바이모달 음성인식의 음성정보와 입술정보 결합방법 비교

Comparison of Integration Methods of Speech and Lip Information in the Bi-modal Speech Recognition

박 병 구*, 김 진 영*, 최 승 호**
(Byung Ku Park*, Jin Young Kim*, Seung Ho Choi**)

※ 이 논문은 한국과학재단의 '98 핵심전문연구 지원에 의해 이루어진 연구결과물 중 하나입니다.

요 약

잡음환경에서 음성인식 시스템의 성능을 향상시키기 위해서 영상정보와 음성정보를 이용한 바이모달(bimodal)음성인식이 제안되어왔다. 영상정보와 음성정보의 결합방식에는 크게 분류하여 인식 전 결합방식과 인식 후 결합방식이 있다. 인식 전 결합방식에서는 고정된 입술파라미터 중요도를 이용한 결합방법과 음성의 신호 대 잡음비 정보에 따라 가변 입술파라미터 중요도를 이용하여 결합하는 방법을 비교하였고, 인식 후 결합방식에서는 영상정보와 음성정보를 독립적으로 결합하는 방법, 음성 최소거리 경로정보를 영상인식에 이용 결합하는 방법, 영상 최소거리 경로정보를 음성인식에 이용 결합하는 방법, 그리고 음성의 신호 대 잡음비 정보를 이용하여 결합하는 방법을 비교했다. 6가지 결합방법 중 인식 전 결합방법인 파라미터 중요도를 이용한 결합방법이 가장 좋은 인식결과를 보였다.

ABSTRACT

A bimodal speech recognition using visual and audio information has been proposed and researched to improve the performance of ASR(Automatic Speech Recognition) system in noisy environments. The integration method of two modalities can be usually classified into an early integration and a late integration. The early integration method includes a method using a fixed weight of lip parameters and a method using a variable weight according to speech SNR information. The 4 late integration methods are a method using audio and visual information independently, a method using speech optimal path, a method using lip optimal path and a way using speech SNR information. Among these 6 methods, the method using the fixed weight of lip parameters showed a better recognition rate.

I. 서 론

기존의 사용하는 키보드와 마우스의 인터페이스에 한계를 느끼는 부분이 많이 생기면서 좀 더 편하게 그리고 손쉽게 기계와 인터페이스를 할 수 있는 방법에 대한 연구가 끊임없이 진행되어오고 있다. 이러한 문제의 해결책으로 얼굴표정, 몸 동작, 사용자 응시방향, 입술모양 등의 정보를 함께 이용하는 멀티모달(mutimodal)인식 분야가 활발히 모색되고 있다[1][2]. 특히, 음성인식에 있어서 잡음환경에서 견인한 음성인식시스템을 위해서 음성신호뿐만 아니라, 음성신호와 가장 밀접한 관계를 가지고 있는 입술정보를 이용하는 바이모달(bimodal) 음성인식이 연구되어왔다[3][4]. 이러한 입술정보를 이용한 바이모달 음성인식에서 크게 고려해야 될 부분은 두 입력에서 받아들인 정보를 어떻게 결합하느냐에 있다. 결합방법은 인식 전 결합방식(early integration)과 인식 후 결합방식(late integration)으로 크게 분류할 수 있다[5]. 본 논문에서는 여러 결합방법들을 비교 검토하고자 한다. 인식 전 결합방법으로 고정된 파라미터 중요도를 이용한 결합방법과 음성의 SNR정보에 따라 파라미터 중요도를 변화시키면서 결합하는 방법을 비교하였고, 인식 후 결합방법으로 영상정보와 음성정보를 시각 가중치를 이용하여 독립적으로 결합하는 방법, 음성 최소거리 경로 정보를 이용하여 결합하는 방법, 영상의 최소거리 경로 정보를 이용하여 결합하는 방법, 그리고 음성의 SNR 정보를 이용하여 결합하는 방법에 대해 설명하고자 한다.

본 논문의 구성은 2장에서 시스템구성 및 음성정보 및 영상정보 추출방법과 두 입력정보의 동기화과정을 설명하고 3장에서는 입술정보와 음성정보의 결합방법으로 6가지 방법을 비교하였고, 4장에서는 각각의 방법을 인식실험을 통하여 성능을 비교하였다.

* 전남대학교 공과대학 전자공학과
** 동신대학교 정보통신공학과
접수일자 : 1999년 1월 5일

II. 음성과 입술정보 추출 및 동기화

입술영상과 음성을 동시에 저장하기 위해 두 대의 컴퓨터를 사용하여 그림 1과 같이 구성하였다. 마스터컴퓨터(Master Computer)에서는 입술이미지를 18프레임/1초의 속도로 100×100크기의 입술이미지를 받아들이고 슬레이브컴퓨터(Slave Computer)에서는 마스터컴퓨터에서 FX케이블을 통해서 보내온 동기신호를 이용해서 동시에 음성신호를 저장하도록 구성하였다. 이미지보드는 Oculus-Tcx 보드를 이용해서 CCTV 컬러카메라인 TMC-7으로부터 이미지를 받아서 저장하고 음성은 ASPI(Atlanta Signal Processors, Inc.)회사의 TMS320C31 DSP칩을 내장한 ELF보드를 이용해서 음성을 저장하도록 구성하였다.

