논문 2010-1-11

자동차 ECU제어를 위한 음성인식 패턴매칭레벨에 관한 연구

A Study on Voice Recognition Pattern matching level for Vehicle ECU control

안종영*, 김영섭**, 김수훈***, 허강인****

Jong-Young Ahn*, Young-Sub Kim, Su-Hoon Kim***, Kang-In Hur****

요 약 자동차 환경에서의 음성인식은 잡음처리가 매우 중요한 요소이다. 하드웨어 및 소프트웨어로 적인 접근방법으로 많은 연구가 되어 지고 있다. 하드웨어적인 방법으로는 Low-pass filter를 기본으로한 잡음처리 필터가 많이 연구되어 가시적인 성과를 보이고 있고, 소프트웨어적으로는 Noise canceler, 신경망 등 패턴인식 알고리듬의 연구가 이루어지고 있다. 본 논문에서는 시계열 패턴인식에 적용 가능한 알고리듬인 DTW(Dynamic Time Warping)를 자동차 잡음환경에 적용하여 그 음성인식을 위한 파라미터 패턴에 대한 매칭 레벨을 분류하여 잡음환경 적합한 패턴 매칭 레벨을 분석 하였다.

Abstract Noise handing is very important in voice recognition of vehicle environment. that has been studying about to hardware and software approach. hardware method that is noise filter circuit design, basically using Low-pass filter. it was shown a good result. and the side of software that has been developing about to algorithm for Noise canceler, NN(neural network), etc. in this paper we have analysis about to classified parameter pattern matting level for voice recognition on car noise environment that use of DTW(Dynamic Time Warping) which is applicable time series pattern recognition algorithm.

Key Words : .Speech recognition, Filter, Car noise environment, pattern recognition

Ⅰ. 서 론

1960년대 이래로 음성인식기술은 많은 발전을 해 왔지만 아직도 기계에 의한 연속음성인식성에는 많은 어려움과 연구해야 될 부분이 많이 남아 있다. 최근에는 고립단어 기반의 상용제품도 등장하고 있어 향후 실용화 단계까지는 얼마 남지 않아서 상당히 고무적이다. 지금의 상용제품이 나오기 까지는 상당한 시간이 걸렸으며 많은 기업들이 더 많은 음성인식 기술을 수용하려고 준비 중이고 다양한 접목이 시도되고 있다.[1]

음성인식에는 특정화자, 불특정화자를 구별하여 화자종속, 화자독립으로 나눌 수 있으며 인식단어에 따른 고립단어인식과 연속단어인식으로 나누어진다. 화자종속에는 화자인증, 핵심어인증으로 나눌 수 있다.

화자독립의 경우 예상되어지는 모든 화자, 즉 나이별 ,지역별, 년령별, 성별 등의 모든 화자데이터 기초 패턴을 만들어 어느 화자가 이야기하든지 인식가능하게 하는 음성인식방법이고 , 화자 종속은 화자 자신의 음성데이터만을 기초로 하여 비교패턴을 만들고 인식 시 화자의 패턴과 비교하여 구성된 패턴과의 일치 여부를 판단하여

*정회원, 한국폴리텍2대 컴퓨터정보과(동아대 박사과정)
**동아대학교 전자공학과 박사과정
***부천대학 모바일통신과
****동아대학교 전자공학과
접수일자 2010.1.15, 수정일자 2010.2.1

인식하게 된다.

음성인식에 있어서의 가장 큰 영향을 미치는 요소 중의 하나가 바로 음성인식 시 환경적으로 발생하는 잡음이다.

특히, 자동차 환경에서는 그 잡음의 강도가 심해 음성인식을 수행하는데 어려움을 초래 한다.[2]

현재 음성인식에서의 잡음처리 기술은 크게 음성향상(speech enhancement), 특징보상(feature compensation), 모델적응(model adaptation)과 같이 세 가지로 구분된다.

음성의 특징을 추출하여 참조 패턴을 만드는 것이 기본이 되는 데 비교 패턴이 주변잡음으로 인해 영향을 받아 인식률 저하를 발생 시킨다. 음성인식 알고리듬은 크게 확률론적인 접근방법인 HMM(Hidden Marcov Model)이 있고 신경세포를 모델링한 NN(Neural Network)이 있다.

