주로 결정된다. 자음의 경우에는 /ㅁ, ㅂ, ㅍ/과 같이 두 입술이 붙어서 발음되는 양순음 외에는 상응되는 입술 모양의 특징이 시각적으로 관찰되지 않는다. 따라서 본 논문에서는 입술 모양의 특징이 시각적으로 명확히 구분되는 모음에 대하여 viseme을 정의한다.

음성학적으로 모음은 입술의 모양 (둥글 또는 안둥글)과 혀의 높낮이, 앞뒤 위치에 의하여 그 음색이 결정된다. 그림 1(a)는 모음을 기술하기 위하여 제안된 모음 사각도 상에 한국어 단모음을 표시한 것이다[9]. 모음 사각도에서 가로축은 혀의 위치를 나타내고 세로축은 입술의 벌어진 정도를 나타내며 단모음마다 그 위치가 서로 다름을 알 수 있다. 발성과정에서 입술이나 혀가 다른 지점으로 움직이는 중모음은 과도음에 따라 /j (ㅣ)/계-/ㅑ, ㅕ, ㅛ, ㅠ, ㅖ, ㅒ/, /w (ㅜ)/계-/ㅘ, ㅝ, ㅙ, ㅞ/, /ɯ (ㅡ)/계-/ㅢ/으로 나눌 수 있다[10]. 그림 1(b)는 /j/계 중모음을 모음 사각도에 나타낸 것이다. 중모음에서의 과도음과 단모음의 음성학적 결합은 시각적으로도 그에 해당하는 입술 모양의 변화로 관찰된다. 그러므로 본 논문에서는 단모음을 시각적으로 구분할 수 있는 기본단위로 판단하고 이에 대한 viseme을 정의하고 분류한다. 그림 2는 한 화자에 대한 모음별 아래와 위 입술의 벌림 정도와 입술 폭의 크기에 관한 연구 결과를 나타낸다[10]. 그림 2(a)에서 /ㅓ/와 /ㅐ/모음은 입술이 벌어진 높이가 서로 근사하더라도 아래와 위 입술의 높이가 서로 다르다. 최종적으로 본 논문에서는 한국어의 음성학적 특성과 입술 모양에 대한 시각적 특징을 근거로 하여 입술독해를 위한 총 7개

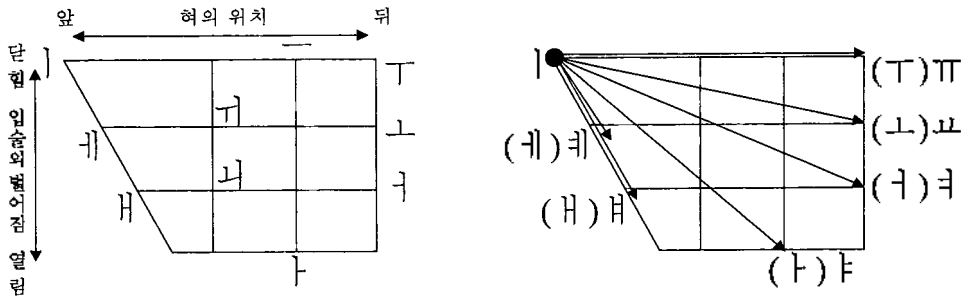


그림 1. 모음 사각도에서의 한국어 모음 (a) 단모음 (b) 모음 사각도에서의 /j/계 중모음

Fig. 1. Korean vowel on vowel chart: (a) Monophthong, (b) /j/-diphthong on vowel chart.

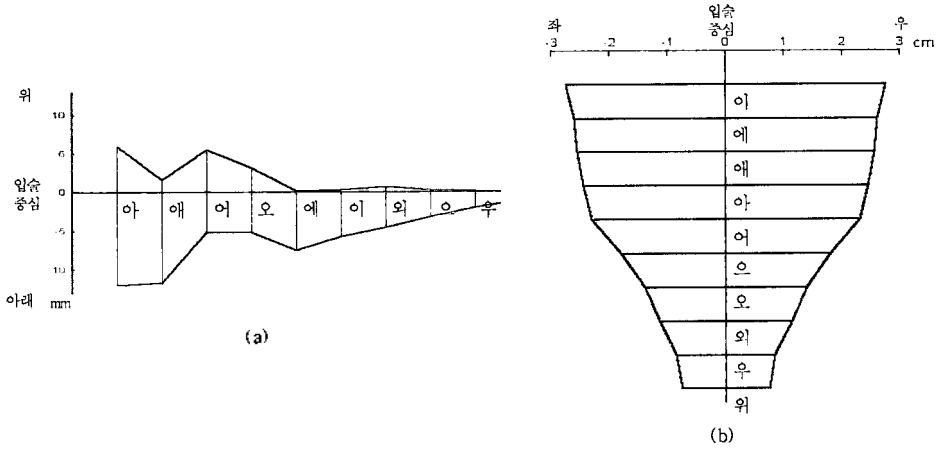


그림 2. 모음에 대한 입술 벌림의 크기 비교 (a) 모음별 아래와 위 입술의 벌림 정도 (b) 모음별 입술 폭의 벌림 정도 비교

Fig. 2. Comparison of lip opening for Koran vowels: (a) Heights of upper and lower lips, (b) Width of lip.

