

## 구문 분석과 One-Stage DP를 이용한 연속 숫자음 인식에 관한 연구

### A Study on the Recognition of Continuous Digits using Syntactic Analysis and One-Stage DP

안 태 옥\*  
(Tae Ock Ann\*)

#### 요 약

본 논문은 음성 다이얼링 시스템 구현을 위한 연속 숫자음 인식에 관한 연구로써, 구문 분석을 이용한 One-Stage DP에 의한 음성 인식 방법을 제안한다.

인식 실험을 위해 우선 구간 구분화 알고리즘을 이용하여 DMS(Dynamic Multi-Section) 모델을 만들며, 제안된 구문 분석을 이용한 One-Stage DP 방법으로 실험 대상의 연속 숫자음 데이터들을 인식하게 하였다.

본 연구에서는, 8명의 남성 화자에 의해 2-3번 발음된 21종의 7자리의 연속 숫자음이 사용되었고, 기존의 One-Stage DP와 제안된 구문 분석을 이용한 One-Stage DP 음성 인식 알고리즘을 사용해서 화자 종속과 화자 독립 실험을 실험실 환경에서 수행하였다.

인식 실험 결과, 기존의 방법보다 제안된 방법이 인식률이 좋은 것으로 나타났으며, 제안된 방법에서는 화자 종속과 화자 독립 실험에서 각각 약 91.7%, 89.7%로 나타났다.

#### ABSTRACT

This paper is a study on the recognition of continuous digits for the implementation of a voice dialing system, and proposes a method of speech recognition using syntactic analysis and One-Stage DP.

In order to perform the speech recognition, first of all, we make DMS model by section division algorithm and let continuous digits data be recognized through the proposed One-Stage DP method using syntactic analysis.

In this study, 7 continuous digits of 21 kinds which is pronounced by 8 male speakers two or three times, are used. The speaker dependent and speaker independent recognition are performed with the above data by way of the conventional One-Stage DP and the proposed One-Stage DP using syntactic analysis under the condition of laboratory environment.

From the recognition experiments, it is shown that the proposed method was better than the established method. And, the recognition accuracy of speaker dependence and independence by the proposed One-Stage DP using syntactic analysis was about 91.7% and 89.7%.

---

\*전북 산업대학교 컴퓨터공학과  
접수일자: 1995년 6월 2일

## I. 서 론

인간은 기계를 만들고 편의를 위해 여러가지 기능을 추가하여왔다. 이에 따라 작동법을 모르는 사람에게 있어서는 기계를 작동하는데 더 큰 어려움이 뒤따르게 되었는데 음성을 이용하면 이런 어려움을 극복할 수 있다. 이와같이 음성을 정보 교환의 수단으로 사용하려면 기계가 음성을 인식하는 것이 선행되어야 하는데, 음성 인식이란 인간의 음성을 기계가 알아듣게 하는 것이다. 이에 반해, 음성 합성이란 기계가 사람처럼 말을 할 수 있게 하는 것을 의미한다.

본 연구는 음성 인식의 일부분인 연속 숫자음 인식에 대한 논문으로, Ney<sup>5)</sup>의 논문에 의하면 One-Stage DP 알고리즘은 Sakoe의 Two-Level DP 알고리즘의 계산 시간의 1/25 걸리고 Myers와 Rabiner의 Level-Building DTW 알고리즘<sup>7,8)</sup>의 계산 시간의 1/4이 소요되면서도 다른 알고리즘에 비해 인식률이 떨어지지 않으므로 One-Stage DP 알고리즘을 이용하며, 연속음이나 연결어의 기준 패턴을 생성하는 DMS model에 의한 DMS/VQ 및 DMS/DP 인식 기법<sup>1)</sup>이 다른 방법에 비해 인식률이 떨어지지 않고 기억 장소 및 계산 시간이 적게 소요되므로 본 연구에서는 음성을 인식함에 있어서는 One-stage DP 방법<sup>4,5)</sup>에 DMS/DP 인식 기법<sup>2,3)</sup>을 적용하였으며 숫자음은 10개가 반복되어 발음된다는 점에 착안한 구문분석의 개념을 도입하여 One-Stage DP 방법 보다 인식률이 향상되리라는 개념에서 구문 분석을 이용한 One-stage DP 방법을 제안하였다.

여기서, DMS(Dynamic Multi-section) 모델<sup>1)</sup>은 VQ(Vector Quantization)<sup>9)</sup> 개념을 확장시킨 MSVQ(Multi-Section VQ)방법<sup>10,12)</sup>이 각 구간을 등간격으로 분할하여 프레임수를 동일한 길이로 잡아주는 것과는 달리 구간의 경계를 동적으로 잡아 각 구간의 프레임수를 가변길이를 잡아 주는 모델이다.

