

실시간 음성 다이얼링 시스템 구현을 위한 연결어 인식에 관한 연구

A Study on Connected Word Recognition for the Implementation of a Real-Time Voice Dialing System

김 천 영*, 양 진 우*, 유 형 근*, 이 형 준*, 홍 진 우*, 이 강 성**, 안 태 옥***, 김 순 협*

(Cheon Young Kim*, Jin Woo Yang*, Hyung Keun Ryu*, Hyung Jun Lee*,
Jin Woo Hong*, Gang Sung Lee**, Tae Ok An***, Soon Hyob Kim*)

◆본 연구는 1992년 통신학술 연구과제의 지원금에 의해 수행되었음◆

요 약

본 논문은 음성 다이얼링 시스템을 구현하기 위한 연결어 인식에 관한 연구이다. 적용된 인식 알고리즘은 기준패턴을 생성할때 DMS(Dynamic Multi-Section) 모델을 이용한 One-Stage DMS/DP 알고리즘이고, 인식 대상어는 광운대학교 부서명 150 단어이다.

연결어 인식을 실시간으로 처리하기 위한 방법으로써 본 논문에서는 음절과 단어 단위의 DMS 템플릿(template)를 구성하여 실험하였고 이 실험 결과로부터 실시간과 인식률을 고려한 최적의 인식은 단어단위 템플릿에서 20 구간의 DMS 모델을 적용하였을때 수행되었고, 이때 다중-화자종속과 화자독립의 인식률은 각각 97.2%, 86.8%이다.

실험된 결과를 이용하여 음성 다이얼링 모델 시스템을 DSP 전용칩인 TMS320C30 프로세서를 내장한 DSP 보오드, 486 PC와 Dial 모델을 이용해서 구현하였고, 전체 다이얼링 시간은 약 7~14초가 소요되었다.

ABSTRACT

This paper is a study on connected word recognition for implementation of real-time voice dialing system. The applied recognition algorithm is One-Stage DMS/DP one using DMS(Dynamic Multi-Section) model in case of reference pattern generation and recognition vocabularies are 150 department names within Kwang Woon university.

As method to process the connected word recognition in real time, this paper consisted DMS template of syllable unit and word unit, and then experimented it. From the results of these experiments, the optimal recognition considering a tradeoff of real-time and recognition rate is obtained by applying DMS model with 20 section's word unit template and recognition rate of the multi-speaker dependent and speaker-independent is 97.2%, 86.8% respectively.

From the results of these experiments, voice dialing model system is implemented by DSP board with TMS320C30 chip, IBM-PC/486, and dial modem. Total dialing time takes about 7~14 sec.

*광운대학교 전자계산기공학과

**광운대학교 부설 전산원

***전북 산업대 컴퓨터공학과

접수일자: 1993년 1월 28일

I. 서 론

연결어 인식은 단어들 인식의 발전된 형태로서 어휘의 양이 제한된 단어가 연결되어 발음될 때 이 연결 단어를 인식해 내는 과정을 의미한다. 이러한 연결어 인식은 그동안 크게 두가지 방식으로 연구되어 왔다 [1][2]. 첫번째 접근 방식은 단어열을 먼저 특성단위에 따라 분할한 다음 이 단위에 의해 분류하는 방식이다. 분할 방법은 각 어휘에 따라 실현상으로 정의되고 최적화된 규칙들에 의해 정해진다. 그러나, 이러한 분할 방식에 근거한 인식 방법은 분할 단계 이후의 분류 단계에 치명적인 영향을 주는 단어 경계 검출 에러(error)를 종종 발생시킬 수 있다. 이러한 부정확한 분할에 의해서 발생하는 에러를 없애기 위해서 두번째 접근 방식인 동적 프로그래밍(Dynamic Programming) 기법이 단어들 인식뿐만 아니라 연결어 인식을 위한 가장 신뢰성있는 대안으로 사용되어 왔으며, 연결어 인식을 단어들 인식 분쇄와 같은 최적화 문제(Optimization problem)로 보고 공식화했다. 이러한 방법의 장점은 인식 시스템이 미리 갖추어야 할 정보와 학습(training)의 양이 적다는 점으로 단지 단어에 대한 기준 패턴만을 알고 있으면 된다. 또 다른 장점은 단어의 경계 검출과 비선형적 시간 정렬(time alignment), 그리고 인식의 세가지 동작이 동시에 수행된다는 것이다. 따라서, 단어 경계의 검출이나 비선형적 시간 정렬에서 비롯되는 인식의 오차는 발생하지 않는다. 이 알고리즘은 완전히 단어에 대해서만 일치하려 하므로 단어의 경계는 자동적으로 결성되어진다. DP를 이용한 연결어 인식 방법은 Two-level DP[1], Level-Building DTW [3], One-Stage DP [4] 등등의 여러 알고리즘들이 있지만, 그 중에서 One-Stage DP를 이용한 방법은 다른 알고리즘들에 비해서 인식률이 떨어지지 않으면서 Sakoe의 Two-Level DP 알고리즘의 계산 시간의 1/25 걸리고 Myers와 Rabiner의 Level-Building DTW 알고리즘의 계산 시간의 단지 1/4이 소요되고, 저장 장소 요구사항에 대해서도 다른 두개의 알고리즘과 비교해서 9배 이상의 감소 요인을 제공한다고 것이 보고되었다[4][5]. 이상의 비교에서 연결어 인식에 One Stage DP 알고리즘이 다른 알고리즘과 비교해서 인식률이 떨어지지 않으면서, 계산시간이 적게 들고, 저장장소가 축소된다는 개념에서 연결어의 실시간 인식에 알맞은 알고리즘이라는 것이 이론적으로 확인되었기 때문에 본 논문에서는 중규모 어휘

의 음성 다이얼링 모델 시스템을 구현하는 인식 알고리즘으로 One-Stage DP 알고리즘을 이용한다.

그리고 연결어 인식의 패턴 매칭 방법으로 인식될 미지의 시험패턴과 비교되어지는 기준 패턴을 생성하는 방법은 아주 중요한 문제이다. 기존에 연결어의 기준패턴을 생성하는 방법은 임의의 선택 방법(Causal method), 평균법(Averaging method), 집단화 방법(Clustering method), 다이نام릭 평균(Dynamic Average), VQ(Vector Quantization) 방법 등이 있지만, 이들 방법들은 어느 정도 한계성이 있기 때문에 새로운 기준 패턴 생성 알고리즘의 필요성이 대두되어지고 있다[2]. 따라서, 본 논문에서는 연결어의 기준패턴을 생성할때 인식률이 보장되면서 기억 장소와 비교 시간에서 효율적인 DMS(dynamic Multi Section) 모델을 이용하여 기준패턴을 작성한다.

본 논문에서는 기준 패턴을 생성할때 연결어 인식에는 처음 적용되어지는 DMS 모델을 적용해서 템플릿을 작성하고 인식할때 One Stage DP 방법을 이용한 One-Stage DMS/DP를 제안한다. 제안된 알고리즘의 효율적인 실험 조건을 찾기 위해서 여러 상황에서 실험해 보고 최적의 실험 결과를 가지고 실시간 음성 다이얼링 모델 시스템을 실험실 환경에서 구현해 본다.

II. DMS 모델을 이용한 기준패턴의 생성방법

2.1 DMS(Dynamic Multi Section) 모델의 작성 과정

DMS 모델은 유사한 특징을 가지는 벡터들을 한 구간으로 만들기 위해 구간을 동적으로 분할하여 대표 특성 벡터를 구함으로써 짧게 발음되는 특성까지도 대표 특성 벡터로 선택될 수 있고, 그 구간의 지속 시간 정보를 갖도록 만든 모델이다. 즉, 단어 패턴의 연속된 대표값(대표특성 벡터와 지속 시간 정보)을 사용하여 음성인식을 수행함으로써 파열음이나 파찰음 등과 같은 짧은 음소의 특성 벡터도 동등한 비중을 가질수 있도록 구간을 동적으로 나누어 준다. 점에서 DMS 모델이라 한다[6][7].

