

# 퍼지 추론을 이용한 음성 명령에 의한 커서 조작

## Cursor Moving by Voice Command using Fuzzy Inference

추 명경 , 손 영선

동명정보대학교 정보통신공학과

Myung-Kyung Chu, Young-Sun Sohn

Communication Engineering , TongMyung University of Information Technology

E-mail : yssohn@tmic.tit.ac.kr

### ABSTRACT

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 음성길이와 비슷한 명령을 입력하였을 때 표준패턴 중 오차 값이 가장 작은 패턴으로 인식하는 것이다. 예를 들면 '아주 많이 이동해' 라는 음성이 입력되었을 때 동일한 음성길이를 가진 '아주 많이 오른쪽' 으로 인식하는 경우가 있다. 이런 오류를 해결하고자 각 패턴의 DTW 오차 값 범위와 표준 패턴의 음성길이를 기준으로 임계값을 퍼지 추론하여 명령으로서의 수락 여부를 결정하였다. 판단이 애매한 부분은 사용자에게 질의를 하여 응답에 따라 수락 여부를 결정하였다.

### 1. 서론

정보화 사회에서는 사람과 기계사이의 정보 교환에 상당한 비중을 차지하고 있다. 이로 인해 인간과 기계간의 자연스러운 정보 교환을 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 이용되고 있다[3, 4, 5]. 본 논문에서는 인간과 컴퓨터간의 interface기술 중 커서의 이동[1]을 음성

언어를 이용하여 구현하였으며 자연어에 내포되어 있는 애매함을 해결하기 위해 퍼지 추론[7]을 사용하였다.

### 2. 커서 이동에 대한 퍼지 추론 알고리즘

상하, 좌우로 구분하여 총 20개의 명령을 입력하여 커서를 이동시키며 컴퓨터 화면상의 픽셀 단위로 멤버십 함수의 범위를 정하였다. 커서 이동은 현재 커서의 위치에 따라서 상대적인 거리 개념을 도입하여 멤버십 함수가 변화하는 추론 방법을 이용하였다[1]. 즉 커서가 화면 중앙에 위치할 때에는 화면 전체를 범위로 목표지점

을 생각하게 되지만 그림 2-1에 보여지듯이 커서 이동 후에는 그 새로운 위치를 새로운 중심으로 생각하고 커서의 위치에 따라 상대적으로 명령을 내리면 된다.

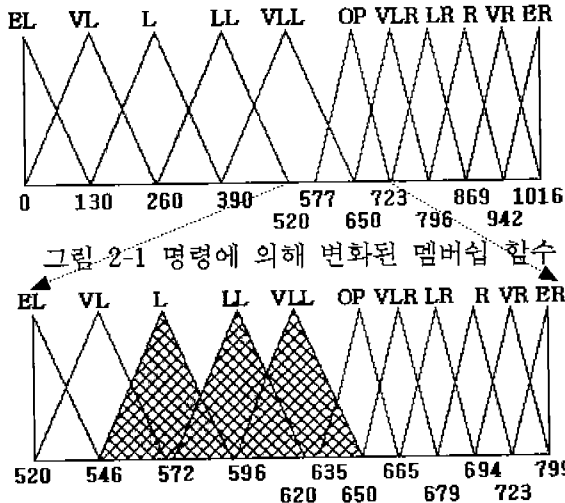


그림 2-2 조금 왼쪽을 선택 할 경우

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 2-2와 같이 좌우 양쪽 범위의 각각 1/5로 이동 범위를 좁혀 추론한다.

### 3. 음성 명령의 인식 알고리즘

음성인식을 처리하는데 패턴 매칭 방법을 사용하였다. 명령에 대한 음성패턴의 특징을 추출하여 표준패턴을 만든 뒤 음성명령이 입력되면 저장된 표준패턴과 비교하여 가장 유사한 패턴을 찾아 인식하는 방법으로서 그림 3-1에서 전체적인 블록도가 보여진다.

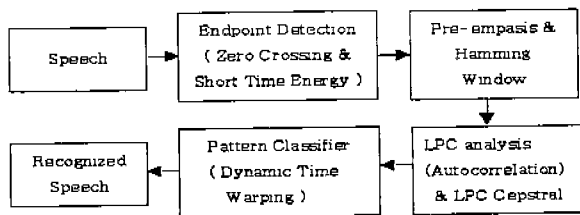


그림 3-1 음성 인식 블록도

### 3.1 실시간에서의 음성구간 추출 [6]

on\_line으로 음성을 받아들이고 있는 상태에서 음성의 유무를 찾기 위해 short time energy 와 zero-crossing 방법을 사용하였다.