화자 독립의 경우 특성상 HMM을 사용하여 참조패턴을 구성하나 화자종속의 경우 신경망을 사용하여 참조패턴을 만들 수 있으나 이 부분에 대해서는 아직도 다변적인 연구가 이루어지고 있다. 그리고, 데이터가 가지는 그 특징대표벡터를 추출하여 참조패턴을 만드는 VQ(vector Quantization), DTW(Dynamic Time Warping) 방법 등이 있다.[3],[4]

특히, DTW는 시간 축 상에서의 비선형 신축을 허용하는 패턴매칭 알고리듬으로 정의를 하는데 수행을 통하여 참조패턴을 생성할 수가 있다. 그리고, 생성된 참조패턴과 입력패턴을 비교하여 인식여부를 결정한다. 여기서 참조패턴에 대한 입력패턴 비교 시 유사도 기준을 분류하여 적용이 가능하다.

본 연구에서는 화자종속방식으로 DTW(Dynamic Time Warping) 알고리듬을 가진 RSC-4128 processor를 활용하여 패턴매칭 레벨을 변화시켜 자동차 소음환경에 적합한 패턴레벨에 관한 연구를 수행 하였다.

II. 음성인식 시스템

본 연구에서 구현한 음성인식 시스템은 RSC-4128기반의 시스템으로 구현하였다. RSC-4128은 음성인식을 위한 전용 MCU인데 8-Bit processor , 16bit ADC, 10bit DAC&PWM의 DSP 기능을 가지고 있다. 특히, preamplifier가 내장되어 있어서 아날로그 입출력이 가능하다. 5개의Timer(3GP,1Watchdog, 1 Multi Tasking)를 가지고 있고 24개의 I/O라인을 사용할 수 있으며 또한 2개의 비교기를 내장하고 있다.

RSC-4128은 별다른 학습과정이 필요 없이 어떤 화자라도 인식이 가능한 화자독립방식과 운용하는 특정화자의 음성만을 인식 하게 되는, 즉 사용자의 학습과정이 필요한 화자종속 및 화자인증의 알고리듬을 제어가능하게 설계 되어 있다. 화자독립의 경우 Hidden Markov Modeling 과 Neural Net 기술과 접목된 Hybrid기반 기술을 적용하고 있고 화자종속 및 화자인증의 경우는 시계열 패턴인식이 가능한 DTW(dynamic time warping) pattern matching 알고리듬을 수행한다.

또한, 20bit address와 8bit 데이터 인터페이스로 외부 메모리를 사용할 수 있다.[5]

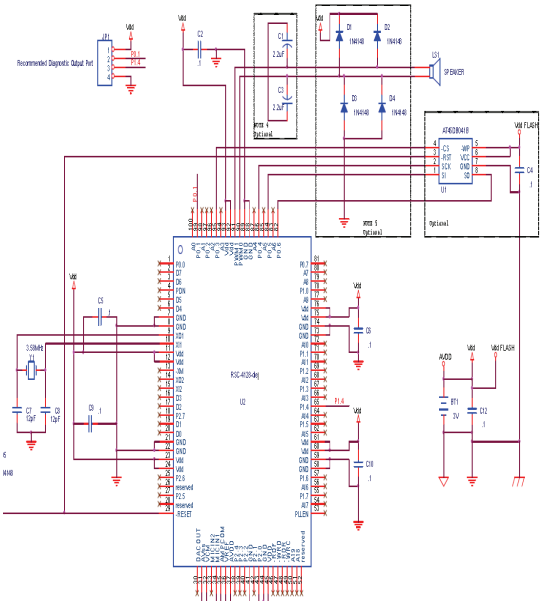

그림 1. 주 시스템
Fig 1. Main System

그림 1는 시스템의 Main Processor부로 입력되어지는 음성신호에 대해서 음성인식과정을 수행하여 인식 여부를 진행하는 부분으로 결과 값을 I/O Port를 통하여 출력하며 패턴인식(DTW) 알고리즘을 수행하는 제어부분으로 시스템에서는 가장 주요한 부분이라고 할 수 있다.