즉, DMS 모델은 비슷한 특징을 나타내는 벡터들을 한 구간에 포함시키고 서로 비슷하지 않은 특징을 나타내는 벡터들은 다른 구간에 분할하는 것으로 음성 패턴의 연속된 시계열 상에서 서로 비슷한 특징을 가지는 벡터들을 모아 한 구간으로 만들어 주는 가변길이를 구간을 구분하는 방법이다.

또한, 본 논문의 구문 분석 개념은 10자로 구성된 연속 숫자음을 대상으로 할 수 있는데 본 연구에서는 실험 대상으로 7자리의 전화 번호를 택했으며, 8명의

20대 남성 화자의 발음을 이용하였다. 본 연구의 대상인 연속 숫자음은 일상 생활에서 가장 많이 말하여지는 것으로 쉽게 응용할 수 있는 분야는 전화기 및 송강기 등이다. 즉, 손으로 단추를 누르는 것보다 목소리를 이용함으로써 잘못되는 일이 적으며, 시각장애자 및 노인들에게도 편리하고 손을 쉽게 움직일 수 없는 사람에게도 도움이 될 수 있다.

## II. 이 론

### 1. 개념 도입

본 연구에서는 전화 번호를 그 인식 대상으로 삼았는데 실험에서 사용되어지는 음성 데이터는 서로 다른 21개 데이터로 20대 남성 화자 8명이 2-3번씩 발음한 데이터이다. 그런데, 본 연구에서 발음된 데이터의 내용은 3자리의 국번과 4자리의 번호로 이루어졌으며, 전화 번호를 발음하는 관계로 국번과 번호 사이에 "에"라는 말을 첨가시켜 발음한 데이터이다.

따라서, 7자리의 숫자음과 "에"라는 발음으로 각각의 데이터가 구성되는 관계로 숫자음의 특성상 서로 다른 7자리의 숫자음이 나올 수 있는 종류는 각 음절마다 10개씩이므로 무려  $10^7$ 개의 단어가 나올 수 있다. 따라서, 이것을 단어 단위의 인식 방법을 도입해서 인식한다면 엄청난 시간을 낭비할 것이다. 따라서, 음소 단위나 음절 단위의 음성 인식 방법을 도입하여 인식하여야 실시간 처리가 가능할 것인데 10개의 단음절로 이루어진 숫자음을 음소 단위로 인식한다는 것은 오히려 인식해야할 음소의 종류가 많아 음절 단위의 인식 방법에 비해 인식 시간이 많이 걸리고, 음소 단위의 인식이 음절 단위의 인식 방법에 비해 인식률이 저하되는 관계로 본 연구에서는 음절 단위의 인식 방법을 선택하였다.

따라서, 본 연구에서는 연결 단어 인식에 좋은 알고리즘으로 평가되고 있는 One-Stage DP 알고리즘을 음절단위의 연속 숫자음 인식에 적용시켰다. 그러나, 이 One-Stage DP 알고리즘을 연속 숫자음에 그대로 적용시키면 상호 조음결합 관계를 고려할 수 없다. 그러므로, 본 연구에서는 연속 숫자음을 3부분으로 나누어 조음 효과를 고려하였다. 즉, 첫 음절로 발음되는 것, 중간 음절로 발음되는 것들과 맨 마지막 음절로 발음되는 것으로 나누어 특징을 추출하여 구문 분석 방법을 적용하는 알고리즘을 도입하여 인식한다.

2. 연속음의 구문 분석법

본 연구에서의 구문 분석은 하향식 방법을 응용하여 사용한다. 이는 어떤 한 상태에서 한개의 입력 심볼에 대해 현재의 다음 상태를 가지는 결정적 유한 오토마타(Deterministic Finite Automata : DFA)를 사용하여 표시하면 식 (1)과 같다.

$$DFA\ M = (\{S_r, S_i, S_e\}, \{a, b, c\}, \delta, S_r, \{S_e\}) \quad (1)$$

$$\delta : \delta(S_r, a) = S_i,$$

$$\delta(S_i, b) = S_i,$$

$$\delta(S_i, c) = S_e$$

식 (1)에 대한 상태전이표(Transition Table)는 다음과 같이 유한 오토마타의 이동함수를 매트릭스 형태로 나타내고 있다. 매트릭스 행과 열은 각각 상태와 입력 심볼을 나타내고 행과 열이 교차하는 위치는 다음 상태를 나타낸다.

식 (1)을 상태천이도로 표시하면 그림 1과 같이 나타낼 수 있다.