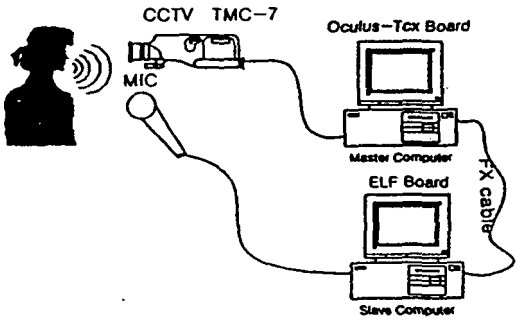

그림 1. 바이모달 음성인식 시스템 구성도
Fig. 1. Bimodal Speech recognition System.

음성 파라미터로 12차 LPC(Linear Predictive Coding) 캡스트럼(Cepstrum) 계수를 이용하였다. 영상과 동기를 맞추면서 8kHz로 녹음된 음성신호를 프레임 크기를 256 샘플로 해석하여 100샘플씩 이동하면서 해밍(Hamming) 윈도우를 사용하였다[6].

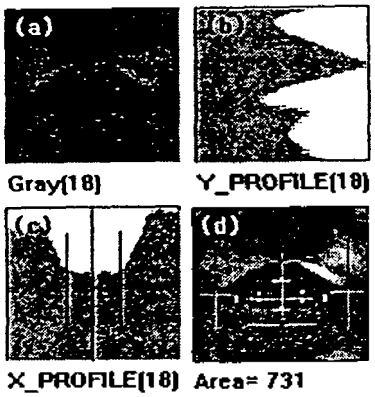

그림 2. 입술파라미터가 추출된 모습
Fig. 2. Lip parameter extraction from gray image.

입술 파라미터 추출 방법으로 이미지의 색상에 근거하여 입술모양을 4개의 파라미터로 만들어서 사용하였다[7][8]. 4개의 파라미터로 바깥 입술의 높이와 폭, 안쪽 입술의 높이와 폭을 이용하였다[9][10]. 실제로 입술파라미터가 추출된 모습은 그림 2.(d)와 같다. 그림 2.(a)는 흑백영상으로 전환된 이미지이고 그림 2.(b)는 바깥입술의 높이를 측정하기 위해서 사용한 y 프로파일(y축 이미지 평균값)이고 그림 2.(c)는 안쪽입술의 폭을 측정하기 위해서 사용한 x 프로파일(x축 이미지 평균값)이다[11][12].

이렇게 추출된 음성 파라미터와 입술 파라미터는 동기화 과정이 필요하다. 카메라와 마이크를 통해서 받은 두 입력의 동기화가 이루어지지 않으면 인식률에 영향을 크게 미치므로 정확한 동기화가 요구된다. 그림 3은 지역이름인 '강화'를 발음한 것으로 입술파라미터와 음성파라미터의 동기화를 나타낸 것이다. 그림 3.(a)와 그림 3.(b)는 추출된 파라미터를 이용하여 입술모양을 나타낸 것이고 그림 3.(c)는 바깥 입술의 높이와 안쪽 입술의 높이의 변화를 각각의 영상프레임별 변화를 나타낸 것이고 그림 3.(d)는 바깥 입술의 폭과 안쪽 입술의 폭의 변화를 각각의 영상 프레임별 변화를 나타내고 그림 3.(e)는 음성 파형을 나타낸 것이다. 그림 3.(c)에서 입술의 바깥입술의 높이(바깥 두 선의 차이)와 안쪽입술의 높이(안쪽 두 선의 차이)를 구할 수 있고 그림 3.(d)에서 입술의 바깥입술의 폭(바깥 두 선의 차이)과 안쪽입술의 폭(안쪽 두 선의 차이)을 구할 수 있다. 우선 음성구간을 에너지(Energy)를 이용하여 구한 다음 이 음성구간에 해당하는 구간만큼 입술의 처음 프레임과 끝 프레임을 추출하였다. 음성과 입술영상의 상호 관계를 살펴보기 위한 실험이므로 입술파라미터를 독립적으로 구하지 않고 음성구간을 이용하여 음성과 영상의 동기화를 맞추었다.

