

論 文

## 음성인식용 DTW PE의 IC화를 위한 MIN 회로의 설계

準會員 鄭 光 載\* 正會員 文 洪 眞\*\* 正會員 崔 圭 勳\*\*\* 正會員 金 鐘 玟\*

## An Integrated MIN Circuit Design of DTW PE for Speech Recognition

Kwang Jae JEONG\*, Hong Chin MOON\*\*, Kyu Hoon CHOI\*\*\*

Chong Kyo KIM\* *Regular Members*

**要 約** 음성 인식에서의 dynamic time warp (DTW)은 반복적 계산을 필요로 하며, 이 계산을 수행하기에 합당한 PE cell의 설계는 매우 중요하다. 따라서 이 연구에서는 실제에 가까운 실시간 어휘 인식을 가능하게 하는 large dictionary의 DTW 알고리즘을 hardware로 구현하기 위한 PE (Processing Element) cell의 설계에 주안점을 두었다.

이 DTW용 PE cell은 크게 세가지의 블록으로 대별된다. 즉 MIN block, ADD block 그리고 ABS block인데, "MIN"은 accumulated minimum distance를 계산하기 위한 블록이고 "ADD"는 이들 minimum distance들의 합을 계산하는 블록, 그리고 "ABS"는 이러한 합에 의한 local distance의 절대값을 구하기 위한 블록이다.

본 연구에서는 이들 세가지 블록중 MIN 회로의 설계 및 검증을 행하였으며, 3 $\mu$ m CMOS N-well 설계 규칙에 따라 MIN블록에 대한 레이아웃을 행한 후 설계 규칙 검사(DRC)를 마쳤다.

**ABSTRACT** Dynamic time warp(DTW) needs for iterative calculations and the design of PE cell suitable for the operations is very important.

Accordingly, this paper aims at the real time recognition design which enables large dictionary hardware realization using DTW algorithm.

The DTW PE cell is separated into three large blocks, "MIN" is the one block for counting accumulated minimum distance, "ADD" block calculates these minimum distances, and "ABS" seeks for the absolute values to the total sum of local distances.

We have accomplished circuit design and verification for the MIN block, and performed MIN layout and DRC(design rule check) using 3 $\mu$ m CMOS N-Well rule base.

\*全北大學校 電子工學科  
Dept. of Electronic Engineering Chonbuk National University

\*\*群山實業專門大學 電子計算科

\*\*\*全州工業專門大學 電子科  
論文番號 : 90-64 (接受 1990. 3. 23)

### I. 서 론

R. Bellman은 1957년에 'Dynamic Programming (DP)' 이라는 수학적 기법을 집대성하여

책으로 출판한 바 있다<sup>9)</sup>. 음성이나 화자 인식에서 화자의 speaking rate 변화로 인한 시간축상에서 음성 패턴의 비선형 변동을 어떻게 시간-규준화 시키느냐 하는 문제는 매우 중요하다.

1970년 V. M. Velichko가 이러한 비선형 변동을 처리하기 위하여 비선형 함수를 DP로 구하여 한 패턴을 다른 패턴축에 모델화시킨 이래 많은 알고리즘이 발표되고 있다. 아울러 DP의 최대 약점인 계산량을 감축시키면서도 warping 함수의 optimality를 보존하는 연구 결과들이 보고되고 있다.<sup>(12, 13)</sup>

이러한 DP 기술의 발달에 힘입어 일본, 미국 및 한국어의 숫자음에 대한 몇가지 보고에서는 대부분 95% 이상의 높은 인식률을 얻었다고 발표하고 있으나 실시간 처리용 실용화 단계에 이르기에는 크게 미흡한 상태에 있다.<sup>(14, 15, 16)</sup>

DP 기법에 의한 Dynamic Time Warp(DTW) 알고리즘은 음성 인식에 있어서 매우 효과적인 방법으로 사용되고 있다. 이 알고리즘은 미지의 발음과 기준 단어 사이의 natural timing variation을 허용함으로써 미지발음과 기준 단어 사이의 비선형 패턴 매칭을 가능토록 하여주는 exhaustive한 탐색기법이다<sup>10)</sup>.

