퍼지 및 지능시스템학회 논문지 2001, Vol. 1, No. 1, pp. 82-87

DTW방식을 이용한 음성 명령에 의한 커서 조작

Cursor Moving by Voice Command using DTW method

추명경 · 손영선

Myung-Kyung Chu and Young-Sun Sohn

동명정보대학교 정보통신공학과

요 약

본 논문에서는 마우스 대신에 음성으로 명령을 입력하여 퍼지 추론을 통해 윈도우 화면상의 커서를 이동시키는 인터페이스를 구현하였다. 입력된 음성이 대체로 짧은 언어이기에 이를 인식하기 위하여 고립단어 인식에 강한 DTW방식을 사용하였다. DTW방식의 단점중인 하나가 음성길이가 비슷한 명령을 입력하였을 때 표준패턴 중 오차 값이 가장 작은 패턴으로 인식하는 것이다. 예를 들면 '아주 많이 이동해'라는 음성이 입력되었을 때 비슷한 음성길이를 가진 '아주 많이 오른쪽'으로 인식하는 경우가 있다. 이런 오류를 해결하고자 각 패턴의 DTW오차 거리 값 과 표준 패턴의 음성길이를 기준으로 임계값을 퍼지 추론하여 명령으로서의 수락 여부를 결정하였다. 판단이 애매한 부분은 사용자에게 질의를 하여 응답에 따라 수락 여부를 결정하였다.

ABSTRACT

In this paper, we realize an interface that moves a cursor on the display of window system by voice command instead of a mouse operation using fuzzy inference. Because of the inputted voice are on the whole short, DTW method that is strong to recognize isolated word is used to recognize inputted voice. When the length of an inputted command is similar to the other command, recognizing the smallest value of error among the standard patterns is one of the demerits of DTW method. For example, if the voice 'move very much' is inputted, then the command 'very much right' having the similar length of voice is recognized in some cases. To solve this sort of error, we inferred the value of DTW error distance and the critical value to the voice length of standard pattern using fuzzy inference, and we decided acceptance and rejection as command. If decision is ambiguous, then our system asks to the user and decided acceptance and rejection to the answer.

Key Words : 퍼지추론, 휴먼인터페이스, 음성인식.

1. 서 론

정보화 사회에서는 사람과 기계사이의 정보교환에 상당한 비중을 차지하고 있다. 이로 인해 인간과 기계간의 자연스러운 정보교환을 위한 인간-기계 interface기술의 중요성이 대두되고 있으며 이에 대한 연구가 계속되고 있다 [1,2,3]. 최근 컴퓨터의 사용이 많아지면서 인간과 컴퓨터 사이에 정보를 전달하는 방법으로 일상 생활에서 사용하는 음성이 다른 수단 없이 정보를 전달할 수 있다는 편리함 때문에 많이 이용 되어지고있다[3,4,5].

본 논문에서는 인간과 컴퓨터간의 interface기술 중 커서의 이동[1]을 음성언어를 이용하여 구현하였으며 자연어에 내포되어 있는 애매함을 해결하기 위해 퍼지추론[8]을 사용하였다.

접수일자 : 2000년 11월 18일
완료일자 : 2001년 01월 15일

2. 전체 시스템 알고리즘

그림 1의 시스템 흐름도에서 알 수 있듯이, 음성명령이 입력되어지면 음성인식 알고리즘을 통하여 명령을 인식한다. 인식된 명령은 DTW 유사도에 대한 퍼지 추론 알고리즘을 통하여 명령으로서의 수락여부를 결정한다. 명령으로서 거절되어지면 사용자에게 명령을 재 입력하도록 알려주며, 판단하기 애매한 경우에는 사용자에게 질의하여 응답여부에 따라 결정하였다. 인지된 명령은 커서이동에 대한 퍼지 추론 알고리즘을 통하여 커서의 새로운 위치를 얻어 커서를 이동시킨다. 사용자가 원하는 아이콘으로 커서가 이동하였을 경우에는 아이콘을 click하여 사용자에게 확인하도록 하였다.