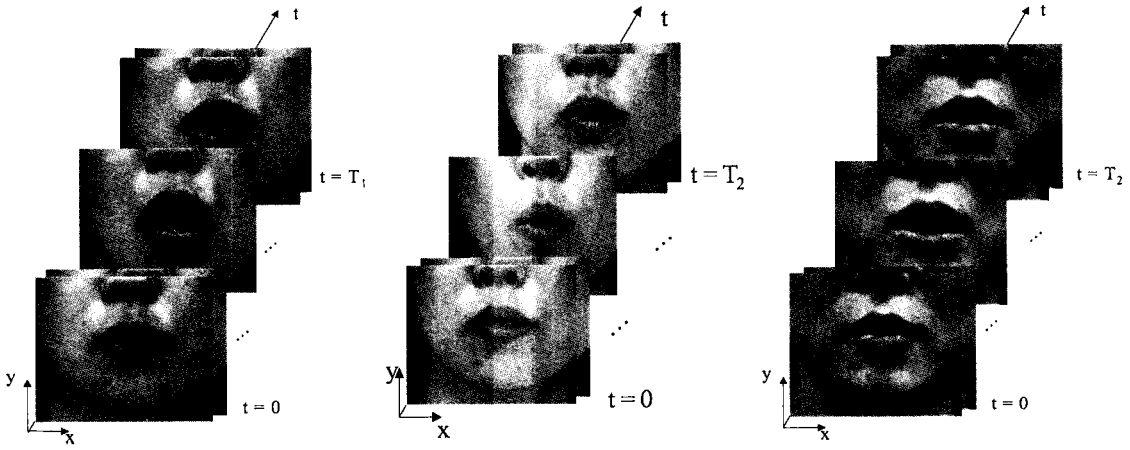


그림 3. 한국어 모음 viseme (/ㅏ/, /ㅗ/, /ㅣ/)에 대한 입술 동영상 데이터베이스의 예

Fig. 3. Lip image sequences of Korean vowels, /a/, /o/, and /i/.

의 한국어 모음 viseme을 /ㅏ/, /ㅐ/, /ㅓ/, /ㅗ/, /ㅜ/, /ㅡ/, /ㅣ/로 정의하였다.

III. 입술 동영상에서 입술독해를 위한 시공간적 특징 벡터 추출

본 논문에서는 전술한 한국어 viseme에 대한 입술독해 실험을 수행하기 위해 대응되는 7개에 모음에 대하여 15

명의 화자가 10번 또는 5번씩 발성하여 획득한 총 875개의 입술 동영상을 데이터베이스로 사용하였다. 동영상 획득시, 각 모음에 대하여 입술이 닫힌 상태에서 발성을 시작하여 평균 2초간 발성하도록 하였으며 320×240 크기의 입술 영상을 초당 30프레임의 속도로 획득하였다. 그림 3은 획득된 동영상 데이터베이스의 예를 나타낸다.

3.1. 입술의 특징점 검출

동영상 데이터를 획득하는 과정에서 화자의 얼굴 움직임으로 인한 입술 영상의 회전과 이동은 입술 특징 추출의 일관성을 저하시킬 수 있다. 본 논문에서는 그림 4에서와 같이 첫번째 프레임의 입술 중심 $(x_c^{(1)}, y_c^{(1)})$ 을 기준점으로 하여 유사변환 (affine transform)을 수행함으로써 이러한 입술 영상의 회전과 이동을 보정한다. 입술독해를 위한 특징 벡터를 추출하기 위해서는 보정된 동영상에서 입술 윤곽선 상의 특징점들을 먼저 검출해야 한다. 최종적인 입술독해 시스템을 구현하기 위해서는 특징점의 검출 과정이 일련의 영상처리 과정을 통해 자동으로 수행되어야 한다. 그러나 본 논문에서는 입술독해에 유용한 특징 벡터를 추출하고 그 특성 및 성능을 정확하게 분석하는 것을 목적으로 하므로 영상처리 과정에서 발생할 수 있는 에러를 배제하고 추출된 특징벡터의 성능 검증에 대한 신뢰도를 향상시키기 위해 각 특징점들을 수동으로 검출하였다. 그림 4는 입술의 안과 바깥 윤곽선 상에서 총 12개의 입술 특징점들을 시계방향 순으로 검출하는 과정을 나타낸다.