그림2는 아나로그 음성 입력부로 Pull-UP저항으로 민감도를 조정할 수 있도록 설계 되어 있다. 저항값이 클수록 민감도가 높아져 조용한 환경에서는 인식률에 효과를

주지만 소음환경에서는 오히려 소음까지 민감하게 받아들여 인식률이 떨어 질 수 도 있다. 따라서 이와 같은 민감도 부분도 무조건적으로 좋다고 해서 좋은 것은 아니다. 오히려 자동차환경에서는 저항치를 줄여서 민감도를 다소 줄이는 것이 오히려 인식률은 더 좋게 나타난다. 음성 입력부는 화자종속 음성인식 시스템에서 음성을 받아들이는 부분으로 중요한 부분의 하나이다. R,C값을 조정하여 민감도 및 입력특성을 조정할 수 있다.

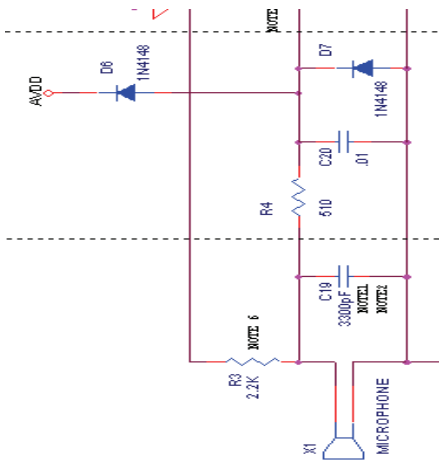

그림 2. 음성 입력부
Fig 2. Speech input circuit

Ⅲ. 패턴 매칭 방법

1. DTW(Dynamic Time Warping)

패턴인식에서 인식의 대상이 되는 패턴은 정적 패턴과 동적 패턴으로 나눌 수 있는데 정적 패턴은 지문, 숫자, 문자와 같이 고정된 영상의 경우 이고 동적패턴은 음성과 같이 시간에 따라서 변하는 패턴에 해당한다.

동적 계획법이 다른 분할-정복(divide & conquer) 알고리즘 등과 구별되는 특징은 메모리에 해당하는 테이블 값과 점화식이 이루는 순환적인 성질을 이용한다는 것이다. 그런데, 모든 문제가 모두 동적 계획법으로 해결될 수 있는 것은 아니다. 어떠한 문제가 동적 계획법을 이용하여 해결 가능하기 위해서는 해당 문제가 최적화의 원리(principal of optimality)가 성립하여야 한다. 최적화의 원리가 적용되는 문제란 「한 문제에 대한 해가 최적이면 그 문제를 이루는 부분 문제들의 해도 최적이다」라는 명제가 성립하는 문제이다. 길이가 다른 두 열에서 어느 한 열을 기준으로 두 열을 비교하기 위해서는 어느 한 열이 신장(늘어남)되거나 축소(줄어듦) 되어야만 한다. 그림 6은 길이가 긴 A열을 길이가 작은 B열을 기준으로 비교하는 경우이다. 이 때 매핑 함수를 통하여 비교가 이루어지는데 이러한 매핑 함수가 직선과 같은 선형적인 경우를 「선형 신축 비교」라고 하고, 곡선과 같은 비선형적인 경우를 「비선형 신축 비교」라고 한다.[6]

DTW 알고리즘을 이용하면 이러한 비선형 매핑 함수를 최적으로 찾아가면서 동시에 비교가 이루어진다. DTW 알고리즘은 일단 두 열의 각 성분에 대한 거리척도 값을 비용으로 설정 한다.