3. 커서 이동에 대한 퍼지 추론 알고리즘

커서를 이동시키기 위한 명령은 상하, 좌우로 구분하여 '아주 많이 위로'에서 '아주 많이 아래로'의 10개의 명령어,'아주 많이 좌로'에서 '아주 많이 우로'의 10개의 명령어, 총 20개의 명령어를 사용하였으며, 멤버쉽 함수는 컴퓨터 화면상의 픽셀 단위로 x좌표를 [0,1016], y좌표를

[0,762]의 범위로 구성하였다. 커서 이동은 현재 커서의 위치에 따라서 사람이 느끼는 거리감의 정도가 다르기 때문에 상대적인 거리 개념을 도입하여 멤버쉽 함수가 변화하는 추론 방법을 이용하였다[1].

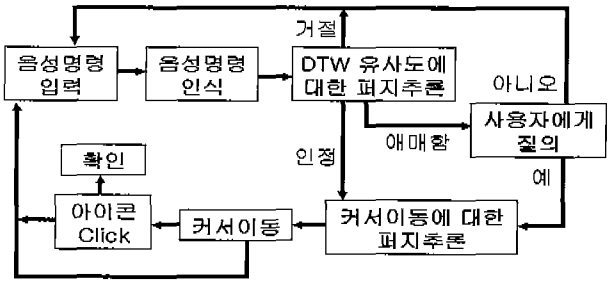


그림 1. 시스템 흐름도

Fig.1 Flow of system

커서가 화면 중앙에 위치할 때에는 화면 전체를 범위로 목표지점을 생각하게 되지만 그림 2에 보여지듯이 커서 이동 후에는 그 새로운 위치를 중심으로 생각하고 커서의 위치에 따라 상대적으로 명령을 내리면 된다. 상대적인 거리 개념을 사용함으로써 아주 좌측 또는 아주 우측에서 역 방향으로 근접한 거리로 이동하거나, 현재 위치에서 조금씩 이동하기는 어렵기 때문에 이를 보완하기 위하여,'조금'이라는 수식어를 입력하여 중심을 기준으로 그림 3과 같이 좌우 양쪽 범위의 각각 1/5로 이동범위를 좁혀 추론하였다. 멤버쉽 함수는 서로 독립적이지 않고 같은 범위를 공유하고 있기 때문에 인접하여 있는 멤버쉽 함수를 함께 선택하여 퍼지 추론하였다.

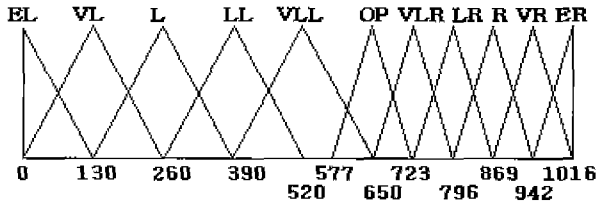


그림 2 명령에 의해 변화된 멤버쉽 함수

Fig.2 Changed membership function by commands

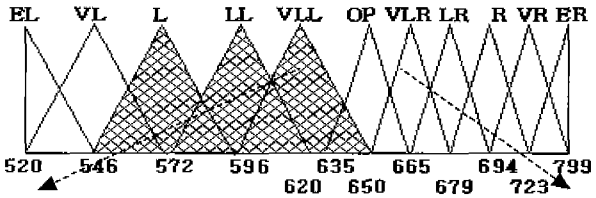


그림 3. 조금 왼쪽을 선택 할 경우

Fig.3 In case of the command 'little left'

4. 음성 명령의 인식 알고리즘

음성인식을 처리하는데는 패턴매칭 방법을 사용하였다. 명령에 대한 음성패턴의 특징을 추출하여 표준패턴을 만든 뒤, 음성명령이 입력되면 저장된 표준패턴과 비교하여 가장 유사한 패턴을 찾아 인식하는 방법으로서 그림 4에서 음성인식 시스템의 개념도가 보여진다.

