웹에서의 저가 음성인식 시스템의 구현

박 용 범† · 박 종 일††

요 약

DTW 알고리즘을 이용한 고립단어 인식은 화자종속이라는 상황에 있어서는 좋은 인식 율을 제공하여 준다. 그러나 DTW 알고리즘은 검색해야 할 단어가 많을 경우 검색시간이 상대적으로 높아지게 되므로 현실적으로 적용하기가 힘들다. 웹에서의 교육용 학습 지와 같이 상황 의존적 단답형 질의 응답을 요구하는 시스템의 경우에 있어서는 주어진 질문에 대한 응답이 비교적 제한되어 있어 검색대상을 줄일 수 있다. 본 논문에서는 이와 같은 상황에서 사용할 수 있는 저가형 음성 인식기를 DTW로 구현하였다. DTW의 단점을 보완 하기위해 검색할 대상을 상황에 따라 줄이는 방법을 이용하였다. 질문에 따라 관심대상을 선정하여 이들만을 검색대상으로 삼았다. 실제적인 구현을 통하여 검색대상을 줄인 결과 높은 인식율을 얻을 수 있었고, 그룹을 실정한 만큼의 빠른 검색시간을 얻을 수 있었다.

The Low Cost Implementation of Speech Recognition System for the Web

Young-Bom Park† · Jong-Il Park††

ABSTRACT

Isolated word recognition using the Dynamic Time Warping algorithm has shown good recognition rate on speaker dependent environment. But, practically, since the searching time of the Dynamic Time Warping algorithm is rapidly increased as searching data is increased, it is hard to implement. In the context-dependent-short-query system such as educational children's workbook on the Web, the number of responses to the specific questions is limited. Therefore, the searching space for the answers can be reduced depending on the questions. In this paper, low cost implementation method using DTW for the Web has been proposed. To cover the weakness of DTW, the searching space is reduced by the context. The searching space, depends on the specific questions, is chosen from interest searchable candidates. In the real implementation, the proposed method shows better performance of both time and recognition rate.

1. 서 론

인터넷의 발달로 WWW은 상당히 보편화 되었고 웹을 이용한 여러 응용분야가 개발 되고 있다. 이들 분야 중 인터넷 학습지와 같이 사용자의 간단한 응답을 요구하는 단순 질의 응답 시스템을 이용하는 시스템이 있다. 이들 시스템은 대부분 정해진 상황에 의존적인 응답만을 처리하는 시스템들이다. 이러한 간단한 상황 의존적 질의 응답은 저가형 음성 인식 시스템을 도입하여 처리 하면 매우 효율적일 것이다. 웹 브라우저등 여러 응용 프로그램을 구동하는 상황에서 음성인식을 위해 많은 컴퓨터 자원을 활용할 수 없는 상황을 고려하면 본격적인 음성인식 시스템을 도입 하기보단 보다 단순한 연산과 간단한 처리로 음성 인식을 이룰

※ 이 연구는 1997년도 단국대학교 대학연구비의 지원으로 연구되었음.

† 정 회 원 : 단국대학교 전자계산학과 교수
†† 준 회 원 : 단국대학교 대학원 전자계산학과

논문접수 : 1998년 10월 7일, 심사완료 : 1998년 12월 12일

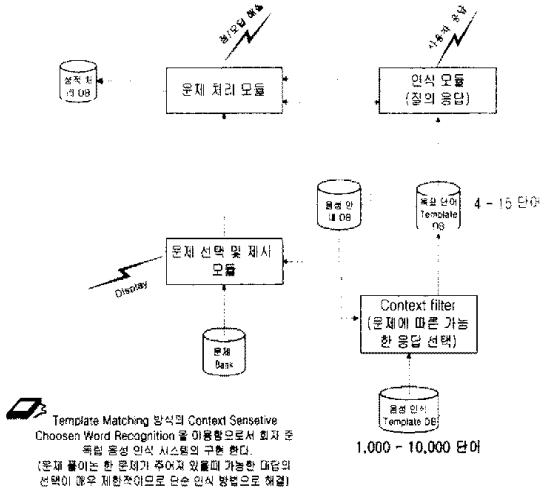
수 있는 방법이 제공 되면 효과적이다. 특히 상황 의존적 특성을 고려 하면 실제 인식의 대상이 되는 단어는 소수에 머물러 복잡한 연산을 요구하는 음성 인식 시스템 보다는 작은 수의 고립단어를 처리하는 DTW (Dynamic Time Warping)같은 방법이 좋을 것이다.