3.2. 한국어 모음 입술독해를 위한 시공간적 특징 벡터

본 논문에서는 한국어 모음 viseme의 인식을 위해 아래에 설명된 총 5개의 시공간적 특징 벡터를 추출하여 그 성능을 분석하였다.

① 입술의 바깥 윤곽선의 높이와 폭: f_a

그림 5에서 나타낸 바와 같이 동영상의 모든 프레임에 대하여 입술의 바깥쪽 윤곽선 상의 특징점으로부터 입술의 중앙 및 좌우의 높이와 입술의 폭을 각각 구하여 특징

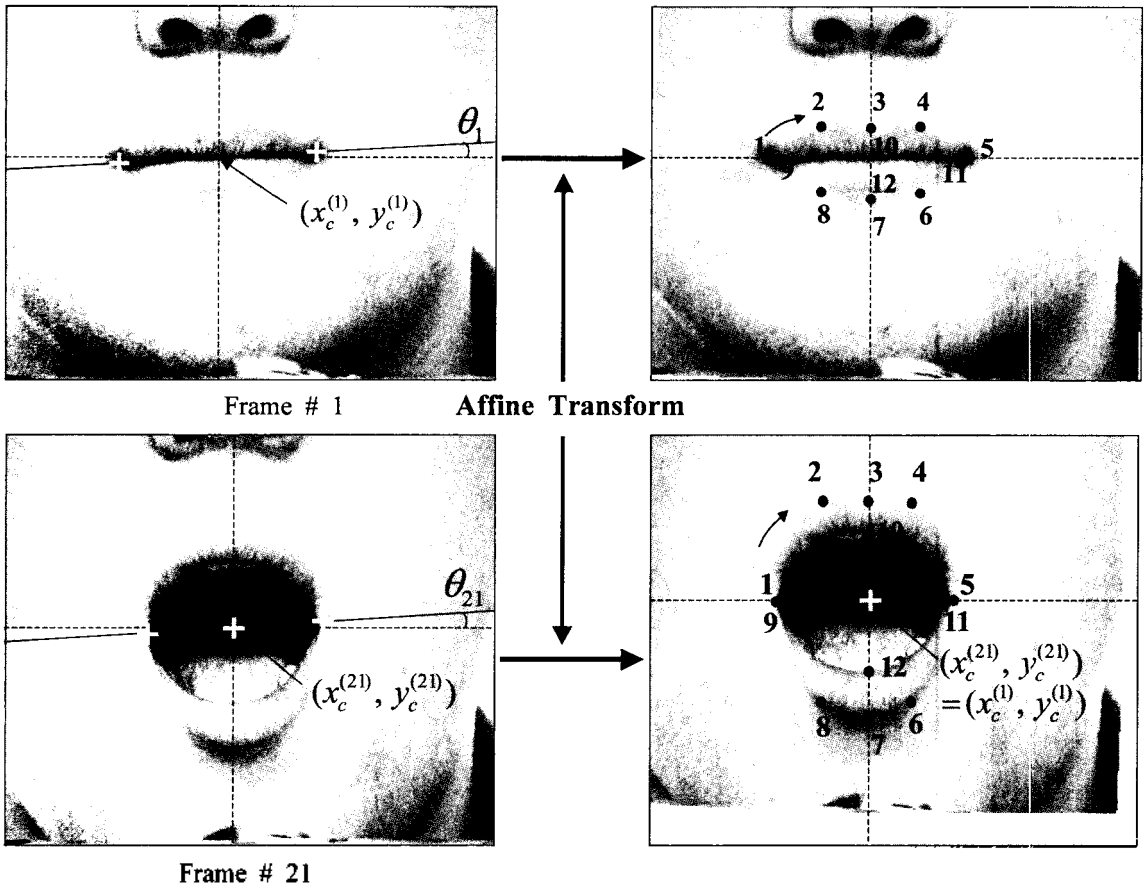


그림 4. 입술영상의 회전과 이동을 보정하기 위한 유사변환과 수동으로 검출된 입술 윤곽선 상의 특징점

Fig. 4. Results of affine transform to compensate rotation and translation of lip images and feature points on lip contours.

벡터로 사용한다. 이 경우 아래 식 (1)에서와 같이 첫번째 프레임에서 산출된 벡터 구성성분의 값들을 이용하여 각 프레임별 성분들의 크기를 정규화 함으로써 화자와 카메라 사이의 거리가 고정된 상태에서 화자의 입술 크기가 서로 다르게 나타나는 문제를 보정한다.

$$\mathbf{f}_a = \left[\mathbf{f}_{a,1}\ \mathbf{f}_{a,2} \cdots \mathbf{f}_{a,N_k} \right]^T,$$
$$\mathbf{f}_{a,i} = \left[h_O^{(i)}\ w_O^{(i)}\ h_{OL}^{(i)}\ h_{OR}^{(i)} \right],$$
$$h_O^{(i)} = H_O^{(i)} / H_O^{(1)},\ w_O^{(i)} = W_O^{(i)} / W_O^{(1)},$$
$$h_{OL}^{(i)} = H_{OL}^{(i)} / H_{OL}^{(1)},\ h_{OR}^{(i)} = H_{OR}^{(i)} / H_{OR}^{(1)} \quad (1)$$

여기서 동영상을 구성하는 프레임 번호 $i=1,2,\cdots,N_k$ 이다.