DMS 모델은 구간을 단어의 특성에 따라 동적인 구간으로 나누고, 각 구간에 대표 특성 벡터와 지속 시간 정보를 가지도록 만들어야 하기 때문에 모델을 작성하는데 좀 더 복잡한 절차가 필요하다. DMS 모델 작성은 크게 두 단계로 분리되어서 수행된다. 먼저, 실행하는 단계는 구간을 동적으로 분할하는 구간 구분화 작업이고, 다음 단계는 구분된 구간들에 대해

각 구간의 대표 특징 벡터와 지속 시간 정보를 구하는 단계이다. 모델 작성은 전체적인 흐름도는 그림 1과 같다.

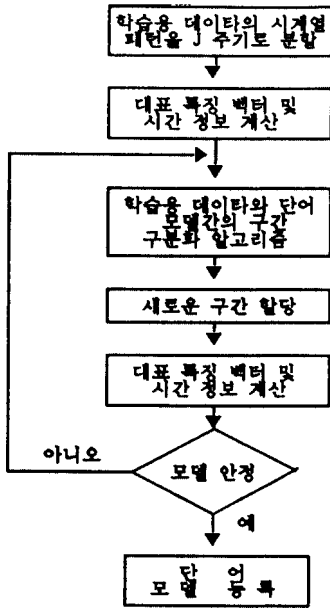


그림. 2.1 단어 모델 작성 과정
Fig. 2.1 The process of word model generation

2.1.1 구간 구분화 알고리즘

각 단어의 DMS 모델 M의 $j(1 \leq j \leq J, J$ 는 DMS 모델의 구간의 수)번째 구간에는 구간정보 $M(j)$ 가 있는데, 구간 정보 $M(j)$ 는 대표 특징 벡터 m_j 와 지속 시간 정보 p_j 로 구성되어 있다.

처음의 단어 모델은 학습용 데이터들을 시간축상에서 동분할하여, 같은 구간에 할당된 각 데이터 프레임들의 특징 벡터들을 한데 모아 중심점을 계산하여 그 구간을 대표하는 대표 특징 벡터를 구했다. 이때 지속 시간 정보는 동간격으로 구간을 분할했다는 것을 고려하여 각 구간의 마지막 프레임수들의 합을 학습용 데이터들의 전체 프레임수로 나누어서 구했다. 지속시간 정보 p_j 는 식(2-1)과 같이 표현 된다.

$$p_j = \frac{\sum_{m=1}^M e_m(j)}{\sum_{m=1}^M l_m}, 1 \leq j \leq J \quad (2-1)$$

여기서 $e_m(j)$ 는 j 구간의 마지막 프레임 번호이고 l_m 은 학습용 데이터들의 전체 프레임 수 이고, M 은 학습용 데이터의 총수이다.

그 다음에 구간 정보를 구한 다음, 구간 구분화 알고리즘을 이용해서 각각의 학습용 데이터와 초기 단어 모델과 DP-매칭을 수행하고[8], 백트래킹에 의해 구간의 경계선을 변경한다.

각각의 학습용 데이터의 DP 매칭의 결과인 총 누적거리를 구하고, 각 구간의 마지막 프레임수를 백트래킹에 의해서 할당된 다음 각각의 학습용 데이터의 새롭게 할당된 구간의 프레임들의 특징 벡터로 부터 중심점을 계산해 대표 특징 벡터로 삼아 단어 모델을 갱신한다. 동시에 각 구간에 할당된 프레임수를 전체 프레임 수로 나눈 비율을 단어 모델의 지속 시간 정보로 등록한다.

이것이 단어 모델을 한번 갱신한 것이 된다. 모델을 갱신한 후 구간 구분화 알고리즘[5][6]을 이용하여 구간 경계를 정한 후 그림 2.1의 흐름도에 따라서 반복하여 단어 모델을 갱신한 후 각각의 학습용 데이터의 DP 매칭의 결과인 각각의 총 누적거리를 합해서 얻어진 새로운 전체거리값이 이전의 거리값 보다 작으면 모델은 더 안정될 가능성이 있으므로 위의 절차를 반복한다. 만일 새로운 거리값이 이전의 거리값 보다 크면 모델은 더 이상 안정될 가능성이 없으므로 가장 작은 거리값을 가졌을때의 모델의 경계를 구간 구분화 알고리즘의 결과로 최종적으로 등록한다.

구간 구분화를 위해 사용되는 DP 알고리즘은 누적 거리 D에 지속 시간 정보에 의한 거리 p_j 를 포함시킨다. 즉

$$D(i, j) = d_e(t, m_j) + \min \begin{cases} D(i-1, j), (1 \leq i \leq j) \\ D(i-1, j-1) + P_{j-1} \end{cases} \quad (2-2)$$

로 해서, 총 누적거리 $DIS(T, M)$ 를 구한다. $P(j)$ 는 대응하는 학습용 데이터의 j 번째 프레임과 단어 모델의 j 번째 구간의 마지막 프레임과의 지속 시간 정보의 차에 의한 거리로서

$$p_j = W \cdot d_3(e(j), i) \quad (2-3)$$

$$d_3(e(j), i) = |(p_j \cdot I) - i| \quad (2-4)$$

또한, W 는 지속 시간 정보의 차에 대한 가중치로서 인식률에 영향을 미치는 요소이다. 식(2-2)에서 학습용 데이터의 $j-1$ 구간까지의 프레임 수의 전체 길이에 대한 비율이 시간 정보에 의한 비율에 가장 가까울 때 시간 정보에 의한 거리값이 가장 적어지므

로 j 구간으로 천이할 확률이 높아지고, 시간 정보에 의한 비율값에서 멀어지면 시간 정보에 의한 거리값이 커져서 j 구간으로 천이할 확률이 낮아짐을 알 수 있다.

2.1.2 모델 작성 방법

구간 구분화 알고리즘의 결과로 등록된 각 학습용 데이터의 단어 패턴(T_1, T_2, \dots, T_N)들의 구간 경계를 이용하여 각 단어 패턴들을 J 개의 구간으로 나누어 주고, 단어 패턴의 j 번째 구간에 할당된 프레임들의 특성 벡터들을 모아 중심점을 계산해 이것을 단어 모델의 j 번째 구간을 대표하는 특징 벡터로 하였다.

지속 시간 정보는 j 번째 구간의 마지막 프레임수들의 합을 단어 패턴들의 전체 프레임수로 나누어서 구한다. DMS 모델의 작성 과정은 그림 2.2와 같다.