DTW(Dynamic Time Warping)을 이용한 음성인식은 고립단어 인식에서 좋은 인식 율을 보이고 있다. DTW는 동적 프로그래밍 방식을 이용하여 기존에 미리 저장해 둔 Template 데이터베이스로부터 자료를 비교하여 가장 유사성이 높은 단어를 검색하는 방법이다[4,6,12]. DTW는 음성인식 모델 분류상 화자종속 인식기로 구분할 수 있으며, 화자종속 인식기는 특정화자에 의해 학습된 모델을 이용하여 학습에 참여한 화자가 사용하는 인식기로서 화자에 대한 인식성능이 우수한 반면에 학습에 참여하지 않은 화자가 사용하는 경우에는 인식성능이 급격하게 저하하는 단점이 있다[7,9,10,13]. Template 데이터베이스의 양이 많아질 경우 검색해야 할 대상의 증가로 검색시간의 증가를 가져오게 된다. 따라서 DTW를 이용한 방식은 제한된 단어를 사용하는 음성인식 모델에 적합하다.

본 논문에서는 웹상에서 사용할 수 있는 DTW를 이용한 상황 의존적 단답형 질의 응답을 위한 저가형 음성 인식기를 구현 하였다. 초,중,고 학생들의 학습용 단 답형 질의응답 시스템과 같은 질문에 응답이 매우 종속되어 있고, 또한 응답될 내용을 미리 알고 있다고 전제하여 각각의 질문에 따른 가능한 응답을 미리 테이블로 구성을 하였을 경우에는 질문에 타당하지 않은 응답을 검색대상에서 미리 제거함으로써 시간적으로는 검색대상을 줄인 효과를 얻을 수 있으며, 정확도 면에 있어서도 답과는 상관없는 유사한 패턴의 단어를 검색하는 작업을 제외시킬 수 있어 시스템의 단순성에 비해 상대적으로 높은 인식율을 얻을 수 있었다.

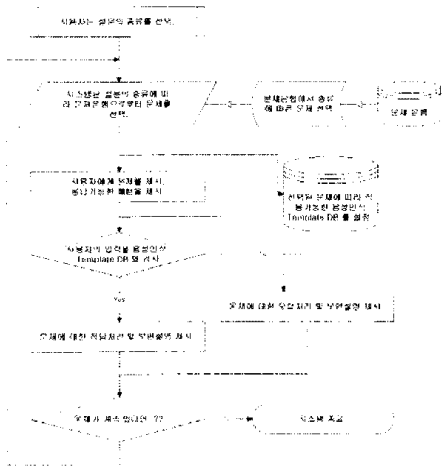
2. 단어 선택형 DTW 인식기

(그림 1)은 질문 의존적 단답형 응답을 가진 질의응답 시스템을 위한 실시간 음성인식 시스템의 구성을 보여준다. 문제은행에는 질문 되어질 예문이 미리 준비되어져 있으며, 음성인식 Template 데이터베이스에는 문제은행의 모든 문제의 가능한 응답이 미리 준비되어 저장된다. 사용자가 본 시스템을 이용하면 먼저 문제 선택 모듈이 문제를 선택하여 사용자에게 제시한다. 이 때 문제 제시 모듈은 선택된 문제를 Context Filter에 알려주어 음성인식 Template 데이터베이스에서 선택된 문제에 따른 가능한 응답들의 음성특질 정보를 뽑아 낼 수 있게 한다.



(그림 1) 단어 선택형 DTW인식기
(Fig. 1) Word selectable DTW recognizer

여기에서 선택된 음성 특질 정보는 즉시 목표단어 Template 데이터베이스를 구축한다. 문제에 대한 응답이 사용자로부터 발생하면 새로 구성된 소규모의 목표단어 Template 데이터베이스를 이용하여 인식모듈이 사용자의 응답을 인식하게 된다. 이 인식결과는 문제처리 모듈에 전달되어 문제처리 모듈은 정오 또는 보조 설명들을 사용자에게 알려준다. 결과는 성적처리 데이터베이스에 저장되어 후에 사용되어 질 수 있도록 한다. (그림 2)는 전체 시스템 작동의 흐름을 보여준다.