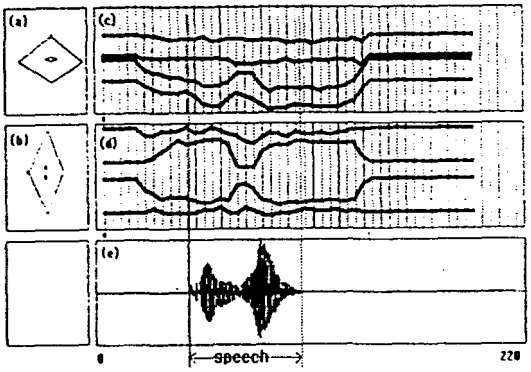

그림 3. 영상과 음성의 동기화 과정
Fig. 3. Synchronization of image and speech signals.

III. 입술정보와 음성정보의 결합 방법

동기화과정을 거쳐서 추출된 음성 파라미터와 입술 파라미터의 결합은 크게 두 가지로 구분할 수 있다. 추출된

음성 파라미터와 입술파라미터를 인식과정 이전에 결합하여 하나의 파라미터 벡터를 만든 다음 인식과정을 수행하는 방법인 인식 전 결합방법과 추출된 음성 파라미터와 입술파라미터를 각각의 인식기에 인식과정을 수행한 다음 각각의 인식결과로 나온 인식스코어(Score)에 가중치를 이용하여 결합하는 방법인 인식 후 결합방법(late integration)이 있다[13][14][15].

1. 인식 전 결합방법

인식 전 결합방법의 구조는 그림 4와 같고 두 가지 방법을 실험하였다. 인식 전 결합 방법으로 고정된 파라미터 중요도를 이용한 결합방법과 음성의 SNR 정보를 이용하여 결합하는 방법이 있다.

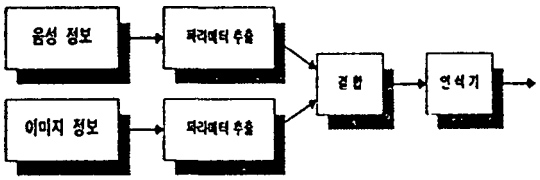

그림 4. 인식 전 결합방법
Fig. 4. Early integration method.

인식 전 결합방법은 인식과정을 거치기 이전에 입술 파라미터와 음성 파라미터를 결합한 후 인식과정을 수행하는 방법이다. 입술 파라미터로 바깥 입술의 높이와 폭, 안쪽 입술의 높이와 폭, 4개의 파라미터를 이용하고, 음성파라미터로 12차 LPC(Linear Predictive Coding) 캡스트럼(Cepstrum) 계수를 이용하여 16개의 영상과 음성 정보가 결합된 파라미터를 계산할 수 있다. 음성프레임과 입술영상프레임의 차이는 입술영상프레임을 내삽(interpolation) 과정을 거쳐서 음성프레임 수에 맞춰서 그림 5와 같이 만들어 낼 수 있다.

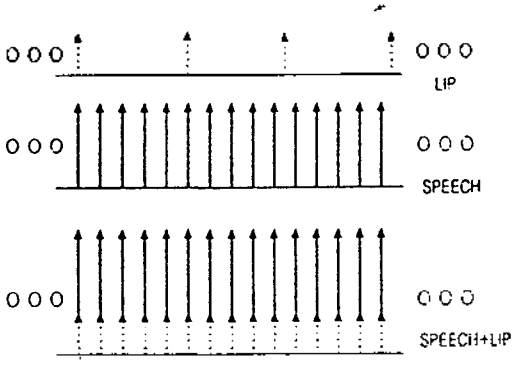

그림 5. 입술파라미터와 음성파라미터 결합
Fig. 5. Integration of Lip and speech parameters.

1.1 고정된 파라미터 중요도를 이용한 결합방법

합쳐진 음성 파라미터와 입술 파라미터의 중요도를 달리하면서 결합하는 방법이다. 그림 6은 한 단어에 대해서 DTW(Dynamic Time Warping)방법을 이용하여 인식 실험한 과정을 나타내고 있다. 인식실험과정에서 파라미터 중요도(λ)를 변화를 시켜가면서 가장 좋은 인식률을 나타내는 인식률을 선택하였다.

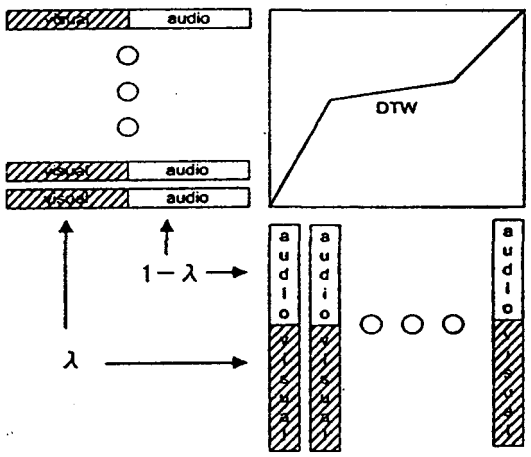

그림 6. 파라미터 중요도(λ)를 이용한 결합방법
Fig. 6. Integration using weighting factor(λ).