표 1. DFA의 상태천이표

Table 1. Transition table of DFA

$\delta$	a	b	c
$S_r$	$S_i$	$\emptyset$	$\emptyset$
$S_i$	$\emptyset$	$S_i$	$S_e$
$S_e$	$\emptyset$	$\emptyset$	$\emptyset$



그림 1. DFA 상태천이도  
Fig 1. Transition diagram of DFA

타낼 수 있다.

연속 숫자음을  $S_r, S_i, S_e$ 과 같이 세 단계로 나눈다. 첫음절, 중간 음절들, 마지막 음절로 나누는데 이는 음절 사이에 발생하는 조음 결합 현상을 고려한 것이다. 첫 음절 다음에 따라오는 중간 음절과 조음 결합을 일으키고, 중간 음절은 앞과 뒤 음절 모두와 조음 결합을 일으킨다. 마지막 음절은 앞 음절하고만 조음 결합 현상을 일으킨다.

식에 나타나는 각각의 기호들은 이들의 위치와 입력되는 음절들을 나타낸다. 즉,  $S_r, S_i, S_e$ 은 각각 첫 음절, 중간 음절, 마지막 음절이고 a, b, c는 그 상태들에 대한 입력 유절들이다. a는 첫 음절에 올 수 있는 {공, 일, 이, ..., 구}를 나타내고, b는 중간 음절에 올 수 있는 {공, 일, 이, ..., 구, 예}를 나타내며, c는 마지막 음절에 올 수 있는 {공, 일, 이, ..., 구}로써 세분하면 음절을 31개로 나누고 있다.

그림 1을 세분하면 그림 2와 같이 나타낼 수 있다.

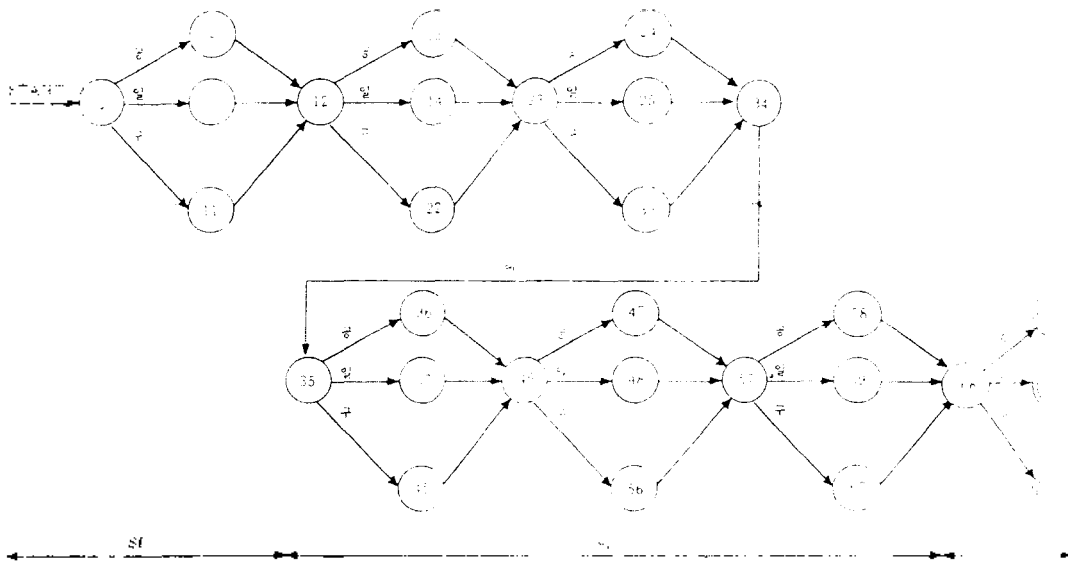


그림 2. 세분된 DFA 상태천이도  
Fig 2. Transition diagram of subdivision DFA

### 3. 제안된 구문 분석을 이용한 One-Stage DP 알고리즘

#### 1) 알고리즘의 선행단계

제안된 구문 분석을 이용한 One-Stage DP 알고리즘을 상세히 설명하기 전에 One-Stage DP 알고리즘의 기본적인 개념들을 정리하면 다음과 같다.

①  $j=1, \dots, N$  개의 시간 프레임으로 구성된 미지의 입력 단어열의 시험 패턴의 특징 추출

② 입력 단어열에 대응되는 기준패턴은 단음절로 구성된  $k$  개의 기준 패턴, 즉, 템플리트의 집합으로 구성되어진다. 즉, 하나의 템플리트마다 DMS 모델링에 의해 각 구간별로 구한 codeword를 포함하고 있고 지속시간 정보는 기준 패턴을 생성할 때에는 이용하지 않는다. 이후로부터 기준패턴 또는 템플리트를 DMS 모델링에 의해 구하기 때문에 DMS 템플리트라고 정의한다.

