

음성인식을 위한 Dynamic Time Warp 소자의 설계

正會員 崔圭勳* 正會員 金鍾玟**

Design of Dynamic Time Warp Element
for Speech RecognitionKyu 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 실시간 어휘 인식을 가능하게 하는 대사전 규모의 대어휘 고립단어인식용 DTW알고리즘을 하드웨어로 구현하기위한 PE(Processing Element)소자의 설계에 주안점을 두었다.

이 DTW용 PE 소자는 크게 세가지의 블록으로 대별된다. 즉, "MIN" block, "ADD" block 그리고 "ABS" block인데, "MIN"은 축적 최소거리를 계산하기 위한 블록이고, "ADD"는 이들 최소거리들의 합을 계산하는 블록, 그리고 "ABS"는 이러한 합에 의한 국부거리의 절대값을 구하기 위한 블록이다.

본 연구에서는 이들 세가지 전체 블록회로의 설계 및 검증을 행하였으며, 1.2 μ m CMOS N-well 설계 규칙에 따라 전체블록에 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 real time recognition design enables large dictionary hardware realization using DTW algorithm.

The DTW PE cell 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.

Circuit design and verification about the three blocks have been accomplished, and performed layout & DRC (design rule check) using 1.2 μ m CMOS N-Well rule base.

* 全州工業專門大學 電子科

** 全北大學校 電子工學科

Dept. of Electronics Engineering, Chonbuk National University

論文番號: 93210

接受日字: 1993年 11月 4日

I. 서 론

인간은 음성을 이용하여 자신의 의사를 이웃에게 전달한다. 이렇게 인간과 인간사이의 전달수단이었던 음성이 이제는 인간과 기계사이의 통신수단(man-machine communication)으로 쓰이게 되었다. 인간과 의사가 소통되는 컴퓨터 내지는 그러한 회로에 의한 하드웨어의 개발은 이제 현실로 다가오고 있다.

R. Bellman이 1957년 'Dynamic Programming(DP)'라는 수학적 기법을 제시한 후, DP 기술에 관한 많은 연구가 진행되고 있다. 특히 음성인식에서 화자의 speaking rate 변화로 인한 시간축 상에서 음성 패턴의 비선형 변동을 어떻게 시간 정규화시켜 인식률을 개선하느냐 하는 문제는 매우 중요하다.[7]

1970년 V.M. Velichko가 이러한 비선형 변동을 처리하기 위하여 비선형함수를 DP로 구하여 한 패턴을 다른 패턴축에 모델화 시킨이래 많은 알고리즘이 발표되고 있다. 아울러 DP의 최대 약점인 계산량을 감축시키면서도 warpin함수의 optimality를 보존하는 연구 결과들이 보고되고 있다.

이러한 DP 기술의 발달에 힘입어 일본, 미국 및 특히 한국의 일부 모음과 숫자음에 대한 몇가지 보고에서는 대부분 95%이상의 높은 인식률을 얻었다고 발표하고 있으나 아직도 실시간 처리를 실용화 단계에 이르기에는 미흡한 상태에 있다.

1970년대 초부터 음성인식과 음성합성에 관한 수많은 연구가 진행되어왔고 디지털 신호처리 기술의 발달에 힘입어 컴퓨터에 의한 인간의 음성인식은 점차 가능하게 되었으나 완벽한 음성인식 시스템은 아직 구현되어 있지 않다. 완벽한 음성인식에서의 난점은 화자간의 발음 차이, 연속음성에 있어서의 조음현상, 대어휘 인식에 따르는 대용량 데이터베이스의 구성, 그리고 제한되어있지 않은 문법등으로 크게 나눌 수 있다.

현재 음성인식은 Dynamic Time Warping(DTW)을 이용하는 방법외에도 Hidden Markov Model(HMM)법, Fuzzy이론을 이용하는 방법, 신경회로망을 활용하는 방법등 여러가지 방면에서 음성의 인식을 시도하고 있다. 이러한 연구에서도 아직은 화자에 무관한 완전한 음성인식시스템을 구현하지 못하고 있으며, 항공사의 좌석 예약시스템이나 날씨 예보, 전화번호 안내등 제한된 범위에서 음성의 인식과 합성이 가능한 제품들이 소개되고 있는 상태이다. 특

히 음성인식에 사용되는 시스템들은 다중프로세서(multiprocessor) 개념을 도입하여 실시간 처리가 가능하도록 하고 있다.

DTW알고리즘을 이용한 음성인식의 구현에 관한 연구는 1980년대에 들어서면서부터 활기를 보이고 있다. 특히 실시간 인식을 위한 알고리즘 개발과 아울러 이를 위한 하드웨어의 구현은 그중에서도 가장 높은 관심거리이다. 이와 같이 실시간 처리를 실현시키기 위하여 이 연구에서는 DTW음성인식용 하드웨어소자(DTW소자)를 설계하고 이를 집적화하기 위하여 레이아웃하였다.

DP기법에 의한 DTW알고리즘은 음성인식에 있어서 매우 효과적인 방법으로 사용되고 있다. 이 알고리즘은 미지의 발음과 기준 단어사이에서의 natural timing variation을 허용함으로써 미지발음과 기준단어사이의 비선형 패턴 매칭을 가능토록 하여 주는 exhaustive한 탐색 기법이다.