(그림 2) 단어 선택형 DTW 인식기 흐름도
(Fig. 2) Flowchart of word selectable DTW recognizer

질문에 따른 응답을 제한하는 context filter는 인식 단이 데이터베이스를 구성할 때 인덱스를 부여하여 인덱스를 제한함으로 구현 할 수 있다. 다음은 인덱스를 이용한 Template 데이터베이스 선정 과정을 보여준다.

```
System initialization
User selection Qi ∈ Question set
Check current template database Tc
IF( i!=c ) select Ti in the database
        Set Tc = Ti
Submit Tc
```

본 시스템을 웹상에 구현할 때 인식기의 위치를 웹 서버에 둘 것인가 혹은 웹 클라이언트에 두어야 하는가를 고려 해야 한다. 인식기의 위치에 따른 고려는 <표 1>과 같다.

〈표 1〉 인식기 위치에 따른 특성
〈Table 1〉 Characteristic depending on position

인식기 위치	웹 서버	웹 클라이언트
음성 입력기 특성	배려 불가능	배려 가능
전송 데이터 양	입력 음성 전체	선정 인덱스
음성 데이터베이스 저장 장소	웹 서버	웹 클라이언트
화자 적응	적용 불가능	적용 가능

인식기의 위치는 전송 되어야 하는 데이터의 양과 밀접한 관계를 가진다. 음성 인식기를 웹 서버에 위치시키는 경우 모든 인식 처리를 서버에서 하여야 함으로 사용자의 음성 데이터를 모두 서버로 전송해야 하는 어려움이 발생한다. 또한 일반 사운드 카드는 고유한 특성을 가지는데 이를 고려하여 주기 어렵다. 반면 인식기를 웹 클라이언트에 위치 시키는 경우 각 사용자에 따라 별도의 적응된 데이터베이스를 구축하여 사용할 수 있으나 인식기의 작동과 인식 대상 단어를 위한 데이터베이스의 운영 등 클라이언트가 처리하여야 할 작업의 양이 늘어나 클라이언트에 상당한 로드가 생겨난다. 그러나 DTW의 특성을 고려하면 인식기를 웹 클라이언트에 위치 시키는 것이 보다 좋은 결과를 얻을 수 있다.

3. DTW(Dynamic Time Warping)

DTW 방식은 음성 패턴의 시간적 변동을 비선형적으로 정규화 시켜서 패턴 정합 방식으로 설계된 고립단어 인식은 DTW만을 사용한 시스템에서도 상당한 인식율을 얻을 수 있다. 그러나 인식환경이 바뀌면 잡음에 의한 영향으로 인식율이 저하되고, 화자 독립 인식의 경우에는 인식 율이 저하되는 단점이 있다[7,8,9,10,13].

음성은 음운이 가지고 있는 고유 특성에서 뿐만 아니라 음성을 사용하는 인간의 시로 다른 특성 때문에 많은 다양성을 가지고 있으며, 이러한 개인의 특성도 말하는 사람의 상태나 기분에 따라 매우 유동적으로 변한다. 따라서 인간이 발성하는 음성은 각 개인차에 의하여 발성 속도가 다르게 되며, 같은 단어를 반복적으로 발성한다고 하더라도 발성 지속 시간의 차가 일어난다. 이러한 변화로 인한 비선형적 시간축의 패턴을 선형적으로 정규화 시켜서 인식에 사용하는 방법이 개발되게 되었는데 이 방법이 DTW이다.