### 3.2 LPC(Linear Prediction Coding)처리 과정[6]

사용자 음성의 특징 정보들을 얻기 위해서 자기 상관 방법을 사용하여 LPC분석을 한다. LPC 분석이란 과거 p개의 sample을 가지고 현재의 sample을 예측하는 방식이다. 음성구간에 LPC를 coding함에 있어서 보다 정확한 양자화를 위해 우선 pre-emphasis 와 windowing 처리과정을 거친다. 구간내의 미세한 파형은 정확한 파형 분석이 어려움으로 pre-emphasis 처리를 통해 저주파 대역에서는 음성신호의 에너지를 감소시키고 고주파 대역에서는 증가시킨다. 연속적인 음성신호의 일부분을 분석하기 위해 Hamming windowing를 사용하여 analysis block의 Boundary 근처의 음성을 감소시켜 신호의 왜곡을 줄인다. 특정 파라미터를 산출하는 과정에서 LPC계수를 얻는데 Durbin's method를 사용하였다. 이 방법은 바로 전의 패턴에 대한 정보만 알면 현재의 패턴정보를 알 수 있는 장점으로써 계산적인 면에서 이점이 있다.

### 3.3 Cepstrum [6]

Cepstrum은 음성이 갖는 정보에서 스펙트럼 포락 정보와 세부구조 정보를 분리하는 기능을 가진다. 본 논문에서는 14차 LPC cepstrum을 통해서 파생된 스펙트럼 포락선을 구하였다. 구하여진 cepstrum 계수의 민감도를 완화시키기 위해 tempered window 기법을 사용하였으며 더 향상되고 확장된 스펙트럼의 표현을 위해 시차를 적용한 delta cepstrum을 사용하였다.

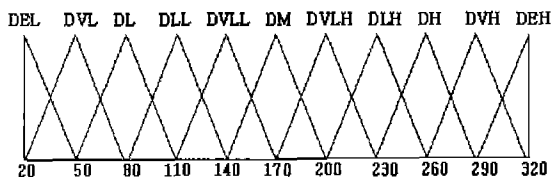
### 3.4 DTW 방식 [6]

DTW방식은 비 선형적인 시간축의 변동 패턴을 선형적으로 정규화 시키는 방식이다. 정규화 거리는 표준 패턴과 입력된 패턴 사이의 오차 거리로부터 계산되어진다. 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.

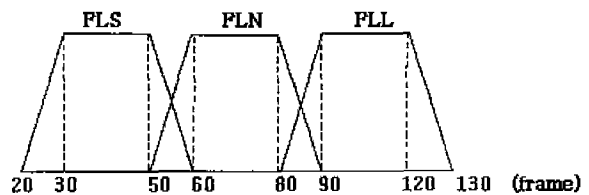
#### 4. DTW 임계값에 대한 퍼지 추론 알고리즘

DTW 방식에서는 입력되어진 표준패턴 이외의 명령이 입력되는 경우라도, 표준패턴과 동일한 프레임 길이를 가지며 임계값의 범위 안에 DTW 오차 값을 가진다면, 입력되어진 명령으로 잘못 인식하는 문제점이 있다. 실험에 의해서도 표준패턴과 동일한 프레임 길이를 가진 다른 음성을 입력한 결과, 표준패턴 중 하나로 인식하는 경우가 있음을 확인할 수 있었다. 이를 해결하기 위해서 본 논문에서는 명령으로써의 수락 여부를 퍼지 추론 알고리즘을 통하여 결정하려고 한다. 표준패턴과 입력된 명령패턴의 오차 거리에 대한 멤버십 함수는 그림4-1에서 알 수 있듯이 20에서 320 사이에 11개로 구성하였다. 커서를 이동시키는 명령들은 ‘아주’, ‘많이’, ‘조금’이란 수식어의 조합에 의해서 구성되어 있으므로 그림 4-2와 같이 명령의 길이를 20에서 130 사이에 3개의 멤버십 함수로 구성하였다. 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 임계값을 추론하여 명령의 실행 가능, 불가능을 판단하는데, 어느 쪽에 포함시켜야 하는지 애매한 경우에는 사용자에게 질의하여 확인하도록 하였다.