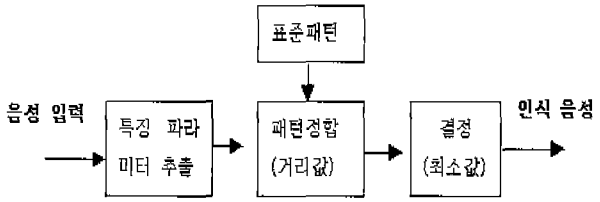


그림 4. 음성인식 시스템의 개념도

Fig. 4 Concept of voice recognition system

음성명령을 인식하기 위한 전체 알고리즘은 그림 5와 같이 구성하였다.

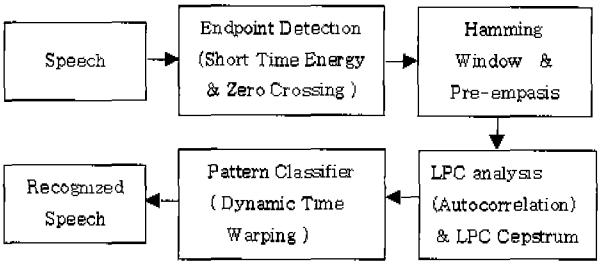


그림 5. 음성 인식 전체 흐름도

Fig. 5 Flow of voice recognition

4.1 실시간에서의 음성구간 추출 [7]

on_line으로 음성을 받아들이고 있는 상태에서 음성의 유무를 찾아야 하기 때문에 계산량이 그리 크지 않는 방법을 사용하여, 현재 많이 이용되어지고 있는 short time energy 와 zero-crossing 방법을 사용하였다.

4.1.1 short time energy

음성 분석 구간에서 절대 에너지 값을 구하여 이들을 기준으로 음성신호의 유무를 판별하는 방법이다. 이때 유성음의 파형은 절대 에너지 크기 값이 크며, 무성음의 파형은 절대 에너지 크기 값이 아주 낮으므로 식(1)에 의해 한 분석구간의 절대 에너지 값이 에너지 상한 값을 넘으면 음성이 있다고 간주한다.

$$E = \sum_{i=0}^{N} |x(i)| \qquad (1)$$

4.1.2 zero-crossing

음성 신호 파형의 위상이 중심 축을 통과하는 횟수를 이용한 방법이다. 무성음의 경우 진폭이 크지 않고 불규칙적인 진동이므로 유성음보다 영 교차율이 크다. 영 교차율은 식(2)로 정의된다.

$$ZCR=\sum_{i=0}^{N-1}|sgn(x(i))-sgn(x(i+1))|*(1/2)$$
$$sgn(x)=1(x\geq 0)$$
$$sgn(x)=-1(x<0) \quad (2)$$

4.2 LPC(Linear Prediction Coding)처리 과정[7]

사용자 음성의 특징 정보들을 얻기 위해서 LPC계수를 사용하였다. LPC는 사람의 발성 기관을 하나의 필터로 가정하고 그 필터의 계수를 음성의 특징 벡터로 사용하는 것이다. 그림 6은 발성기관을 수학적으로 모델링 한 것이다.

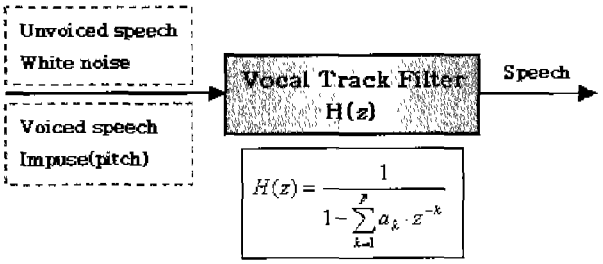


그림 6. 발성기관의 수학적 모델링
Fig.6 Mathematical model of a vocal organ

음성구간에 LPC를 coding함에 있어서 보다 정확한 양자화를 위해 우선 windowing 과 pre-emphasis 처리과정이 필요하다.