1.2. 음성의 SNR정보에 따른 가변 파라미터 중요도를 이용한 결합방법

음성의 SNR 정보를 이용하여 잡음이 많이 섞인 음성 프레임은 입술 파라미터의 중요도(λ)를 크게 하고, 깨끗한 음성 프레임에서는 음성 파라미터의 중요도(1-λ)를 크게 하여 결합하는 방법이다.

그런데 여기서 먼저 고려해야 될 사항은 각각의 음성의 SNR 정도에 따라서 얼마의 파라미터 중요도를 줄 것인가의 문제가 발생한다. 그래서 음성의 SNR 정보에 따른 기준 시각 파라미터 중요도를 계산하기 위해서 먼저 음성신호에 SNR이 일정하도록 백색 가우시안(Gaussian) 잡음을 섞은 음성신호를 만든다. 이렇게 해서 만들어진 잡음이 일정하게 섞인 음성신호를 이용해서 파라미터 중요도를 달리하면서 인식실험을 하여 각각의 SNR에서 가장 인식률이 좋게 나오는 시각 파라미터 중요도 값을 그림 7과 같이 계산해 낼 수 있었다.

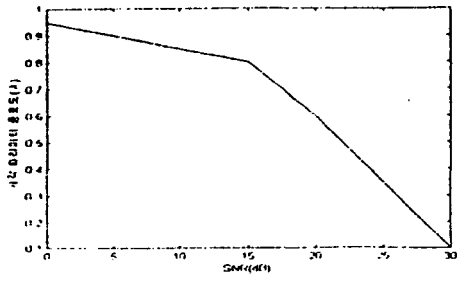

그림 7. 기준 시각 파라미터 중요도(λ)
Fig. 7. Visual parameter weighting value with SNR.

이러한 기준 시각 파라미터 중요도 값을 이용하여 음성 및 입술 파라미터 벡터에서 각각의 음성의 SNR 정보를 이용하여 각각의 음성 프레임마다 시각 파라미터의 중요도를 달리하면서 결합한 모습은 그림 8과 같다.

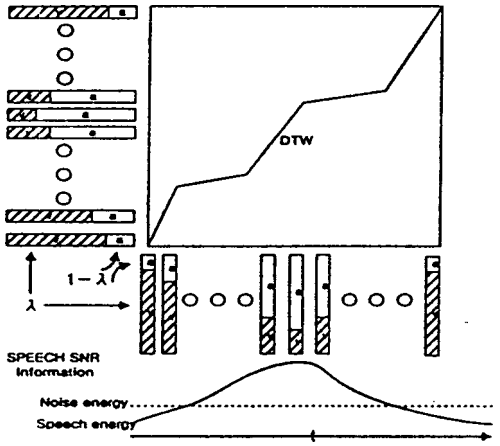

그림 8. 음성의 SNR 정보를 이용한 결합방법
Fig. 8. Integration using speech SNR information.

2. 인식 후 결합방법

인식 후 결합방법(late integration)의 구조는 그림 9와 같고 실험한 방법은 4가지 방법이 있다. 첫 번째 방법은 음성정보와 입술정보를 독립적으로 결합하는 방법, 두 번째 방법은 음성의 최소거리 경로정보를 이용한 방법, 세 번째 방법은 영상의 최소거리 경로정보를 이용한 방법 그리고 마지막 방법으로 음성의 SNR 정보를 이용하여 결합하는 방법을 실험하였다.

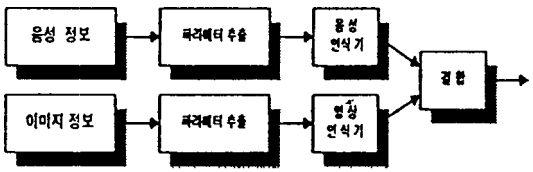

그림 9. 인식 후 결합 방법
Fig. 9. Late integration method.

2.1 영상정보와 음성정보를 독립적으로 결합

그림 10은 인식스코어를 시각스코어와 음성스코어를 가중치를 이용하여 독립적으로 결합하는 방법을 나타내고 식1과 같이 표현이 가능하다. 인식 전 결합방법은 모든 파라미터 정보를 결합하므로 인식과정에서 인식정보의 손실이 적으나 파라미터 중요도에 따라서 최소거리가 달라지므로 모든 경우에 대해서 인식과정을 수행하다보면 인식시간이 인식 후 결합방법과 비교할 때 많이 걸린다는 단점이 있다.

영상은 1초에 18프레임을 획득하므로 좌측의 영상프레임이 우측의 음성프레임보다는 적게 나타남을 알 수 있다. 각각의 DTW방법을 이용하여 결합한 스코어를 아래 식 1에 의해 가중치를 주어서 독립적으로 결합하였다.