②입술의 안팎 윤곽선의 높이와 폭: $\mathbf{f}_b$

입술이 크게 벌어진 경우에 입술의 안쪽 윤곽선은 시각적으로 바깥 윤곽선에 비해 좀더 뚜렷하고 안정적인 특징을 보여준다. 이러한 특성을 반영하기 위해 그림 5에서와 같이 입술의 안쪽 윤곽선 상의 특징점으로부터 높이 H_I 와 폭 W_I 를 구하여 전술한 특징벡터 $\mathbf{f}_a$ 의 구성 성분과 결합함으로써 새로운 특징벡터 $\mathbf{f}_b$ 를 생성한다. 이때 $\mathbf{f}_b$ 의 각 구성 성분은 $\mathbf{f}_a$ 와 같이 첫번째 프레임에서 구한 초기 값들로 각각 정규화 된다. 또한 입술이 닫힌 경우에는 입술의 안쪽 높이와 폭은 0으로 정의된다.

$$\mathbf{f}_{b,i} = \left[h_O^{(i)}\ w_O^{(i)}\ h_{OL}^{(i)}\ h_{OR}^{(i)}\ h_I^{(i)}\ w_I^{(i)} \right] \quad (2)$$

③ 시간에 따른 입술 특징점의 이동 궤적: $\mathbf{f}_c$

Viseme을 발성하는 과정에서의 입술 움직임의 변화를 보다 정확하게 표현하기 위해 첫번째 프레임의 12개 특징점 좌표를 기준으로 하는 프레임별 각 특징점의 상대 좌표를 계산하여 특징 벡터 $\mathbf{f}_c$ 를 구성한다.

$$u_j^{(i)} = \left(x_j^{(i)} - x_j^{(1)}\right) / W_O^{(1)},\ v_j^{(i)} = \left(y_j^{(i)} - y_j^{(1)}\right) / H_O^{(1)}$$
$$\mathbf{f}_{c,i} = \left[u_1^{(i)}\ v_1^{(i)} \cdots u_{12}^{(i)}\ v_{12}^{(i)} \right] \quad (3)$$

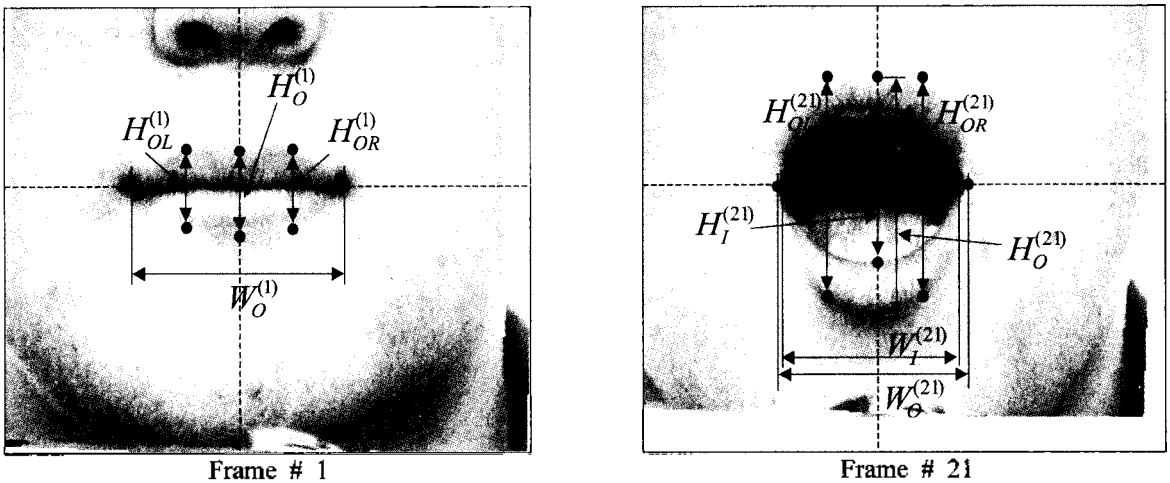
여기서 $(x_j^{(i)}, y_j^{(i)})$ 는 i번째 프레임에서 j번째 특징점의 x, y 좌표값을 나타내며 $(u_j^{(i)}, v_j^{(i)})$는 그 특징점의 정규화된 상대 좌표를 나타낸다. 따라서 특징벡터 $\mathbf{f}_c$ 의 각 구성원은 입술 움직임에 따른 각 특징점들의 시간적 이동 궤적을 나타낸다. 그림 6은 한 화자로부터 7개 모음 viseme에 대하여 입술의 바깥쪽 윤곽선에서 검출한 5개 특징점들의 시간에 따른 이동 궤적을 나타낸 그래프이다. 입술 특징점들의 시간에 따른 이동 크기와 방향은 viseme마다 고유한 특징을 나타내며 화자마다 이와 같은 고유한 이동 궤적 특성이 유사하게 관찰된다.