DTW프로세서용 주문형 VLSI 제작을 위한 Processing Element(PE)소자의 설계는 약 200,000단이에 대해 미지 발음의 매치를 수행하는데 1초 미만의 시간을 소요하는 실시간 인식이 목표로 되어 있다.

본 연구에서는 이러한 목표를 수행하기 위하여 우선 이에 부합되는 PE소자를 설계하고자 하여 PE의 세가지 하위 블록회로를 설계하였다. 이 PE들은 종래의 systolic DTW 방식과는 다르게 digit pipelined 방식으로 배열됨으로써 1) 모든 데이터들이 base4 format signed-digit로 표현되고 2) 이러한 데이터들은 digit serial 방식으로 PE사이를 통과하게 되어서 3) 패턴 매칭이 digit 레벨에서 pipeline방식으로 이루어지게 된다. 이러한 방식이 도입됨에 따라서 아주 높은 논리실과 데이터율을 유지하게한다.[12]

II. DTW 프로세서

고립단어인식 프로세서는 DTW알고리즘에 근거를 두고 있는데 이 알고리즘은 미지 음성과 기준 단어들 사이에서의 자연 시간 변화(natural timing variations)를 허용함으로써 사전대 단어들과 미지 음성사이에서의 비선형 패턴 매칭을 가능토록 하여주는 반복적인 탐색 기법이다.

4 프레임 음성을 위한 간단한 DTW 프로세서가 그림1에 제시되어 있다.

수평축은 사전으로 부터의 기준 단어 출력 패턴을 의미하는데, 이 패턴은 연속적 시간 간격을 제시하는

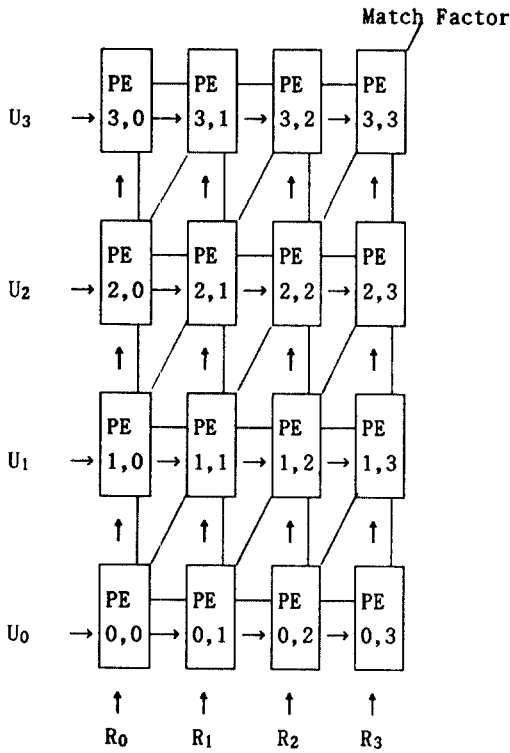


그림 1. 간단한 DTW 프로세서(: I = 4 frames)
 Fig. 1. A simple DTW processor(: I = 4 frames)

R₀부터 R₃까지 4프레임으로 시간 정규화되어 있다. 각 프레임에 관계된 계수들은 그 시점에서의 단어를 특징지우는 것이다.[3]

마찬가지로 수직축은 미지 음성을 나타내는데, 이 음성 패턴도 4 프레임(U₀부터 U₃) 계수들로써 이미 시간 정규화된 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 통신 능력에 인하여 제한을 받는다. 경로 이동상 기율기에 제약울 가함으로써 연결 문제는 단지 최근접한 이웃 PE들 상호간의 연결만 가능하다. 병렬 계산은 그림1의 PE 격자상에서 systolic 파형들을 아래 왼쪽으로

부터 위 오른쪽으로 나아가게 하는데, 여기서의 동일 대각선 상의 모든 PE들은 동일 파형으로 동시적으로 실행된다.[4]

경로 PE(0,0)로부터 PE(I-1, I-1)까지의 최소 국부 거리의 합인 축적된 최소 거리는 두 패턴 사이의 매치가 얼마나 근접하게 이루어 졌는가를 말하여 주는 매치 인자(match factor)를 생성한다. 이러한 방식으로 매치인자는 사전내 모든 단어에 따라 생성된다. 그림1의 간단한 DTW 프로세서로 부터 단일 PE 하나를 구체화하여 제시하면 그림2와 같다. 이와 같은 한 개의 프로세싱 소자를 'PE소자'라고 할 수 있다.

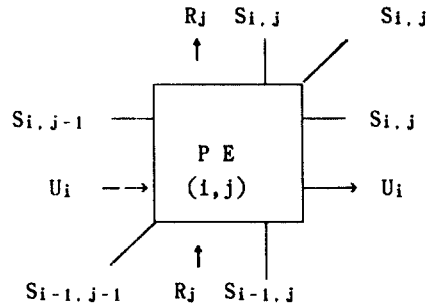


그림 2. 한 개의 DTW 프로세싱 소자
 Fig. 2. One DTW processing element(PE)

그림2의 표시 '-', /, |'는 이웃 PE들 사이를 통과 하면서 계산된 값들의 진행 경로를 나타내며, '→, ↑' 표시는 이웃 PE사이를 통과하는 계수들의 진행 방향을 나타내고 있다. 일반적으로 각 PE는 다음과 같이 계산될 수 있다.

