

[ 연구실 소개 ]

# 휴먼인터페이스 연구실 (Human Interface Lab.)

(한남대학교 정보통신공학과)

김 경 태 ( Kyung Tae KIM )

「Since you are the end-user, our purpose is to research the human-oriented system that is ease-to-use.」

인간 상호간의 통신은, 정보통신의 발전에 의하여 시간과 공간을 초월한 보다 풍요롭고 자유로운 통신으로 발전되어 가고있다. 이는, 고도의 네트워크와 컴퓨터가 필요하지만, 사용자(인간)는 이들의 존재를 의식하지않고, 원하는 정보를 사용하기에 편한 형태와 수단을 자유롭게 선택할수 있어야 한다. 즉, 인간과 기계사이의 원활한 통신을 위하여 자연스러운 휴먼 인터페이스의 기술확립이 필요하다. 이를 위하여, 음성언어, 시각패턴(문자, 영상, 도면등)의 정보를 공학적으로 생성, 처리, 통합처리/구현하는 기술을 확립하고 인간통신의 종합적 연구를 수행한다. 이의 주요 연구내용은 다음과 같다.

- 음성정보 처리 기술 : 인식(음성을 문자로 변환) 및 이해, 합성(문자를 음성  
을 변환)
- 문자 및 도형처리 기술
- 신호의 정보압축기술 (부호화)
- 미디어(음성, 영상등)를 통합, 처리하는 멀티미디어 기술
- 미디어의 평가기술

상기의 연구내용을 수행하기 위하여, 지난 '92년 3월부터 지금까지 연구수행결과를 본 논문에서 소개한다.

- 디지털 신호처리의 기초를 위한 교육용 소프트웨어의 개발
- 디지털 IC Trainer 개발
- 음성정보처리에 필요한 음성데이터 수집을 위한 발성내용제시 시스템
- 음성구간 검출 알고리즘 개발

## 1. 디지털 신호처리를 위한 교육용 소프트웨어의 개발(INTROSP Version 1.0)

- 목적 : 대학 학부생이, Digital Signal Processing (DSP)의 기초를 눈으로 익히고, 그의미를 확실히 터득하기 위한 도움을 주는데 목적이 있다.

- 내용 :

## 1. ftprpt

푸리에 변환의 성질을 설명한다.

### 1) Linearity (선형성)

임의의 상수 a, b, 에 대하여

$$af_1(t) + bf_2(t) \langle \text{----} \rangle aF_1(w) + bF_2(w)$$

### 2) Scaling : 시간축의 신축성

$$f(at) \langle \text{----} \rangle (1/a)F(w/a)$$

### 3) Time Shift : 시간축의 이동

$$f(t-t_0) \langle \text{----} \rangle F(w)e^{-j\omega t_0}$$

\* 진폭 스펙트럼과 위상 스펙트럼을 볼수있다.

즉, 진폭스펙트럼은 변함이 없고, 다만 위상특성은 다르게 됨을 알수있다.

### 4) Symmetry : 대칭성

시간함수와 주파수함수의 교환이 이루어진다.

$$f(-t) \langle \text{----} \rangle F(-w)$$

$$F(-t) \langle \text{----} \rangle 2\pi f(w)$$

## 2. sinc

신호처리에서 가장 기본적으로 사용하는 sinc 함수를 보여준다.

$$f(x) = (\sin x)/x, \quad (f(x))^2$$

## 3. recfs

구형파 (Rectangular Wave)를 푸리에 시리즈로 변환하는 과정을 보여준다.

## 4. invsmp

불연속적인 이산치가 Sampling 함수(저역통과 필터의 역할)를 이용하여 원래의 연속적인 파형으로 복원됨을 보여준다.

## 5. window

Rectangular, Baretlett, Hanning, Hamming, Black-man Window 를 sine파에 적용하여 원래 파형의 변형된 모습을 보여주고, 각 window의 이산 푸리에 변환을 보여준다.

## 6. expft

지수함수 곡선 (Exponential curve)의 진폭 및 위상스펙트럼을 보여준다.

## 7. convol

두개의 구형파  $f_1(t)$ ,  $f_2(t)$ 가 convolution 되는 모양을 나타내 주고 각각의 푸리에 변환이

$$f_1(t)*f_2(t) \langle \text{----} \rangle F_1(w)F_2(w)$$

가 됨을 나타낸다.

### 8. convol2

구형파  $u(x)$ 와 단위스텝파형  $h(x)$ 가 convolution되는 과정을 나타낸다.

- 1) Folding
- 2) Displacement
- 3) Integration

### 9. dirich

푸리에 변환의 계수가 sinc 함수와 Dirichlet's Kernel의 곱으로 됨을 나타내는 물리적 의미를 보여준다.

### 10. dftcalc

이산신호를 이산푸리에 변환(DFT)계산 과정을 나타낸다.

- 1) DFT의 실수 부분
- 2) DFT의 허수부분
- 3) 위상 스펙트럼
- 4) 진폭스펙트럼
- 5) 역 변환 (원래의 신호가 됨을 확인 할수 있음)

### 11. dftanaly

이산푸리에 변환의 과정을 해석하여, DFT를 이해한다.

