

적교변환을 이용한 음성인식

이병희\* 김태원\*\* 이광형\*\* 은재강\*\*

대전직업훈련원\* 송실대학교 전자공학과\*\*

Speech Recognition Using Orthogonal Transformation

Lee Myung-Hui, Kim Tae-Won, Lee Kwang-Hyung, Youne Jae-Gang

Tae Jon Vocational Training Institute Soong Sil Univ.

ABSTRACT

In this paper, we describes the design of Speech Recognition System of isolated words using a low cost personal computer and analyzes the characteristics of the speech recognition. Hadamard Transformation is defined and several of its properties are studied.

Using matrix partitioning, fast algorithms for efficient computation of Hadamard Coefficients and power and phase spectra are developed.

1. 서론

Man-Machine interface로 음성처리의 역할은 지대하다. 음성인식 system은 각기 단어인식, 연결 단어인식과 연속음성인식으로 이루어지며 특징화자와 불특정(독립)화자인식으로 연구되고 있다. 음성 파라미터를 추출하는 방법에는 시간영역에서의 분석(대수에너지 분석, 영교차파분석, 단기간 자기상관 분석), 파수영역에서의 분석(BFF bank 분석, 직교분석(Karhunen-Loeve 변환), Slant 변환, Fourier 변환, Hadamard 변환), Cepstrum 분석, LPC 분석 등이 있다. 직교변환에 의한 음성인식 연구는 L.Bolan-Sky[1]은 적응인산방식들에 의해 design된 음성인식 절차를 개발하여 최적적교근거 함수들의 집합을 결정하는데 큰 기여를 했다. Hanley[2]와 O'Neill[3]은 음성인식을 위해 복수지수 적교함수용 computer로 실험하였다. Young과 Huggins[4]는 람음의 짧은 시분들을 표현하기 위한 근거함수들의 직교적집합을 구성하였다. Hadamard 변환은 양상부호화, digital system의 스펙트럼 분석, 음성처리와 Sequency 분석등에 다양하게 이용된다. HI는 단지 덧셈과 뺄셈만으로 이루어지므로 디지털 신호 처리 H/W 실험을 간단하게 하여, 특히 소형 컴퓨터로도 처리가 용이한 장점[5]이 있다. FHI(fast Hadamard transformation)은 행렬분해로 연산속도를 줄인다.

2. 하다마르 변환

Time 대 Frequency의 변환종류는 FI, FL, LI가 있는데 적교변환들은 time domain과 frequency domain [6] 사이의 식들에서 서로 orthogonal pair를 형성한다. 심수 연속함수의 집합

$$\{U_n(t)\} = \{U_0(t), U_1(t), \dots\} \text{이 } (t_0, t_0+T) \text{ 범위에서}$$

$$\int_T U_m(t) \cdot U_n(t) dt = \begin{cases} C & (m=n \text{ 인 경우}) \\ 0 & (m \neq n \text{ 인 경우}) \end{cases} \quad (1)$$

이런 '직교(orthogonal) 한다' 라고 말한다. 이때,  $\int_T$ 는  $\int_{t_0}^{t_0+T}$ 이다. C=1인 경우에는  $\{U_n(t)\}$ 를 정직교 집합(Orthonormal Set)이라고 한다[7]. 직교변환의 이용함수는 Rademacher, Haar, Walsh 함수가 있으며 Walsh 함수로 얻는 배열은 Sequency 또는 Paley 배열과 natural 또는 Hadamard 배열이 있다. Walsh 함수는 식(2)와 같이 표시한다.

$$U_0(t) = 1$$

$$U_{m,1}(t) = \cos(2\pi m t) \quad (2)$$

$$U_{m,2}(t) = \sin(2\pi m t)$$

Fourier 정현파와 Walsh 함수의 관계를 그림1에서 보인다.[5]