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

(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)부터 (i,j)까지의 최적 경로상에서의 부분 합의 거리를 뜻한다. 이 연구에서 국부 거리 D_{i,j}는 R_j와 U_i차의 절대값으로 측정된다. 즉,

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

이 밖의 국부 거리 측정법으로는 평방형 유클리드 식 거리 측정과 대수분할 거리 측정(: Itakura measure)법등이 있다. 더욱 일반적인 형태로 (2)식을 표현하여 보면 (3)과 같다.[5]

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

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 PE소자는 앞에서 설명한 (1)식의 계산을 수행할 수 있도록 설계되어야 한다. (1)식을 계산 수행 알고리즘에 합당하게 다시 표현하면 (4)식과 같게 된다.

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

(4)식을 수행하기 위한 PE의 내부구조는 그림3과 같다. 그림3과 같은 PE셀은 (4)식에 의하여 동일한 절대값을 반복하여 계산하기 위한 한개의 ABS블록, 네 번의 가산을 위한 네 개의 ADD블록, 그리고 가산 결과를 최소화하고 이 가산 결과와 절대값의 합을 최소화하기 위한 두 개의 MIN블록으로 구성된다. 레지스터로는 s_{out} 디지털 값을 래칭하기 위한 한 개의 3비트 레지스터와 r_{out} 과 u_{out} 디지털 값을 래칭하기 위한 두 개의 2비트 레지스터가 있다. 또한 세어 신호 W_{in} 은 래치되지않은 상태로 모든 PE셀 사이를 동일한 역 대각선 방향으로 전파되어 나간다.

본 논문에서 채용된 수치 데이터 형태는 base4 signed digit을 이용한다. 즉, 심볼 집합은 $\{-3, -2, -1, 0, 1, 2, 3\}$ 와 같이 7개의 디지털로 이루어진다. 편의상 음수는 3, 2, 1과 같이 표시한다. 예로서 십진수 63₁₀을 0333₄, 1133₄, 1013₄, 또는 1001₄로 나타낼 수 있다.

DTW프로세서의 입력데이터, 즉 기준패턴계수 및 미지패턴계수는 base4 digit, 즉 $\{0, 1, 2, 3\}$ 의 심볼을 이용한다. 이 집합은 signed digit 집합의 부분집합

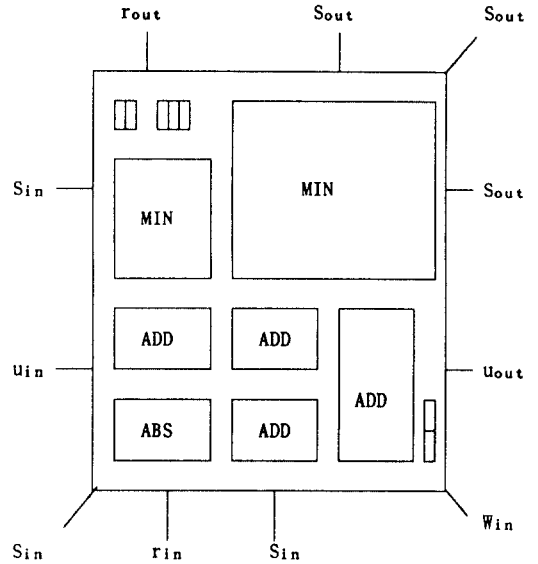


그림 3. DTW PE소자 내부구성
Fig. 3. DTW PE structure

이므로 입력데이터를 따로 변환할 필요가 없다. 각 PE에서는 매치인자는 signed digit을 이용한다. 한 개의 디지털은 세 개의 2진수를 이용하되, 음수의 경우 2의 보수 변환을 사용한다. 따라서 패턴들의 계수 표시에는 2비트, PE에서 입출력되는 값은 3비트가 각각 이용된다.

예를 들어서 1013₄의 4디지털 값은 001 000 111 011₂의 12비트로 encode된다. 그런데 base 4 signed 디지털 집합은 $\{-3, -2, -1, 0, 1, 2, 3\}$ 이어서 encoding "111₂"은 7이되어 집합에 포함되지 않는다. 즉, 'NAN'(Not a Number)조건에 해당되므로 첫 번째 비트 라인을 이러한 경우등에 대비하여 오버플로우용이나 에러검출용으로 이용하게 된다. 따라서 실제 설계의 경우 두번째와 세번째 비트 라인이 base 4 디지털 집합 $\{0, 1, 2, 3\}$ 의 비지 음성이나 기준 단어 계수로써의 패턴 계수입력 라인으로 사용된다.[1-6]

오버플로우 검출 또는 하드웨어의 에러 검출용으로 사용할 수 있다. 본 연구에서 필요한 모든 데이터는 4개의 디지털을 사용하고 각 디지털이 순차적(digit serial)으로 PE에 입력된다.