(지수함수를 예로들어 설명)

- 1) 지수함수  $h(t)$ 의 시간, 주파수영역 표시
- 2) 펄스파형  $g_1(t)$  : Sampling 용
- 3)  $h(t)g_1(t)$  : 신호의 이산화
- 4)  $x(t)$  도입 : 해석대상영역 결정용 window
- 5)  $h(t)g_1(t)x(t)$  : 해석 대상영역 결정
- 6)  $(h(t)g_1(t)c(t))*g_2(t)$

### 12. wdsptm

Rectangular, Bartlett, Hanning, Hamming, Black-man Window의 진폭스펙트럼을 나타내고, 각window를 상호비교한다.

#### - 컴퓨터 환경 및 사용방법

- . IBM PC/AT 이상
- . Graphic monitor : VGA
- . INTROSP로 모든것이 동작되며, 일반적으로, 탈출은 ESC, 다음화면, 혹은 계속은 ENTER를 사용한다. (사용 커맨드들이 화면에 디스플레이 됨)

## 2. 디지털 전자회로 실험을 위한 Digital IC Trainer 개발

- 실험항목 :

Gate회로 실험 : AND Gate, OR Gate  
Not Gate, NAND Gate  
NOR Gate, EX-OR Gate

조합 Gate회로의 실험 :

2중 반전의 실험  
NAND, NOR로 NOT을 만들  
De-Morgan 정리 1, 2  
NAND, NOR로써 AND, OR을 만들

Gate로 구성하는 발진회로의 구성 :

NOT회로로 구성  
NAND 회로로 구성  
발진주파수의 변경

Flip/Flop회로의 실험 :

RS FF  
FF의 동작과 응용  
D FF  
JK FF  
T FF과 분주동작

응용회로의 실험 :

Counter IC에 의한 BCD Code  
7 Segment Driver 회로  
간단한 가산회로

\* 건전지 1.5V 4개로 동작되고, 크기가 작아( 14.4 cm x 9 cm) 가지고 다니면서 학습하기에 편리하도록 만들어 졌음.

### 3. 음성 데이터 수집을 위한 발성내용 제시시스템 구현[1]

음성정보처리 분야에서, 음성 데이터베이스 구축의 필요성 및 중요성이 더욱 강조되고 있다. 음성연구를 하기위하여는, 음성데이터가 필요함은 물론, 가능하면 많은 종류, 많은 량(성별, 나이, 방언, 화자수등)의 데이터가 필요한데 이는, 음성연구가 고도화되고 고신뢰성이 있는 연구가 되기위하여는 더욱 절실히 요구되는 사항이다. 이러한 많은량의 음성을 수집함에 있어서, 가장 중요한점은 정상시대로의 자연스럽게 발성된 음성데이터를 수집하는 일이다. 일반적으로 음성을 수집하기 위하여 녹음을 한다고 하면 정상시의 자연스러운 발성이 나오지 않고 긴장하게 되고 조심스럽게 발성이 되어 실제적인 효과를 얻는데는 상당히 어렵다. 따라서, 어떻게 하면 보통때의 발성과 똑같이 발성된 음성을 얻을수 있을까? 가 중요한 연구의 대상이 된다. 즉, 음성데이터 수집을 위한 발성내용 제시방법에

대한 연구이다. 보통 생각될수 있는 방법으로는 첫째, 발성단어 혹은 문장의 리스트를 만들어 두고 발성자로 하여금 읽게 하는 방법이다. 이는 가장 간단한 방법이지만 녹음중에 불필요한 종이를 넘기는 소리라든가 복수회 발성을 할때 몇번째 발성인지 혹은 방금 발성한 단어가 어느단어인지 잊어버리기 쉬운 단점이 있다. 두번째, 발성해야할 단어를 헤드폰을 통하여 음성으로써 알려주는 방법이 있다. 이는 리스트를 보고 발성하는 방법보다는 좋지만 들고 발성하므로써 발성자 나름대로 발성하기 보다는 인토네이션, 발성방법, 소리의 크기등이 헤드폰에서 나오는대로 발성하려는 경향이 있어 개인성이 상실될 염려가 있다. 세번째, 발성해야할 대상단어, 또는 문장을 개인용 컴퓨터등의 화면에 디스플레이 시켜 발성자가 읽게하는 방법이 있다. 이 방법은 발성자에게 발성에 대한 불안과 부담을 적게 주며, 발성등의 잘못이 있을때는 미리 구현되어 있는 컴퓨터의제어기능을 이용하여 수정작업이 쉽게 이루어 지는 장점이 있다. 또한 발성시간, 나아가서는 발성의 크기등을 적절히 조정할수 있어 매우 좋은 방법이다. 그러나, 이는 발성 데이터들을 화면에 디스플레이 시키고 발성자와 운영자를 위한 각종 기능을 제어할수있는 소프트웨어구현등의 발성을 위한 도구가 필요하다. 네번째, 가장 좋은 방법은, 화면등을 이용하여 원하는 발성이 나오도록 유도 질문하는 방법이다. 즉, 화면상에 원하는 대답(발성)이 나올수 있도록 질문하는 방법이다. 외국의 경우를 보더라도 주로 첫번째 방법인 발성리스트를 이용하는것이 가장 보편적이고 세번째의 방법도 소프트웨어의 구현문제 때문에 널리 이용되고 있지는 않은 실정이다.