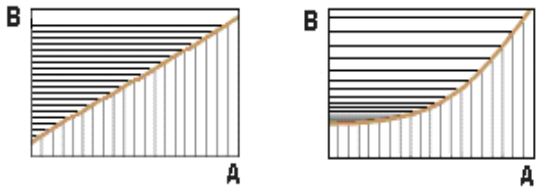

그림 3. 선형과 비선형 매핑
Fig 3. Liner and Non-Liner Mapping

그리고, 두 열이 이루는 격자(lattice)상에서 각 열의 시작 성분에서 시작하여 끝 성분에 이르기까지 비용 테이블에 최소 비용을 순환적으로 택하여 저장하는 점화식을 이용하는 동적 계획법으로 매핑 함수를 찾아가면서 두 열을 비교하는 알고리즘이다. 최종적으로 끝 성분에서 비용 테이블에 저장되는 비용 값이 두 열에 대한 유사도가 된다. 한편, 매핑 함수의 궤적은 앞의 동적 계획법의 최적 탐색패스를 찾는 것과 같이 탐색 과정에서 최소 비용을 택하는 경로를 별도의 경로 테이블에 매 단계마다 저장하고 끝 성분에서 최종 최소 비용을 구한 후에 역추적(backtracking)하여 찾게 된다. 그러므로 DTW 알고리즘은 열의 길이가 일치하지 않는 두 열의 유사도를 측정하는 매칭 알고리즘으로 안성맞춤이라고 할 수 있다. DTW 알고리즘은 주로 음성 인식에서 많이 이용한다. 비교 루틴이 아주 간결하고 단순하여 단어 단위의 간단한 음성 인식기에 적용 가능한 알고리즘이다.

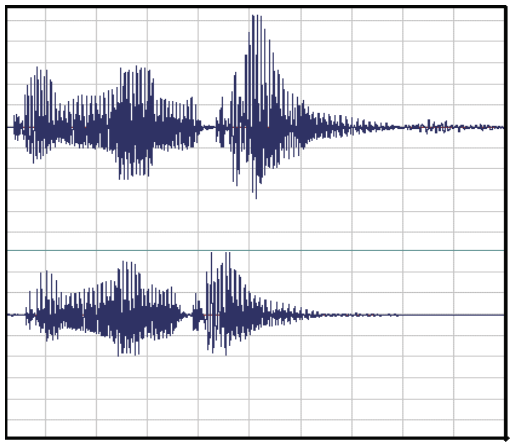

그림 4. 발성이 다른 동일한 음성신호
Fig 4. The same word but different utterance

DTW를 이용한 음성 인식은 그림4의 PCM 디지털 데이터를 그대로 사용하는 것이 아니라, 음성이 10-20ms 지속 시간 동안은 정상적(stationary)인 구간이라고 가정하고 행하는 단 구간 분석에 의하여 프레임 단위로 음성 특징벡터를 추출하는 전처리를 거친 후에 이루어진다. 그리고 인식 과정에서는 음성 인식 후보 단어 각각을 이러한 특징 추출 과정을 거쳐 기준 벡터 열로 미리 준비하여 두고, 인식할 단어에 대한 특징을 추출하여 시험 벡터 열을 각 후보 단어와 DTW 알고리즘을 이용하여 비교하여 최소가 되는 후보 단어 카테고리를 인식 결과로 결정하는 비교적 간단한 음성 인식 알고리즘이다.[7]

2. 인식레벨분류 패턴매칭 방법

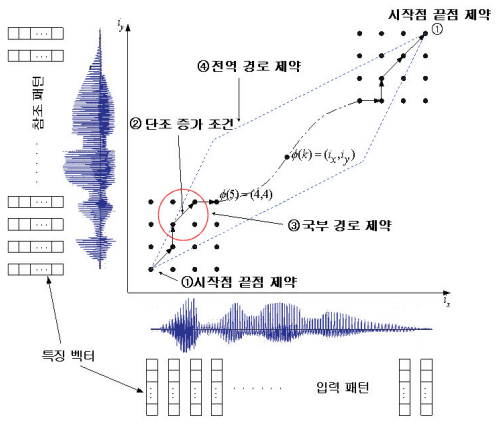

그림 5. DTW를 이용한 음성인식
Fig 5. Voice recognition using DTW

그림5와 같이 DTW는 음성인식을 수행할 수 있는데 그림에서 알 수 있듯이 참조패턴과 입력패턴의 유사도를 측정하여 인식을 수행한다.