III. DTW PE소자의 설계

1. 디지털 파이프라인 DTW PE

VLSI 칩을 설계하는데 있어서 I/O하기위한 핀의 갯수를 제한하는 것이 요구되는데 이때 사용하는 데이터 전송방식에는 비트 직렬(bit serial)과 디지털 직렬(digit serial)이 있다.

본 연구의 DTW 신호처리용 PE사이의 전송방식으로는 방대한 데이터를 고속으로 처리하여야 하므로 지연시간(latency)이 분해되는 비트 직렬 방식 대신 디지털 직렬 방식을 채택하였다.[4],[9]

디지털 직렬 전송에서 MSD를 먼저 보낼 것인가, LSD를 먼저 보낼 것인가를 결정하는 문제는 표1을 근거로 하여 MSD우선(MSD-first)방식으로 정하였다.

연산자 중 DP에서 많이 사용되는 덧셈기, 최대·최소 비교기등의 연산은 데이터 크기에 관계없이 일정 시간내에 수행되는 덧셈기, 최대·최소 비교기등의 연산은 데이터 크기에 관계없이 일정시간내에 수행가능하도록하는 MSD우선 방식에서 가능함을 알 수 있다. 그리고 MSD우선 방식을 이용하여 연산할 경우 표와 같게 되기 위해서는 부호화 디지털(signed digit) 표현을 이용하여야 한다.

예를 들어서 본 논문의 가산기 ADD는 디지털 직렬 가산기인데, 그림4는 이 가산기를 부호화 디지털, base 4, MSD우선, 디지털 파이프라인 덧셈기로 구체화 한 것이다. LSD우선의 경우에는 내부 올림수 c_j 를 저장하지만 MSD우선에서는 내부에 합을 저장한다.

그림4를 보면 덧셈기의 처음 단 1s-1c에서는 중간합(interim sum, s_j)과 중간 올림수(interim carry, c_j)를 계산하고, 두번째 단 2s에서는 중간 올림수와 이전의 중간합(s_{j-1})을 합하여 출력을 내게 된다.

표 1. 신호처리기 연산자

Table. 1. Signal processor operators

연산자	LSD first	MSD first
Boolean	Yes	Yes
덧셈(보수)	Yes	*Yes
최대·최소 비교	No	*Yes
좌로 이동	Yes	No
우로 이동	No	Yes
곱셈·나눗셈	No	Yes

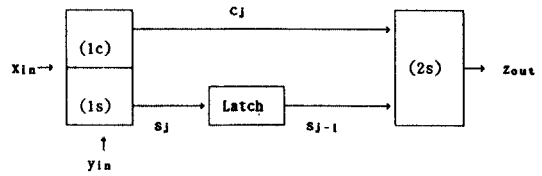


그림 4. MSD우선, 디지털 파이프라인 덧셈기

Fig. 4. MSD first, digit pipeline adder

여기에서 s_j 와 c_j 는 두개의 값을 합했을 때 -3과 3 사이의 제한된 값을 가지므로 두번째 단 2s 덧셈에서는 올림수가 생기지 않아 올림수의 직렬 전파(serial propagation)가 없어지게 된다. 이를 'carry free'라고 하며, 이러한 상태하에서의 매핑을 'round down' ADD 매핑 표로 소개한 바 있다. 즉, 이 표에서는 디지털 동작에서 MSD가 생성될 경우 모든 올림수를 'zero'로 조정하고 있다.[12]

DTW 프로세서의 구성요소인 PE소자는 앞에서 설명한 (4)식의 계산을 수행할 수 있도록 설계되어야 한다. 알고리즘에 합당하게 수행하기위한 PE소자는 (4)식에 의하여 동일한 절대값을 반복하여 계산하기 위한 한 개의 ABS블록, 네 번의 가산을 위한 네개의 ADD블록, 그리고 가산 결과를 최소화하고 이 가산결과와 극부거리인 절대값과의 합을 최소화하기 위한 두 개의 MIN블록으로 구성된다.[11-12]

레지스터로는 s_{out} 디지털 값을 래칭하기위한 한 개의 3비트 레지스터와 r_{out} 과 u_{out} 디지털 값을 래칭하기위한 두 개의 2비트 레지스터가 있다. 또한 제어 신호 W_{in} 은 래치되지않은 상태로 모든 PE소자 사이를 동일한 역 대각선 방향으로 전파되어 나간다.

이 연구에서의 데이터는 모든 PE소자 사이를 base 4 signed 디지털 방식의 직렬 상태로 전파되도록 설계된다. 디지털 데이터를 멀리 전송시키기 위해서는 세 개의 배선이 요구되지만 계수 디지털을 전송하는 데는 단지 두선만이 필요하고, 한 선은 신호 오버플로우용이나 하드웨어적인 에러 검출용으로 쓰이게 된다.

각 PE의 하위 블록회로에 대한 클럭킹 개요(clocking scheme)는 그림5에서 제시하고 있다. PE를 통하여 한 개의 디지털이 완전하게 처리되는데는 two-phase 클럭으로 네 개의 펄스가 필요로 된다.