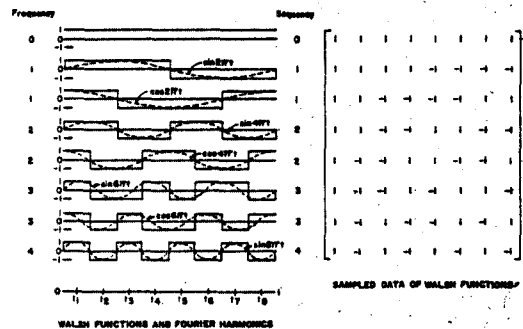


Fig.1. Fourier harmonics, Walsh functions and Hadamard matrix for n=8

Walsh 함수의 표본파는 +1 또는 -1의 8 \* 8 배열이 되는데 이를 Hadamard에 의한 특별한 재배열을함으로 그림 2와 같이  $[H(3)] = 2^3 * 2^3$ 의 행렬을 얻는다. 이를 일반적 수식으로 나타내면 식(3)이 된다.

$$[H(0)] = [1]$$

$$[H(k+1)] = \begin{bmatrix} [H(k)] & [H(k)] \\ [H(k)] & -[H(k)] \end{bmatrix} \quad (3)$$

$$H_8(g) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{bmatrix}$$

Fig 2. Hadamard matrix  
( $m=8$ )

$k = 0, 1, 2, \dots, g : g = \log_2 n$   
 $H(g)$  는 대칭적이고  
 직교하므로

$$[H(g)]^T [H(g)] = M [I(g)] \quad (6)$$

가 된다. 여기서  $[I(g)]$  는  
 $n \times n$  항등행렬이다.

$n$  주기적 계열은 식(5)로 표시된다.

$$\{x(k)\} = \{x(0), x(1), \dots, x(n-1)\} \quad (5)$$

$\{x(k)\}$  는 식(6)을 일으켜  $n$ -vector  $x(g)$  에  
 의해 표현된다.

$$\{x(g)\}^T = \{x(0), x(1), \dots, x(n-1)\} \quad (6)$$

$\{x(k)\}$  의 HT는 식(7)로 정의한다.

$$\{Bx(g)\} = 1/n [H(g)] \{x(g)\} \quad (7)$$

여기서  $Bx(\cdot)$  는 HT의 계수가 된다.

### 3 고속 하다아르 변환(FHT)

FHT는 HT를 효과적으로 계산하는 연산방식이다.  
 FHT는 matrix factoring 또는 matrix  
 partitioning의 기술을 이용하여 HT의 빠르고  
 효율적인 계산을 위한 algorithm을 개발한다.  
 HT를 계산하기 위한 수학적 동작은  $n=8$ 인 경우흐  
 림도로 표시하면 그림3과 같이 표시할 수 있다.  
 그림3에 의해서 식(8)을 얻는다. 모든 변환 계  
 수들을 계산하는데 수학적 동작의 총순는 근사적  
 으로  $n \log_2 n$  [6]이다.

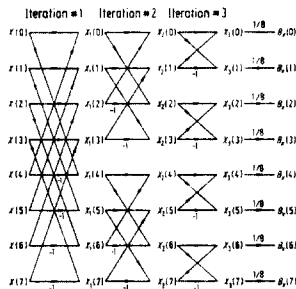


Fig3 Signal flow graph illustrating the  
 computation of HT coefficients for  $n=8$

$$\begin{bmatrix} B_x(0) \\ B_x(1) \\ B_x(2) \\ B_x(3) \\ B_x(4) \\ B_x(5) \\ B_x(6) \\ B_x(7) \end{bmatrix} = \frac{1}{8} \begin{bmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} & 1 & \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} & & \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} & & 1 & \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \\ x(4) \\ x(5) \\ x(6) \\ x(7) \end{bmatrix} \quad (8)$$

식(7)을 이항하여 식(8)을 얻는다.  
 $\{Bx(g)\}^T = 1/n \{x(g)\}^T [H(g)]^T \quad (9)$

식(7)과 (9)에 의해서

$$\{Bx(g)\}^T \{Bx(g)\} = 1/n \{x(g)\}^T [H(g)]^T [H(g)] \{x(g)\} \quad (10)$$

이는 식(11)을 이룬다.

$$\frac{1}{n} \sum_{h=0}^{n-1} x^2(h) = \sum_{h=0}^{n-1} Bx^2(h) \quad (11)$$

그래서 Sequence  $\{x(k)\}$  는 전유나 전압신호  $x(t)$   
 의 표본화된 값으로 표현한다.  
 그림3의 신호 흐름도로 부터 식(12)을 얻는다.