DTW processor용 주문형 VLSI 제작을 위한 Processing Element (PE) cell의 설계는 약 200,000단어에 대해 미지 발음의 매치를 수행하는데 1초 미만의 시간을 소요하는 실시간 인식이 목표로 되어 있다. 본 연구에서는 이러한 목표를 수행하기 위하여 우선 이에 부합되는 PE cell의 "MIN" block을 회로 설계하였다. 이들 각 PE들은 종래의 systolic DTW 방식과는 다르게 digit pipelined 방식으로 array 됨으로써 1) 모든 데이터들이 base-4 format signed-digit로 표현되고 2) 이러한 데이터들은 digit serial 방식으로 PE 사이를 통과하게 되어서 3) 패턴 매칭이 digit 레벨에서 pipeline 방식으로 이루어지게 된다. 이러한 방식이 도입됨에 따라서 아주 높은 동시성과 데이터율을 유지하게 한다<sup>11)</sup>.

## II. DTW 프로세싱 소자

격리 단어 인식 프로세서는 DTW 알고리즘에 근거를 두고 있는데 이 알고리즘은 미지 음성과 기준 단어들 사이에서의 자연 시간 변화(natural timing variations)를 허용함으로써 사전내 단어들과 미지 음성사이에서의 비선형 패턴 매칭을 가능토록 하여주는 반복적인 탐색 기법이다<sup>11, 7)</sup>.

4 프레임 음성을 위한 간단한 DTW 프로세서가 그림 1에 제시되어 있다.

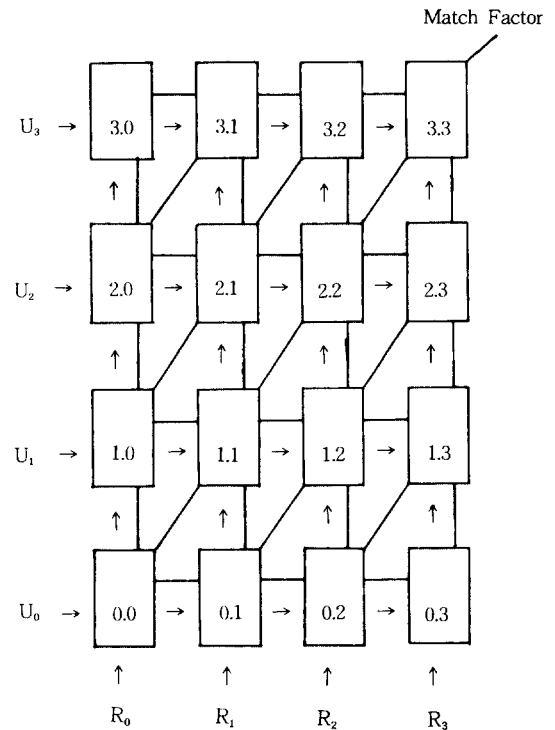


그림 1. 간단한 DTW 프로세서 (: I=4 frames)

수평축은 사전으로 부터의 기준 단어 출력 패턴을 의미하는데, 이 패턴은 연속적 시간 간격을 제외하는 R<sub>0</sub>부터 R<sub>3</sub>까지 4 프레임으로 시간 규준화되어 있다. 각 프레임에 관계된 계수들은 그 시점에서의 단어를 특정지우는 것이다<sup>12)</sup>.

마찬가지로 수직축은 미지 음성을 나타내는

데, 이 음성 패턴도 4 프레임(;  $U_0$ 부터  $U_3$ ) 계수들로써 이미 시간 기준화된 상태이다. 이 간단한 예의 DTW 프로세서는 PE(0, 0)로부터 PE(3, 3)까지 16개의 프로세싱 소자 (Processing Element ; PE)들로써 격자 모양을 이루고 있다.

각 PE는 가장 근접한 이웃(; 서, 남, 남서 방향) 입력을 받아서 이웃한 동, 북, 북동 방향으로 출력을 공급하는 등으로 주변과 연결되어 있다. DP는 PE(0, 0)로부터 PE(I-1, I-1)에 이르는 최단 거리의 경로를 찾기위해 이용된다. 경로의 최소화는 통신능력에 의하여 제한을 받는다. 경로 이동상 기울기에 제약을 가함으로써 연결 문제는 단지 최근접한 이웃 PE들 상호간의 연결만 가능할 뿐이다.