파형 제어 펄스 W_k 는 다음 번째 계수의 MSD가 도착될 때 다시급 발생할 것이다. 한개의 디지털은

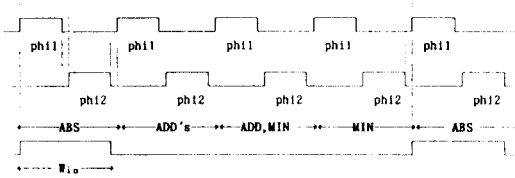


그림 5. DTW PE소자의 클럭킹
Fig. 5. DTW PE clocking

네 개의 펄스를 필요로 하므로 8개의 디지털들을 처리하려면 32개의 클럭 펄스가 요구된다. 즉, 8개의 디지털 오퍼랜드는 32개의 펄스로 처리되며, 이 경우 \$W_k\$는 32개의 펄스마다 다시 발생된다고 할 수도 있다.

한 디지털당 소요되는 펄스의 수(4 클럭)는 처리 속도의 개선과 관계 되는데, 이러한 개선은 단일 PE 내의 원래 동작을 'pipelining'함으로써 가능하다.

즉, 개선율의 증가는 각 PE 내부의 네 단계 스텝(\$\Leftarrow\$ ABS \$\Rightarrow\$ \$\Leftarrow\$ ADD's \$\Rightarrow\$ \$\Leftarrow\$ ADD, MIN \$\Rightarrow\$ \$\Leftarrow\$ MIN \$\Rightarrow\$)과 디지털 레지스터들을 하드웨어적으로 분리시킴으로써 가능하다고 할 수 있다.

그림5에서 한 개의 게이트 지연시간이 2ns인 경우, 각 phase(ph1 또는 ph2)는 네개의 게이트 동작을 가지므로 8ns의 시간이 걸린다. 예를 들어서 그림4의 ADD 로직 블록 다이어그램에서 첫 번째 디지털 동작(1c-1s)은 phi1에서, 두 번째 디지털 동작(2s)은 phi2 중의 적절한 클럭 사이클동안에 각각 수행된다.

이 연구에서의 대수적 성분 블록들은 모두 디지털 라이프라인 되기 때문에 현재의 PE 사이클 수행 결과가 이웃한 PE로 전달될 동안 이어지는 다음의 PE 사이클 시간중의 다음 디지털 오퍼랜드가 현재의 PE 사이클 시간에 의하여 처리된다.[1]

2. 계층적 구성

표준 셀 방식의 설계를 하기 위하여 DTW PE 전체적인 구조를 여러 계층으로 나누고, 각 계층별로 부분블록(sub-block)을 두었다. 부분블록중 일부는 다시 하위 계층을 두는 방식을 적용하였으며 최종적으로는 최하위 계층의 셀들은 우선 workstation상에서 DTW 알고리즘에 합당하도록 게이트레벨 회로를 설계하였다.

그림6은 전체 시스템의 계층적 구성을 보여주고 있다. 그림에서의 DTW PE는 부분블록으로 ADD, ABS,

그리고 MIN 블록으로 구성되어 있으며, 다시 각 블록별로 ADD블록은 하위 계층에 MPWADD회로등 5개의 회로, ABS블록은 하위 계층에 BLB회로등 8개 회로로 이루어져 있다. MIN블록은 하위 계층에 또 다시 3개의 ST \odot , ST \ominus , ST \oplus 부분블록을 두고 있으며, 이들 각 부분블록들은 COM4EQ등 3개의 하위 블록과 MONE회로등 10개의 회로로 조합을 이루고 있다.

각 블록별 회로 설계는 표준 셀 라이브러리(standard cell library)의 셀들을 주로 활용하였다.

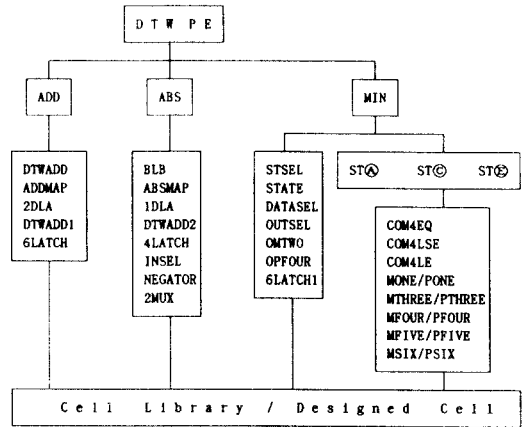


그림 6. 전체 시스템의 계층적 구성
Fig. 6. Hierarchical structure of the system

그림7은 회로설계를 마친 후, 완성된 최상위 계층의 전체 DTW PE소자 블록 다이어그램 회로이다. [10-14]

3. DTW PE소자의 지연시간과 레이아웃

본 논문은 11개의 표준 셀을 사용하였으며, 각 셀에 대한 셀 지연방정식은 (5) 및 (6)식과 같다.