유사도 비교 시 어느 정도 유사도로 볼 것 인지가 바로 레벨 매칭이 되는데 그 방법은 참조패턴이 8bit로 구성되게 설계 되어 있다. 즉, 1개의 인식 단어는 256개의 특징 벡터 열로 구성되고, 2번 학습 한다면 1개의 인식단어는 16bit 즉, 2개의 256개 특징 벡터열로 구성될 수 있다는 것이다. 따라서, 2번 학습된 단어는 8bit ,256개의 백터를 2개의 8bit, 256개의 백터와 비교하여 인식을 수행한다는 것이다.

그렇다면 입력벡터와 참조벡터의 비교 시 0에서 255까지 비교를 행할 수 있는데 여기서 유사도의 값을 줄 수 있다면 그 레벨값 역시 0에서 255의 값을 가질 수가 있다는 것이다.

따라서, 유사도를 S라고하면 S= 50 이면 레벨1, 50<S<100 이면 레벨2, 100<S<150 이면 레벨 3, 150<S<200 이면 레벨4, 200<S<255 레벨로 그 인식 매칭 값을 조정할 수 가 있다.

Ⅳ. 실험 및 결과

유사도의 레벨을 5스텝으로 분류하여 음성인식을 수행하였다. 이론적으로는 레벨 1, S=50 이하의 경우는 가장 쉽게 인식할 수 있는 레벨로 오인식이 가능한 레벨이다. 그리고 레벨3의 경우 통상 화자종속 인식에서 많이 쓰이는 레벨이다. 레벨5의 경우 완벽한 참조벡터일치로 동일화자 일지라도 학습 시 발음이나 환경 소음이 틀리다면 인식이 어려운 화자인증 레벨로 볼 수가 있다.

실험환경은 차량 시동 후 30Km 이하의 속도에서 진행 하였다. 본 논문서는 인식률 향상을 위한 시도가 아닌 자동차 환경에서의 패턴인식 레벨의 최적화를 위한 연구이므로 많은 화자의 테스트 보다는 동일화자의 레벨테스트 중심으로 실험 하였다. 그 결과는 표 1과 같다.

결과를 보면 통상 3레벨에서 인식률을 진행 했으나 차량환경에서는 오히려 2레벨에서 인식률이 다소 높음을 알 수 있다. 그리고 1레벨에서도 인식률이 나쁘지는 않으나 오 인식률이 예상대로 높게 나왔다. 그리고, 레벨 5에서는 인식률이 현저하게 줄었다. 이론적으로도 레벨 5에서는 인식률이 나쁘게 나오는게 당연하지만 차량 소음

환경에서 역시 인식하기가 힘든 레벨임이 분명하다고 추측한다. 실험은 각 인식 단에 대해서 2번 학습 시키고 각 10번 발성하여 그 인식률을 나타내었다.

레벨2 에서는 잡음환경이지만 패턴매칭 스코어가 50에서 100이므로 기존 패턴도 어느 정보 loose한 상황이라고 볼 때 노이즈도 같이 무시되어 오히려 인식률이 높게 나왔다고 사료된다.

표 1. 인식률(%)
Table 1. Recognition rete(%)

인식단어	연계ECU	레벨 1	레벨 2	레벨 3	레벨 4	레벨 5
창문열어	Door Module	80	90	90	50	30
닫어	Door Module	100	100	100	80	50
창문올려	Door Module	70	80	80	50	50
정지	Door Module	80	90	80	50	50
비상등	BCM	90	100	80	60	60
와이퍼	BCM	80	90	80	70	60
실내등	BCM	70	90	80	70	40
미등	BCM	70	100	80	70	50
좌측 깜박이	BCM	60	70	60	50	30
우측 깜박이	BCM	50	80	60	40	30
오디오	Audio	80	80	70	50	40
라디오	Audio	80	70	70	50	60
CD	Audio	80	90	90	40	50
DVD	Audio	80	90	80	50	50
볼륨크게	Audio	70	70	70	50	50
볼륨작게	Audio	60	70	70	40	30
전조등	BCM	80	90	80	60	50
상향들	BCM	80	80	80	60	40
에어컨	HVAC	80	90	70	50	50
히터	HVAC	70	80	90	50	50
네비게이션	Navigation	80	100	80	60	40
핸즈프리	Hands free	80	80	70	60	40

