(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) 。Int. Cl. [7]
G10L 15/20

(11) 공개번호 특2002-0005205
(43) 공개일자 2002년01월17일

(21) 출원번호 10-2000-0035286
(22) 출원일자 2000년06월26일

(71) 출원인 한국과학기술원
윤덕용
대전 유성구 구성동 373-1

(72) 발명자 이수영
대전광역시유성구구성동한국과학기술원전자전산학과
김창민
대전광역시유성구구성동한국과학기술원전자전산학과

(74) 대리인 이종일

심사청구 : 있음

(54) 인간 청각 모델을 이용한 효율적인 음성인식 장치

요약

본 발명은 음성인식장치에 관한 것이다.

종래의 음성 인식 장치는 ASIC 구현이 어렵기 때문에 소프트웨어만 처리하든가 아니면 DSP를 이용하여 복잡한 시스템을 구성해야 하므로 실생활에 응용하기가 어려웠다.

본 발명에서는 아날로그 음성 신호를 디지털 신호로 변환하는 A/D 변환기와 상기 A/D 변환기에서 변환된 디지털 신호를 powers-of-two 변환을 이용하여 다수의 채널로 필터링하는 FIR 필터 처리부와, FIR 필터 처리부의 출력 신호로부터 잡음에 강한 음성특징을 메모리를 추가로 사용하지 않고 빠른 시간 안에 추출하는 특징추출부, 상기 특징추출부에서 추출된 음성특징에 기초하여 음성신호의 시작점과 끝점 정보를 판별하는 워드 경계검출부, 특징추출부로 부터 제공되는 음성특징과 워드 경계검출부로부터의 음성신호의 시작과 끝점 정보에 기초하여 시간 정규화 및 RBF 신경 회로망을 이용한 분류 과정을 거쳐 최종의 결과를 코딩하여 출력하는 인식부를 구비함으로서 외부 잡음에 둔감하며, 실생활에 응용이 가능한 음성 인식 장치가 제시된다.

대표도
도 2

명세서

도면의 간단한 설명

도 1은 종래의 음성 인식 장치를 개략적으로 도시한 블록 구성도이다.

도 2는 본 발명의 실시예에 따른 음성 인식 장치를 도시한 블록 구성도이다.

도 3은 도 2의 동작을 설명하기 위한 타이밍도이다.

도 4는 특징 추출 방법의 설명을 위한 블록 구성도 이다.

< 도면의 주요 부분에 대한 부호의 설명 >

2 : 특징추출부 4 : 인식기

10 : FIR 필터 처리부 20 : 특징 추출부

30 : 워드 경계 검출부 40 : 정규화 및 인식부

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 음성 인식 장치에 관한 것으로, 특히 외부 잡음에 둔감하고 효율적인 계산을 함으로서, 실생활에 응용이 가능한 음성 인식 장치에 관한 것이다.

최근 음성인식 분야가 발달하면서 그 용도가 다양화되고 있다.

도 1은 종래의 음성 인식 장치를 개략적으로 도시한 블럭 구성도이다.

도 1에 도시된 바와 같이, 음성 인식 장치는 크게 특징 추출부(2)와 인식기(4)로 구성된다. 즉, 입력 음성 신호는 기존의 특징 추출 방법으로 LPC(Linear Prediction Coding Analysis)등과 같은 방법을 이용하였고, 인식기는 HMM(Hidden Markov Model)과 같은 방법을 많이 이용하였다.

또한, 실제 전자제품 등에 응용이 가능한 음성인식장치로써, 사람의 청각 모델과 신경회로망을 이용한 것이 개발되었다. 이와 같은 종래의 음성인식장치는 특허 제10-0180651호(음성인식장치)로 1998년 12월 2일에 특허등록되어 공지되어 있다.