$$\left. \begin{aligned} t_{PLH} &= INT_{t_{PLH}} + EXT_{t_{PLH}} * CL \\ t_{PHL} &= INT_{t_{PHL}} + EXT_{t_{PHL}} * CL \end{aligned} \right\} \quad (5)$$

t_{PLH} : the propagation delay for low to high transition

t_{PHL} : the propagation delay for high to low transition

$INT_{t_{PLH}}$: the intrinsic delay represents the

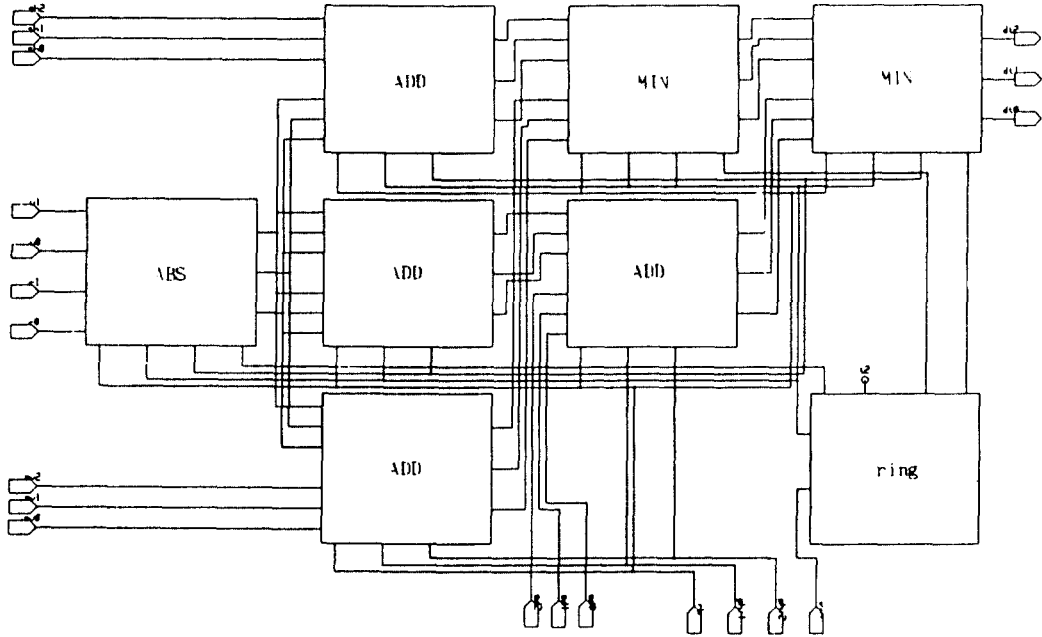


그림 7. 전체 DTW PE소자 블록다이아그램
Fig. 7. The total DTW PE block diagram

zero-load delay of the cell at t_{PLH}
 EXT_{tPHL} : the extrinsic delay related to the cell's output impedance (loading capacitance CL , etc.) at t_{PHL}
 etc.

$$tP = t_{PLH} + t_{PHL} \quad (6)$$

tP : total propagation delay

표2의 블록별 전달지연시간은 표준 셀 CMOS 데이터의 스위칭 특성표를 참조하여 정상조건 (typical case, $T = 25^\circ C$) 하에서의 tP 를 추출하였으며, 이를 각 블록회로의 최악경로 (the worst path)에 적용하여 계산하였다.

각 블록별 지연시간은 온도 $T = 25^\circ C$, 부하커패시턴스 $C_L = 0.5pF$, $V_{DD} = 5V$ 를 기준하여 추출하였다. [13]

그림8을 통하여 최종적인 상위 블록사이에서의 최악 경로를 따르는 전체 전달 지연시간을 알 수 있다.

그림에 의하여 DTW PE의 전체 전달 지연시간은

176.56 nsec. (= 58.25 + 30.99 + 43.66 + 43.66)임을 알 수 있다.

DTW PE 시스템 각 블록에 대한 계층적 구조에 따라 상향식 (bottom up method)으로 표준 셀 설계 방식을 채택하여 레이아웃하였다.

또한 표2와 그림7로 부터 알 수 있는 바와 같이 DTW PE 한 개의 소자는 ADD블록 4개, ABS블록 1개, MIN블록 2개 그리고 주변회로 104개의 트랜지스터를 합하여 총 7,280개의 MOS 트랜지스터로 구성 되어 있다.

DTW PE의 계층적 구성은 앞에서의 그림6과 같으며, 이러한 시스템을 $1.2\mu m$ CMOS N-well 설계 규칙에 따라 표준 셀 방식으로 최종 레이아웃 하였다. [11-14]

그림9는 전체 DTW PE소자 시스템의 레이아웃에 대한 그림이다.