V. 결 론

본 연구에서는 음성인식 DTW알고리듬을 수행하는 패턴매칭 레벨에 따른 자동차 환경에서의 인식실험을 수행 하였다. 레벨에 따른 인식률을 고찰해 보면 레벨 2에서 비교적 좋은 인식률이 나왔다. 그리고 , 레벨 1에서도 그리 나쁘지 않은 인식률을 나타내고 있다. 그 이유는 패턴의 인식 벡터는 줄었지만 그에 따른 잡음벡터 역시 그 스케일에 맞게 줄어들게 된다. 결국 8bit, 256개의 벡터에서 100개 이하의 벡터만 참조벡터와 비교하게 되므로 노이즈 성분이 있는 비교벡터도 100개 이하의 벡터만 비교하게 되어 나머지 잡음성분도 같이 줄어드는 결과로 추정된다.

상기 실험 결과에서 보면 차량환경에서는 레벨 2에서 인식결과가 좋음을 확인하였다. 그것은 인식벡터가 차량환경에서는 레벨2 즉 100내외의 스코어에서 적합한 인식레벨임을 확인 하였다.

그러나 , 잡음환경은 보다 더 복잡하고 고려해야 할 사항이 많다. 마이크와 화자간의 거리 역시 주요한 요인으로 작용될 수 도 있으며 매칭 오차범위 또한 중요한 인자로 작용 될 수 있을 것으로 판단된다. 향후 상기와 같은 파라미터의 변위에 대한 연구가 필요 하다고 사료된다.

참 고 문 헌

[1] Leo L. Beranek, "Acoustics", Acoustical Society of America, 1993.

[2] Peter V. Loeppeart, " Advanced Microphone Technology", AVIOS, 1999.

[3] Rabiner L. R., Juang B. H., "Fundamentals of Speech Recognition", Englewood Cliffs, Prentice-Hall, 1993.

[4] 안종영,김영섭,허강인 “차량용 BCM을 위한 음성인식 시스템 설계” 한국인터넷방송 학회 추계학술대회 논문집, pp169-171, 2009,12

[5] Sensory Speech 7 Technology, " RSC-4x Evaluation Manual, 2003.

[6] 안종영, 김주성, 김수훈, 허강인 “RBFN을 이용한 음소인식에 관한 연구” 제12회 음성통신 및 신호처리 워크샵 논문집, pp.239-242, 1995.6

[7] 한학용 저 “패턴인식 개론”, 한빛미디어

※ 본 논문은 동아대학교 학술연구비 지원에 의하여 연구되었음.

저자 소개

안 종 영(정회원)

- 1993년 : 동아대학교 전자공학과 공학사
- 1996년 : 동아대학교 전자공학과 공학석사
- 1996-2000 ;현대오토넷 전임연구원
- 2001-2003:한국폴리텍 아산캠퍼스 영상매체과 교수
- 2004-2006 : (주)대성전기 선임연구원
- 현 : 동아대학교 대학원 전자공학과 박사과정
  한국폴리텍II 인천대학 컴퓨터정보과 초빙교수

<주관심분야 : 음성신호처리, 임베디드 시스템, DSP, 전장 ECU>

김 영 섭

- 2005년 : 동명정보대학교 컴퓨터공학과 공학사
- 2007년 : 동아대학교 전자공학과 공학석사

2009년~현: 동아대학교 전자공학과 박사과정

<관심분야 : 패턴인식, 음성/영상처리, DSP application>

김 수 훈

- 1990년: 동아대학교 전자공학과 공학사
- 1992년 : 동아대학교 전자공학과 공학석사
- 1999년 : 동아대학교 전자공학과 공학박사

2001년~현: 부천대학 모바일통신과 부교수

<주관심분야 : DSP, 음성인식, 모바일콘텐츠>

허 강 인

- 1980년 : 동아대학교 전자공학과 공학사
- 1982년 동아대학교 전자공학과 공학석사
- 1990년 경희대학교 전자공학과 공학박사
- 1998년 9월~1989년 8월 일본 쓰쿠바대학 객원연구원
- 1992년 9월~1993년 8월 일본 도요하시대학 객원연구원
- 1984년-현: 동아대학교 전자공학과 교수

<주관심분야 : DSP, 음성인식, 음성합성, 신경회로망>