DTW는 음성인식에 동적 프로그래밍(Dynamic Programming)을 적용한 방법으로, 동적 정합 법(Dynamic Warping)은 기준 패턴과 입력되는 음성간의 길이가 다른 경우 두 패턴 사이의 거리를 측정하기 위해서 기준 음성 패턴의 각 프레임과 그에 대응하는 입력 음성 패턴의 프레임간을 동적 프로그래밍 기법을 사용하여 왜곡 도(Distortion)를 구한다.[8] 기준 음성 패턴과 입력 음성 패턴의 발음시간의 차이가 있을 경우, 두 패턴 사이의 왜곡 도를 측정하기 위해서 우선 기준 음성 패턴의 각 프레임과 그에 대응하는 입력 음성 패턴의 프레임을 비교하는 방식을 택하며, 이때 두 패턴의 비교하는 순서는 Warping 함수를 이용하여 구하고, 이때 동적 프로그래밍 기법이 사용된다. 동적 프로그래밍 기법에 따르게 되면 Warping 함수에 의하여 구해진 경로는 모든 경로에 대한 최소비용 거리로 정해지게 된다.

3.1 Time Alignment and Normalization

Time Normalization은 단어들의 시간적 변화가 있는 특징들에 대한 비교를 동일선상에서 처리할 수 있다는 것을 의미한다. 고전적인 방법으로는 비교하는 대상을 같은 길이가 되도록 늘리거나 줄이는 간단한 방법을 사용한다. 이것은 Front-End detection이 정확히 될 수 있다는 가정하에서 작동하는데 보통 Front-End detection을 정확하게 하는 것은 어렵다[4,6]. 즉, 단어 내에서 시간적인 변화를 정확히 찾지 못한다는 것이다. 이런 경우 대부분 정확히 맞추기에는 충분하지 않지만 시간 축에 대하여 확장하거나 축소하는 방법을 사용한다.

$(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_{Tx})$와 $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \cdots, \mathbf{y}_{Ty})$의 두 개의 자료들로 구성이 되어 있을 때, X와 Y에 대하여 시간적인 순서를 각각 i_x, i_y라고 한다면, X와 Y의 비유사성의 정도는 다음과 같은 수식에 의하여 표현되어 질 수 있다.

$$d(X, Y) = \sum_{i_t=1}^{T} d(i_x, i_y) \tag{1}$$

이때 i_x, i_y는 다음과 같이 정의되어 진다.

$$i_t = \frac{T_y}{T_x} i_x \tag{2}$$

X와 Y의 비율에 따라서 비교를 수행한다는 것은 시간 축에 있어서 모두에게 동일한 조건일수 없다. 따라서 X와 Y에 유사도를 고려한 비교가 이루어져야 한다. 이를 위해 각각의 i_x와 i_y에 warping function을 적용하여 다름과 같은 수식을 만든다.

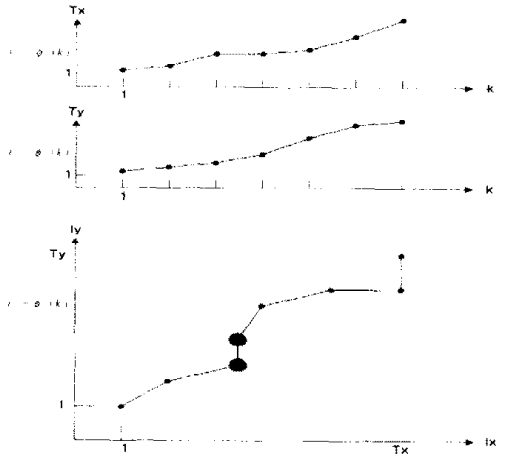
$$i_x = \phi_x(k) \quad k = 1, 2, \ldots\ldots, T \tag{3}$$

$$i_y = \phi_y(k) \quad k = 1, 2, \ldots\ldots, T \tag{4}$$

즉 X와 Y축에 대하여 동일한 비율로 검사를 수행하는 것이 아니라 각각의 좌표축에 대하여 유동적으로 움직일 수 있는 함수의 형태를 제공하는 것이다. 이와 같이 했을 때 위에서처럼 에러 율을 계산하여 보면 다음과 같다.

$$d_\phi(X, Y) = \sum_{k=1}^{T} d(\phi_x(k), \phi_y(k)) m(k) / M_\phi \tag{5}$$

여기에서 $m(k)$는 중요도계수이며, M_ϕ는 정규화 계수이다. 일반적으로 이를 도식화하여 보면 다음과 같다.