$$Bx(0) = X_3(0)/8$$

$$Bx(1) = X_3(1)/8$$

$$\begin{bmatrix} Bx(2) \\ Bx(3) \end{bmatrix} = \frac{1}{8} [H(1)] \begin{bmatrix} X_2(2) \\ X_2(3) \end{bmatrix} \quad (12)$$

$$\begin{bmatrix} Bx(4) \\ Bx(5) \\ Bx(6) \\ Bx(7) \end{bmatrix} = \frac{1}{8} [H(2)] \begin{bmatrix} X_1(4) \\ X_1(5) \\ X_1(6) \\ X_1(7) \end{bmatrix}$$

식(9)에서 식(11)을 유도하여  $[H(g)]$  는 직교하  
 므로 식(13)을 얻는다.

$$\sum_{k=2}^3 Bx^2(k) = \frac{2}{8^2} \sum_{m=2}^3 X_1^2(m) \quad (13)$$

$$\sum_{k=4}^7 Bx^2(k) = \frac{4}{8^2} \sum_{m=4}^7 X_1^2(m)$$

전력 스펙트럼을 계산기 위한 FHT 번조의 일반  
 적인식(14)를 만든다.

$$P_0 = X_0^2(0)/M^2$$

$$P_s = \frac{2^{s-1}}{m^2} \sum_{m=2^{s-1}}^{2^s-1} X_{g^{s-1}}^2(m) \quad (14)$$

$s = 1, 2, \dots, g$

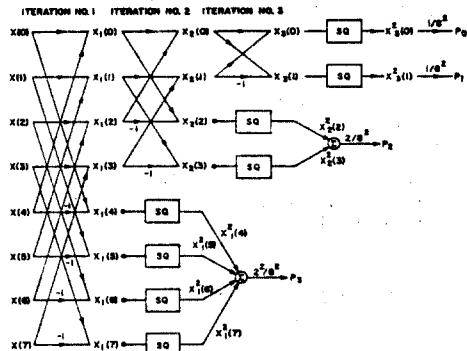


Fig4. Signal flow graph illustrating the  
 computation of FHT power spectrum

식(14)에 고속algorithm에 대응되는 신호흐름도  
 는 그림4와 같다. FHT의 전력 스펙트럼은 두가지  
 의의를 가진다. 스펙트럼 포인트  $P_s$ 는  $(g+1)$ 개의  
 sequence 집합의 평균전력을 표현하고, 또한 각  
 spectral point는 하나의 주파수의 전력보다 오  
 히려 주파수 그룹의 전력 내용을 표현한다.  $n$  주  
 기적 sequence  $\{x(k)\}$ 는 다음과 같이  $n/2$ -주기적  
 seq.  $\{f_r(k)\}$ 와  $n/2$ -반주기적 seq.  $\{g_r(k)\}$ 로 분해  
 되어 식(15)를 얻는다.

$$\{x(k)\} = \{f_r(k)\} + \{g_r(k)\} \quad (15)$$

주어진  $n$ -주기적 seq.  $\{x(k)\}$ 가  $(g+1)$ 개의 subseq.  
 들로 분해되면 식(16)이 된다.

$$\{x(k)\} = \{f_r(k)\} + \{g_r(k)\} + \{g_{g-r}(k)\} + \dots + \{g_r(k)\} \quad (16)$$

여기서  $\{f_r(k)\}$ 는 1-주기적 seq.이고,  $\{g_{g-r}(k)\}$   
 는 2<sup>r</sup>-반주기적 seq.이다.  $r=0, 1, \dots, (g-1)$ 이다.  
 이러한 분해과정은 그림5에서 설명되어진다.

$n=8$ 인 경우 seq.  $\{x(k)\}$ 는 식(17)이 된다.

$$\{x(k)\} = \{f(k)\} + \{g(k)\} + \{g(k)\} + \{g(k)\} \quad (17)$$

식(17)을 vector들에 의해서 표현하면 식(18)과  
 같다.