상기 공지된 특허 발명을 개략적으로 살펴보면, 아날로그 음성신호를 디지털신호로 변환하기 위한 A/D 변환기와, 상기 A/D 변환기에서 변환된 12비트의 디지털신호를 소정개의 채널로 필터링하는 필터처리부와, 상기 필터처리부의 출력신호로부터 잡음에 강한 음성특징을 추출하고 이 추출된 결과를 출력하는 특징 추출부와, 상기 디지털신호로 변환된 음성신호에 기초하여 이 음성신호의 시작점과 끝점 정보를 판별하는 워드 경계검출부와, 상기 특징 추출부로부터 제공되는 음성특징과 상기 워드경계 검출부로부터의 음성신호의 시작과 끝점 정보에 기초하여 시간 정규화 및 신경회로망을 이용한 분류과정을 거쳐 최종의 결과를 코딩하여 출력하는 분석처리부로 구성되어 있다.

발명이 이루고자 하는 기술적 과제

그러나 상기 종래의 음성인식장치는 특징추출방법으로 LPC 등과 같은 방법을 사용하고, 인식기로써 HMM 방법을 사용함으로써 ASIC구현이 어렵기 때문에 소프트웨어만 처리하든가 아니면 DSP를 이용하여 복잡한 시스템을 구성해야만 함으로 실생활에 응용하기 어려웠다.

또한, 상기 종래의 공지된 특허 발명인 음성인식장치는 상기 필터처리부의 A/D변환기에서 처리된 디지털신호가 다수의 채널로 필터링되어 파워소모가 크고, 또한 단어의 경계검출을 먼저하고 음성특징을 나중에 추출함으로써 그 효율성이 떨어지는 문제점이 있다.

따라서, 본 발명은 상기한 문제점을 해결하기 위한 것으로써, 본 발명의 목적은 사람의 청각 모델과 신경회로망을 이용하여 외부 잡음에 둔감하고, 실제 생활에서도 응용이 가능하고, FIR 필터를 이용하고, powers - of - two 변환을 이용하여 덧셈과 쉬프트 연산만으로 필터뱅크를 구축한 필터링시 파워 소모가 적고, 효율성이 좋은 음성 인식 장치를 제공하는데 있다.

상기 본 발명의 목적을 달성하기 위하여, 본 발명에 따른 음성 인식 장치는 아날로그 음성 신호를 디지털 신호로 변환하는 A/D 변환기와, A/D 변환기에서 변환된 12비트의 디지털 신호를 16개의 채널로 필터링하는 powers-of-two 변환을 이용한 FIR 필터 처리부와, 상기 FIR 필터 처리부의 출력신호로부터 잡음에 강한 음성 특징을 추출하고 이 추출된 결과를 출력하는 특징 추출부와; 상기의 특징 추출부에서 추출된 잡음에 강한 음성 특징에 기초하여 음성 신호의 시작점과 끝점 정보를 판별하는 워드경계 검출부와; 상기 특징 추출부로부터 제공되는 음성 특징과 상기 워드경계 검출부로부터의 음성 신호 시작과 끝점 정보에 기초하여 시간 정규화 및 신경 회로망을 이용한 분류과정을 거쳐 최종의 결과를 코딩하여 출력하는 정규화 및 인식부를 구비하는 음성인식장치가 제시된다.

발명의 구성 및 작용

이하, 본 발명의 실시예의 구성 및 그 작용에 대하여 첨부한 도면을 참조하면서 상세히 설명하기로 한다.

도 2는 본 발명에 따른 음성 인식 장치의 실시예를 도시한 블럭 구성도이다.

도 2를 참조하면, 본 발명에 따른 음성 인식 장치는 A/D 변환기로부터 신호를 입력받는 FIR 필터처리부(10)와, 상기 FIR 필터처리부(10)와 접속되어 있는 특징추출부(20)와, 상기 FIR 필터처리부(10) 및 특징추출부(20)에 클럭을 출력시키기 위한 특징추출부(20)와, 상기 특징추출부(20)와 접속되어 있는 워드경계검출부(30)와, 상기 워드경계검출부(30)와 접속되어 있는, 정규화 및 인식부(40)와, 상기 워드경계검출부(30) 및 정규화 및 인식부(40)와 접속되는 SRAM을 포함하여 구성된다.