(그림 3) Warping function을 이용한 time normalization
(Fig. 3) Time normalization using warping function

즉 (그림 3)에서 볼 수 있는 것처럼 꼭 좌표축에 대하여 정해진 비율대로 비교를 수행하는 것이 아니라 Warping 함수에 의하여 위와 같이 비교대상이 같은 비율이 아닌 유동적으로 움직일 수 있다는 것을 의미한다. 이런 가정하에서 위의 에러 율을 줄이기 위한 방법으로 Dynamic Time Warping이라는 방법이 사용되며, least cost pass 방법을 이용하여 가장 적합하다고 생각되는 경로의 에러 율을 계산한다.

3.2 DTW 알고리즘

초기화

$$D_A(1,1) = d(1,1)\, m(1). \tag{6}$$

재귀호출

$$D_A(i_x, i_y) = \min[D_A(i_x', i_y') + \zeta((i_x', i_y'), (i_x, i_y))]$$
$$\text{for } 1 \le i_x \le T_x,\ 1 \le i_y \le T_y \tag{7}$$

여기에서 $\zeta((i_x', i_y'), (i_x, i_y))$는 두 점간의 적용 가능한 경로의 지역적 거리를 나타낸다.

$$\zeta((i_x', i_y'), (i_x, i_y)) =$$
$$\sum_{l=0}^{L} d(\phi_x(T'-l), \phi_y(T'-l))\, m(T'-l) \tag{8}$$

where $\phi_x(T') = i_x$, **and** $\phi_y(T') = i_y$

검색 공간을 줄이기 위해 다음 같은 검색 조건을 이용하여 검색 범위를 제한하였다.

```
Check = Total_Xlen / Total_Ylen * Cur_YPos;          (9)
Range = Total_Xlen / 2
        - fabs(Total_Xlen / 2-Check);                (10)
Range = Range < (Total_XLen / 8)
        ? (Total_Xlen / 8) : Range;                  (11)
If  (abs(Check ? Cur_XPos) > (Range)/(1.5))
      then ignore check point                        (12)
```

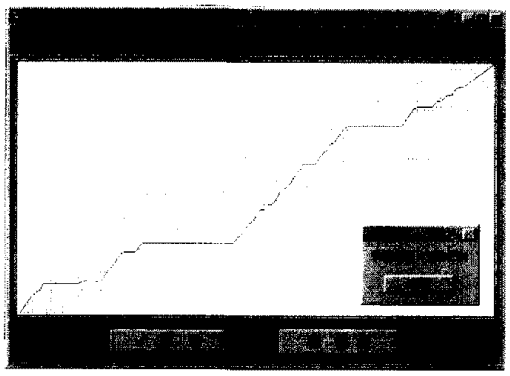
종료조건

$$d(X, Y) = \frac{D_A(T_x, T_y)}{M_\phi} \tag{13}$$

4. 실험 및 결과 분석

본 연구에서는 음성 구간을 검출 하기 위하여 Zero

Crossing과 Power 값을 이용하였다. 본 실험에서는 블록의 크기를 음성이 특성을 가질 수 있다는 최소 크기인 10ms로 설정하였다[6]. 또한 겹치는 구간에서의 자료의 유실을 위해 각각의 구간을 반씩 겹치는 방법을 이용하여 실제의 음성이 자료가 있음에도 불구하고 의미가 없는 블록으로 분류되는 것을 막았다. 음성은 일반 사운드카드를 이용하여 11khz샘플링을 16비트 모노로 녹음한 데이터를 이용하였다.



(그림 4) DTW 결과
(Fig. 4) Test result using DTW

실험은 전체 인식 대상을 하나의 검색 집합으로 보고 인식하였을 때와 검색 인덱스에 의한 Template 데이터베이스를 구성하여 인식하였을 경우를 비교하였고 훈련에 이용된 데이터만을 이용한 경우와 훈련에 참여되지 않은 데이터를 포함한 경우에 대하여 비교하였다. 마지막으로 웹 서버에 인식기를 위치시킨 경우와 웹 클라이언트에 인식기를 위치 시킨 경우의 인식 율을 비교 하여 보았다.

DTW는 Template Matching 방식이기 때문에 비교를 해야 할 대상의 수가 인식의 속도를 좌우한다. 따라서 비교를 해야 할 대상의 수를 가능하면 줄여야 인식에 필요한 시간을 절약하고 인식 율을 높일 수 있다.