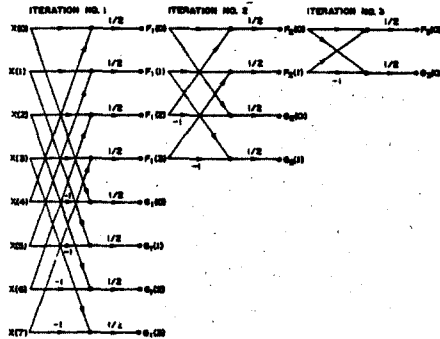


Fig 5. Flow graph illustrating the decomposition of  $\{x(k)\}$  into subsequences

$$\begin{aligned}
 \{F_1(3)\}^* &= \{b_0(0) \quad b_0(0) \quad b_0(0) \quad b_0(0) \quad b_0(0) \quad b_0(0) \quad b_0(0) \quad b_0(0)\} \\
 \{G_1(3)\}^* &= \{b_0(1) - b_0(1) \quad b_0(1) - b_0(1) \quad b_0(1) - b_0(1) \quad b_0(1) - b_0(1)\} \quad (18) \\
 \{G_2(3)\}^* &= \{b_0(2) \quad b_0(3) - b_0(2) - b_0(3) \quad b_0(2) \quad b_0(3) - b_0(2) - b_0(3)\} \\
 \{G_3(3)\}^* &= \{b_0(4) \quad b_0(5) \quad b_0(6) \quad b_0(7) - b_0(4) - b_0(5) - b_0(6) - b_0(7)\}
 \end{aligned}$$

여기서 계수  $b_n(k)$ 는 가장 낮은 stage의 Hadamard 계수  $Bx(k)$ 의 term으로 표현된다. 식(18)에서 vector의 조차는 그들 모두가 상호직교하여 다음 식이 된다.

$$\|\{x(3)\}\|^2 = \|\{F_1(3)\}\|^2 + \|\{G_1(3)\}\|^2 + \|\{G_2(3)\}\|^2 + \|\{G_3(3)\}\|^2 \quad (19)$$

$\{x(k)\}$ 의 평균전력은

$$\begin{aligned}
 P_{av} &= b_n^2(0) + b_n^2(1) + \frac{1}{2} \sum_{k=2}^3 b_n^2(k) + \frac{1}{4} \sum_{k=4}^7 b_n^2(k) \\
 &= P_0 + P_1 + P_2 + P_3 \quad (20)
 \end{aligned}$$

여기서  $P_0$ 는 1-3주기적 seq.  $\{F_1(k)\}$ 에서의 평균전력으로 표현하고  $P_1, P_2, P_3$ 는 각각 1, 2, 4-방주기적인  $\{G_1(k)\}, \{G_2(k)\}, \{G_3(k)\}$ 에서의 평균전력을 표시한다. 일반적으로  $P_r = (r=1, 2, \dots, g)$ 는  $\{G_{g+1-k}(k)\}$ 에서의 평균전력으로 표현한다.

#### 4. 인식 시스템

##### 4-1. H/W 구성

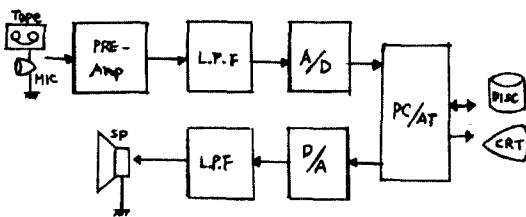
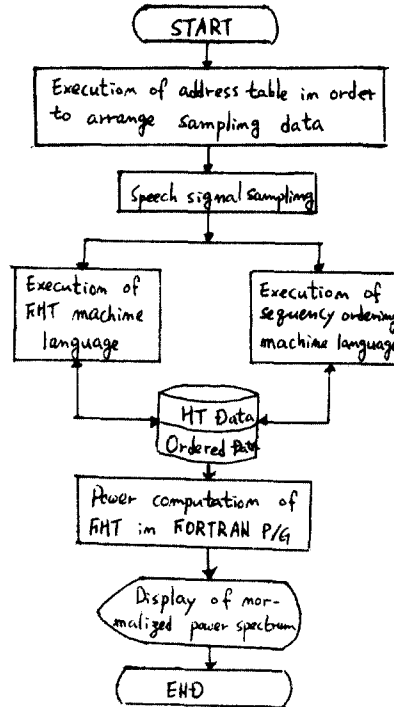


Fig6. Block diagram of the recognition system

마이크로폰 또는 recoder tape 을 거쳐 입력된 음성은 증폭기를 거쳐 음성신호의 Dynamic range 를 조정하고 음성분석의 동작범위를 줄이고 LPF를 통해 200~3400[HZ] 사이의 음성주파수를 8[KHZ]로 음성신호를 샘플링하여 주파수 스펙트럼이 주기적으로 반복되도록 하였다. A/D 변환으로 40초 이내의 음성 데이터를 12bit로 변환하여 disk 상에 file 형태로 저장한다. 한편 정규화된 power spectrum은 고해상도로 CRT 상에 그려진다. 저장된 음성 데이터는 입력 음성신호와 비교된 후 인신된 신호를 D/A 변환한 후 출력을 얻게 된다.