④ 시간에 따른 입술 특징점의 이동 크기: $\mathbf{f}_d$

특징 벡터 $\mathbf{f}_c$ 에서 입술 특징점의 이동 방향은 배제하고 이동 크기만을 계산하여 특징벡터 $\mathbf{f}_d$ 를 구성한다.

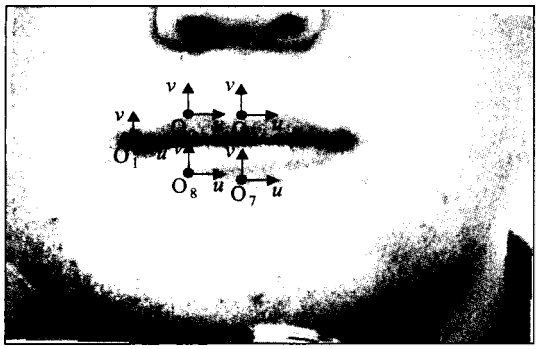
$$M_j^{(i)} = \sqrt{\left(u_j^{(i)}\right)^2 + \left(v_j^{(i)}\right)^2}$$
$$\mathbf{f}_{d,i} = \left[M_1^{(i)}\ M_2^{(i)} \cdots M_{12}^{(i)} \right] \quad (4)$$

⑤ 입술 특징점의 시간에 따른 움직임 속도: $\mathbf{f}_e$

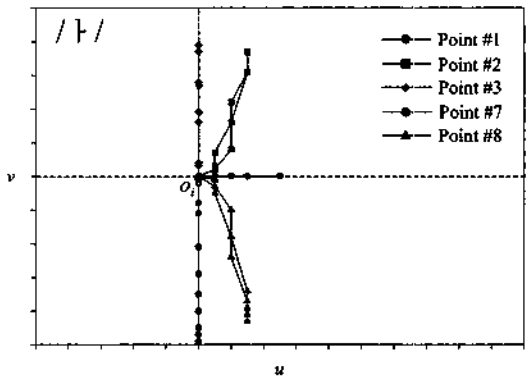


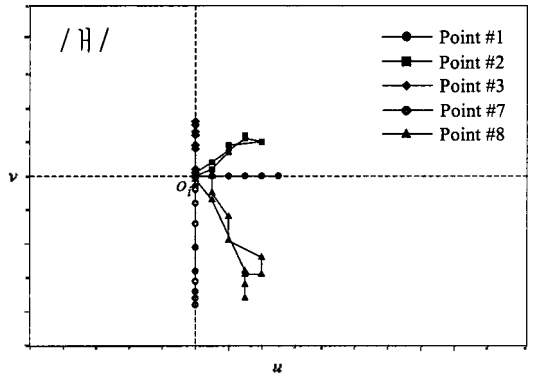
Frame # 1 Frame # 21

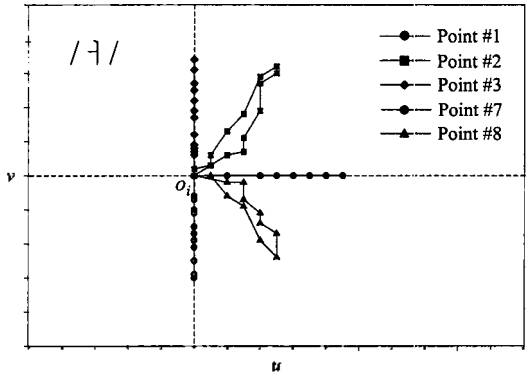
그림 5. 모음 /ㅏ/에 대한 입술의 안팎 높이와 폭
Fig. 5. Outer and inner heights and widths of lip for vowel /a/.