③ DMS 템플리트  $k$  의 시간 프레임은  $j=1, \dots, J$  로서 표시되고 이때  $J$ 는 DMS 템플리트의 길이(구간수)를 의미한다.

④ 인식목표는 입력 단어열과 가장 잘 매칭되는 DMS 템플리트들의 접속, 즉,  $q(1), \dots, q(R)$ 을 결정하는 것이다. 여기서,  $q(1), \dots, q(R)$ 을 "super" 기준 패턴이라고 한다.

#### 2) 제안된 알고리즘

One-Stage DP 알고리즘의 기본적인 개념은 앞절에서 설명되었고, 이를 바탕으로한 연속 숫자음의 구문 분석을 이용한 One-Stage DP 음성 인식 알고리즘은 다음과 같다.

알고리즘을 작성하는데 있어서, 레이블링된 데이터를 살펴본바 가장 짧은 데이터가 19 프레임이었고 가장 긴 데이터가 45 프레임이었으므로 이를 감안하여 실험에서 가장 짧은 시험 데이터의 첫음절이 15 프레임 이하이고 50 프레임 이상인 것은 없다는 가정하에 실제적인 알고리즘에서는 15 프레임에서부터 50 프레임까지 첫 음절과 중간음절이 공존 할 수 있는 것으로 생각하여 작성하였으며, 중간 음절과 마지막 음절은 총 프레임보다 50프레임 적은 프레임에서 총 프레임보다 15프레임 적은 프레임까지 공존할 수 있는 것으로 생각하여 작성하였다.

본 논문에서 사용한 알고리즘은 다음과 같다

#### <알고리즘>

단계 1) 첫 음절 부분의 초기화를 행한다.

a)  $m=1$ 로 설정한다.

b)  $i=1$ 로 설정한다.

c)  $k=1, \dots, K(m)$ 에 대해서, 단계 1d-1e를 행한다.

d)  $j=1$ 에 대해서, 다음을 행한다.

$$D(i, j, k, m) = d(i, j, k, m) \quad (2)$$

e)  $j=2, \dots, J$ 에 대해서, 다음을 행한다.

$$D(i, j, k, m) = d(i, j, k, m) + D(i, j-1, k, m) \quad (3)$$

단계 2) 첫음절 부분의 최소 총누적 거리를 계산한다.

a)  $x=1$  및  $y=1$ 을 설정한다

b)  $i=2, \dots, 15$ 에 대해서, 단계 2c를 행한다.

c)  $k=1, \dots, K(m)$ 에 대해서, 단계 2d-2e를 행한다.

d)  $j=1$ 일 때,

$$D(i, j, k, m) = d(i, j, k, m) + \min \begin{cases} D(i-1, j, k, m^*) \\ D(i-1, J, k^*, m^*) \end{cases} \quad (4)$$

$m^* = x, \dots, y$  반복

$k^* = 1, \dots, K(m)$  반복

e)  $j=2, \dots, J$ 에 대해서, 다음을 행한다.

$$D(i, j, k, m) = d(i, j, k, m) + \min \begin{cases} D(i-1, j, k, m) \\ D(i-1, j-1, k, m) \\ D(i, j-1, k, m) \end{cases} \quad (5)$$

단계 3) 중간 음절 부분을 초기화 한다.

a)  $m=2$ 로 설정한다.

b)  $i=15$ 으로 설정한다.

c)  $k=1, \dots, K(m)$ 에 대해서, 단계 3d-3e를 행한다.

d)  $j=1$ 에 대해서, 다음을 행한다.

$$D(i, j, k, m) = d(i, j, k, m) + \min \{ D(i, j-1, k^*, m-1) \} \quad (6)$$

$k^* = 1, \dots, K(m-1)$  반복

e) 단계 1e를 행한다.

단계 4) 첫음절 부분과 중간 음절 부분의 최소 총누적 거리를 계산한다.

a)  $i=16, \dots, 50$ 에 대해서, 단계 4b-4c를 행한다.

b)  $m=1, x=1$  및  $y=1$ 로 설정하고 단계 2c를 행한다.

c)  $m=2$ ,  $x=1$  및  $y=2$ 로 설정하고 단계 2c를 행한다.

단계 5) 중간 음절 부분만의 최소 총누적 거리값 계산한다.

- a)  $m=2$ ,  $x=2$  및  $y=2$ 를 설정한다
- b)  $i=51, \dots, N-50$ 에 대해서, 단계 2c를 행한다.

단계 6) 마지막 음절 부분을 초기화 한다.

- a)  $m=3$ 로 설정한다.
- b)  $i=N-50$ 으로 설정하고 단계 3c를 행한다.

단계 7) 중간 음절 부분과 마지막 음절 부분의 최소 총누적 거리를 계산한다.