표 2. 블록별 전달 지연시간 및 트랜지스터 수

Table. 2. Propagation delay and the number of transistors

분류	부 분류	전달 지연시간[nsec.]	트랜지스터 수
ADD	6LATCH	4.39	180
	DTWADD	8.82	98
	ADDMAP	2.30	54
	2DLA	7.96	152
	DTWADD1	7.52	86
합 계		30.99	570
ABS	4LATCH	4.39	120
	BLB	6.39	64
	ABSMAP	2.30	44
	IDLA	8.14	152
	DTWADD2	8.89	86
	INSEL	22.51	84
	2MUX (NEGATOR)	*	40
합 계		58.25	624
MIN	6LATCH1	4.39	180
	STC	16.27	510
	STSEL	2.76	102
	STATE	10.20	254
	DATASEL	4.99	82
	OUTSEL	5.05	96
	(OMTWO)	*	32
	(OPFOUR)	*	4
	(ST ^A)	*	398
(ST ^B)	*	478	
합 계		43.66	2136
전체 합계		132.90	3330

(부블럭) * 표는 최약경로에 해당되지 않는 블록

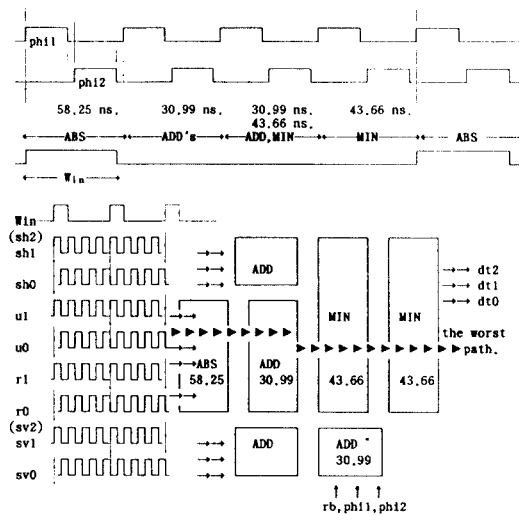


그림 8. DTW PE의 데이터 흐름과 최악 경로

Fig. 8. Data flow and the worst path of DTW PE

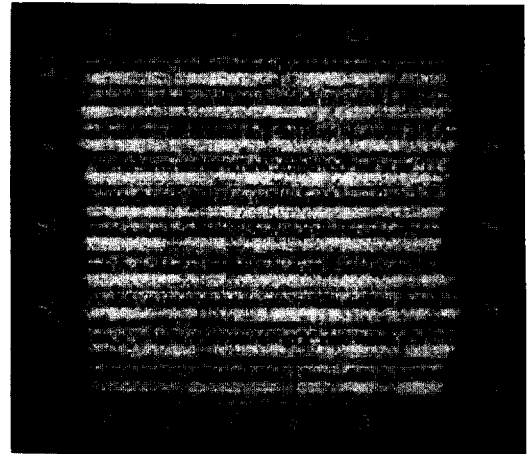


그림 9. DTW PE소자 레이아웃

Fig. 9. DTW PE layout

IV. 결 론

본 연구에서는 DTW를 이용한 음성 인식 알고리즘을 이용한 DTW PE소자에 대하여 고찰하였으며, 특히 고립단이 인식분제에 중점을 두었다. 인식에서의 중요한 관점은 실시간 처리의 신속성, 환용성 재고를 위한 경제성, 그리고 편리한 이용을 위한 기능성 등을 들 수 있다. 이러한 현실적 문제의 구현을 위해서는 DTW 프로세서의 개발이 필요한데, 실시간 처리용 음성인식 프로세서를 위한 PE소자 설계는 문제 해결의 가장 중요한 부분이라고 할 수 있다.

본 연구에서의 DTW용 PE소자는 크게 세가지의 블록으로 분류된다. 즉 MIN 블록, ADD 블록 그리고 ABS 블록인데, "MIN"은 축적된 최소거리들 계산하기 위한 블록이고, "ADD"는 이들 최소거리들의 합을 계산하는 블록, 그리고 "ABS"는 소자 자체 내부에서의 국부거리 절대값을 구하기 위한 블록이다.

DTW프로세서 각 PE사이에서의 데이터 전달은 인접 PE끼리만 허용되도록 되어 있어서 통신선로에 의한 비용을 절약할 수 있고 프로세서의 모듈화가 가능하게 된다. 즉, 이렇게 시스톨릭(systolic)한 배열을 하고 있는 DTW 프로세서는 각 PE소자가 규칙적으로 연산을 하고, 인접한 PE로 데이터를 전달하는 PE들의 집합체인 셈이다.

음성인식에서의 DTW는 반복적 계산을 필요로 하며, 이 계산을 수행하기에 적합한 PE소자의 설계는

매우 중요하다. 실제의 경우 기준단어와 미지발음사 이에서는 비선형 패턴 매칭이 수행되므로 이는 인식 어휘수가 많은 고립 단어들을 실시간 인식으로 처리 하는데 가장 큰 장애 요소가 되고 있다.

따라서 본 연구에서는 실제에 가까운 실시간 어휘 인식을 가능하게 하는 대사전 어휘 규모의 DTW 알고리즘을 하드웨어로 구현하기 위한 PE소자의 설계에 주안점을 두었으며, 실시간 고속처리 및 간결한 하드웨어의 설계를 위하여 이들 각 PE들이 종래의 시스톨릭 DTW방식과는 다르게 디지털 파이프라인(digit pipelined)방식으로 배열됨으로써 1) 모든 데이터들이 base4 format signed digit로 표현되고 2) 이러한 데이터들은 직렬 디지털 방식으로 PE사이를 통과하게 되어서 3) 패턴 매칭이 디지털 레벨에서 파이프라인 방식으로 이루어질 수 있도록 설계하였다.