상기 A/D 변환기는 아날로그 신호인 음성 신호를 입력받아 이를 12비트의 디지털 음성 신호로 변환하여 후술하는 필터처리부(10)로 출력하도록 구성된다.

상기 필터 처리부(10)는 A/D 변환기에서 변환된 12비트의 디지털 신호를 16개의 채널로 필터링하여 후술하는 특징추출부(20)에 출력하도록 이루어지며, 상기 필터 처리부(10)는 16채널을 갖는 필터 뱅크로 구성된다.

상기 각 채널의 주파수 특징은 포유류의 귀로부터 얻어지는 데이터에 기초하여 만들어진다.

상기 필터는 100단의 FIR 필터로 구현되며, powers-of-two 변환을 이용하여 덧셈과 쉬프트 연산만으로 필터 뱅크를 구현한다. 여기에서 powers-of-two 변환은 수를 표현함에 있어서 다음 식과 같은 형태로 나타내는 것이다.

수학식 1

$$\sum_{n} c_n 2^n,\ c_n = -1,0,+1$$

상기 수학식 1과 같은 특징을 이용하면 곱셈기 없이 덧셈기와 쉬프트기만을 이용하여 FIR 필터를 구현할 수 있다. 상기 수학식 1과 같은 과정에서c_n중 '0'이 가장 많은 변환을 찾으면 면적과 동작속도를 동시에 줄일 수 있는 FIR 필터를 설계할 수 있다.

또한 본 발명의 실시예에서 필요한 상기 FIR 필터는 제한된 계수를 갖는 와우각 FIR 필터이다. 일반적인 powers-of-two 변환은 계수의 수가 증가하지만 본 발명에서 필용한 와우각 필터의 특징을 이용하여 상기 powers-of-two 변환을 사용하지 않는 필터와 유사한 수준의 계수를 차지하는 명령어를 설계 구현하였다.

상기 특징 추출부(20)는 필터 처리부(10)의 출력신호로부터 잡음에 강한 음성 특징을 추출하고, 이 추출된 결과를 후술하는 워드 경계 검출부(30)와 정규화 및 인식부(40)로 출력되도록 구성된다.

상기 특징 추출부(20)는 인간의 청각 모델링을 기초로 해서 음성 특징을 추출하였으며, 특징 벡터 자체를 버퍼링하여 실시간으로 특징 벡터를 추출하도록 구현된다.

상기 워드 경계 검출부(30)는 상기 특징 추출부로부터의 잡음에 강한 음성 특징에 기초하여 이 음성 신호의 시작점과 끝점 정보를 판별하여 후술하는 정규화 및 인식부(40)로 출력되도록 구성되며, 상기 워드 경계 검출부(30)는 음성신호의 각 채널의 특징벡터로부터 신호의 시작점과 끝점 정보를 판별한다.

상기 정규화 및 인식부(40)는 특징 추출부(20)에서 추출된 50개 단어중에서 선별하며, 상기 워드 경계 검출부(30)로부터 음성 신호시작과 끝점 정보에 기초하여 시간 정규화를 수행하게 되는데, 이 때 사용되는 정규화 방법은 비선형 방법인 트레이스 세그먼트(Trace Segment) 방법을 이용한다.

정규화 블럭은 끝점 추출 블럭으로부터 시작점과 끝점의 주소를 받아 16채널 64개의 프레임으로 정규화한 후, 정해진 에너지를 갖도록 정규화 한다. 또한, 정규화된 데이타를 RBF(Radial Basis Function)라는 신경 회로망 입력으로 가해서 50개의 기준 단어에 대한 각각의 출력값을 구한 다음에 각각의 출력값중 가장 큰 값을 갖는 단어를 6비트로 코딩하여 출력하게 된다.