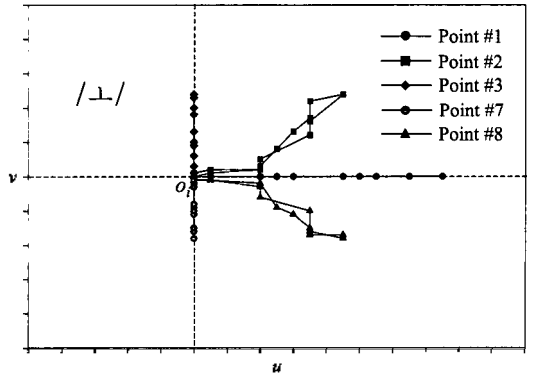


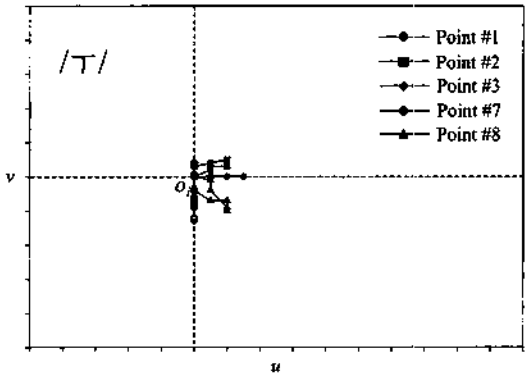
Frame#1

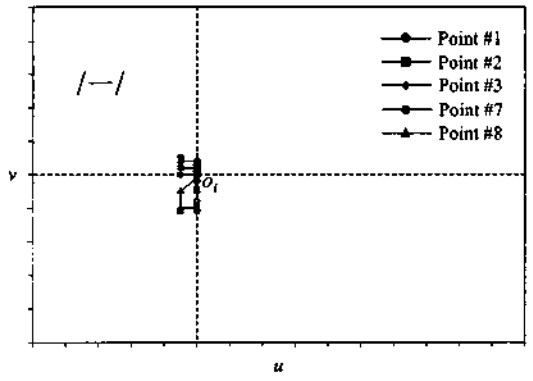












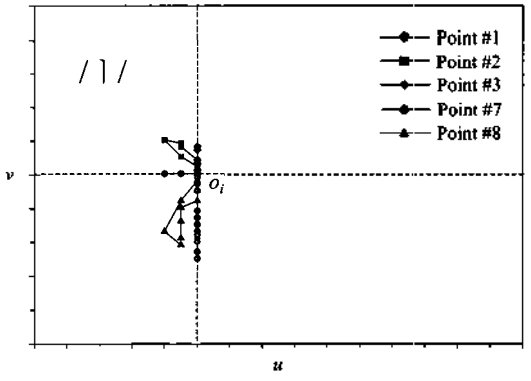


그림 6. 한 화자로부터 획득된 7개 모음 viseme에 대한 시간에 따른 입술 특징점들의 이동 궤적

Fig. 6. Trajectories of eight feature points located on left side of lip over time for Korean vowels spoken by a speaker.

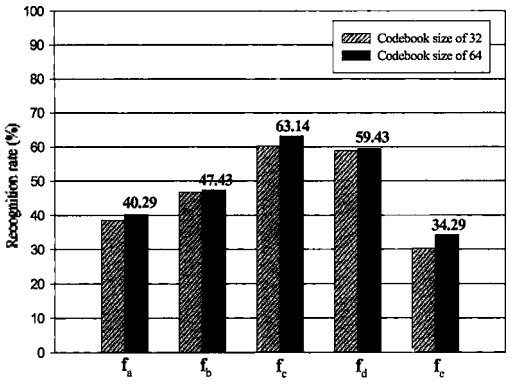


그림 7. 5개의 입술 특징 벡터를 이용한 입술독해 인식률
Fig. 7. Comparison of highest recognition rates of five sets of feature vectors.

마지막으로, 특징점의 이동 궤적을 나타내는 특징벡터 $\mathbf{f}_c$의 각 구성 성분에 대해 아래 식 (5)와 같이 현재 프레임과 이전 프레임 사이의 차를 이용한 일차 미분 과정을 수행하여 각 특징점별 속도 벡터를 구한다.

$$v_{x,j}^{(i)} = u_j^{(i)} - u_j^{(i-1)} \; v_{y,j}^{(i)} = v_j^{(i)} - v_j^{(i-1)} \tag{5}$$

산출된 속도 벡터의 크기 $\|\mathbf{v}_j^{(i)}\|$와 방향 $\alpha_j^{(i)}$으로 구성된 특징 벡터 $\mathbf{f}_e$를 식 (6)과 같이 정의하고 입술 움직임의 속도가 각 viseme의 인식에 미치는 영향을 평가한다.

$$\|\mathbf{v}_j^{(i)}\| = \sqrt{\left(v_{x,j}^{(i)}\right)^2 + \left(v_{y,j}^{(i)}\right)^2}, \quad \alpha_j^{(i)} = \tan^{-1}\left(v_{y,j}^{(i)} / v_{x,j}^{(i)}\right)$$
$$\mathbf{f}_{e,i} = \left[\|\mathbf{v}_1^{(i)}\| \; \|\mathbf{v}_2^{(i)}\| \; \cdots \; \|\mathbf{v}_{12}^{(i)}\| \; \alpha_1^{(i)} \; \alpha_2^{(i)} \; \cdots \; \alpha_{12}^{(i)}\right] \tag{6}$$