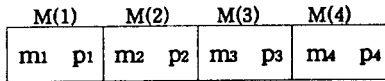
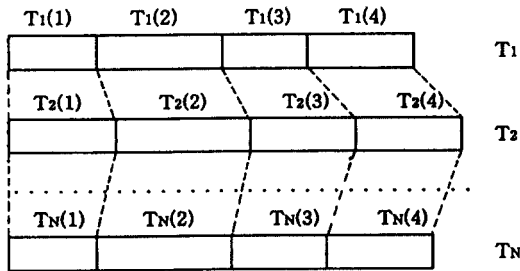


그림. 2.2 완성된 DMS 모델의 예 (a) 학습용 데이터 (b) DMS 모델

Fig. 2.2 An example of completed DMS model (a) training data (b) DMS model

III. 제안된 연결어 인식 방법

3.1 기존의 One-Stage DP 방법

One-Stage DP 알고리즘[4]은 하나의 지수에 의해서 시간 워핑 경로를 파라미터화하여, 최적화 기준을 직접 워핑 경로의 함수로 취급한다. 이 방법에서는 두 가지 전이 규칙(transition rule) 즉, 내부 단어에 대한 전이 규칙과 단어 경계에서의 전이규칙을 사용하여 워핑 경로를 제한한다. 이러한 경로 제약조건과

DP를 사용하여 최적화를 시키면 하나의 stage를 갖는 알고리즘이 되며, 여기에서는 다른 연결어 알고리즘들에서와 같은 여러단계의 최적화가 필요없게 된다. One-Stage DP 알고리즘은 Level-Building DTW 방법과 달리 입력 문자열의 최대값을 제한할 필요가 없으며, 더우기 이 방법은 단독어 인식에 소요되는 정도의 계산량 만으로 충분하다.

One-Stage DP 방법의 기본적인 개념이 그림 3.1에 묘사되어진다. 시험패턴의 시간 프레임 j 와 각각의 템플릿 k 의 시간 프레임 i 는 격자점(i, j, k)의 집합으로 정의되어진다. 연결어 인식의 문제는 시험패턴과 템플릿들의 비시의 열사이의 가장 최적의 매칭 경로를 제공하는 격자점(i, j, k)의 집합을 발견하는 것이다.

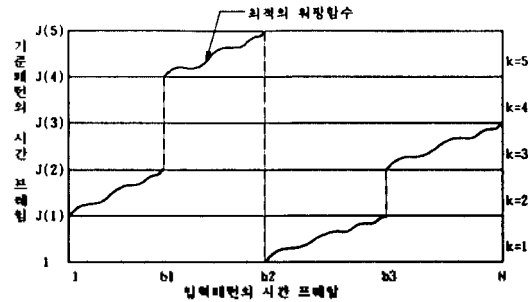


그림. 3.1 One-Stage DP 알고리즘에서 비시의 입력 패턴과 일련의 기준 패턴

Fig. 3.1 Optimal warping path between unknown input pattern and reference pattern in One-Stage DP algorithm

일반적으로, 경로 W 는 격자점들의 열로써 다음과 같이 주어진다.

$$W = (w(1), w(2), \dots, w(l), \dots, w(L)) \quad (3-1)$$

여기서, $w(l) = (i(l), j(l), k(l))$ 과 l 은 경로 원소(path element)들의 순서화된 집합을 지수화하기 위한 경로 파라미터이다. 매칭 절차(matching procedure)에 대한 범주는 주어진 경로상에 있는 소구간 거리들의 합인 전체 거리(global distance)이다. 연결어 인식의 문제는 다음과 같은 최소화 문제로써 정의되어질 수 있다.

$$\min_W \sum_l d(w(l)) \quad (3-2)$$

즉, 모든 허용된 경로들을 간주해서 전체 거리를 최소화하는 것이다. 가장 최적의 경로를 달성하는 템플리트들의 열은 그림 3.1에서 보는 것처럼 단일하게 발견되어진다.

전체 거리를 최소화하는 과정에서 시간 워핑 경로는 매칭되어지는 패턴들의 물리적인 특성에 의해서 연속성(continuty) 제약조건들이 만족되어야 한다. 이 제약조건들은 경로의 연속적인 점들에 적용되어지고, 이 연속성 제약조건들 주어진 경로 (i, j, k) 에 대해서 선행하는 가능한 점들을 결정하게 되는데, 여기서는 이것을 전이규칙(transition rule)이라고 정의한다.

전이규칙의 2가지 형태가 존재하게 된다. 템플리트 내부의 전이 규칙은 within-template 전이 규칙이라고 불리우고 템플리트 경계에서의 전이 규칙은 between-template 전이 규칙이라고 불리운다. 전이 규칙의 이 두가지 형태들은 그림 3.2에 묘사되어진다.

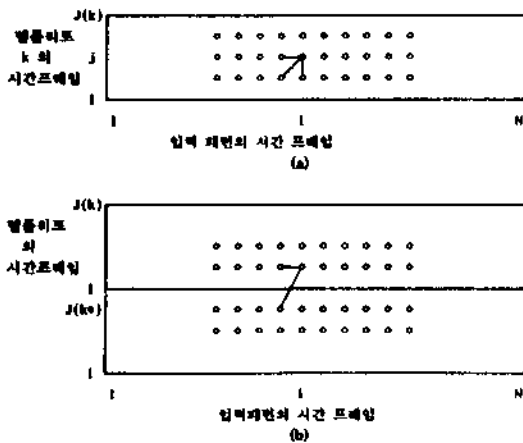


그림. 3.2 One-Stage DP 알고리즘에서 두가지 전이 규칙의 예
(a) 템플리트 내부의 전이 규칙
(b) 템플리트 경계의 전이 규칙

Fig. 3.2 Example of two transition rule in One-Stage DP algorithm
(a) With-template transition rules
(b) Between-template transition rules

템플리트 내부에서 즉, $j > 1$ 에서 within-template 전이 규칙은 다음에 따르는 공식에 의해서 정의되어진다.

$$D(i, j, k) = d(i, j, k) + \min \begin{cases} D(i-1, j, k), \\ D(i-1, j-1, k), \\ D(i, j-1, k), \end{cases} \quad (3.3a)$$

$j=1$ 을 가지는 템플리트 경계에서는 between-template 전이 규칙이 다음과 같이 정의되어진다.

$$D(i, 1, k) = d(i, 1, k) + \min \begin{cases} D(i-1, 1, k), \\ D(i-1, j(k^*), k^*), \end{cases} \quad (3.3b)$$

시험 패턴의 시작 프레임의 격자점들에서, 시험 패턴의 시간 축상에서 선행하는 프레임이 없으므로, 전이 규칙은 수정되어야 한다. 즉, 격자점 $(1, j, k)$ 는 격자점 $(1, j, -1, k)$ 로부터만 달성되어진다.

식 (3.3a)와 (3.3b)은 동적 프로그래밍의 전형적인 재귀 관계식이다. 이 재귀 관계식을 사용하는 것에 의해서 누적 거리 $D(1, j, k)$ 는 격자점과 격자점을 반복적으로 전이하면서 측정되어진다. 그래서, 구해진 누적 거리들의 배열 $D(1, j, k)$ 를 사용해서 최소 총 누적거리를 가지는 종단 프레임의 템플리트를 구하고 이 격자점에서 부터 백트래킹을 해서 최상의 경로를 구한다.

3.2 제안된 One-Stage DMS/DP 인식 알고리즘

3.2.1 알고리즘의 기초단계

One-Stage DMS/DP 알고리즘을 상세히 설명하기 전에 기본적인 개념들을 정리하면 다음과 같다.

- 1) $i=1, \dots, N$ 개의 시간 프레임으로 구성된 미지의 입력 단어열의 시험패턴
- 2) 입력 단어열에 대응되는 기준패턴은 단음절 또는 2 음절이상의 단어로 구성된 k 개의 기준 패턴, 즉, 템플리트의 집합으로 구성되어진다. 이 템플리트는 DMS 모델링에 의해 각 구간별로 구한 대표 특징벡터를 포함하고 있고 지속시간 정보는 포함하지 않는다. 이후로부터 기준패턴 또는 템플리트를 DMS 모델링을 해서 구했기 때문에 DMS 템플리트라고 정의한다.
- 3) DMS 템플리트 k 의 시간 프레임은 $j=1, \dots, J(k)$ 로서 표시되고 이때 $J(k)$ 는 DMS 템플리트의 길이를 의미한다.
- 4) 인식목표는 입력 단어열과 가장 잘 매칭되는 DMS 템플리트들의 집합, 즉, $q(1), \dots, q(R)$ 을 결정하는 것이다. 여기서, $q(1), \dots, q(R)$ 을 "super"

기운 패턴이라고 한다.