병렬 계산은 그림 1의 PE 격자상에서 systolic 파형들을 아래 왼쪽으로부터 위 오른쪽으로 나아가게 하는데, 여기서의 동일 대각선 상의 모든 PE들은 동일 파형으로 동시에 실행된다<sup>9)</sup>

경로 PE(0, 0)로부터 PE(I-1, I-1)까지의 최소 국부 거리의 합인 축적된 최소거리는 두 패턴 사이의 매치가 얼마나 근접하게 이루어졌는가를 말하여 주는 매치 인자(match factor)를 생성한다. 이러한 방식으로 매치 인자는 사전내 모든 단어에 따라 생성된다. 그림 1의 간단한 DTW 프로세서로부터 단일 PE 하나를 구체화하여 제시하면 그림 2와 같다.

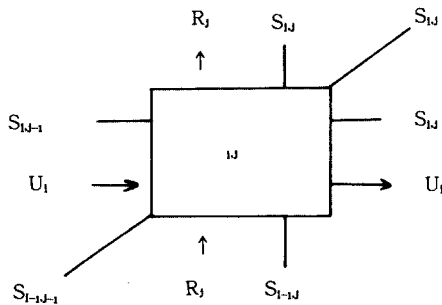


그림 2 한 개의 DTW 프로세싱 소자

그림 2의 표시 ‘-, /, |’는 이웃 PE들 사이를 통과하면서 계산된 값들의 진행경로를 나타내며, ‘→, ↑’ 표시는 이웃 PE사이를 통과하는 계수들의 진행 방향을 나타내고 있다. 일반적으로 각 PE는 다음과 같이 계산될 수 있다<sup>10)</sup>.

$$S_{i,j} = \min(S_{i-1, j} + D_{i,j}, S_{i-1, j-1} + D_{i,j}, S_{i, j-1} + D_{i,j}) \quad (1)$$

(1)식의  $D_{i,j}$ 는 미지 음성 프레임 i의 계수와 기준 단어 프레임 j의 계수 사이에서의 국부 “distance”이며,  $S_{i,j}$ 는 인접 PE들의 경로를 따라 축적된 최소 “distance”이다. 이는 (1, 1)로부터 (i, j)까지의 최적 경로상에서의 부분 합의 거리를 뜻한다. 이 연구에서 국부 거리  $D_{i,j}$ 는  $R_j$ 와  $U_i$ 차의 절대값으로 측정된다.

즉,

$$D_{i,j} = |R_j - U_i| \quad (2)$$

이밖의 국부 거리 측정법으로는 평방형 유클리드식 거리 측정과 대수 눈금거리 측정(; Itakura measure)법 등이 있다. 더욱 일반적인 형태로 (2)식을 표현하여 보면 (3)과 같다<sup>10)</sup>.

$$S_{i,j} = \min(S_{p,q} + W_{p,q} D_{i,j}) \quad (3)$$

여기서  $1 \leq p \leq i, 1 \leq q \leq j$  그리고  $(p, q) \neq (i, j)$ 이다.  $W_{p,q}$ 는 (i, j) 점에 대한 (p, q)점의 상대적 함수로 ‘weighting factor’라고 한다.

DTW 프로세서의 구성 요소인 PE 셀은 앞에서 설명한 (1)식의 계산을 수행할 수 있도록 설계되어야 한다. (1)식을 계산수행 알고리즘에 합당하게 다시 표현하면 (4)식과 같게 된다.

$$\begin{aligned} S_{i,j} &= \min(S_{i-1,j} + D_{i,j}, S_{i-1, j-1} + 2D_{i,j}, S_{i, j-1} + D_{i,j}) \\ &= \text{MIN}(\text{MIN}(\text{ADD}(S_{i-1, j}, \text{ABS}(u_i, r_j)), \text{ADD}(S_{i, j-1}, \text{ABS}(u_i, r_j))), \text{ADD}(S_{i-1, j-1}, \text{ADD}(\text{ABS}(u_i, r_j), \text{ABS}(u_i, r_j)))) \end{aligned} \quad (4)$$