분류하는 음성에 따라 변하게 되는 신경회로망의 가중값 데이타는 외부 메모리를 이용하여 저장하고, 메모리 데이타를 바꿈으로서 다른 음성에 대해서도 쉽게 적용할 수 있다.

상기와 같이 구성된 본 발명의 실시예의 작용을 상세히 설명하기로 한다.

상기 필터 처리부(10)는 A/D 변환기에서 변환된 12비트의 디지털 신호를 16개 채널로 필터링하여 특징 추출(20)에 출력한다. 여기에서 FIR_out과 nOUT은 12비트 신호로 샘플링 주파수인 Clk1(11.056KHz)와 칩의 연산에 필요한 CLKin(9MHz)에 동기되며, 타이밍은 제 3도(a)와 같다.

상기 특징추출부(20)는 필터 처리부(10)의 출력 신호로부터 잡음에 강한 음성 특징을 추출한 신호를 워드 경계검출부(30)와 정규화 및 인식부(40)로 제공한다. 즉, FEX_out은 특징 추출부(20)의 출력인 주파수 빈값이 상기 워드 경계 검출부(30)로 전송되는 신호이다. 이 신호는 8비트이며, 매 10ms(110샘플) 마다 clk1 에 동기되며, 타이밍은 제 3도(b)와 같다.

상기 특징추출부(20)의 신호는 FIR 필터 처리부(10)로부터의 nOUT 신호와 clk1에 동기되는 신호이다.

nOUT은 FIR 필터 처리부로부터 FIR_out이 출력될때마다 라이징 에지에서 활성화 되는 제어 신호이다.

반면 nBusy 신호는 특징추출부(20) 내부의 제어 신호로서, 폴링 에지에서 활성화 되어진다.

S0는 nBusy와 nOUT신호가 활성화 되기 전의 초기 상태를 의미한다. S1은 nOUT과 nBusy가 활성화 되었을 때 영교차점 사이의 에너지의 합을 계산하고, FIR 필터 처리부(10)로부터의 출력을 저장한다.

S2 상태에서는 영교차점을 찾아서 교차점 사이의 교차율을 계산한다.

S3는 누적될 특징 벡터 채널을 선택하고, 선택된 채널이 유효한지를 확인한다.

S4에서는 S3에서 선택된 채널에 특징을 누적한다. Case-1에서는 영교차점이 찾지 못할 경우 S2상태에서 다시 초기화 상태인 S0 상태가 됨을 보여주고 있으면, case-2는 선택된 채널에 특징 벡터를 누적하는 과정을 보여주고 있다.

도 4는 특징 추출 방법의 설명을 위한 블록 구성도 이다.

도 4에 도시한 바와 같이, 음성특징 벡터 자체를 버퍼링하여 실시간으로 특징벡터를 추출하였다. 음성특징 벡터는 각 채널의 영 교차점으로부터 시간 정보와 누적 에너지 값을 산출하여 음성 특징 벡터를 산출하게 된다. 이때 각 채널마다 영 교차점의 빈도수가 다르게 되는데 이를 일정하게 유지하기 위해 각 채널마다 다른 길이의 윈도우를 이용하여 특징벡터를 추출하게 된다.

도 4에서 종래의 방법은 각 채널에서의 필요한 FIR 필터 출력을 메모리에 저장한 후 이를 이용하여 다시 영 교차점을 검출하고 특징 벡터를 추출하게 된다. FIR 필터부의 출력을 메모리에 저장하고 이를 이용하여 특징 벡터를 추출하게 되면, 많은 메모리가 필요할 뿐아니라 많은 연산을 필요로 하게 된다.

특징 벡터를 추출하기 위해서 필요한 정보는 최대값과 영 교차점의 시간 간격이다. 이는 신호가 영점을 교차하게 될 때 바로 계산할 수 있다. 따라서 FIR 필터의 출력을 저장하지 않고 영 교차점에서의 정보를 이용하여 특징 벡터를 계속 누적해 나간다면 필요로 하는 메모리를 크게 줄일 수 있고 또한 연산 량도 대폭 줄어들게 된다.