3.2.2 제안된 One-Stage DMS/DP 알고리즘

One-Stage DP 알고리즘의 기본적인 개념은 앞 절에서 설명되었고, 이를 바탕으로 One Stage DMS/DP 알고리즘의 상세한 알고리즘은 다음과 같이 간단하게 기술되어진다.

단계 1) $k=1, \dots, K$ 에 대해서 다음과 같은 초기화를 행한다.

$$D(1, j, k) = \sum_{i=1}^j d(1, j, k) \quad (3.4)$$

단계 2)

a) $i=2, \dots, N$ 에 대해서, 단계들 2b-2c를 행한다.

b) $k=1, \dots, K$ 에 대해서, 단계들 2c-2e를 행한다.

$$c) D(i, 1, k) = d(i, 1, k) + \min \begin{cases} D(i-1, 1, k), \\ D(i-1, j(k^*), k^*) \end{cases} \quad (3.5)$$

$$k^* = 1, \dots, K$$

d) $j=2, \dots, J(k)$ 에 대해서, 단계 2c를 행한다.

$$e) D(i, j, k) = d(i, j, k) + \min \begin{cases} D(i-1, j, k), \\ D(i-1, j-1, k), \\ D(i, j-1, k) \end{cases} \quad (3.6)$$

단계 3) 누적 거리들의 배열 $D(i, j, k)$ 를 사용해서 최소 총 누적 거리값을 가지는 DMS 템플릿의 종단 프레임이 있는 격자점으로부터 가장 최적의 경로를 역으로 추적한다.

이 알고리즘의 단계 3은 각각의 격자점에서 "minimum" 연산자에 의해서 달성되어진 결정들을 역으로 추적하는 것에 의해서 입력 패턴에 있는 단어들의 미지의 열을 발견하는 것이다.

3.2.3 알고리즘의 실제적인 구현

One-Stage DMS/DP 알고리즘의 최종적인 목표는 입력 단어들의 미지의 열을 결정하는 것이다. 시간 프레임 i 에 대해서 동적 프로그래밍을 재귀적으로 수행하기 위해서, 누적 거리 배열 $D(i, j, k)$ 의 단지 작은 배열 집합이 필요하다. 즉, 시간 프레임 i 에 이전하는 배열인 $\{D(i-1, j, k) : k=1, \dots, K; j=1, \dots, J(k)\}$ 이다. 이 배열들과 격자점들은 그림 3.1의 시간평면상에서 수직 열을 형성한다. 저장 장소의 이 열(column)은 누적 거리들의 열배열로써 간단하게 언

급되어지고, $D(i, k)$ 로써 표시되어진다. 이와같이, 배열 $D(i, j, k)$ 는 단지 하나의 열만을 사용해서, 동적 프로그래밍의 재귀 관계식 (3-5)와 (3-6)이 시험 패턴의 시간축을 따라서 수행되어지고 격자점과 격자점을 따라서 누적 거리의 열 배열을 갱신한다.

저장 장소 감축에 대한 이 기술은 단독어 인식의 경우에서 유추되어지고, 목표는 시험패턴과 DMS 템플릿사이에서 최소 총 누적거리를 구하는 것이다. 하지만, 필수적인 차이점은 연결어에서 백트래킹이 입력음성 단어들의 미지의 열을 발견하기 위해서 One-Stage DMS/DP 알고리즘에 적용되어진다는 점이다. 백트래킹 정보는 동적 프로그래밍의 재귀식이 평가되어지는 동안에 저장되어진다.

그래서, 각각의 격자점에 대해서 백포인터(backpointer) $B(i, j, k)$ 는 격자점 (i, j, k) 에 경로를 가지는 선행하는 단어의 종단 프레임(ending frame)의 값을 저장한다. 본 논문에서는 배열 $D(i, j, k)$ 의 경우에서처럼 백포인터 $B(i, k)$ 열 배열(그림 3.1에서)로서 백포인터 $B(i, j, k)$ 의 배열을 저장할 수 있고, 가장 최적의 경로의 격자점과 격자점에서 백포인터의 값을 갱신한다.

또한, "from DMS template" 배열 $T(i)$ 가 시험 패턴의 각각의 프레임에 대해서 DMS 템플릿의 종단 프레임(ending frame) $J(k^*)$ 에서 최소 누적 거리를 가지는 DMS 템플릿의 지수 k^* 를 저장하기 위해서 사용되어진다. 일반적으로, 이것은 다음식 (3-7)처럼 표현되어진다.

$$T(i) = k^* = \operatorname{argmin} D(i, J(k), k) \doteq D(J(k), k), k=1, \dots, K \quad (3-7)$$

여기서, $\operatorname{argmin}\{f(x), x=x_0, \dots, x_n\}$ 은 $f(x)$ 를 최소화하는 최적의 인수 x 를 구하는 연산자이다.

덧붙여서, "from frame" 배열 $F(i)$ 는 시험 패턴 프레임 i 에서 선행하는 단어의 종단 프레임에 백포인터 $B(j(k^*), k^*) \doteq B(i, j(k^*), k^*)$ 를 유지하기 위해서 소개되어진다. 이와같이, "from frame" 배열 $F(i)$ 는 격자점 $(i, J(k^*), k^*)$ 에 가장 최적의 경로로 오는 시험 패턴의 시간 프레임의 track을 저장한다. 다시 말하면, "from frame" 배열은 단어 경계들의 track을 저장하고 "from DMS template" 배열은 인식된 단어의 지수를 저장한다.

"from DMS template"와 "from frame" 배열들의 개념은 DMS 경계의 전이 규칙들과 연관되어진다. 각각의 시간 프레임 i 에 대해서, 가장 최적의 템플리

트의 열 즉, DMS 템플리트의 중단 프레임에서 최소 누적 거리를 가지는 DMS 템플리트의 지수와 단어 경계를 결정하는 배열은 최적의 전체 경로를 결정해 준다.

그림 3.3은 그림 3.1의 워핑 경로에 대한 백트래킹 절차의 예를 보여준다. 시험 패턴의 마지막 시간 프레임에 대해서, "from DMS template" 배열은 각각의 DMS 템플리트의 중단 프레임에서 최소 총 누적 거리를 가지는 템플리트의 지수를 결정한다. "from frame" 배열은 시험 패턴의 시간축상에서 선행하는 단어의 중단 프레임을 결정한다. 이 단어 또는 음절의 중단 프레임에서, 다시 "from DMS template" 배열은 가장 최적의 선행하는 DMS 템플리트의 지수를 결정하고 대응하는 "from frame" 배열이 선행하는 단어의 중단 프레임을 다시 역추적해서 결정한다. 이 백트래킹은 시험 패턴의 0번째 프레임에 도달할때까지 계속되어진다. 이와같이, 최적의 단어경계들에 대해서, 음절 또는 단어의 중단 프레임의 열(역순에서)을 달성한다.

$$N, F(N), F(F(N)), \dots \quad (3-8)$$

그리고, 최적의 단어 열에 대해서, 음절 또는 단어의 지수 열(역순에서)을 달성한다.

$$T(N), T(F(N)), T(F(F(N))), \dots \quad (3-9)$$

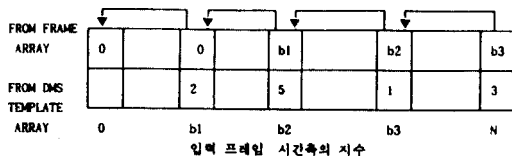


그림. 3.3 백트래킹 절차
Fig. 3.3 The backtracking procedure

이제까지 설명한 One-Stage DMS/DP를 구현하기 위한 배열들은 다음과 같이 정의되어진다.