4.2.1 Windowing

windowing은 연속적인 음성신호의 일부분을 분석하기 위한 수식으로 원하는 길이 만큼의 데이타를 일정하게 분리하는 과정이다. 여기서 이용한 Hamming windowing은 분석 구간의 가장자리로 갈수록 신호 데이터의 크기를 자연스럽게 점차 줄여감으로써, analysis block의 Boundary 근처의 신호의 왜곡을 줄여 잡음 성분의 발생을 방지하도록 하는 것으로서 식(3)과 같이 정의되어진다.

$$W(n)=0.54-0.46\cos\left(\frac{2\pi n}{N-1}\right)0\leq n\leq N-1 \quad (3)$$

4.2.2 Pre-emphasis

분석구간내의 중심 축 가까이의 미세한 고주파 파형은 저주파에 비해 현저히 작아 정확한 파형 분석이 어렵다, 따라서 pre-emphasis 처리를 통하여 기존 주파수 이외의 성분을 강화시켜 파형의 정보를 추출한다. 즉, 저주파 대역에서는 음성신호의 에너지를 감소시키고 고주파 대역에서는 증가시키는 처리과정으로 식(4)과 같이 정의되어진다.

$$\bar{x}(n)=x(n)-\alpha\bar{x}(n-1) \quad 0.9\leq\alpha\leq 1.0 \quad (4)$$

4.2.3 LPC analysis

LPC 분석이란 과거 p개의 sample을 가지고 현재의 sample을 예측하는 방식으로 식(5)와 같은 관계를 가진다.

$$y(n)=x(n)+\sum_{k=1}^{p}a_k\cdot y(n-k) \quad (5)$$

LPC분석을 위하여 안정성이 좋은 자기 상관 방법을 사용하였다. 자기 상관 방법은 식(6)에서 알 수 있듯이 어느 일정의 거리만큼 떨어져 있는 두 신호 표본의 유사한 정도를 거리와 유사성의 관계의 함수로서 표현한 것이다.

$$R(i)=\sum_{k=1}^{p}a_kR(|i-k|) \quad (6)$$

LPC 계수를 산출하는 과정에서 Durbin's method를 사용하였는데 이 방법은 바로 전의 패턴에 대한 정보만 알면 현재의 패턴정보를 알 수 있는 장점으로써 계산적인 면에서 이점이 있다.

4.3 Cepstrum [6]

Cepstrum은 주파수 영역의 함수를 역 변환 한 것이기 때문에 시간영역의 함수라 할 수 있다. 가장 큰 특징은 음성이 갖는 정보에서 스펙트럼 포락 정보와 세부구조 정보를 분리하는 기능을 가진다는 것이다. 구하여진 LPC 계수는 식(7)에 의해 LPC cepstrum 계수를 얻을 수 있다.

$$\hat{C}_1=-\alpha_1$$
$$\hat{C}_n=-\alpha_n+\sum_{m=1}^{n-1}\frac{m}{n}\alpha_m\hat{C}_{n-m}\ (1<n\leq p)$$
$$\hat{C}_n=\sum_{1}^{n}\frac{m}{n}\alpha_m\hat{C}_{n-m}\ (p<n) \quad (7)$$

구하여진 cepstrum 계수의 민감도를 완화시키기 위해 식(8)에 보여지는 tempered window 기법을 사용하였으며 더 향상되고 확장된 스펙트럼의 표현을 위해 식(9)에 보여지는 시차를 적용한 delta cepstrum을 사용하였다.