- a)  $i=N-50, \dots, N-15$ 에 대해서, 단계 7b-7c를 행한다.
- b)  $m=2$ ,  $x=2$  및  $y=2$ 로 설정하고 단계 2c를 행한다.
- c)  $m=3$ ,  $x=2$  및  $y=3$ 로 설정하고 단계 2c를 행한다.

단계 8) 마지막 음절 부분만의 최소 총누적 거리를 계산한다.

- a)  $m=3$ ,  $x=3$  및  $y=3$ 를 설정한다
- b)  $i=N-15, \dots, N$ 에 대해서, 단계 2c를 행한다.

단계 9) 누적 거리들의 배열  $D(i, j, k, m)$ 를 사용해서 최소 총누적 거리값을 가지는 DMS 템플리트의 중단 프레임이 있는 격자점으로부터 가장 최적의 경로를 백트래킹한다.

<기 호>

- $m$ : 음절의 위치를 나타내며, 1이며 첫 음절, 2이며 중간 음절들, 3이며 마지막 음절을 나타낸다.
- $i$ : 시험 패턴의 프레임수를 나타낸다
- $k$ : 음절 위치별 DMS 템플리트의 번호
- $j$ : DMS 템플리트의 구간수
- $d(i, j, k, m)$ :  $m$ 번째 음절 위치의  $k$ 번의 DMS 템플리트의  $j$ 번째 구간의 특징 벡터와  $i$ 번째 프레임의 특징 벡터간의 거리값
- $D(i, j, k, m)$ :  $m$ 번째 음절 위치의  $k$ 번의 DMS 템플리트의  $j$ 번째 구간의 특징 벡터와 1번째 프레임까지의 최소 총누적 거리값

$x$ 와  $y$ : 알고리즘에서 음절의 위치가  $x$  음절에서  $y$  음절까지의 총누적 거리값을 다 구해야 할 경우 사용되는 기호

제안된 방법에 의한 알고리즘의 최종적인 목표는 입력 단어들의 비지의 열을 결정하는 것이고 이것은 위에서 설명된 알고리즘의 단계 9에서 백트래킹을 수행하는 것에 의해서 결정되어진다.

여기서, 위의 기호에서 설명한 것처럼  $m(m=1, \dots, 3)$ 은 음절의 위치를 나타내며 1인 경우가 첫 음절을 의미하고 2인 경우가 중간 음절들, 3인 경우가 마지막 음절을 의미한다.  $i(i=1, \dots, N)$ 는 입력(시험) 패턴의 특징벡터를 의미하고,  $j(j=1, \dots, J)$ 는 각 DMS 템플리트의 구간을 의미하며,  $k(k=1, \dots, K[m])$ 는 각 부분(첫 음절, 중간 음절 및 마지막 음절 부분)의 어떤 숫자음 음절을 의미한다. 여기서,  $K[1]$ 과  $K[3]$ 는 첫음절에 숫자음만 오는 관계로 10이고,

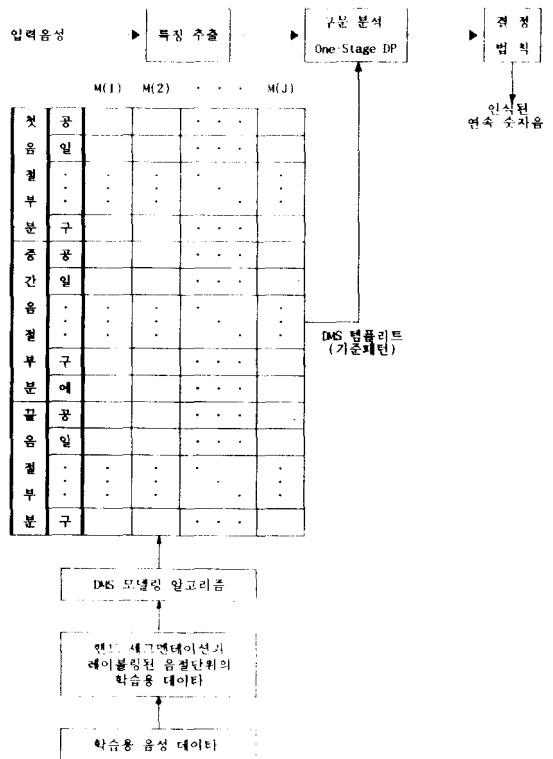


그림 3. 구분 분석과 One-Stage DP를 이용한 전체 인식 시스템의 블록도.

Fig 3. Block Diagram of overall recognition system using One-Stage DP and syntactic analysis.

KI2]는 숫자음 10개와 '에'가 오는 관계로 11이다.