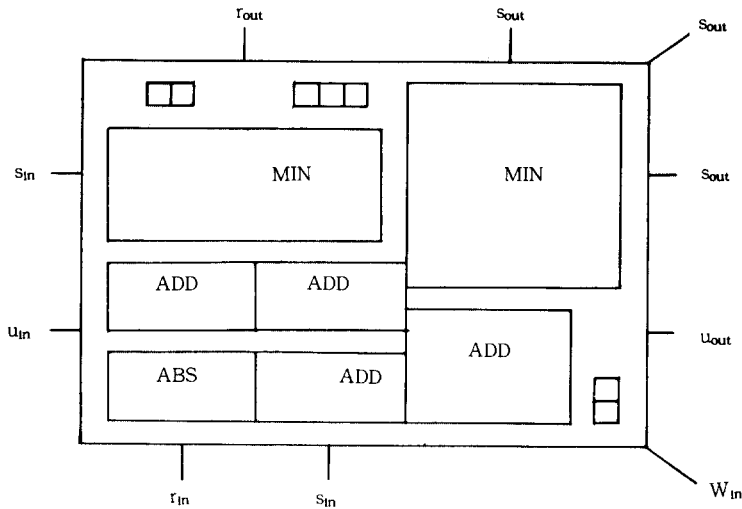


그림 3. DTW PE 내부 블록다이아그램

(4)식을 수행하기 위한 PE의 내부 블록다이아그램은 그림 3과 같다. 그림 3과 같은 PE셀은 (4)식에 의하여 동일한 절대값을 반복하여 계산하기 위한 한개의 ABS블록, 네번의 가산을 위한 네 개의 ADD블록, 그리고 가산 결과를 최소화하고 이 가산 결과와 절대값의 합을 최소화하기 위한 두 개의 MIN 블록으로 구성된다. 레지스터로는 Sout 디지털 값을 래칭하기 위한 한 개의 3비트 레지스터와 Rout 과 Uout 디지털 값을 래칭하기 위한 두 개의 2비트 레지스터가 있다. 또한 제어 신호 Win은 래치되지 않은 상태로 모든 PE 셀 사이를 동일한 역대각선 방향으로 전파되어 나간다.

이 연구에서의 데이터는 모든 PE 셀 사이를 base-4 signed 디지털 방식의 직렬 상태로 전이 되도록 설계된다. 디지털 데이터를 멀리 전송시키기 위해서는 세개의 배선이 요구되지만 계수 디지털을 전송하는 데는 단지 두 선만이 필요하고, 한 선은 신호 overflow 용이나 하드웨어적인 에러 검출용으로 쓰이게 된다.

예를 들어서 1013<sub>4</sub>의 4 디지털 값은 001 00 0 111 011<sub>2</sub>의 12비트로 encode 된다. 그런데 base-4 signed 디지털 set은 {-3, -2, -1,

0, 1, 2, 3}이어서 encoding "111<sub>2</sub>"은 7이 되어 set에 포함되지 않는다. 즉, 이 경우는 첫번째 비트 라인이 'NAN'(Not a Number) 조건에 해당되므로 이 라인을 overflow용이나 에러 검출 용으로 이용하게 된다. 따라서 실제 설계의 경우 base-4 디지털 set{0, 1, 2, 3}이 미지 음성이나 기존 단어 계수로써의 입력 값으로 표시된다<sup>1), 5)</sup>

### III. MIN 블록의 설계

그림 4는 두 입력값의 최소치를 계산하는 digit serial minimum 성분 블록이다. 두 오퍼랜드 사이에서 최소치를 찾는 일은 conventional digit set 보다 signed-digit set을 이용하여 동작 상태를 파악하는 편이 더욱 이해하기 쉽다.

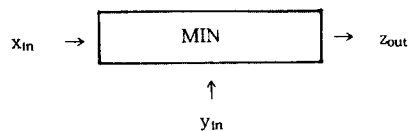


그림 4. MSD digit serial minimum 블록

그림 4에서,

$$z_{out} = \text{MIN}(x_{in}, y_{in}) \quad (5)$$

그림 5는 MIN 성분을 더욱 완벽하게 정의하기 위한 전체적인 다섯가지 상태의 MIN 동작을 보여주는 흐름선도이다.

파형 제어 신호  $W_{in}$ 은 A상태로 상태 흐름을 초기화 시키기 위하여 이용된다. 그림 5에서 알 수 있는 바와 같이 전체적인 MIN동작이 수행되려면 비교기 (comparator), 가산기 (adder), 그리고 감산기 (subtractor)의 기능이 이루어져야 한다<sup>(1,6)</sup>.

그림 6은 전체 MIN회로에 대한 블록다이아그램이다. 이 그림은 그림 5의 알고리즘에 근거하

여 설계한 회로를 계층적으로 블록화한 후, 최상위 계층의 전체를 제시한 것이다. 그리고 그림 7은 MIN 회로에 대한 기능적 블록다이아그램이다.

전체 MIN 회로는 크게 나누어 상태 수행 부분, 상태 선택 처리 부분, 그리고 데이터 선택 출력 부분 등으로 구분된다.