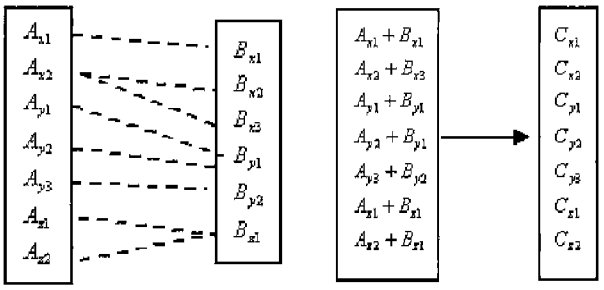
$$W_m=\left[1+\frac{\Omega}{2}\sin\left(\frac{\pi m}{\Omega}\right)\right],1\leq m\leq\Omega \quad (8)$$

$$\frac{\partial}{\partial w}[\log|S(e^{jw},t)|]=\sum_{m=-\infty}^{\infty}\frac{\partial C_m(t)}{\partial t}e^{-jwm}$$
$$\frac{\partial C_m(t)}{\partial t}=\Delta C_m(t)\approx\mu\sum_{k=-K}^{K}kC_m(t+k) \quad (9)$$

4.4 DTW 방식 [6]

동일인이 동일한 단어를 발성하여도 발성 할 때마다 단어의 시간적 길이가 변화한다. 이를 표준 패턴과 단순히 비교하면 시간 축이 고르지 않기 때문에 오류나 인식 불

능이 생기게 된다. 이를 해결하기 위하여 비 선형적인 시간축의 변동 패턴을 선형적으로 정규화 시켜 패턴을 분류한다. DTW 방식은 그림 7에서 볼 수 있듯이 표준 패턴과 입력한 패턴을 서로 정합하여 새로운 다른 패턴을 만들어 인식하는 방법이다.



(a) 프레임 길이가 서로 다른 패턴 (b) 새로운 표준 패턴

그림 7. 패턴의 정합
Fig.7 Pattern matching

유사도를 측정하는 방식으로는 식(10)에서 알 수 있듯이 표준 패턴과 입력된 패턴 사이의 오차 거리로부터 계산되어지는 정규화 거리 방식을 사용하였다. 계산 결과, 유사한 프레임일수록 오차거리는 작은 값을 가진다.

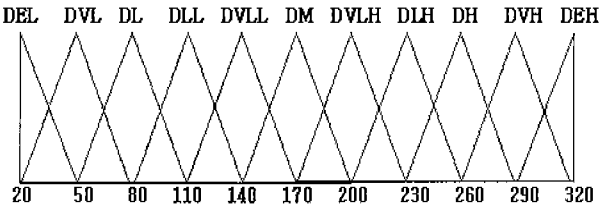
$$D(T_x, T_y) = \min \sum_{k=1}^{T} d(\varphi_x(k), \varphi_y(k)) m(k) \qquad (10)$$

따라서 DTW는 연속단어 보다는 고립단어에 많이 사용되며 고립단어에서 더욱 인식률이 높다. 커서조작을 위한 명령은 고립단어에 가깝게 인식되기 때문에 본 논문에서는 DTW방식을 사용하였다. 총 27개의 명령 중 입력되어진 명령패턴과 27개의 표준패턴의 비교에 의해서 오차 거리가 가장 작은 패턴을 해당 명령으로 인식한다.

5. DTW 임계값에 대한 퍼지 추론 알고리즘

DTW 방식에서는 입력되어진 표준패턴 이외의 명령이 입력되는 경우라도, 표준패턴과 동일한 프레임 길이를 가지며 임계값의 범위 안에 DTW오차 값을 가진다면, 입력되어진 명령으로 잘못 인식하는 문제점이 있다. 뿐만 아니라 표준패턴의 명령을 너무 길게 발음하는 경우에도 표준패턴 중 다른 명령으로 인식하는 경우가 있다. 실험에 의해서도 표준패턴과 동일한 프레임 길이를 가진 다른 음성을 입력한 결과, 표준패턴 중 하나로 인식하는 경우가 있었으며, 명령을 너무 길게 발음하여 다른 명령으로 인식하는 경우가 있음을 확인할 수 있었다. 이를 해결하기 위해서 본 논문에서는 명령으로써의 수락 여부를 퍼지 추론 알고리즘을 통하여 개선하였다.