시간 프레임에 대해서 동적 프로그래밍을 재귀적으로 수행하기 위해서, 누적 거리의 배열  $D(i, j, k, m)$ 의 집합이 필요하다. 격상 장소의 이 열은 누적 거리의 열 배열로써 간단하게 언급되었고,  $D(i, k, m)$ 로서 표시되어진다. 이와 같이, 배열  $D(i, j, k, m)$ 는 단지 하나의 열만을 사용해서 동적 프로그래밍의 재귀 관계식이 시험 패턴의 시간축을 따라서 수행되어지고 격자점과 격자점을 따라서 누적 거리의 열 배열을 갱신한다.

시험 패턴과 DMS 템플릿 사이에서 최소 총 누적 거리를 구하는데 있어 연속어에서는 백트래킹이 입력 음성 단어들 사이의 열을 발견하기 위해서 One-Stage DP 알고리즘에 적용되어진다. 백트래킹 정보는 동적 프로그래밍의 재귀식이 평가되어지는 동안에 저장되어지고, 격자점  $(i, j, k, m)$ 를 경유하는 경로상에서 DMS 템플릿 k에 있는 j 프레임은 단일한 시작점에서 시작되어진다.

그래서, 각각의 격자점에 대해서 백포인터(back pointer)는 격자점  $(i, j, k, m)$ 에 경로를 가지는 선행하는 단어의 종단 프레임(ending frame)의 값을 저장한다.

앞에서 기술한 것처럼, 본 논문에서는 배열  $D(i, j, k, m)$ 의 경우에서처럼 백포인터 배열을 저장할 수 있고, 가장 최적인 경로의 격자점과 격자점에서 백포인터의 값을 갱신한다. 그림 3은 제안된 구분분석과 One-Stage DP를 이용한 전체 인식 시스템을 나타낸다.

### III. 실험 결과 및 고찰

#### 1. 실험 결과

본 논문에서 사용되어지는 대상 어휘는 숫자음절 상호간에 연속해서 일어날 수 있는 모든 경우의 수를 조합해서 구성한 Balanced list이다. 데이터 목록은 표 2과 같은 21개의 대상 어휘이다.

실험에 사용되어지는 음성 데이터는 21개의 연속 숫자음을 20대 남성 화자 8인에 의해서 2-3번씩 발성된 음성 데이터를 DMS 템플릿과 화자 종속의 실험을 수행하기 위해서 사용하였고 3명의 화자가 2번 발성된 음성 데이터는 화자 독립의 실험을 수행하기 위해서 사용되어졌다. 그리고, 전체적으로 음성

데이터들은 보통(일상 생활에서 사용되어지는 사람의 대화를 행하는 정도의 발생 속도)의 발생 속도에 서 녹음되어 저장되어졌다.

표 2. 연속 숫자음 데이터 목록.

Table 2. Data list of the continuous digits.

순번	국번	전화 번호	순번	국번	전화 번호
1	512	0257	12	270	9483
2	630	1349	13	396	0011
3	745	6780	14	408	6281
4	826	9318	15	689	6542
5	904	0371	16	209	1921
6	910	2388	17	147	3324
7	843	4616	18	986	5066
8	729	5522	19	569	1775
9	607	7641	20	795	9785
10	358	8736	21	448	1231
11	153	0599			

#### 1) 인식 시스템의 구성

DMS 모델을 이용한 DMS 템플릿 작성 과정"은 그림 4에 나타내었으며, 제안된 인식 실험을 위한 음성 인식 시스템은 그림 5와 같이 구성하였다.

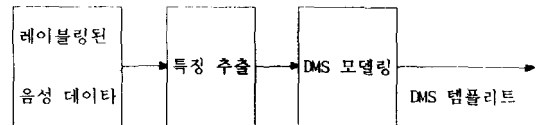


그림 4. 레이블링된 학습 데이터의 DMS 템플릿 추출 과정 블록도.

Fig 4. The Block diagram of the DMS template extraction process of labeled training data

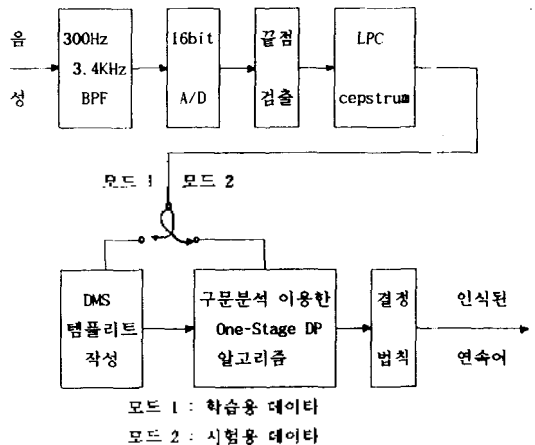


그림 5. 제안된 인식 시스템의 구성도

Fig 5. The block diagram of the proposed recognition system

2) 인식 실험

가) 기존의 One-Stage DP 방법에 의한 인식 실험

8명의 화자 중 5명의 화자가 발성한 2회 발음으로 모델을 만들고 나머지 발음으로 실험을 행한 화자 종속의 실험 결과가 표 3에 나타내어졌다. 실험 결과 12 구간으로 나누었을 때 가장 좋은 82.3%의 인식률을 보였다.