상태 수행 부분은 두 입력 데이터를 그림 5의 흐름에 따라 비교, 가산, 감산하기 위한 하위 계층(STA, STC, STE)을 두고 있다. 즉, STA (state A)는  $y_{in} = x_{in}$ ,  $y_{in} \leq x_{in} - 1$ ,  $y_{in} \geq x_{in} + 1$ 의 기능을 수행하기 위한 4 bit adder / subtractor, 4 bit comparator 등을 필요로 하며 표 1과 같이 출력을 낸다. 또한 STC, STE도 그림 5의 흐름에 따라 STA와 비슷한 기능을 수행하

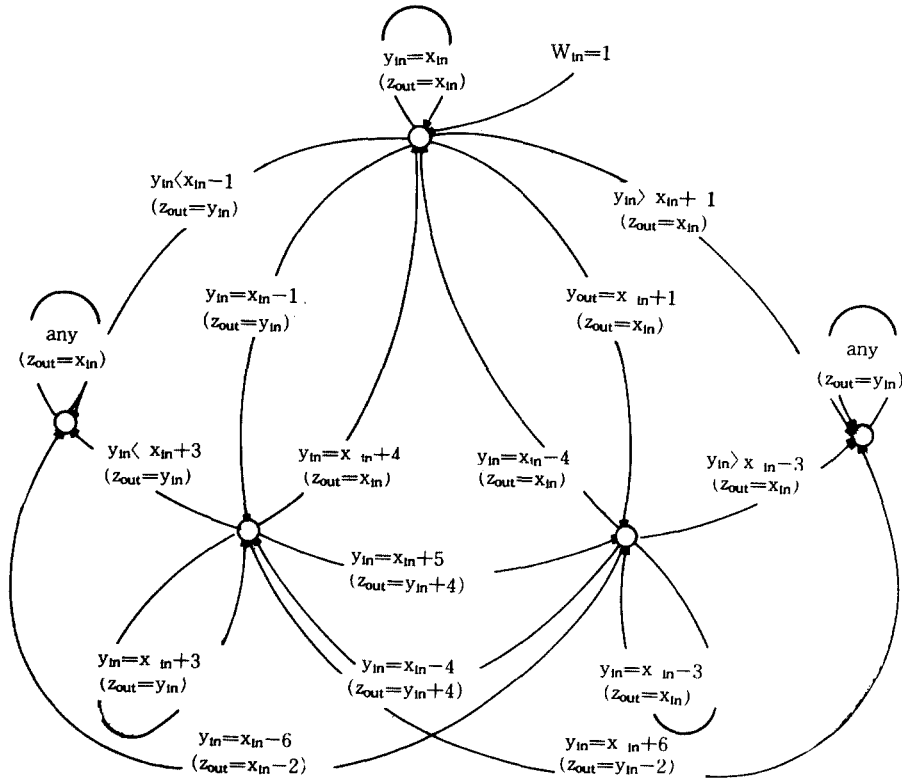


그림 5. 전체적인 MIN 동작 상태 흐름선도

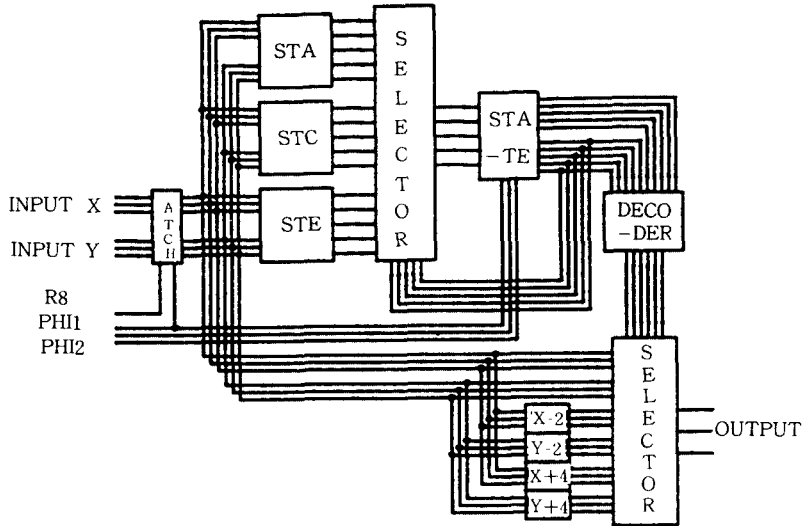


그림 6. 전체 MIN 회로 블록다이아그램

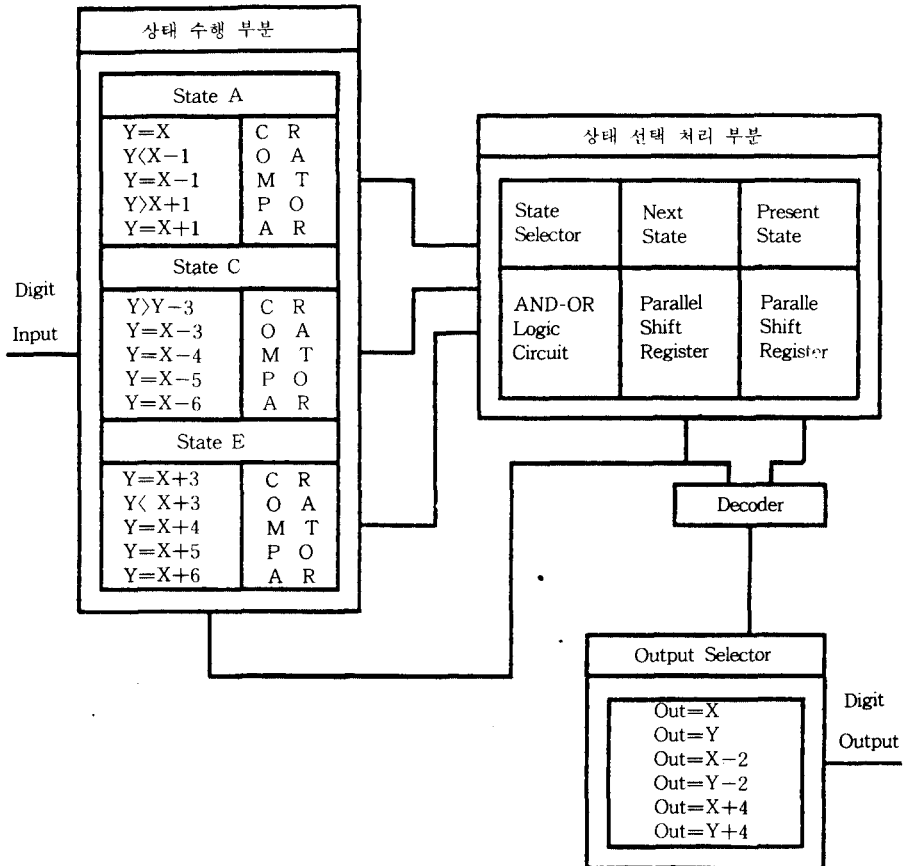


그림 7. MIN 회로에 대한 기능적 블록다이아그램

며 STB(state ㉔)와 STD (state ㉕)는 상태 선택 처리 부분의 SELECTOR에서 단지 state code만을 발생시키도록 설계되어 있다.

표1. State code

state	state code
A	1 0 0 0 0
B	0 1 0 0 0
C	0 0 1 0 0
D	0 0 0 1 0
E	0 0 0 0 1

상태 선택 처리 부분은 상태 수행 부분으로부터의 출력을 받아들여 next state를 선택하는 SELECTOR, next state와 present state를 latching 하기 위한 5 bit 2-stage parallel shifter 등으로 구성되어 있다. SELECTOR는 AND-OR

로직으로 구성되어 있으며 레지스터는 edge-triggered D f/f으로 구성되어 있다.

데이터 선택 출력 부분은 next state code와 present state code로 부터 선택 신호를 발생시키는 DECODER, DECODER의 출력에 의해 digit data( $x_{in}$ ,  $y_{in}$ ,  $x_{in-2}$ ,  $y_{in-2}$ ,  $x_{in+4}$ ,  $y_{in+4}$ )로부터 한 개의 최종적인  $z_{out}$ 을 선택하는 SELECTOR 등으로 구성되어 있다.