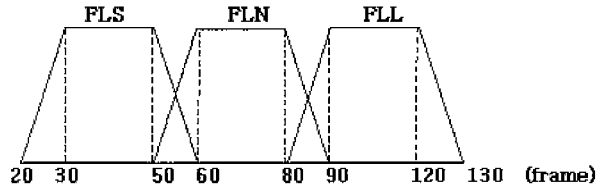
표준패턴과 입력된 명령패턴의 오차 거리에 대한 멤버쉽 함수는 그림 8에서 알 수 있듯이 20에서 320사이에 11개로 구성하였다.



DEL (DTW Extremely Low) = DTW 거리 오차 값이 아주 많이 낮다.
DVL (DTW Very Low) = DTW 거리 오차 값이 많이 낮다
DL (DTW Low) = DTW 거리 오차 값이 낮다.
DLL (DTW Little Low) = DTW 거리 오차 값이 조금 낮다.
DVLL (DTW Very Little Low) = DTW 거리 오차 값이 아주 많이 낮다
DM (DTW Middle) = DTW 거리 오차 값이 보통이다
DVLH (DTW Very Little High) = DTW 거리 오차 값이 아주 많이 높다
DLH (DTW Little High) = DTW 거리 오차 값이 조금 높다
DH (DTW High) = DTW 거리 오차 값이 높다.
DVH (DTW Very High) = DTW 거리 오차 값이 많이 높다.
DEH (DTW Extremely High) = DTW 거리 오차 값이 아주 많이 높다.

그림 8 DTW 거리 오차 멤버쉽 함수
Fig.5-1 Membership function of the DTW distance error

커서를 이동시키는 명령들은 '아주','많이','조금'이란 수식어의 조합에 의해서 구성되어져 있으므로 그림 9와 같이 명령의 길이를 20에서 130사이에 3개의 멤버쉽 함수로 구성하였다. 수식어로 이루어진 명령이라서 같은 멤버쉽 함수에 포함되는 명령어라도 프레임 길이가 다르며, 사용자가 명령을 입력할 때 평소와 크게 차이가 나지 않게 발음한다고 가정하여 최대 허용 가능한 범위까지 인정하기 위해 사다리 모양의 멤버쉽 함수로 만들었다.



FLS (Frame Length Short) = 프레임 길이가 짧다.
FLN (Frame Length Normal) = 프레임 길이기 보통이다.
FLL (Frame Length Long) = 프레임 길이가 길다.

그림 9. 프레임 길이 멤버쉽 함수
Fig. 9 Membership function of the frame length

여러 번의 실험을 통해서 프레임 길이에 따른 DTW 오차 값의 허용 가능한 범위를 표 1과 같이 구성하였다.

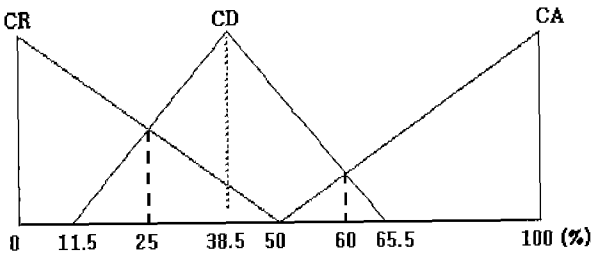
입력된 패턴의 프레임 길이와 계산된 거리 오차 값을 통해 추론되어질 명령으로서의 수락여부 멤버쉽 함수는 그림 10과 같이 명령의 실행 가능(CA), 불가능(CR)과 어느 쪽에 포함시켜야 하는지 판단하기 애매한 경우(CD)로 구성하였다. 불가능(CR)의 경우를 줄이고, 애매한(CD) 경우로 대처함으로써 오인률을 줄이도록 하기 위하여 CD 멤버쉽 함수는 CR 멤버쉽 함수 쪽으로 치우쳐 구성하였다.