본논문에서는, 상기의 세번째 방법을 위한 소프트웨어를 구현함에 있어서 고려해야 할 발성내용 제시 시스템의 요구사항을 정의하고, 범용 시스템으로써의 기능과 구현에 대하여 기술한다.

#### 4. 4연속 숫자음에 대한 음성구간 검출 알고리즘 개발[2]

현재 음성 인식 합성에 관한 연구는 하드웨어 및 소프트웨어 등 관련 기술의 진보에 따라 한정 어휘에 대한 인식 합성 시스템이 실용화되고 있으며, 임의 어휘의 음성인식 합성 시스템에 대한 연구도 활발히 진행되고 있다. 특히 불특정 다수어의한 연속음성을 대상으로 하는 인식 시스템의 개발에는 고도의 인식 기술 개발이 필수적이지만, 연속 음성중의 음소의 특징은 발성자에 따라 혹은 단어내의 음소의 조음결합에 의해 크게 변화하기 때문에 인식시스템의 설계에 어려움이 많다. 따라서 이러한 개인차 및 조음결합의 현상등을 파악하는 연구를 수행하기 위해서는 많은 사람이 발성한 대량의 음성 데이터가 필요하다. 또, 음성관련 시스템이 점차 실용화 되어감에 따라 시스템의 객관적인 성능 평가에도 이용할 수 있는 표준 음성 데이터베이스도 필요하다. 이를 위하여 외국에서는 이미 공통으로 이용 가능한 음성 데이터베이스를 구축하고 이를 배포하고 있는 단계에 있다. 대표적인 음성 데이터베이스로는 미 국방성의 주도에 의해 제작한 연속 음성 인식을 위한 DARPA 음성 데이터베이스, 일본의 경우 ETL 음성 데이터베이스등이 구축되어 공동 사용되고 있다. 우리나라의 경우는 아직 외국처럼 체계적으로 만든 음성 데이터베이스가 보급되어 있지는 않

으며, 대학 및 연구소에서 실험에 필요한 소규모의 음성 데이터를 연구자 각자가 자체 제작하여 사용하고 있는 정도이다.

한편 이러한 음성 데이터베이스를 구축하여 음성인식에 이용하고자할 때 수록된 음성 신호로부터 무음(silence)부분과 음성부분을 정확하게 구분하는 것이 필요하다. 이는 정확한 음성구간의 검출이 데이터베이스의 구축時 불필요한 데이터의 양을 줄임으로서 메모리를 절약하고, 인식時에는 인식시간을 단축시킬 수 있을 뿐만 아니라 인식률 향상에도 기여하기 때문이다. 그러나 이러한 음성구간의 검출은 음성의 발생 또는 기록환경, 전송시스템 등에서 야기되어질 수 있는 자연적 인공적 잡음의 영향으로 인하여 그리 용이하지만은 않다. 또 현재까지 발표되고있는 인식시스템의 대부분은 이러한 음성구간의 검출을 觀察에 의하여 手動으로 행하고 있는 실정이며 자동으로 행하고 있는 경우도 대략적인 검출을 거친후 인식시 처리하는 경우가 대부분이다. 따라서, 불필요한 여분의 데이터가 첨가되게 됨으로 인해 음성 데이터베이스의 규모가 방대해지며 인식시에도 많은 시간이 필요하게 된다. 따라서, 이러한 문제점을 해결하기위해 음성 자동다이얼링을 위한 데이터베이스 구축을 위해 한국어 4연속 숫자음을 대상으로 음성 구간 검출방법을 검토하였다. 특히, 정확한 음성구간 추출을 위하여 가능한 새로운 알고리즘을 개발하여 음성 데이터베이스의 규모를 줄이는 한편 인식률 제고에도 기여 하고자 한다. 실험결과 숫자음을 음소군별로 세분화하여 각각의 음소군에 적합한 검출 알고리즘을 적용했을 경우 검출오차를  $\pm 20\text{ms}$ 로 하였을 때 각 음소군별 끝점검출율은 시점의 경우 평균 95.3% 종점의 경우 평균 91.5%의 검출율을 얻었다.

- 참고문헌

- . E. Oran Brigham, "The fast Fourier Transform", Prentice-Hall, 1988
- . Robert A. Gabel et al, "Signals and Linear System" 3ed.,
- . 松尾博, "やさしいフ-リエ變換", (1986)
- [1] 김 경 태 외, "음성데이터 수집을 위한 발성내용 제시 시스템", 제9회 음성통신 및 신호처리 워크샵, 1992.8
- [2] 김 경 태 외, "음성편집 시스템을 위한 4연속숫자음의 음성구간 검출", 제9회 음성통신 및 신호처리 워크샵, 1992.8