IV. 실험 및 고찰

실험에서는 전술한 5개의 시공간적 특징 벡터들의 성능을 평가하기 위하여 먼저 이산 HMM에 기반한 인식기를 구현한다. 이를 위해 15명의 화자로부터 획득된 총 875개의 동영상에서 525개의 영상에 대해 특징 벡터를 추출하여 해당하는 HMM의 구현을 위한 학습용 데이터로 사용하였다. 또한 나머지 350개의 동영상은 검사용 특징 벡터를 추출하기 위해 사용된다. HMM의 각 파라메터 값들은 Baum-Welch 알고리듬을 사용하여 추정되었으며, 학습된 HMM에 기반한 검사용 데이터에 대한 인식 실험은 Viterbi 알고리듬을 이용하여 수행된다. 그림 7은 5개의 시공간적 특징 벡터에 대한 인식 실험 결과를 나타낸다.

입술독해에 관련된 표준 데이터베이스가 아직 구축되어 있지 않음으로 인해 본 논문에서 제안한 특징 벡터의 인식 성능을 객관적으로 평가하는 데는 어려움이 있으나, 입술독해에 관한 기존의 연구사례[3-6]에서 나타난 30~60%의 인식률과 비교하여 볼 때 본 논문에서 제안한 특징 벡터 $\mathbf{f}_c$와 $\mathbf{f}_d$의 성능이 상대적으로 우수함을 그림 7의 결과에서 확인할 수 있다. 또한 그림 7의 결과에서 입술의 안과 바깥 윤곽선에서 추출한 특징 벡터 $\mathbf{f}_b$가 안쪽 윤곽선에서만 추출한 $\mathbf{f}_a$의 결과보다 약 7%정도 우수한 성능을 나타낸다. 이와 같은 결과는 입술의 안과 바깥 윤곽선에서 추출된 특징이 입술의 시각적인 정보를 적절히 나타내기 위해 상호 보완적으로 모두 중요하게 사용됨을 의미한다. 그러나 특징 벡터 $\mathbf{f}_a$와 $\mathbf{f}_b$는 발성 과정에서 입술의 벌어진 높이와 폭만을 나타내므로 그 크기가 유사한 모음들에 대하여 오인식이 다수 발생한다. 특히, /ㅏ/, /ㅐ/, /ㅓ/모음들과 /ㅡ/, /ㅣ/ 모음에서 많은 인식 오류가 발생함을 확인할 수 있었다.

특징벡터 $\mathbf{f}_c$는 입술 윤곽선 상에서 검출된 특징점들의 시간에 따른 모음마다의 고유한 이동 방향과 크기를 나타

표 1. 특징 벡터 $\mathbf{f}_c(\mathbf{f}_b)$의 인식 실험 결과에 대한 혼동표
Table 1. Confusion matrix for recognition result of feature vector $\mathbf{f}_c(\mathbf{f}_b)$.

모음	/ㅏ/	/ㅐ/	/ㅓ/	/ㅗ/	/ㅜ/	/ㅡ/	/ㅣ/	인식률(%)
/ㅏ/	32 (19)	12 (14)	0 (8)	0 (0)	0 (0)	2 (3)	4 (6)	64.0 (38.0)
/ㅐ/	10 (9)	26 (21)	4 (7)	0 (0)	0 (0)	2 (2)	8 (11)	52.0 (42.0)
/ㅓ/	4 (8)	4 (8)	34 (16)	0 (0)	0 (0)	6 (11)	2 (7)	68.0 (32.0)
/ㅗ/	0 (0)	0 (0)	0 (4)	31 (28)	18 (18)	1 (0)	0 (0)	62.0 (56.0)
/ㅜ/	0 (0)	0 (0)	0 (0)	13 (13)	36 (36)	1 (1)	0 (0)	72.0 (72.0)
/ㅡ/	0 (0)	2 (1)	6 (10)	0 (0)	0 (2)	40 (25)	2 (12)	80.0 (50.0)
/ㅣ/	4 (0)	9 (6)	1 (10)	1 (0)	0 (0)	13 (13)	22 (21)	44.0 (42.0)
평균								63.1 (47.4)

내며 다른 특징벡터와 비교하여 가장 우수한 인식 성능을 나타낸다. 또한 표 1에 나타낸 f_c와 f_b의 인식 결과 혼동표 (confusion matrix)를 살펴보면 f_b에서 다수 발생되는 /ㅏ/, /ㅐ/, /ㅓ/모음과 /ㅡ/, /ㅣ/ 에서의 인식 오류가 f_c의 경우에 현저히 감소함을 알 수 있다. 입술의 움직임 속도를 나타내는 특징 벡터 f_e의 경우에는 동일한 음에 대해 화자마다 그 발음 속도가 차이가 나며 동일한 화자에 대해서도 화자의 감정 상태에 따라서 그 변화가 크게 나타나므로 가장 낮은 인식 성능을 보여준다. 결과적으로 한국어 모음 viseme에 대한 인식에는 입술 움직임의 이동 방향과 크기를 반영한 특징벡터 f_c가 가장 적절하게 사용될 수 있음을 확인하였다.