- 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 4-1 DTW 거리 오차 멤버십 함수



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

그림 4-2 프레임 길이 멤버십 함수

표 1. 퍼지 추론 규칙

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

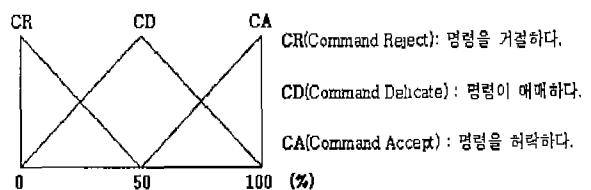


그림 4-3 명령의 수락여부 멤버십 함수

예를 들어, ‘오른쪽’이란 명령을 48 프레임 크기로 입력하였을 때 150 DTW 오차 값을 얻었다면 표 1에 의해서 CA(인정하다), CD(애매하다)의

멤버십 함수가 선택된다. 퍼지 추론 과정을 통하여 중심 값에 의해 비 퍼지값을 얻어낸다. 얻어진 비 퍼지값이 25%이하인 경우에는 입력한 명령을 실행할 수 없다는 메시지 박스를 띄운다. 25%에서 60%사이일 때는 사용자에게 명령이 맞는지의 여부를 질의하여 사용자의 응답 여부에 따라서 명령의 실행을 결정한다. 60%이상인 경우에는 명령을 실행한다. 똑같은 명령이 3번 연속으로 입력되어지면 시스템이 잘못 인식했음을 판단하여 사용자에게 명령을 확인시키도록 하였다.

### 5. Interface 화면

초기화면은 그림5-1에 보여지듯이 정렬된 icon을 대신하여 실행되어지는 버튼들을 random하게 배치시켰고 커서는 화면 중앙에 위치한다. 수회의 상하, 좌우 명령에 의해 커서가 원하는 버튼 위치까지 이동되어지면 버튼의 색깔이 변화되어지고, click이란 명령에 의해 해당 버튼의 기능이 실행되어진다. 화면상에 나타난 메시지 박스들은 '확인', '예' 또는 '아니오'라는 명령에 의해 닫히게 되고, 커서를 화면 중앙에 위치시킬 때는 'screen center' 명령, 프로그램을 종료하고자 할 때는 'screen close' 명령을 사용한다.

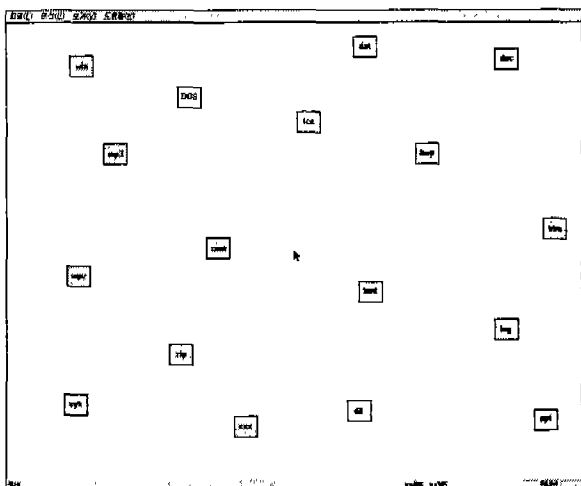


그림 5-1 구현된 초기 화면

### 6. 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. 실험 결과 음성이 입력되었을 때 제안된 방법이 기존의 DTW방식보다 오류률이 적었으며, 판단하기 애매했던 부분에 대해서는 사용자에게 질의를 함으로서 오류도 줄일 수 있었다. 향후 과제로는 잡음이 있는 공간에서 사용하기 위해 오차값 허용범위가 유동적으로 변해야하며, 누구나 사용할 수 있는 화자독립 시스템을 구축하는 것이라 고려되어진다.

### 참고문헌

- [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, Vol II, pp.841-846, 1999
- [3] CHORAYAN, "CONTRIBUTION TO FUZZY ALGORITHM CONSTRUCTION FOR NATURAL INTELLIGENCE", IFSA '97, Vol IV, pp.79-81, 1997
- [4] 유하진, 오영환, "불균일 인식 단위를 이용한 연속음성인식 퍼지 전문가 시스템", 韓國情報科學會論文誌, pp.128-135, 1995
- [5] 장성령, 공성근, "LPC와 RBF 신경망을 이용한 화자인식", 한국 퍼지 및 지능시스템 학회 1999년도 추계학술대회 학술논문발표논문집, Vol.9, No.2, pp.102-107, 1999
- [6] 오영환, 음성언어정보처리, 흥릉과학출판사, 1998
- [7] 本多中二, 大理有生, "フアジイ 工學入門", 1998