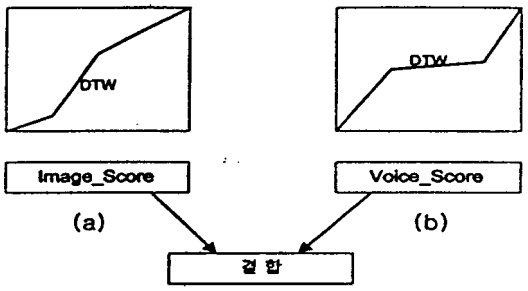

그림 10. 영상정보와 음성정보를 독립적으로 결합
Fig. 10. Independent integration of image and speech informations.

$$S = \alpha S_v + (1-\alpha) S_a \quad (1)$$

S : 인식스코어

S_v : 시각스코어, S_a : 음성스코어

α : 시각가중치

2.2 음성 경로정보를 영상인식에 이용하여 결합

음성의 기준패턴과 테스트패턴과의 비교를 DTW를 통해서 그림 11.(b)와 같이 최소거리 경로를 구할 수 있는데, 이 최소거리 경로정보를 저장하여 그림 11.(a)에서 보듯이 영상인식 때 이 음성경로를 따라가게 하는 방법이다. 그림 11.(a)를 보면 먼저 점선으로 음성의 경로가 그려져 있고 그 경로 위에 큰 점이 그려져 있다. 큰 점은 각 이미지 프레임별로 음성 최소거리 경로에 해당되는 위치에 표시를 하였고 식 2와 같이 표현된다. 만약 기준패턴과 테스트패턴이 같은 발음이라면 최소거리 경로가 음성과 영상이 비슷한 경로를 갖게 되므로 이때 더 신뢰성이 높은 정보를 선택하여 최소거리 경로를 이용하려고 하였다.

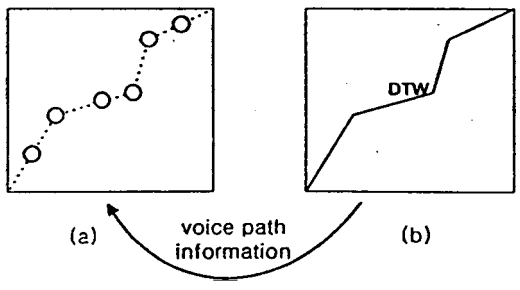

그림 11. 음성 최소거리 경로정보를 영상인식에 이용하는 방법
Fig. 11. Integration method using speech optimal path.

$$S = \alpha S_{v\cdots a} + (1-\alpha) S_a \quad (2)$$

S : 인식스코어

$S_{v\cdots a}$: 음성의 경로정보를 이용한 시각스코어

S_a : 음성스코어

α : 시각가중치

2.3 영상 경로정보를 음성인식에 이용하여 결합

그림 12.(a)와 같이 영상 파라미터를 DTW방법을 이용하여 최소거리경로를 찾고 나서, 영상의 최소거리경로를 저장한 후 음성 파라미터를 DTW방법을 이용하여 비교할 때 그림 12.(b)에서 보여지듯이 영상의 최소거리경로를 따라가게 하는 방법이다. 이전 결합방법인 음성 최소거리 경로정보를 이용한 결합방식에서는 음성의 프레임수가 많고 영상의 프레임수가 적기 때문에 세밀한 음성 최소 경로정보를 이용할 때 영상 부분에서 문제가 없었으나 이번 결합방식은 영상 프레임수가 적으므로 바로 영상 최소거리 경로를 이용할 수 없는 문제점이 생긴다. 이 문제점을 해결하기 위해서 음성 최소거리 경로정보를 저장한 후 그림 12.(b)의 그림과 같이 음성인식부분에서 이 경로를 이용하되 그림 12.(b)에서 점선과 같이 영상경로에 여유를 두도록 하였다. 영상의 최소거리경로에서 제한된 여유경로(음성의 10프레임)를 두어서 그 한정된 경로에서 DTW방법을 이용 음성최소거리경로를 찾도록 하는 방법이며 식 3과 같이 표현된다.

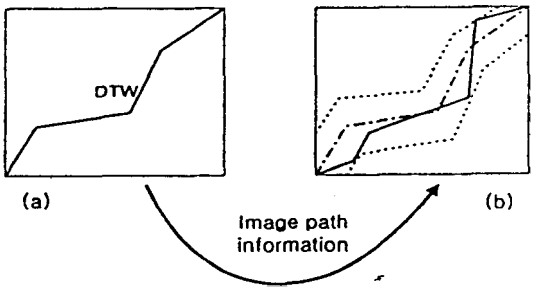

그림 12. 영상 최소거리 경로정보를 음성인식에 이용하는 방법
Fig. 12. Integration method using image optimal path.