표 3. 구간수 변화에 따른 실험 결과

Table 3. Recognition result according to the section variation (단위:%)

구간수	6	8	10	12	14	16
화자 A	72.8	73.5	74.1	77.6	77.2	77.2
화자 B	73.5	73.5	73.5	78.9	74.1	72.1
화자 C	93.1	93.1	93.9	95.2	95.2	94.5
화자 D	77.5	77.1	80.2	80.2	79.6	76.9
화자 E	73.4	77.5	78.9	79.6	77.5	79.6
평균	78.1	79.0	80.1	82.3	80.7	80.1

또한, 나머지 3명의 화자의 발음으로 실험을 행한 화자 독립의 실험 결과가 표 4에 나타내어졌다. 실험 결과 12구간으로 나누었을 때 가장 좋은 80.6%의 인식률을 보였다.

표 4. 구간수 변화에 따른 실험 결과

Table 4. Recognition result according to the section variation (단위:%)

구간수	6	8	10	12	14	16
화자 F	59.2	74.1	74.8	78.9	79.6	79.6
화자 G	76.4	76.9	78.6	81.6	81.2	82.4
화자 H	76.9	78.3	79.8	81.2	78.6	79.3
평균	70.8	76.4	77.7	80.6	79.8	79.8

나) 제안된 구문분석을 이용한 One-Stage DP 방법에 의한 인식 실험

8명의 화자 중 5명의 화자가 발성한 2회 발음으로 모델을 만들고 나머지 발음으로 실험을 행한 화자 종속의 실험 결과가 표 5에 나타내어졌다. 실험 결과 12 구간으로 나누었을 때 가장 좋은 91.7%의 인식률을 보였다.

또한, 나머지 3명의 화자의 발음으로 실험을 행한 화자 독립의 실험 결과가 표 6에 나타내어졌다. 실험 결과 12구간으로 나누었을 때 가장 좋은 89.7%의 인식률을 보였다.

표 5. 구간수 변화에 따른 실험 결과

Table 5. Recognition result according to the section variation (단위:%)

구간수	6	8	10	12	14	16
화자 A	83.3	85.2	88.1	89.1	86.9	86.9
화자 B	85.7	86.3	88.1	91.7	88.7	86.9
화자 C	95.8	91.7	93.9	95.8	93.5	95.8
화자 D	83.9	83.3	82.7	89.1	85.7	82.7
화자 E	86.3	92.9	92.7	92.9	92.9	91.7
평균	87.0	87.9	89.1	91.7	89.5	88.8

표 6. 구간수 변화에 따른 실험 결과

Table 6. Recognition result according to the section variation (단위:%)

구간수	6	8	10	12	14	16
화자 F	82.7	81.6	84.5	86.9	83.3	82.7
화자 G	85.7	87.5	87.5	89.3	89.3	87.5
화자 H	84.5	89.3	91.1	92.9	91.7	89.3
평균	84.3	86.1	87.7	89.7	88.1	86.5

2. 고 찰

위의 실험 결과에서 보여지듯 본 실험은 두가지로 나누어서 비교 실험하고 있다. 기존의 방법과 제안된 방법 둘다에서 공히 DMS의 구간수를 증가하면 할수록 인식률이 증가하다가 떨어지는 경향이 있다. 기존의 방법에서나 제안된 방법에서나 화자 종속 및 화자 독립의 경우 모두 구간수 12에서 인식률이 가장 좋았다.

실험 데이터에서, 오인식되는 숫자들은 '이'가 '일'과 '칠'로, '오'가 '구'로, '삼'이 '팔'로, 그리고, '에'가 '육' 등으로 오인식되고 있다.

각각의 인식율은 기존의 방법인 경우 화자 종속인 데이터는 82.3%, 그리고 화자 독립인 데이터는 80.6%의 인식률을 보이고 있다. 기존의 방법에 비해 구문 분석 기법을 도입한 방법에서는 화자 종속인 데이터는 91.7%의 인식률을 나타내고 화자 독립인 데이터는 89.7%를 나타낸다. 기존의 방법과 제안된 방법과의 차이는 약 9% 정도로 제안된 방법이 인식률의 증가를 가져왔다.