1) 누적 거리들의 열 배열

$$D(j, k) \cong D(i, j, k) \quad (3-10)$$

2) 백포인터들의 열 배열

$$B(j, k) \cong B(i, j, k) \quad (3-11)$$

3) "from DMS template" 배열

$$T(i) = \arg \min \{D(i, j(k), k), k = 1, \dots, K\} \quad (3-12)$$

4) "from frame" 배열

$$F(i) = B(j(T(i)), T(i)) = B(i, j(T(i)), T(i)) \quad (3-13)$$

그림 3.4는 One-Stage DMS/DP 알고리즘의 상세한 흐름도를 나타낸다.

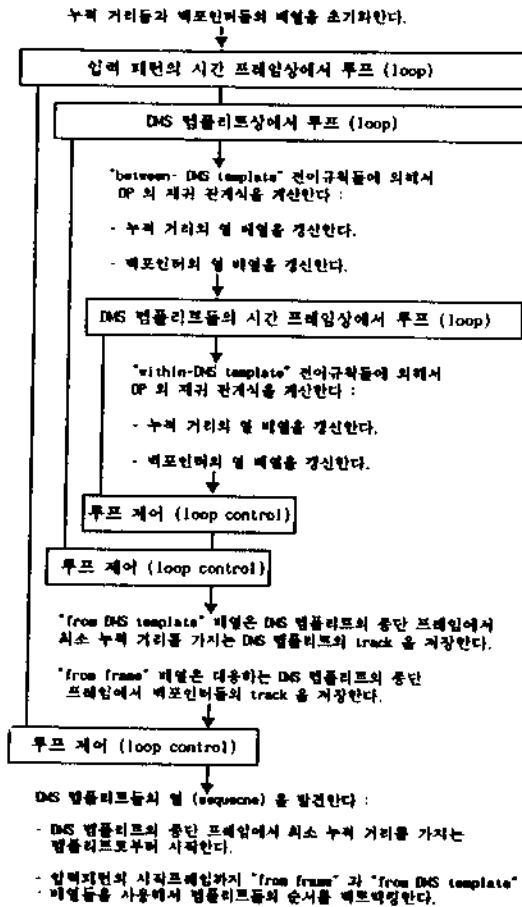


그림. 3.4 One-Stage DMS/DP 알고리즘의 흐름도
Fig. 3.4 Flow chart of One-Stage DMS/DP algorithm

IV. 실험 결과 및 고찰

4.1 음성 데이터 베이스와 DMS 템플리트 작성 과정

4.1.1 특징 추출 과정

음성 데이터는 300Hz~3.4KHz BPF를 통과해서 8KHz 샘플링과 16bit A/D 변환이 되고 pre-empha-

sis, Hamming window를 거쳐 128 샘플마다 10차의 자기상관계수, LPC 계수가 구해지고 최종적으로 특징 벡터인 10차의 LPC 캐스트럼 계수가 구해진다[6][7].

4.1.2 음성 DB 구축과 DMS 모델의 생성과정

4.1.2.1 음성 데이터 베이스 구축

음성 데이터 베이스는 광운대학교 교내 부서명 150 단어(표 4.1)를 20대의 남성 10명의 화자가 발성한 2 음절~9음절의 단어로 구성되어있다. 이 중 5명(A, B, C, D, E 화자)이 3번 발성한 데이터는 DMS 템플릿(첫번째와 두번째 발성음)와 화자중속의 실험(세번째 발성음)을 수행하기 위해서 사용되었고, 또 다른 5명(F, G, H, I, J 화자)이 두번 발성한 데이터는

표 4.1 광운대학교내 150개 부서명
Table 4.1 150 department name in Kwang Woon university

번호	단어명	번호	단어명	번호	단어명
1.	총장실	25.	기술관리과	49.	전기과 학생회
2.	비서실	26.	공과대 학생실	50.	생묘과 학과장실
3.	부총장실	27.	공과대 교학과	51.	생묘과 사무실
4.	기획실장실	28.	공과대 P.C실	52.	생묘과 학생회
5.	기획과	29.	공과대 교수휴게실	53.	농신과 학과장실
6.	홍보과	30.	공과대 학생회	54.	농신과 사무실
7.	출판부	31.	이과대 학생실	55.	농신과 학생회
8.	역사관관실	32.	이과대 교학과	56.	농신과 학과장실
9.	교부처장실	33.	이과대 P.C실	57.	화공과 사무실
10.	교무과	34.	이과대 교수휴게실	58.	화공과 학생회
11.	학석과	35.	이과대 학생회	59.	제이과 학과장실
12.	학생처장실	36.	인사대 학생실	60.	제이과 사무실
13.	학생과	37.	인사대 교학과	61.	제이과 학생회
14.	상학과	38.	인사대 P.C실	62.	환경과 학과장실
15.	취업정보실	39.	인사대 교수휴게실	63.	환경과 사무실
16.	양호실	40.	인사대 학생회	64.	환경과 학생회
17.	예비교본부	41.	전자과 학과장실	65.	건축과 학과장실
18.	총부처장실	42.	전자과 사무실	66.	건축과 사무실
19.	총부과	43.	전자과 학생회	67.	건축과 학생회
20.	경리과	44.	전자과 학과장실	68.	전과과 학과장실
21.	관리처장실	45.	전자과 사무실	69.	전과과 사무실
22.	관리과	46.	전자과 학생회	70.	전과과 학생회
23.	전기실	47.	전자과 학과장실	71.	전산과 학과장실
24.	전화보수실	48.	전자과 사무실	72.	전산과 사무실

표 4.1 계속
Table 4.1 continued

번호	단어명	번호	단어명	번호	단어명
73.	전산과 학생회	99.	법학과 사무실	125.	내학신문사
74.	수학과 학과장실	100.	법학과 학생회	126.	영자신문사
75.	수학과 사무실	101.	신방과 학과장실	127.	교육방송국
76.	수학과 학생회	102.	신방과 사무실	128.	웅원단
77.	물리과 학과장실	103.	신방과 학생회	129.	무선국
78.	물리과 사무실	104.	산업심리과 학과장실	130.	체육부
79.	물리과 학생회	105.	산업심리과 사무실	131.	학군단
80.	화학과 학과장실	106.	산업심리과 학생회	132.	교학실
81.	화학과 사무실	107.	경영정보과 학과장실	133.	숙직실
82.	화학과 학생회	108.	경영정보과 사무실	134.	경미실
83.	이문과 학과장실	109.	경영정보과 학생회	135.	수위실
84.	이문과 사무실	110.	대학원 원장실	136.	노동조합
85.	수문과 학생회	111.	대학원 교학과	137.	우체국
86.	영문과 학과장실	112.	산업대학원 원장실	138.	농향회 사무실
87.	영문과 사무실	113.	산업대학원 교학과	139.	총학생회
88.	영문과 학생회	114.	경영대학원 원장실	140.	총여학생회
89.	경영과 학과장실	115.	경영대학원 교학과	141.	총대의원회
90.	경영과 사무실	116.	전산대학원 원장실	142.	복지위원회
91.	경영과 학생회	117.	전산대학원 교학과	143.	명예위원회
92.	무역과 학과장실	118.	관정실	144.	교직원집일
93.	무역과 사무실	119.	사시과	145.	동아리 연합회
94.	무역과 학생회	120.	경교열람실	146.	전산원 원정실
95.	행정과 학과장실	121.	내출실	147.	전산원 교학과
96.	행정과 사무실	122.	전자계산소	148.	전산원 학생회
97.	행정과 학생회	123.	전자계산소 소장실	149.	전산원 교수실
98.	법학과 학과장실	124.	전산교육실	150.	안내

화자중속의 실험과 발성음의 변화에 따른 인식률의 변화를 측정하기 위해서 사용되었다. 음성 데이터 베이스에서 첫번째 발성된 음성 데이터는 느린 발성속도에서 정확하게 조음되어서 발성된 음성 데이터이고 두번째 또는 세번째 발성된 음성 데이터는 빠른 발성속도에서 자연스럽게 조음되어서 발성된 음성 데이터이다.