#### IV. 레이아웃

설계된 MIN 회로를  $3\mu\text{m}$  CMOS N-Well 설계 규칙에 따라 설계한 레이아웃 결과는 그림 8과 같다. 레이아웃 설계는 standard cell 방식의 기법을 활용하였으며, 이와 같은 레이아웃 설계 및 설계 규칙 검사(design rule check)는 한국전자통신 연구소의 자동 설계 tool을 이용하였으며

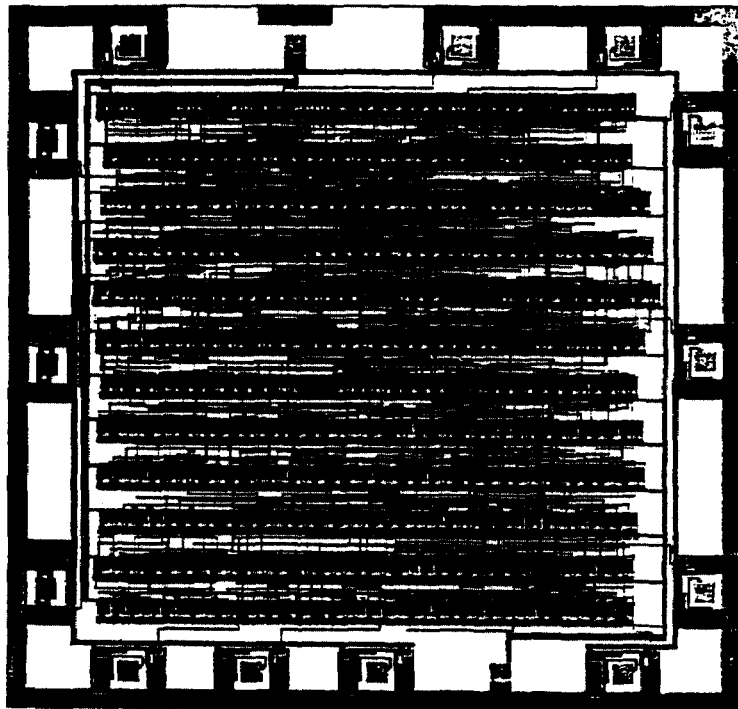


그림 8. MIN 회로의 레이아웃

SUN workstation을 활용하여 도면을 plot하였다<sup>(11)</sup>.

## V. 결 론

이 연구에서는 DTW를 이용한 음성인식 알고리즘에 대하여 고찰하였으며 특히 격리 단어 인식 문제에 중점을 두었다. 인식에서의 중요한 관점은 실시간 처리의 신속성, 활용성 계고를 위한 경제성, 그리고 편리한 이용을 위한 기능성 등을 들 수 있다. 이러한 현실적 문제의 구현을 위해서는 DTW processor의 개발이 필요한데, 이 프로세서의 Processing Element(PE) 설계 및 one chip layout은 문제 해결의 가장 중요한 부분이라고 할 수 있다.

PE 회로는 크게 세가지 블록으로 대별되어 ADD, ABS, 그리고 MIN으로 나뉜다. 본 논문에서는 이들 각각의 블록중 MIN 블록을 DTW 알고리즘에 합당하도록 회로 설계하고, 이를 레이아웃 및 검증하였다.

Base-4 signed digit pipeline 방식을 이용함으로써 종래의 parallel processing 방식과는 달리 하드웨어가 간단하여지며, 이 결과 VLSI화가 용이하게 되어 고속처리가 가능하게 됨으로써 이를 이용한 DTW processor가 완성될 경우 실시간 음성인식이 기대된다.

계속하여 나머지 블록의 회로설계 및 레이아웃을 진행중에 있으며, 이 결과를 이용하여 DTW processor용 PE 전체를 완성할 예정이다.