IV. 결 론

본 논문은 구문 분석을 이용한 One-Stage DP에 의한 연속 숫자음 인식에 관한 연구로써, 기존의 One-Stage DP 알고리즘에서 문제가 제기된 바 있는 상호 조음결합 현상을 해결하기 위해 구문 분석의 개념을

도입하였다.

실현한 데이터로는 20대 남성 화자 8명이 2-3회 발성한 7자라 연속 숫자음을 사용하였다. 이 중 5명의 화자가 2번 발음한 데이터로 모델을 만들고, 나머지 1번 발음한 데이터로는 화자 종속의 실험을 하였고, 나머지 3명의 화자가 2번 발음한 데이터로는 화자 독립의 실험을 수행하였다.

본 실험은 두가지로 나누어서 비교 실험하였는데 기존의 방법과 제안된 방법 둘다에서 공허 DMS의 구간수를 증가하면 활수록 인식률이 증가하다가 떨어지는 경향을 보이고 있는데, 기존의 방법에서나 제안된 방법에서나 화자 종속 및 화자 독립의 경우 모두 구간수 12에서 인식률이 가장 좋았다.

실험 데이터에서, 오인식되는 숫자들은 살펴보면 '이'가 '일'과 '칠'로, '오'가 '구'로, '삼'이 '팔'로, 그리고, '에'가 '육' 등으로 오인식되는 것을 알 수 있다.

각각의 인식율은 기존의 방법인 경우 화자 종속인 데이터는 82.3%, 그리고 화자 독립인 데이터는 80.6%의 인식률을 보이고 있다. 기존의 방법에 비해 구문 분석 기법을 도입한 방법에서는 화자 종속인 데이터는 91.7%의 인식률을 나타내고 화자 독립인 데이터는 89.7%를 나타낸다. 기존의 방법과 제안된 방법과의 차이는 약 9% 정도로 제안된 방법이 인식률의 증가를 가져와 제안된 시스템의 우수성이 입증되었다.

연속 숫자음 인식이 현실 생활에 적용되면 ARS(Audio Response System), 장애자 시스템, 음성 다이얼링, 승강기 운용 등 응용할 수 있는 분야가 무수히 많다. 그리고 인간의 가장 간편한 통신 수단인 음성을 이용하면 편리한 점이 많다. 그러나, 아직도 생활에 적용하기 위해서는 더 높은 인식율을 요구한다.

상호 조음결합 현상으로 인한 오인식율을 줄이고, 인식률 향상을 목적으로한 이번 실험은 실험 결과 약 9%의 인식률 향상이 있었다. 그러나, 앞으로 더 인식율을 높이는 데는 상호 조음 효과를 완화시킬 수 있는 더 새로운 알고리즘의 연구와 기준 패턴을 정하는 방법의 연구가 우선되어야 할 것이다.

## 참 고 문 헌

1. Tae Ock Ann and Young Kyu Byun, "A Study on Speech Recognition using DMS Model," Acoustical Society of Korea, Vol. 13, No 2E, pp. 41-50, 1994. 12.

2. 이항섭, 실시간 음성 다이얼링 시스템 구현을 위한 단음어 인식에 관한 연구, 광운 대학교 대학원, 1991. 2.
3. 안태옥 외, 실시간 음성 다이얼링 시스템 구현을 위한 연결어 인식에 관한 연구, 한국 음향학회지, 제12권 제3호, pp. 13-25, 1993. 6.
4. Hiroaki Sakoe and Seibi 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.
5. 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
6. 안태옥, 변용규, 김순협, 구문 분석과 Level Building을 이용한 한국어 연속음 인식, 한국 음향학회지, 제5권 제4호, pp. 27-36, 1986. 12.
7. C. S. Myers, R. R. Rabiner, "Connected Digit Recognition Using a Level-Building DTW Algorithm", IEEE Trans. Acoust. Speech, signal processing, Vol. ASSP-29, pp 351-363, June 1981.
8. 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
9. Y. Linde, A. Buzo, and R. M. Gray "An algorithm of Vector Quantizer Design", IEEE Trans, Comman, Vol. COM-28 pp. 84-95, Jan 1980
10. D. K. Burton, J. E. Shore, J. T. Buck "Isolated-Word Speech Recognition Using Multisection Vector Quantization Codebooks" IEEE Trans. of Acoustics, Speech, Signal Processing Vol ASSP-33, No 4, August 1985
11. 이성권 외, 시간 정보와 VQ를 이용한 DDD 지역명 인식에 관한 연구, 한국 음향학회지, Vol. 8, No. 5, pp. 102-112, 1989. 10.
12. 안태옥, 김순협, TSVQ를 이용한 Computer 자동 음성 인식, 대한 전자 공학회지, 제 27권, 제 7호, pp. 157-165, 1990. 7.

▲안 태 옥: 제 13권 2E호 참조