특징 벡터를 계산하기 위해서 특징 벡터를 계속 누적 시켜서 다음 레지스터에 버퍼링하게 되며, 110샘플 사이에서 특징 벡터를 누적하기 위한 레지스터와 각 채널의 유효 시간 동안에만 특징 벡터를 누적하는 레지스터, 그리고 총 시간 구간(110 샘플)의 특징 벡터를 저장하기 위한 버퍼링 레지스터가 필요하다.

도 4에서 RR은 유효레지스터를 나타내고 R0는 다음 누적 레지스터에 버퍼링되게 될 특징 벡터 값을 누적시키는 레지스터이다. 그러므로 시간 t에서 구하는 특징 벡터는 위의 레지스터에 저장된 것을 순차적으로 더해서 구할 수 있게 되어, 필터 뱅크의 출력을 저장할 수 있는 메모리를 줄일 수 있다.

특징 벡터는 110 샘플 사이마다 추출하게 되는데, 최종 특징 벡터는 유효 누적 레지스터(RR)과 버퍼링된 레지스터(R1, R2, R3)의 합으로 덧셈기를 이용하여 쉽게 계산해 낼 수 있고 RR < = 0이 되고 R3 < = R2 < = R1 < = R0의 순으로 버퍼링되게 된다.

상기 정규화 및 인식부(40)는 특징추출부(20)로부터 제공되는 음성 특징과 상기 워드 경계검출부(30)의 음성 신호 시작과 끝점 정보에 기초하여 분류 과정을 거쳐 최종의 결과를 코딩한다. 여기에서 시작 태그, 종료 태그는 워드의 시작과 끝을 찾았음을 의미하는 신호로서 이 두 신호는 항상 정규화 및 인식부(40)가 CLKin의 라이징 에지에서 체크되도록 전후 1 동기 클럭의 여유를 두면된다.

정규화 방법인 트레이스 세그먼트 방법을 이용하여 나눗셈기를 곱셈기를 이용하여 구현함으로서 메모리 연산과 클럭을 줄여 주었다.

워드 경계검출부(30)와 규준화 블록은 특징 벡터를 참조해야 하므로 메모리 연산을 하게 된다. 따라서 주로 메모리 연산을 하게 되는 RBF 네트웍과 함께 구성되어 있다.

반면, 상기 정규화 및 인식부(40)는 현재의 특징 메모리(Feature Memory)(특징 추출로부터의 FEX_out을 메모리한다.)의 위치를 CLKin 과 동기되어 있는 SRAM(Feature Memory Address) 신호를 통하여 알 수 있다. 즉, 정규화 및 인식부(40)의 내부 메모리가 완전히 차더라도, 워드 경계검출부(30)에서 주어지는 신호는 계속해서 올라가므로, 정규화 및 인식부(40)는 메모리가 풀(Full) 인가를 스스로 점검해서 같은 곳에 데이터가 덮어쓰여지지 않도록 되어 있다. 그리고 워드경계 검출부(30)에서 시작태그를 넘어서는 종료 태그를 넘겨주어서는 안되며, 이때는 항상 시작 태그-1을 넘겨준다.

제 3도(c)에서 보는 바와 같이 끝점 추출후 18.7ms 후에 인식 결과가 출력되므로, 실시간 구현에 적합하다.

상기 A/D변환기로부터 인가되는 12 bit의 디지털 음성 데이터는 Clk1의 라이징 에지에서 필터 처리부(10)와 특징 추출부(20)에 의해서 읽혀진다. 그러므로 음성 신호를 디지털 신호로 변환하는 외부의 12비트 디지털 신호는 적어도 Clk1의 라이징 에지에서 한 시스템 클럭 이전에 변환을 마쳐야 한다.