본 실험에서도 이와 같은 점을 고려하여 객관식 문제 유형 같은, 사용자로 하여금 입력되어질 음성이 어떤 제한된 데이터들로 이루어진 상황이라고 가정을 하고 실험을 하였다. 즉 응답군을 G1, G2, G3의 세 개의 그룹으로 실험 집단을 나누고 각각의 그룹들을 대표할 수 있는 데이터들을 각각 4개씩 선정을 하였다. 응답군에 따른 자료는 인덱스에 의해 구분되어 Context filter가 문제에 따라 응답군을 선정하도록 하였다. 실험 데이터는 훈련 자료를 제공한 사용자와 실험 데이터를 제공하지 않은 사용자를 포함한 두 개의 집단으로부터 구하였다. 먼저 인식기를 웹 서버에 놓고 웹 클라이언트에는 입력만을 받아 전송 시켜 웹 클라이언트의 로드를 최소화 하여 실험하였다.

실험에 사용된 데이터베이스는 음성데이터이름, 그룹번호, 결과 표현 인덱스, 각 실험에 있어서의 왜곡도들을 저장할 필드로 나누어 진다. 여기에 그룹번호는 그룹 인덱스로 사용되어 특정그룹만을 검색하는 것이 가능해지며, 결과 표현 인덱스를 이용하여 사용자가 입력한 음성을 수치화 함으로써 우리가 원하는 결과인지 판단할 수 있도록 하였다.

〈표 2〉 인식기가 웹 서버에 구현된 결과
〈Table 2〉 Test result of the recognizer on Web server

(인식 율/인식시간비율)

	응답군 선택	응답군 비 선택
학 습 참 여 자	0.90 / 0.37	0.83 / 1.00
비 학 습 참여자 포 함	0.77 / 0.28	0.73 / 1.00

<표 2>에서 보여지는 봐와 같이 응답군을 선택하여 인식을 하는 단어 선택형 시스템이 훈련에 자료를 제공한 학습 참여자의 경우와 훈련 자료제공을 하지않은 비 학습 참여자를 포함한 경우 모두 상대적으로 높은 인식율을 얻을 수 있었다.

다음은 인식기를 웹 클라이언트에 구현하여 각 사용자별 별도의 데이터베이스를 구축한 후 같은 실험을 행하였다.

〈표 3〉 인식기가 웹 클라이언트에 구현된 결과
〈Table 3〉 Test result of the recognizer on Web client

(인식 율/인식시간비율)

	응답군 선택	응답군 비선택
학 습 참 여 자	0.97 / 0.35	0.89 / 1.00
비 학 습 참여자 포 함	0.80 / 0.38	0.74 / 1.00

<표 3>에서도 <표 2>에서와 같이 응답군을 선택하여 인식을 하는 단어 선택형 시스템이 훈련에 자료를 제공한 학습 참여자의 경우와 훈련 자료 제공을 하지 않은 비 학습 참여자를 포함한 경우 모두 상대적으로

높은 인식율을 얻을 수 있었다. 그러나 인식 시간에 경우에 있어서는 웹 클라이언트에서 인식을 하는 경우와 웹 서버에서 인식을 하는 경우 모두 비슷한 비율로 응답군을 선택하는 것이 응답군을 미 선택하는 경우보다 약 40%의 시간 만이 사용되었다. 데이터 전송 시간은 통신 상황에 따라 달라 질 수 있으므로 여기에는 통신 데이터 전송에 걸리는 시간을 제외 시켰다.

5. 결 론

본 논문에서는 양방향 통신이 가능한 환경에서의 기존의 입력 매체 보다 손쉽게 사용할 수 있는 입력 매체인 음성을 이용하는 DTW를 이용한 음성 인식기를 설계하여 보았다. 이것을 통하여 제한된 어휘를 사용하는 음성인식 분야에서 적용이 가능한 시스템의 구성을 보였다. 그리고 DTW에서 검색시간에 가장 큰 영향을 미치는 Template 데이터베이스의 수직인 크기에 있어서도, 단어를 검사하는데 전체를 대상으로 하는 것이 아니라 검색할 대상을 제한하는 방법을 사용함으로써 검색시간을 줄여 부분적이지만 실시간 사용에 대한 가능성을 확인하였다.