V. 결 론

본 논문에서는 한국어 발성 시 입술 모양을 결정하는데 핵심이 되는 각 모음별 시각적 특성을 음성학에 기반하여 분석함으로써 한국어 입술독해에 적합한 음성의 시각적 기본 단위인 viseme을 정의하였다. 또한 입술 윤곽선 상의 특징점을 검출하고 동영상에서의 이들 특징점으로부터 입술의 안팎 높이 및 폭, 그리고 입술의 움직임 궤적, 크기, 및 방향 등의 정보를 포함하는 시공간적 특징 벡터들을 추출하여 각 viseme에 대한 입술독해 실험을 수행하였다. 이산 HMM에 기반한 인식 실험 결과로부터 입술의 안과 바깥 윤곽선의 정보가 입술의 시각적인 정보를 적절히 나타내기 위해 상호 보완적으로 모두 중요하게 사용됨을 확인하였다. 또한 시간에 따른 입술 움직임의 이동 방향과 크기를 나타내는 특징 벡터가 한국어 모음 viseme에 대한 인식에 가장 적절하게 사용될 수 있음을 확인하였다.

본 논문에서는 특징 벡터의 추출과 성능 검증의 신뢰성을 위하여 입술 특징점을 수동으로 검출하였으나 향후 연구에서는 이를 자동으로 검출함으로써 자동 입술독해 시스템을 구축하고 입술독해 기술이 결합된 바이모달 음성인식 시스템을 구현하여 잡음 환경하에서의 음성인식 성능의 향상을 검증하고자 한다.

감사의 글

본 연구는 한국과학재단 목적기초연구 (R01-1999-00233) 지원으로 수행되었습니다.

참 고 문 헌

1. H. Kaplan, C. J. Bally and C. Garretson, *Speechreading: A Way to Improve Understanding*, Gallaudet University Press, Washington D.C., 1999.
2. H. McGurk and J. MacDonald, "Hearing Lips and Seeing Voices," Nature 264, pp. 746-748, 1976.
3. P. Duchnowski, M. Hunke, D. Busching, U. Meier and A. Waibel, "Toward Movement-invariant Automatic Lip-reading and Speech Recognition," *IEEE Proc. In. Conf. Acoustics, Speech and Signal Processing*, 1, pp. 109-112, 1995.
4. G. Potamianos, H. P. Graf and E. Cosatto, "An Image Transform Approach for HMM Based Automatic Lipreading," *IEEE In. Conf., Image Processing*, 3, Chicago, USA, pp. 173-177, 1998.
5. M.E. Hennecke, K.V. Prasad and D.G. Stork, "Automatic Speech Recognition System Using Acoustic and Visual Signals," *Proc. of 29th Asilomar Conf. Signals, Systems and Computers*, 2, pp. 1214-1218, 1995.
6. 박병구, 김진영, 임재열, "입술 파라미터 선정에 따른 바이모달 음성인식 성능 비교 및 검증," 한국음향학회지 제18권 제3호, pp. 68-72, 1999.
7. A. J. Goldschen, O. N. Garcia and E. Petajan, "Continuous Optical Automatic Speech Recognition by Lipreading," *Proc. of 28th Asilomar Conf. Signals, Systems and Computers*, 1994.
8. A. Rozen and P. Deleglise, "Visible Speech Modeling and Hybrid Markov Models/Neural Networks Based Learning for Lipreading," *IEEE In. Joint Symposia on Intelligence and Systems*, pp. 336-342, 1998.
9. 구현옥, 국어 음운학의 이해, 한국문화사, 1999.
10. 김영송, 우리말 소리의 연구, 샘문화사, 4, 1975.

저자 약력

● 오 현 화 (Hyun Hwa Oh)

1998년 2월: 경북대학교 전자공학과 (공학사)
2000년 2월: 경북대학교 대학원 전자공학과 (공학석사)
2000년 3월~현재: 경북대학교 대학원 전자공학과 (박사과정)
※ 주관심분야: 컴퓨터비전, 패턴인식, 입술독해, 바이모달 음성인식

● 김 인 철 (In Cheol Kim)
한국음향학회지 제19권 제2호 참조

● 김 동 수 (Dong Su Kim)

1992년 2월: 경북대학교 전자공학과 (공학과)
1996년 2월: 경북대학교 대학원 전자공학과 (공학박사)
1996년 3월~현재: 경북대학교 대학원 전자공학과 (박사과정)
※ 주관심분야: 컴퓨터비전, 패턴인식, 바이모달 음성인식

● 진 성 일 (Chien Sung Il)
한국음향학회지 제19권 제2호 참조