표 1. 퍼지 추론 규칙

Table1. The rule table of fuzzy inference

DTW / frame	DEL	DVL	DL	DLL	DVLL	DM	DVLH	DLH	DH	DVH	DEH
VLS	CD	CA	CA	CA	CA	CD	CD	CR	CR	CR	CR
VLM	CR	CD	CA	CA	CA	CA	CA	CD	CR	CR	CR
VLL	CR	CR	CD	CA	CA	CA	CA	CA	CA	CD	CD

얻어진 결과 값이 25%이하인 경우에는 입력한 명령을 실행할 수 없음을 알려 사용자가 명령을 재 입력하도록 하였으며, 25%에서 60%사이일 때는 사용자에게 명령이 맞는지의 여부를 질의하여 사용자의 응답 여부에 따라서 명령의 실행을 결정하도록 하였다. 60%이상인 경우에는 명령을 실행시킨다. 똑같은 명령이 3번 연속으로 입력되어지면 시스템이 잘못 인식했음을 판단하여 사용자에게 명령을 확인시키도록 하였다.



CR(Command Reject): 명령을 거절하다.

CD(Command Delicate) : 명령이 애매하다.

CA(Command Accept) : 명령을 허락하다.

그림 10. 명령의 수락여부 멤버쉽 함수

Fig.10 Membership function of the command acceptance and rejection

본 시스템을 실험해 본 결과, 표준패턴 이외의 다른 명령인 '많이 이동해'란 명령을 65frame으로 입력했을 경우, 223 DTW거리 오차 값을 얻었다. 기존의 DTW 방식을 이용하였을 경우에는 66 frame으로 저장된 '많이 오른쪽'패턴으로 인식하여 명령을 실행 시켰으나 본 논문의 알고리즘을 이용하였을 경우에는 49% 유사도를 얻어 사용자에게 질의를 하여 응답 여부에 따라서 실행시킴으로써 오인률을 줄일 수 있었다.

'왼쪽'이란 명령을 62 frame 크기로 아주 길게 발음하였을 때 288 DTW 오차 값을 얻었다. 기존의 DTW 방식을 이용하였을 경우에는 57 frame으로 저장된 '많이 위로' 패턴으로 인식하여 명령을 실행시켰으나 본 논문의 알고리즘을 이용하였을 경우에는 18% 유사도를 얻어 명령으로서 거부하여 사용자가 명령을 재 입력 할 수 있도록 함으로써 오인률을 줄일 수 있었다.

6. Interface 화면

초기화면은 그림 11에 보여지듯이 정렬된 icon 대신에 실행되어지는 버튼들을 random하게 배치시켰고 커서는 화면 중앙에 위치하도록 하였다. 수회의 상하, 좌우 명령에 의해 커서가 원하는 버튼위치까지 이동되어지면 버튼의 색깔이 변화되어지고, click이란 명령에 의해 해당 버튼의 기능이 실행되어진다. 화면상에 나타난 메시지 박스들은 '확인','OK' 또는 'NO'라는 명령에 의해 동작되어지며, 커서를 화면 중앙에 위치시킬 때는 'screen center' 명령, 프로그램을 종료하고자 할 때는 'screen close'명령을 사용한다.