## 參 考 文 獻

1. M. J. Irwin, "A Digit Pipelined Dynamic Time Warp Processor", IEEE Trans, Acoust., Speech, Signal Processing, vol. 36, No. 9, pp. 1,412~1,422, Sept. 1 988.
2. N. Weste, D. J. Burr, and B. D. Ackland, "Dynamic Time Warp Pattern Matching Using an Integrated Multiprocessing Array", IEEE Trans, on Computers, vol. C-32, No. 8, pp. 731~744, Aug. 1983.
3. C. Myers and L. R. Rabiner, "A Levele Building Dynamic Time Warping Algorithm for Connected Word Recognition", IEEE Trans., Acoust., Speech, Signal Processing, vol. ASSP-29, No. 2, pp. 284~2 97, April 1981.
4. D. J. Burr, B. D. Ackland, and N. Weste, "Array Configurations for Dynamic Time Warping", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP- 32, No. 1, pp. 119~127, Feb. 1984.
5. L. R. Rabiner, A. E. Rosenberg, and S. E. Levinson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, No. 6, pp. 575~582. Dec. 1978.
6. H. Murveit and R. W. Brodersen, "An Integrated Circuit Based Speech Recognition System", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP- 34, No. 6, pp. 1,465~1,472, Dec. 1986.
7. C. Myers, L. R. Rabiner, and A. E. Rosenberg, " Performance Tradeoffs in Dynamic Time Warping Alogorithms for Isolated Word Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP- 28, No. 6, pp. 623~635, Dec. 1980.
8. F. Charot, P. Frison, and P. Quinton, "Systolic Arch- itectures for Connected Speech Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP- 34, No. 4, Aug. 1986.
9. R. Bellman, "Dynamic Programming", Princeton University Press, Princeton, New Jersey, 1957.
10. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recogni- tion", IEEE Trans. Acoust., Speech, Signal Proces- sing, vol. ASSP-26, No. 1, pp. 194~200, Feb, 197 8.
11. N. Weste and K. Eshraghian, "Principles of CMOS VLSI Design", Addison Wesley Publishing Company, 1985.
12. C. C. Tappert and Subraba K. Das, "Memory and Time improvements in a Dynamic Programming Algorithm for Matching Speech Patterns", IEEE Trans., Acoust., Speech, Signal Processing. vol. ASSP-26, pp. 583~586, Dec. 1978.
13. Fumitada Itakura, "Minimum Prediction Residual Principle Applied to Speech Recgonition", IEEE

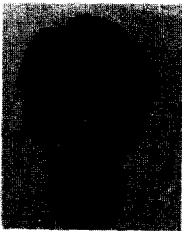


Trans., ASSP-23, No. 1, pp. 67~72, Feb. 1975.

14. L. R. Rabiner, "A Speaker-independent digit recognition System", The BSTJ. vol. 54, No. 1, pp. 81~102, Jan. 1975.
15. H. Sakoe and S. Chiba, "A Similarity Evaluation of Speech Patterns by Dynamic Programming",

presented at the Dig. 1970 Nat. Meeting, Inst. Electron. Comm. Eng. Japan, pp. 136, July 1970.

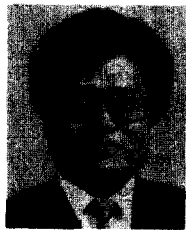
16. 한희, 김순협, 박규태, "특정 대역 에너지를 이용한 한국어 기본 숫자 음성의 자동인식에 관한 연구", 대한전자공학회지, 제19권, 제3호, 6월. 1982년.



**鄭 光 載(Kwang Jae Jeong)** 準會員  
1968年 4月18日生  
1989年 2月: 全北大學校 電子工學科 卒業  
1989年 3月~1990年 現在: 全北大學校 電子工學科 碩士課程  
※主關心分野: 디지털 신호처리 및 CAD



**文 洪 眞(Hong Chin Moon)** 正會員  
1960年 7月3日生  
1983年 2月: 全北大學校 電子工學科 卒業  
1986年 8月: 全北大學校 大學院 電子工學科 工學碩士  
1990年 2月: 全北大學校 大學院 電子工學科 博士課程 修了  
1988年10月~現在: 郡山實業專門大學 電子計算課 全任講師  
※主關心分野: 디지털信號處理 및 컴퓨터 그래픽스



**崔 圭 勳(Kyu Hoon Choi)** 正會員  
1950年 7月11日生  
1974年 2月: 光云工科大学 電子工學課 卒業  
1976年 2月: 延世大學校 産業大學院 電子工學科 工學碩士  
1989年 8月: 全北大學校 電子工學科 博士課程 修了  
1981年 3月~現在: 全州工業專門大學 電子科 助教授  
※主關心分野: 디지털信號處理 및 필터링



**金 鐘 莪 (Chong Kyo Kim)** 正會員  
1944年 8月22日生  
1966年 2月: 全北大學校 電氣工學科 卒業  
1977年 8月: 全北大學校 大學院 電氣工學科 工學碩士  
1983年 8月: 全北大學校 大學院 電氣工學科 工學博士  
1979年 4月~現在: 全北大學校 電子工學科 副教授  
1966年 7月~1972年 5月: 遞信部 勤務 (장거리 通信擔當)  
1976年 3月~1979年 4月: 全州工業專門大學 電子科  
1983年 8月~1984年 8月: 美國 일리노이 工科大学 (IIT) 客員教授  
※主關心分野: 디지털信號處理, CAD 및 컴퓨터 그래픽스