4.1.2.2 DMS 모델을 이용한 DMS 템플릿 작성 과정

연결어 인식은 기준 패턴을 정하는 인식 단위에 의해서 많은 영향을 받기 때문에 기준패턴을 정하는 문

제는 연결어 인식에서 중요한 문제이다[9]. 그래서, 본 논문에서는 네가지의 인식단위에서 DMS 템플릿을 정하는 방식을 설정하여 인식 실험을 수행하고 중규모 어휘의 인식에 효율적인 기준 패턴의 단위를 실험적으로 증명해 본다. DMS 템플릿을 생성하는 네가지 형태는 다음과 같다.

- 1) DMS 템플릿의 형태 I : 단음절로만 구성된 DMS 템플릿(99개의 DMS 템플릿)
- 2) DMS 템플릿의 형태 II : 단어의 분맥에 의해서 음절의 위치를 고려한 음절 DMS 템플릿(133개의 DMS 템플릿)
- 3) DMS 템플릿의 형태 III : 음절과 단어들로 분리된 DMS 템플릿(95개의 DMS 템플릿)
- 4) DMS 템플릿의 형태 IV : 단어들로만 구성된 DMS 템플릿(96개의 DMS 템플릿)이 각각의 DMS 템플릿은 유성의 단어와 단어사이의 묵음(silence) 부분을 처리하기 위해서 묵음의 DMS 템플릿도 포함하고 있다.

화자 5명 (A, B, C, D, E 화자)의 첫번째와 두번째 발생된 단어들은 우선 음성 파형, ZCR(Zero Crossing Rate)과 스펙트로그램상에서 핸드-세그멘테이션(hand segmentation)에 의해 레이블링되어지고, 그 다음에 그림 4.1과 같은 절차로 각각의 형태의 DMS 템플릿들이 작성되어진다.

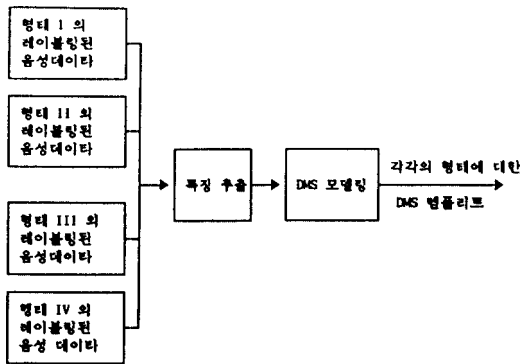


그림. 4.1 레이블링된 학습용 데이터의 DMS 템플릿 추출과정 블록도

Fig. 4.1 The Block diagram of the DMS template extraction process of labeled training data

4.2 인식 시스템의 구성

인식 실험을 위한 음성 인식 시스템은 디지털 신호 처리 전용 프로세서인 TMS320C30을 이용한 TMS-

320C30 시스템 보드 [10][11]와 IBM PC/486을 이용하여 그림 4.2와 같이 구성하였다.

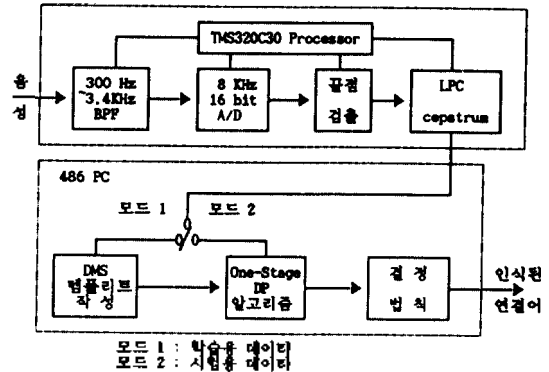


그림. 4.2 인식 시스템 구성도

Fig. 4.2 The organization diagram of recognition system

4.3 인식 실험

전이규칙의 가중치계수와 인식단위의 변화에 의한 예비 실험에서 대칭형의 가중치계수와 단어 단위의 DMS 템플릿에 기초한 결과가 가장 높은 인식률을 나타내었고, 이 결과를 바탕으로 단어 단위의 DMS 템플릿에 대해서 10명의 화자에 대해서 화자 종속, 화자독립 실험과 발생률에 따른 인식 실험을 수행하였다. 실험된 결과는 다음 표 4.2, 4.3, 4.4에 나타나 있다.

다중-화자 종속의 실험에서 구간수를 증가시킨수록 인식률이 상승되어지는 결과를 얻었고, DMS 템플릿의 구간수를 20 구간으로 정하였을때 평균 97.2%의 가장 좋은 인식률을 나타내었다. 화자독립 실험에서 느린 발생률인 경우의 인식 정확도는 DMS 템플릿의 구간수를 20 구간으로 정하였을때 79.2%를 나타내었고 빠른 발생률인 경우의 인식 정확도는 86.8%를 나타내었다. 여기서, 20 구간 이상의 인식 실험 결과는 실시간적인 개념에서 생각할때 무의미하므로 무시해 버렸다.

4.4 실시간 음성 다이얼링 모델 시스템의 구현

음성 다이얼링 모델 시스템[7][12-14]은 TMS320C30, PC-486 그리고 다이얼 모델로 이루어지고 크게 두개의 부분으로 구성되어진다. 하나의 부분은 인식 부분이고 또다른 부분은 다이얼링 부분이다. 인식 부분은 TMS320C30 시스템 보드와 PC-486으로

표 4.2 DMS 템플리트 형태 IV의 구간수 변화율에 대한 인식 결과(다중 화자 종속인 경우의 정확도)

Table 4.2. Recognition result according to the section variation of DMS template type IV(In case of multi-speaker dependent)

(단위: %)

화자	구간수	10	11	12	13	14	15	16	17	18	19	20
화자 A		58.0	62.0	66.7	71.3	74.7	86.7	87.3	90.0	92.0	93.3	94.0
화자 B		75.3	76.7	79.3	83.3	86.7	90.7	92.0	93.3	94.7	94.7	95.3
화자 C		84.0	86.7	89.3	93.3	94.7	97.3	97.3	98.0	98.0	98.7	99.3
화자 D		75.3	79.3	84.7	88.0	90.0	94.7	96.0	96.0	96.7	98.0	99.3
화자 E		64.7	71.3	80.0	82.7	84.7	89.3	91.3	93.3	96.7	97.3	98.0
평균인식률		71.5	75.2	80.0	83.7	86.2	91.7	92.8	94.1	95.6	96.4	97.2

표 4.3 DMS 템플리트 형태 IV의 구간수 변화율에 대한 인식 결과(화자독립의 느린 발생률인 경우의 정확도)

Table 4.3 Recognition result according to the section variation of DMS template type IV(In case of speaker-independent and low speaking rate)

(단위: %)