$$S = \alpha S_v + (1-\alpha) S_{a\cdots v} \quad (3)$$

S : 인식스코어

S_v : 시각스코어

$S_{a\cdots v}$: 영상의 경로정보를 이용한 시각스코어

α : 시각가중치

2.4 음성의 SNR 정보를 이용한 결합

음성의 SNR정보를 이용한 결합방법은 음성정보와 입술정보를 결합하는 과정에서 음성의 SNR을 보고 음성에 잡음이 많이 섞여 있으면 시각 가중치(α)를 크게 하고 음성이 깨끗하면 음성 가중치($1-\alpha$)를 크게 하면서 각 프레임을 결합하는 방법이다. 인식 전 결합방식에서 음성과 입술 파라미터가 결합되어서 파라미터 가중치를 달리하면서 결합하면 되지만 이 경우는 파라미터가 각기 따로 존재한다. 그래서 2.2절의 음성 최소거리 경로정보를 이용한 결합방법과 비슷하게 우선 음성부분에서 DTW방법을 이용하여 음성의 최소거리 경로를 찾고 찾아진 경로정보에 따라서 입술 파라미터를 비교하도록 한다. 이 과정에서 음성의 SNR 정보를 이용하여 시각 가중치를 달리하면서 음성정보와 입술정보를 결합한다. 그림 13는 이러한 과정을 나타내었다. 그림에서처럼 시각 가중치를 계산하기 위해서 음성 파라미터에 삼각윈도우(식 4)를 이용하여 평균 음성의 SNR(식 5)를 구하고 그에 해당하는 시각 가중치를 구하도록 하였다.

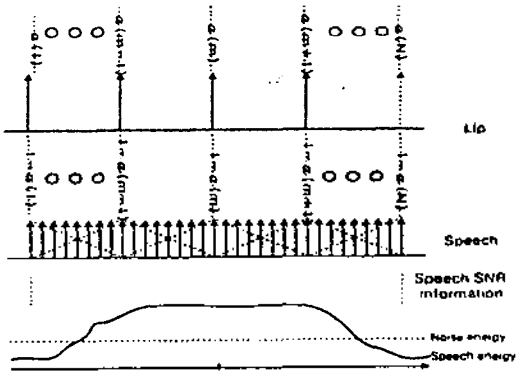

그림 13. 음성의 신호 대 잡음비 정보를 이용한 결합 알고리듬
Fig. 13. Integration using speech SNR information.

$$W_i = \begin{cases} \dfrac{2i}{n-1} & ; \quad 0 \le i \le \dfrac{n-1}{2} \\ 2 - \dfrac{2i}{n-1} & ; \quad \dfrac{n-1}{2} \le i \le n-1 \\ 0 & ; \quad otherwise \end{cases} \quad (4)$$

$$\text{평균음성} SNR_k = 10 \log \frac{\sum_{i=1}^{n} 10^{\frac{SNR_i}{10}} * W_i}{\sum_{i=1}^{n} W_i} \quad (5)$$

k : 영상프레임

n : 삼각 윈도우 샘플 수

이 결합방법도 인식 전 결합방법과 마찬가지로 음성의 SNR에 따라서 얼마의 시각 가중치를 줄 것인가의 기준 시각 가중치 값이 필요하므로 SNR이 일정하도록 백색 가우시안(Gaussian)잡음을 섞은 신호에서 각 잡음별로 가장 좋은 인식률을 나타내는 기준 시각 가중치 값들을 2.2절의 입술 최소거리 경로정보를 이용한 결합방법을 이용하여 구하면 그림 14와 같이 실험 값을 얻을 수 있다. 이렇게 구해진 값을 이용하여 음성과 입술정보를 결합할 때 각각의 프레임마다 음성의 SNR 값에 따라서 시각 가중치를 달리하여 결합한다.

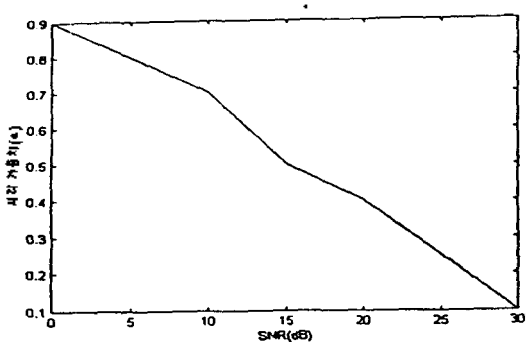

그림 14. 기준 시각 가중치 값(a)
Fig. 14. Reference visual weighting factor(a).

IV. 실험 및 결과

실험 데이터로 단일화자에 대해서 시외 전화 지역 이름 160개 자명을 이용하여 160×4=640개의 데이터를 만들고 그중 160개를 기준 패턴으로 하고 나머지 160×3=480개의 데이터를 테스트 패턴으로 하여 DTW(Dynamic Time Warping) 패턴 비교를 하였고 화자종속 고립단어 인식실험을 수행하였다. 음성은 8kHz로 저장되었고 입술영상은 18frame/1sec로 저장되었다. 먼저 인식 전 결합 방법(early integration)으로 고정된 파라미터 중요도(λ)를 이용한 결합 방법과 음성의 SNR 정보에 따라 가변 파라미터 중요도를 이용한 결합방법의 실험결과를 그림 15에 나타냈다. 음성의 SNR 정보에 따라 가변 파라미터 중요도를 이용한 결과가 전체적으로 인식률이 고정된 파라미터 가중치를 이용하여 결합한 방법보다 1.04-7.71%정도 낮게 나왔다.