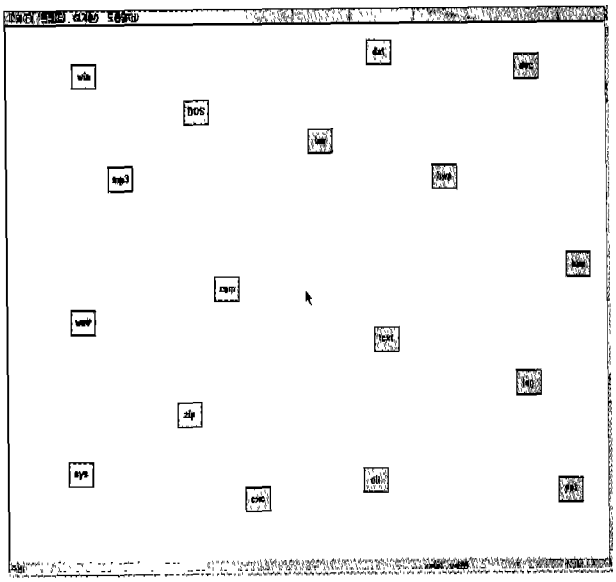


그림 11. 구현된 초기 화면

Fig.11 A realized initial screen

7. 결론 및 향후과제

기존의 DTW방식은 임의의 임계값 범위 안에서 두 패턴의 거리 오차가 최소인 패턴을 인식한다. 본 연구에서는 DTW방식을 이용하여 단어를 인식하는 과정 중에 비슷한 음성길이의 다른 단어가 입력되었을 때 잘못 인식하는 오류를 줄이고자 음성길이에 따라 오차 허용 범위를 퍼지 추론을 통해 음성인식 결과를 결정하였다. 실험결과 음성이 입력되었을 때 제안된 방법이 기존의 DTW 방식보다 오류률이 적었으며, 판단하기 애매했던 부분에 대해서는 사용자에게 질의를 함으로서 오류도 줄일 수 있었다. 향후 과제로는 잡음이 있는 공간에서 사용하기 위해 오차 값 허용범위가 유동적으로 변해야하며, 누구나 사용할 수 있는 화자독립 시스템을 구축하는 것이라 고려되어진다.

참 고 문 헌

[1] 손영선 외 4,"퍼지 추론에 의한 커서의 조작", 한국 퍼지 및 지능시스템 학회 1999년도 추계학술대회 학술논문발표논문집, Vol.9, No.2, pp.239-242, 1999
[2] Kabsuk OH, Geuntaek Kang, Kaoru HIROTA, "A Support System Construction for Multimedia Information Data Acquisition", FUZZ-IEEE '99, VolⅡ, pp.841-846, 1999
[3] CHORAYAN, "CONTRIBUTION TO FUZZY ALGORITHM CONSTRUCTION FOR NATURAL INTELLIGENCE", IFSA'97, VolⅣ, pp.79-81, 1997
[4] 유하진,오영환,"불균일 인식 단위를 이용한 연속음성 인식 퍼지 전문가 시스템", 韓國情報科學會論文誌, pp.128-135, 1995
[5] 장성령, 공성곤, "LPC와 RBF 신경망을 이용 한 화자 인식",한국 퍼지 및 지능시스템 학회 1999년도 추계학술대회 학술논문발표논문집, Vol.9, No.2, pp.102-107, 1999
[6]오영환, 음성언어정보처리, 홍릉과학출판사, 1998
[7]박경법, 음성의분석및합성과그응용, 그린, 1997
[8] 本多中二, 大理有生, "ファジイ 工學入門", 1998

저 자 소 개

추 명 경 (Myung-Kyung Chu)
2001년 동명정보대학교 정보통신공학과 졸업(공학사).
관심분야 : 휴먼인터페이스, 퍼지이론
Phone : 018-765-7370
Fax : 02-449-3028
E=mail : mirror1228@hanmail.net

손 영 선 (Young-Sun Sohn)
1981년 동아대학교 전자공학과 졸업 (공학사)
1983년 동대학원 졸업(공학석사).
1990-1998년 한국전자통신연구소 선임연구원
1998년 쯔쿠바대학 졸업(공학박사).
1998년-현재 동명정보대학교 정보공학부 조교수
관심분야 : 휴먼인터페이스, 퍼지 측도·적분, 평가
Phone : 051-629-7232
Fax : 051-629-7249
E=mail : yssohn@tmic.tit.ac.kr