화자	구간수	10	11	12	13	14	15	16	17	18	19	20
화자 F		28.7	36.7	38.0	46.7	47.3	47.3	50.0	58.0	59.3	64.7	66.7
화자 G		38.7	43.3	59.3	65.3	67.3	70.0	71.3	73.3	74.7	76.7	78.0
화자 H		52.0	54.7	56.7	60.0	63.3	68.7	72.0	74.7	76.7	82.0	82.7
화자 I		34.7	45.3	48.0	70.0	73.3	74.0	76.7	79.3	84.0	84.7	88.0
화자 J		40.7	53.3	52.7	54.7	58.7	60.0	69.3	78.7	78.7	80.0	80.7
평균인식률		39.0	46.7	50.9	59.3	61.9	64.0	67.9	72.8	74.7	77.6	79.2

표 4.4 DMS 템플리트 형태 IV의 구간수 변화율에 대한 인식 결과(화자독립의 빠른 발생률인 경우의 정확도)

Table 4.4 Recognition result according to the section variation of DMS template type IV(In case of speaker-independent and high speaking rate)

(단위: %)

화자	구간수	10	11	12	13	14	15	16	17	18	19	20
화자 F		31.1	42.0	60.7	80.0	83.3	88.7	90.0	90.7	91.3	91.3	93.3
화자 G		58.0	65.3	68.7	70.0	72.7	80.0	80.0	80.0	80.7	80.7	81.3
화자 H		53.3	58.7	60.7	71.3	74.0	76.0	80.0	80.0	82.7	84.0	85.3
화자 I		61.3	66.7	72.0	74.7	83.3	90.0	90.0	90.7	90.7	91.3	92.0
화자 J		34.0	40.0	50.0	59.3	64.7	70.7	74.7	78.7	80.0	81.3	82.0
평균인식률		47.6	54.5	62.4	71.7	75.6	81.8	82.9	84.0	85.1	85.7	86.8

구성되어지고 입력 음성이 들어오면 TMS320C30 시스템 보오드에 의해서 음성 인식의 전처리 과정이 수행되어 특징이 추출되고, 이 파라미터를 PC-486에 넘겨주게된다. 음성 인식의 실시간 처리를 위해서 음성의 입력과 동시에 TMS 보오드에 의해서 실시간에 특징이 추출되어진다. 입력 패킷을 넘겨받은 PC-486은 인식전에 PC 상에서 학습(training)되어서 메모리에 저장되어 있는 DMS 템플리트와 One-Stage DMS/DP 알고리즘을 수행해서 총 누적거리값이 계산되어지고 이 격자점에서 백트래킹을 수행하여 인식된 단어열을 찾아낸다. 이 인식된 단어열은 PC-486상에 메모리에 저장되어있는 전화번호표(즉, 부서명과 전화번호로 구성된 단어표)와 비교를 하여 매칭되어지는 부서명의 전화번호를 찾아내어 음성 다이얼링 모델 시스템의 다른 부분인 다이얼링 부분에 전달하여 실질적으로 다이얼링을 수행한다. 전체적인 음성 다이얼링 모델 시스템의 간단한 블록도는 그림 4.3에 보여진다.

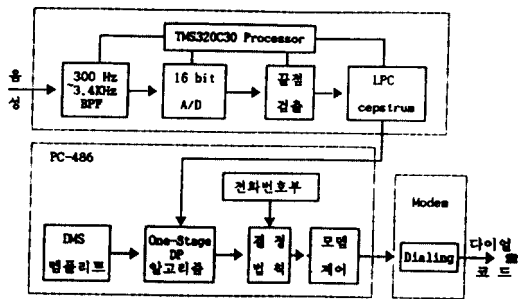


그림. 4.3 음성 다이얼링 모델 시스템의 구성도
Fig. 4.3 Organization diagram of Voice dialing model system

실제로 음성 다이얼링 모델 시스템을 구현할때 음절단위 DMS 템플리트를 가지고 다이얼링시스템을 구현할때에는 2~4초안에 다이얼링이 되는 반면에 인식률이 저조해서, 이 모델을 가지고 모델링을 할때에는 많은 문제가 발생되어지는 불합리성 때문에 인식률이 매우 높은 단어단위 DMS 템플리트를 가지고 음성 다이얼링 모델 시스템을 구현한다. 그러나 단어단위로 음성 다이얼링 모델 시스템을 구현할때에는 10~20초의 다이얼링 시간이 걸리기 때문에 이 문제점을 해결하기 위한 하나의 방안으로 150 단어를 7개의 subsection(즉, 공파대, 이과대, 인사대, 대학원, 행정부서, 학생단체, 기타)으로 나누어서 인식을 수

행하는 방법과 인식부분도 DSP 전용 칩인 TMS320C30 시스템 보오드에서 실시간으로 수행하는 방법들이 연구되어졌다.

개선 방법중 인식 부분도 TMS 시스템 도오드에서 수행하는 방법은 TMS 보오드의 메모리가 부족한 이유로 구현상에 문제가 있기 때문에 두번째 개선 방법인 150개의 부서명을 7개의 subsection으로 나누어서 다이얼링하는 방법을 구현하였다. 이 방법은 7개의 subsection으로 구분하여 우선 subsection명을 발생하고 다음에 이 section에 포함되지 않는 부서명을 발생하는 것에 의해서 실시간에 인식을 수행한다. 그래서, subsection으로 구분되어있기 때문에 DMS 템플리트의 수가 감소되어서 비교시간과 인식시간이 감소되어지는 효과가 생겨 전체 다이얼링 시간은 약 7~14초가 소요되어진다.

V. 고 찰

본 논문에서는 기준패턴의 템플리트를 단음절단위, 위치를 고려한 단음절단위, 음절과 단어를 포함하는 단위, 단어단위로 나누어서 실험을 수행하여 인식 결과를 얻었다. 이 인식결과로부터 음절단위와 위치를 고려한 단음절단위는 하나의 입력패턴을 인식하는데 약 2~4초의 적은 소요시간이 필요하지만 인식률이 저조하다는 결과를 얻었다. 이 이유는 음절과 음절과 상호조음효과를 적절히 처리해주는 기법의 부재에서 기인한다. 그리고, 위치를 고려한 음절 단위는 위치를 고려하지 않는 음절 단위보다 인식률이 향상되었는데 이것은 음절의 위치를 고려한 음절단위가 음절 상호간에 어느 정도 상호조음효과를 고려해 주었기 때문이다. 이 음절의 상대적인 위치를 고려한 음절 단위는 보통의 음절 단위보다 DMS 템플리트의 수가 많기 때문에 비교시간의 증가로 인식시간은 약 3~5초가 소요된다. 그리고, 음절과 단어를 포함하는 DMS템플리트 형태 3은 형태 1, 형태 2 보다는 인식률이 향상되지만 여전히 음절과 음절 또는 단어와 음절사이에 상호조음 효과의 영향을 완전히 배제하지 못하는 현상이 있으며 인식시간은 보통 6~7초 정도가 소요되는 것으로 나타났다. 마지막으로 단어단위에 입각한 DMS 템플리트 형태 4는 다중화자중속 실험, 화자독립 실험에서 각각 97.2%, 86.8%의 높은 인식률을 보이지만 인식하는데 걸리는 시간은 약 10~20초 정도 소요되는 결과를 얻었고, 개선방법을 적용하였을때 약 7~14초가 소요된다. 그리

고, 발성률에 따르는 실험은 DMS 템플리트를 사용했을 때 너무 느린 발성률을 가지고 발성한 음성 데이터는 인식률에 분체가 있기 때문에 이를 보상하기 위해서 좀 더 효율적인 시간 정규화 방법과 상호 조음 효과를 고려해 줄 알고리즘의 연구가 필요하다.

이상으로부터 얻은 일반적인 실험 결과의 고찰 다음과 같다.