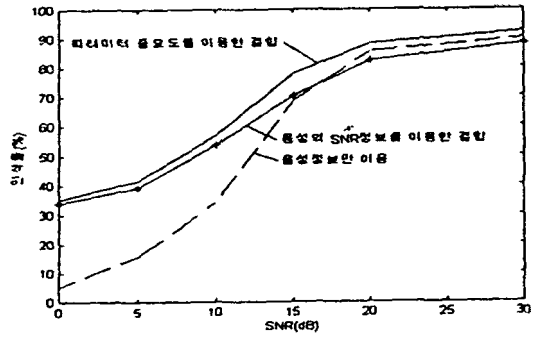

그림 15. 인식 전 결합방법의 인식률
Fig. 15. Recognition rate using early interation method.

인식 후 결합 방법(late integration)으로 음성 정보와 입술 정보를 독립적으로 결합한 방법, 음성 경로정보를 이용한 방법, 영상 경로 정보를 이용한 방법 그리고 음성의 SNR 정보를 이용한 방법을 실험하였고 그림 16에서 보는바와 같이 음성 정보와 입술 정보를 독립적으로 결합한 방법이 인식률이 가장 좋게 나왔다.

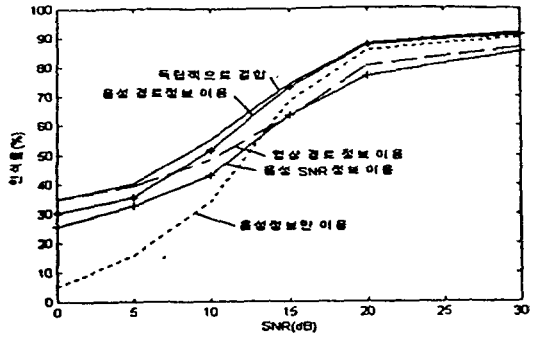

그림 16. 인식 후 결합방법의 인식률
Fig. 16. Recognition rate using late integration method.

마지막으로 인식 전 결합 방법과 인식 후 결합방법에서 가장 좋은 결과를 나타낸 고정된 파라미터 중요도(λ)를 이용한 결합방법과 음성 정보와 입술 정보를 시각 가중치를 독립적으로 이용하여 결합한 방법의 실험 결과를 그림 17에 나타내었다. 전체적으로 인식 전 결합 방법이 인식 후 결합 방법보다 인식률이 더 좋게 나타났고 5dB 이하에서는 인식 후 결합 방법이 약간 높은 인식률을 보였으나 거의 비슷하였다.

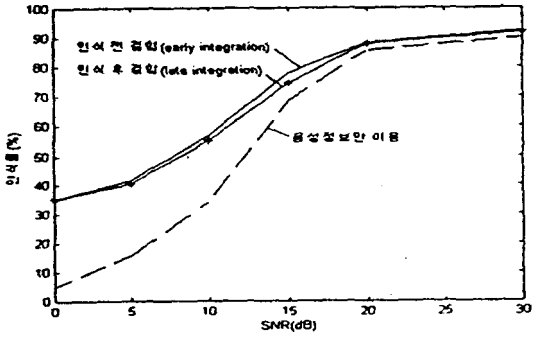

그림 17. 인식 전 결합방법과 인식 후 결합방법
Fig. 17. Comparisons of early and late integration methods.

바이모달 음성인식 실험을 통하여 인식 전 결합방법과 인식 후 결합방법을 비교하면 다음과 같다. 인식 전 결합 방법은 모든 파라미터 정보를 결합하므로 인식과정에서 인식정보의 손실이 적으나 파라미터 중요도에 따라서 최소거리 경로가 달라지므로 모든 경우에 대해서 인식과정을 수행하여야 하므로 인식시간이 많이 걸린다는 단점이 있다. 반면, 인식 후 결합방법은 음성 파라미터와 입술 파라미터가 각각 존재하므로 각각의 파라미터 벡터가 크기가 작고 인식스코어를 시각 가중치를 달리하면서 결합하는 방법이므로 인식시간이 적게 걸린다는 장점이 있으나 음성과 입술의 인식률을 통하여 결과로 나온 인식스코어를 결합하므로 인식 전 결합과정과 비교할 때 결합과정에서 정보의 손실이 있다는 단점이 있다.