표준 셀 방식으로 설계한 결과 각 셀의 전달 지연시간의 특성으로 인하여 전체 전달 지연시간은 최악경로로 계산한 결과 176.56 nsec.로 나타났다. 따라서 한 개의 PE소자가 데이터를 처리하여 이웃 PE로 전달하는데 쓰는 시간은 최소한 약 177 nsec. 이상임을 알 수 있었다.

완성된 한 개의 PE소자는 약 7,300여개의 트랜지스터로 집적되어 있으며, 이러한 셀이 차지하는 DTW PE 회로 레이아웃 면적은 1.2 μ m CMOS N-well 설계규칙에 따라 표준 셀 방식으로 설계한 결과 3.6mm \times 3.3mm으로 나타났으나 미세가공 기술의 발달로 설계규칙도 점점 더 submicron화 되어가는 추세이므로 더욱 축소될 수 있을 것이다.

더욱 연구·개발하여야 할 부분으로는 DTW PE소자의 전달 지연시간의 정확한 분석을 위한 보다 효율적인 소프트웨어 개발이 필요하다고 하겠으며, 본 논문에서 제시한 레이아웃 결과가 하드웨어로 IC화 된 후에는 보다 실질적인 전기적 특성 고찰을 포함한 분석이 수반되어야 할 것이다.

또한 앞으로 이러한 PE를 조합하여 하나의 완성된 DTW 프로세서로 활용하기 위해서는 pre-processor 및 post-processor의 연구·개발이 함께 이루어져야 할 것이다.

참 고 문 헌

1. M. J. Irwin, "A Digit Pipelined Dynamic Time Warp Processor," IEEE Trans. Acoust., Speech, Signal Processing, Vol. 36, No. 9, pp. 1412-

1422, Sept. 1988.
 2. C. Myers, L. R. Rabiner, and A. E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-28, No. 6, pp. 623-635, Dec. 1980.
 3. 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.
 4. F. Charot, P. Frison, and P. Quinton, "Systolic Architectures for Connected Speech Recognition," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-34, No. 4, pp. 765-779, Aug. 1986.
 5. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-26, No. 1, pp. 194-200, Feb. 1978.
 6. 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.
 7. R. Bellman, "Dynamic Programming," Princeton University Press, Princeton, New Jersey, 1957.
 8. N. Weste and K. Eshraghian, "Principles of CMOS VLSI Design," Addison-Wesley Publishing Company, 1985.
 9. 이황수, 김진을, 조윤석, 윤순영, 고영훈, "고속 디지털 신호처리를 위한 알고리즘 및 구조연구와 VLSI Processor의 설계 개발," 과학기술처 '89특정 연구과제 논문집, 1990년 8월.
 10. 한국전자통신연구소 자동설계기술 개발부, "ETRI 표준셀 기술서-CMOS CS 1030," 한국전자통신연구소, 1987년.
 11. 정광재, 문홍진, 최규훈, 김종교, "음성인식용 DTW PE의 IC화를 위한 MIN 회로의 설계," 한국통신학회지, 제 15 권, 제 8 호, 1990년 8월.

12. 정광재, 문홍진, 최규훈, 김종교, "음성인식용 DTW PE의 IC화를 위한 ADD 및 ABS 회로의 설계," 한국통신학회지, 제 15 권, 제 8 호, 1990년 8월.
 13. SAMSUNG Electronics Co., "SEC KG40000

SOG CELL LIBRARY-Data Book," Jan, 1991.
 14. Seodu Logic, ASIC Division, "SCMOS DATA BOOK SCMOS Cell Library," Seodu Logic, Inc., Feb, 1993.

※이 논문은 1992년도 교육부지원 한국학술진흥재단의 자유공모(지방대학 육성)과제 학술연구조성비에 의하여 연구되었음.



崔圭勳(Kyu Hoon Choi) 正會員
 1950年 7月 11日生
 1974년 2월 : 광운대학교 전자공학과 공학사
 1976년 2월 : 연세대학교 산업대학원 전자공학과 공학석사
 1991년 2월 : 전북대학교 대학원 전자공학과 공학박사

1981년 3월 ~ 현재 : 전주공업전문대학 전자과 부교수
 1978년 3월 ~ 1981년 2월 : 인덕공업전문대학 공업디자인과
 ※주관심분야 : 신호처리 및 집적회로 설계



金鍾玦(Chong Kyo Kim) 正會員
 1944年 8月 22日生
 1966년 2월 : 전북대학교 전기공학과 공학사
 1977년 8월 : 전북대학교 대학원 전기공학과 공학석사
 1983년 8월 : 전북대학교 대학원 전기공학과 공학박사

1979년 4월 ~ 현재 : 전북대학교 전자공학과 교수
 1966년 7월 ~ 1972년 5월 : 제철부 금부(장기리 통신당방)
 1976년 3월 ~ 1979년 4월 : 전주공업전문대학 전자과
 1983년 8월 ~ 1984년 8월 : 미국 일리노이 공과대학(IIT) 객원교수
 ※주관심분야 : 신호처리, CAD 및 컴퓨터그래픽스