- 1) 인식 단위에 상관없이 DMS의 구간수를 증가시키면 인식률은 증가한다. 그러나, 구간수의 증가에 비례하여 인식시간이 증가하기 때문에 구간수와 인식률에, 그리고 인식시간에 대한 tradeoff를 결정할 필요가 있다.
- 2) One-Stage DMS/DP 알고리즘의 가중치 계수는 단독어 인식에서 상대적으로 높은 인식률을 나타낸 대칭형의 가중치 계수가 연결어 인식에서도 비대칭 계수보다 높은 인식률을 보인다.
- 3) 인식 단위의 변화에 따른 실험에서 부단어 단어 중 단위로 인식이 진행되어질 때 인식률과 인식시간은 증가한다. 실시간에 인식을 수행하기 위해서는 부단어 단위에 입각한 인식 단위로 기준패턴을 정하는 것이 합리적이지만 이것을 실현하기 위해서는 상호조음효과와 시간 정규화에 대해서 더 많은 연구가 선행되어야만 하겠다. 본 논문에서 제안하고 싶은 상호조음효과의 해결책은 부단어 단위의 전이부분에 해당되는 음성의 특성을 제거하여 부단어 단위의 안정부분만 One-Stage DMS/DP 알고리즘을 적용하여 총 누적 거리를 구하는 방법과 다중-DMS 템플리트를 사용해서 상호조음효과를 해결하는 방법들을 제안한다. 그리고, 시간 정규화 분체를 해결하기 위해서는 기울기 제약조건의 연구가 선행되어야만 하겠다.
- 4) 실험에 의해서 단어 단위에 입각한 DMS 템플리트는 화자종속과 화자독립의 모든 실험에서 높은 인식률을 나타내었다. 여기에서, 주목할 사항은 DMS 템플리트를 생성할 때 화자의 수를 증가시킬수록 인식률이 증가하는 결과를 얻었다. 이 결과는 DMS 템플리트의 고유 특성(즉, 여러 화자의 특성벡터의 평균 값인 대표 특성벡터를 음성의 기준패턴으로 정한다는 특성)과 One-Stage DMS/DP 알고리즘의 우수성에 기인한 결과이지만 앞으로 화자의 수를 더 증가시켜서 일반적인 사실을 유추해내는 실험이 수행되어 이론적으로 이 사실을 증명하는 방법이 모색되어야

만 하겠다.

V. 결 론

본 논문에서는 실시간으로 중규모 어휘의 연결어를 인식하는 음성 다이얼링 모델 시스템을 구현해 보았다.

적용된 연결어 인식 알고리즘은 DP 기법중 One-Stage DP 방법을 응용한 알고리즘으로써 기준패턴의 템플리트 직성에 DMS 알고리즘을 혼합한 One-Stage DMS/DP 방법이다. 기존의 연결어 인식 알고리즘보다 인식시간, 저장장소와 인식률에서 효율적인 제안된 One-Stage DMS/DP 방법은 음성 다이얼링 모델 시스템을 구현하기 위한 최적의 실험조건을 정하기 위해서 인식 실험에 적용되었다. 우선, 광운대학교 교내 부서명 150 단어를 인식 대상 어휘로 선정하여 기준패턴의 DMS 템플리트의 단위들을 다양하게 변화시키고, One-Stage DMS/DP 알고리즘의 전이규칙의 가중치 계수를 변화시키면서 인식 실험을 수행하였다. 실험 결과는 대칭형의 가중치 계수를 적용한 20 구간의 단어 단위에 입각한 DMS 템플리트에 대해서 화자 종속, 화자 독립 실험에서 각각 평균 97.2%, 86.6%의 매우 높은 인식 결과를 얻었고 이 결과를 토대로 음성 다이얼링 모델 시스템을 구현했을 때 전체 전체 다이얼링 시간은 약 7~14초가 소요되어진다.

앞으로, 실시간으로 연결어를 인식하는 실질적인 음성 다이얼링 모델 시스템을 구현하기 위해서는 부단어 단위에 입각한 DMS 템플리트를 기준패턴으로 정하는 방법의 연구가 진행되어야 한다. 따라서, 부단어 단위에서 효율적인 시간 정규화 방법과 부단어 사이의 상호조음 효과를 처리해 줄 수 있는 실질적인 알고리즘의 연구가 선행되어야만 하겠다.

참 고 문 헌

1. H. Sakoe, "Two-level DP-matching-A dynamic programming based pattern matching algorithm for connected word recognition," IEEE Trans. on Acoust, Speech, Signal Processing, vol. ASSP-27, pp.558~595, Dec.1979
2. 윤종관, "한국어 음성 인식 시스템 개발 연구," 한국 과학 기술원, 최종 보고서, 1988.1.
3. C. S. Myers and L. R. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected

Word Recognition," IEEE Trans. on ASSP, Vol. ASSP-29, No.2, pp.284~296, April 1981

4. H. Ney, "The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition," IEEE Trans. on ASSP, Vol. ASSP-32, No.2, pp.263~271, April 1984
5. C.S. Myers and L. R. Rabiner, "A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected Word Recognition," The Bell System Technical Journal, Vol.60, No.7, pp.1389~1409, September 1981
6. 변용규, "DMS 모델을 이용한 단독어 인식에 관한 연구," 박사학위 논문, 광운대학교 대학원, 1991, 2.
7. 이합섭, "실시간 음성 다이얼링 시스템 구현을 위한 단독어 인식에 관한 연구," 광운대학교 대학원, 1991.12.
8. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP 26, No.1, pp.43-49, Feb. 1978.
9. M.K. Brown, M.A. McGee, L.R. Rabiner, and J.G. Wilpon, "Training Set Design for Connected Speech Recognition," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-39, No.6, pp. 1268~1281, June 1991
10. Third Generation TMS320 User's Guide, Texas Instrument, Inc., Houston, 1988
11. TMS320C30 PC SYSTEM BOARD USER MANUAL, Loughborough Sound Images Ltd. The Technology Centre, England, Ver. 1.0, Jan. 1990.
12. M. Immendorfer, "Voice Dialer," Electrical Communication, Vol.59, No.3, 1985
13. A. Fukui, Y. Fujihashi and F. Nakagawa, "Signal Processor Application to Voice Dialing Equipment," Proceedings ICASSP 86, TOKYO, pp. 337~340, 1986
14. S.M. Miller, D.P. Morgan, H.F. Silverman, M.N. Karam and N.R. Dixon, "A Real-Time Evaluation System for a Real-Time Connected-Speech Recognizer," Proceeding ICASSP-87, Dallas, TEXAS, pp.801~804, 1987

▲김 천 영(학생회원)

1967년 6월 7일생



1991년 2월 : 광운대학교 전자계산기공학과 졸업(공학사)

1993년 2월 : 광운대학교 대학원 전자계산기공학과 졸업(공학석사)

1993년 3월~현재 : 국방과학연구소(ADD)

※주관심분야 : 음성신호처리, Object-Oriented DB, 실시간 처리 등

▲양 진 우(동신회원)

1959년 9월 30일생



1982년 2월 : 원광대학교 전자공학과 졸업(공학사)

1985년 2월 : 광운대학교 대학원 전자공학과 졸업(공학석사)

1991년 8월~현재 : 광운대학교 대학원 전자계산기공학과 박사과정

1993년 7월~현재 : 광운대학교 전자계산기공학과 전임교수

※주관심분야 : Speech Signal Processing, Neural Networks, Speech DataBase 등임

▲유 형 근 : 제 10권 1호 참조

▲이 형 준 : 제 10권 1호 참조

▲홍 진 우 : 제 11권 1E호 참조

▲이 강 성 : 제 11권 1E호 참조

▲안 태 옥 : 제 10권 1호 참조

▲김 순 협 : 제 10권 1호 참조