V. 결 론

본 논문에서는 바이모달(bimodal) 음성인식에서 입술정보와 음성정보를 결합하는 6가지 방식을 비교하고 실험을 통하여 성능을 비교해 보았다. 인식 전 결합(early integration) 방식으로 2가지와 인식 후 결합(late integration)방법으로 4가지를 구현하였다. 인식 전 결합방법에서는 파라미터 중요도를 이용한 결합방법이 좋은 인식률을 보였고, 인식 후 결합방법에서는 입술정보와 음성정보를 독립적으로 결합하는 방법이 4가지 방법 중 가장 좋은 인식률을 보였다. 그리고 인식 전 결합방법에서 파라미터 중요도를 이용한 결합방법과 인식 후 결합방법에서 독립적으로 결합하는 방법을 비교하면 고정된 파라미터 중요도를 이용한 결합방법(인식 전 결합방법)이 전체적으로 좋은 인식률을 보였고 5dB이하에서만 입술정보와 음성정보를 독립적으로 결합한 방법(인식 후 결합방법)이 좋은 인식률을 보였다.

참 고 문 헌

1. Ronald A. Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, Victor Zue, "Survey of the State of the Art in Human Language Technology," Center for Spoken Language Understanding, Oregon Graduate Institute, p.329-362, 1995.
2. Rajeev Sharma, Vladimir I, Pavlovic, Thomas S. Huang, "Toward Multimodal Human-Computer Interface," p.853-869, IEEE, 1998.
3. Paul Duchnowski, Martin Hunke, Dietrich Büsching, Uwe Meier and Alex Waibel, "Toward Movement-Invariant Automatic Lip-Reading and Speech Recognition," Processing of ICASSP '95, 1995.
4. A. Waibel and P. Duchnowski, "Connectionist Models in Multimodal Human-Computer Interaction," Proceedings of the Government Microcircuit Applications Conference(GOMAC), San Diego, 1994.
5. Peter L. Silsbee and Alan C. Bovik, "Computer Lipreading for Improved Accuracy in Automatic Speech Recognition," IEEE Trans. on Speech and Audio Processing, Vol. 4, No. 5, pp 337-351, 1996.
6. Lawrence Rabiner, Biing-Hwang Juang "Fundamentals of Speech Recognition," PTR Prentice-Hall, 1993.
7. Rao, R.R. & Mersereau, "Lip Modeling for Visual Speech Recognition," 28th Annual Asilomar Conference on Signals, Systems, and Computers, 1994.
8. Marcus E. Hennecke, K. Venkatesh Prasad, David G. Stork, "Using Deformable Templates to Infer Visual Speech Dynamics," CRC-TR-9430, 1994.
9. M. Vogt, "Interpreted Multi-State Lip Models for Audio-Visual Speech Recognition," Proceedings of the AVSP'97, ESCA ISSN #10184554, 1997.
10. K. Venkatesh Prasad, David G. Stork and Gregory J. Wolff, "Preprocessing video images for neural learning of lipreading," Ricoh California Research Center, Technical Report CRC-TR-93-26, 1993.
11. Juergen Luettin, Neil A. Thacker and Steve W. Beet, "Speechreading Using Shape and Intensity Information," Processings of the 4th International conference on Spoken Language processing, ICSLP'96 , 1996.
12. Earl Gose, Richard Johnsonbaugh, Steve Jost "Pattern recognition and Image analysis," Prentice Hall, 1996.
13. Tsuhan Chen and Ram R. Rao, "Audio-Visual Integration in Multimodal Communication," Proceedings of the IEEE, Vol. 86, No. 5, pp 837-852, May 1998.
14. Paul Duchnowski, Uwe Meier and Alex Waibel, "See Me, Hear Me: Integrating Automatic Speech Recognition and Lip-Reading," Proceedings of the International conference on Spoken Language Processing, Yokohama Japan, Setember 1994.
15. Silsbee, P. L., "Sensory Integration in Audiovisual Automatic Speech Recognition," 28th Annual Asilomar Conference on Signals, Systems, and Computers, 1994.

▲박 병 구(Byung Ku Park) 1971년 5월 19일생

1997년 2월 : 전남대학교 전자공학과(공학사)
1999년 2월 : 전남대학교 전자공학과(공학석사)
※주관심분야 : 음성인식 및 신호처리

▲김 진 영(Jin Young Kim) 1962년 4월 26일생

1986년 2월 : 서울대학교 전자공학과(공학사)
1988년 2월 : 서울대학교 전자공학과(공학석사)
1994년 8월 : 서울대학교 전자공학과(공학박사)
1993년 3월 ~ 1994년 12월 : 한국통신 소프트웨어연구소 전임연구원
1995년 ~ 현재 : 전남대학교 공과대학교 전자공학과 조교수
※주관심분야 : 음성인식 및 음성합성, 멀티모달 MMI

▲최 승 호(Seung Ho Choi) 1955년 8월 24일생

1981년 2월 : 전북대학교 물리학과(이학사)
1984년 8월 : 명지대학교 전자공학과(공학석사)
1992년 2월 : 명지대학교 전자공학과(공학박사)
1992년 3월 ~ 일재 : 동신대학교 정보통신공학과 교수
※주관심분야 : 음성인식, 멀티모달 MMI