인식율에 있어서도 사용자에 대한 적응과정을 거친다면 보다 높은 인식율을 얻을 수 있다는 점에서 입력의 수가 제한되어 있고 미리 학습이 가능하다면 보다 높은 인식율을 얻을 수 있어 입력에 대한 정확성을 높일 수 있다. 즉 기존의 입력 매체인 키보드나 마우스 같은 매체보다 현저히 떨어질 수밖에 없는 음성의 정확성을 조금이라도 높여 현실적으로 사용이 가능하다는 것을 확인할 수 있었다.

음성인식에 있어서 화자독립과 연속음성 인식은 아직 계속 연구되고 있는 분야이고 아직은 어려움이 많은 분야이다. 반면에 고립단어 인식은 현실적으로 구현이 되고 있는 분야이며, 사용이 가능한 분야이다. 우리는 본 논문에서 고립단어 인식을 통하여 현실에 적용 가능한 여가의 음성 인식 시스템을 구현 할 수 있음을 확인 하였다.

또한 현재의 인터넷이 대중화되어 있는 환경에서는, 이와 같은 시스템을 광범위한 클라이언트 서버 개념으로 인터넷 환경하에 클라이언트에 ActiveX Control 같은 기술을 이용하면 사용자 자신의 음성에 대한 특징 성분들을 클라이언트에서 따로 관리함으로써 서버에 대한 처리의 부담을 줄일 수 있으며, 또한 통신에 대한 트래픽 문제와 함께 속도면에서도 탁월한 성능을 보일 수 있을 것이라고 사료된다.

참 고 문 헌

[1] Sadaoki Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, INC, 1992.

[2] Jae S.Lim, *Speech Enhancement*, Prentice-Hall, INC, 1983.

[3] H. Ney, The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-32(2) : 263-271, April, 1984.

[4] Thomas Parsons, *Voice and Speech Processing*, McGraw-Hill Book Company, 1986.

[5] Michael R. Portnoff, Short-Time Fourier Analysis of Sampled Speech, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-29(3) : 364-373, June, 1981.

[6] L.R. Rabiner and Biing-Hwang Juang, *Fundamentals of speech recognition*, Prentice-Hall International, INC, 1993.

[7] L.R. Rabiner and S.E. Levinson, Isolated and Connected Word Recognition-Theory and Selected Applications, *IEEE Transactions on Communications* COM29(5) : 621-659, May, 1981.

[8] L.R. Rabiner, S.E. Levinson, A.E. Rosenberg, and J.G. Wilpon, Speaker-Independent Recognition of Isolated Words Using Clustering Techniques, *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-27(4) : 336-349, August, 1979.

[9] H.Sakoe, Two-Level DP-Matching-A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-27(6) : 588-595, December, 1979.

[10] H. Sakoe and S.Chiba, Dynamic Programming Algorithm Optimization for Spoken Word Recog-

nition, *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-26(1) : 43-49, February, 1978.

[11] Shuzo Saito, *Speech Science and Technology*, IOS Press, 1992.

[12] Alex Waibel and Kai-Fu Lee, *Reading in Speech Recognition*, Morgan Kaufmann Publishers, INC, 1990.

[13] 김현호, "DTW와 HMM을 결합한 적응형 한국어 음성 인식 시스템의 설계 및 구현", 건국대학교 대학원 논문, 1996.

박 용 범

e-mail : ybpark@cs.dankook.ac.kr

1985년 서강대 전자계산학과 학사(B.S)

1987년 N.Y. Polytechnic University(M.S)

1991년 N.Y. Polytechnic University(ph.D)

1992년 현대전자 산전연구소 선임연구원

1993년~현재 단국대학교 전자계산학과 조교수

관심분야 : Speech Recognition, 신경회로망 등

박 종 일

e-mail : always@cs.dankook.ac.kr

1996년 단국대학교 전자계산학과(학사)

1997년~현재 단국대학교 일반대학원 전자계산학과(석사과정)

관심분야 : 음성인식, 신경회로망