##### 4-2. S/W 구성

본 시스템의 S/W는 메인루틴, 기계어 서브루틴과 단어 power spectrum 분석루틴으로 구성된다. 그림7은 S/W의 전체적인 흐름도를 보인 것이다.



특히 FHT와 Sequency ordering 과정에서 machine language 를 사용한 것은 보다 빠른 시간으로 수행시키기 위함이다.

##### 5. 결과 및 검토

숫자음 /영/, /일/, /이/, /삼/, /사/, /오/, /육/, /일/, /팔/, /구/, /십/ 을 남자 6명, 여자 4명의 독립화자로 부터평균적으로 얻은 data를 disk에 저장시킨 후 새로운숫자음이 입력되면 disk의 data와 비교하여 인식을 한다. Window size 를 8ms로 하여 125 Hz의 분해능이 얻어진다. S.J. Campanella, Guner S. Robinson 8 에의해서는 fourier, Hadamard 그리고 k-L 변환을정보압축시키는식험에서 k-L 변환 13.5 Kb/s, Fourier 변환 10 Kb/s 그리고 Hadamard 7.5 Kb/s의 압축율이 있음을 알았다. Moyett T. Clark 9 가 행판 단어 인식을한 결과 71.6% 96.6%의 불안정한 인식을 했는데 이는 시간길이를 맞추지 않아 인식이 낮았다.

##### 6. 결론

숫자 음성 인식을 위해서 Hadamard 변환을 적용하는 시스템을 구성하였다. 이는 상당히 우수한 성능에비해서 인구가 부진하며 영상처리 혹은 많이 연구되고있으므로 음성 인식으로 연구한다. 신호 데이터의 창길이를 관계잡을 수록 분해능이 향상된다. Hadamard 계수를 증가 시키면, 심한 일하를 받기 쉬운 음소의 수는 감소됨을 알 수 있다. 비음음 인식을 하는 데는 적어도 5개의 계수를 필요로 한다. 인식을올 제고시키기 위해서는 파라메타의 분석을 정확히 하고, 각종 파라메타의 우월한 음성 분류를 명확히 하는 연구가 필요하다.

## Reference

- [1] L. Dolansky, "Choice of base signals in speech signal analysis," IRET Trans. Audio, Vol. AU-8, pp. 221-229, November-December 1960.
- [2] H. J. Manley, "Analysis-Synthesis of connected speech in terms of orthogonalized exponentially damped sinusoids", J. Acoust. Soc. Am., Vol. 35, pp 464-473, April 1963.
- [3] L. A. O'Neill, "The representation of continuous speech with a periodically sampled orthogonal basis", IEEE Trans. Audio. EA., Vol. AU-17, pp. 14-21, March 1969.
- [4] T. Y. Young & W. H. Huggins, "Representation & analysis of signals, part 3, Discrete orthonormal exponentials", Proc. Nat. Electronics conf., Vol. 18, pp. 10-18, October 1962.
- [5] Nasir Ahmed, M. R. Rao, A. L. Abdussattar, "BIFORE or Hadamard Transform", IEEE Trans. Audio Electroacoust., Vol. AU-19, No 3, pp. 225-234, sep. 1971.
- [6] C. X. Fan & S. H. Wang, "A Fast Fourier Transform Algorithm using Hadamard Transform", ICASSP 86, Tokyo, pp. 225-228.
- [7] N. Ahmed & K. R. Rao, "Orthogonal Transforms for Digital signal processing", Springer-Verlag, Berlin, Heidelberg, New York 1975.
- [8] S. J. Campanella & Guners Robinson, "A Comparison of Orthogonal Transformations for DSP", IEEE Trans. communication, Vol. COM-19, No 6, pp. 1045-1050, Dec. 1971.
- [9] Moryett T. Clark, "Word recognition by means of orthogonal functions". IEEE Trans.