본 발명의 실시예에서는 비동기 SRAM을 사용하였으며, 특징추출부(20)에서의 특징을 저장하고 있다가 정규화 및 인식부(40)에서만 읽어들인다. 읽기 동작은 계속해서 수행할 수 있지만, 쓰기 동작의 경우는 주소값을 설정한 후 쓰기 신호를 생성해 주어야 하므로 읽기와 함께 수행된다.

이상에서 본 발명의 실시예를 설명했지만, 본 발명의 보호범위는 상기 실시예에 한정되는 것이 아니고, 본 발명의 기술적 사상에 포함되는 당업자가 예측할 수 있는 다양한 실시예를 포함한다 하겠다.

발명의 효과

이상에서 설명한 바와 같이, 본 발명에서는 보다 적은 메모리 연산에 의한 빠른 특징추출방법을 사용하여 특징추출과정에서의 파워 소모를 줄이고,

또한 음성특성을 먼저 추출한 후, 이를 이용하여 단어 경계를 검출할 수 있어서 음성인식 장치는 외부 잡음에 둔감하고 계산이 효율적이며 하드웨어 구현이 용이한 음성인식장치를 제공함으로써, 실생활에서도 응용이 가능한 효과가 있다.

(57) 청구의 범위

청구항 1.

아날로그 음성신호를 디지털 신호로 변환하기 위한 A/D 변환기와,

상기 A/D 변환기에서 변환된 12비트의 디지털 신호를 powers-of-two 변환을 이용하여 소정개의 채널로 필터링하기 위한 FIR 필터처리부와,

상기 FIR 필터처리부의 출력 신호로부터 잡음에 강한 음성특징을 추출하기 위한 특징추출부와,

상기 특징추출부에 추출된 특징에 기초하여 음성 신호의 시작점과 끝점 정보를 판별하기 위한 워드 경계검출부와,

상기 특징추출로부터 제공되는 음성특징과 상기 워드 경계검출부로부터의 음성신호 시작점과 끝점 정보에 기초하여 시간 정규화 및 RBF (radial basis function) 신경 회로망을 이용한 분류 과정을 거쳐 최종의 결과를 코딩하여 출력하기 위한 인식부를 포함하는 것을 특징으로 하는 음성 인식 장치.

청구항 2.

청구항 1에 있어서,

상기 특징추출부는 FIR 필터부의 출력을 저장하지 않고 FIR 필터 출력의 영교차점에서 바로 특징 벡터를 계산하여 누적하는 것을 특징으로 하는 음성인식장치.

청구항 3.

청구항 2에 있어서,

상기 계산된 특징벡터응 누적하기 위해 각 채널마다 레지스터를 구비하고 있는 것을 특징으로 하는 음성인식장치.

청구항 4.

청구항 3에 있어서,

상기 레지스터는 총시간 구간(110 샘플) 사이에서 특징 벡터를 누적하기 위한 레지스터와, 각 채널의 유효 시간 동안에만 특징 벡터를 누적하는 레지스터와, 그리고 총 시간 구간(110 샘플)의 특징 벡터를 저장하기 위한 버퍼링 레지스터를 포함하는 것을 특징으로 하는 음성인식장치.

청구항 5.

청구항 1에 있어서,

상기 FIR 필터는 제한된 계수를 갖는 와우각 FIR 필터인 것을 특징으로 하는 음성인식장치.

청구항 6.

청구항 5에 있어서,

상기 FIR 필터는 상기 와우각 FIR 필터의 특성을 이용하여 powers-of-two 변환의 계수를 제한하는 명령어를 포함하는 것을 특징으로 하는 음성인식장치.

청구항 7.

청구항 1에 있어서,

상기 FIR 필터는 powers-of-two 변환을 이용하여 덧셈과 쉬프트 연산만으로 필터 뱅크를 구현하는 것을 특징으로 하는 음성인식장치.

청구항 8.

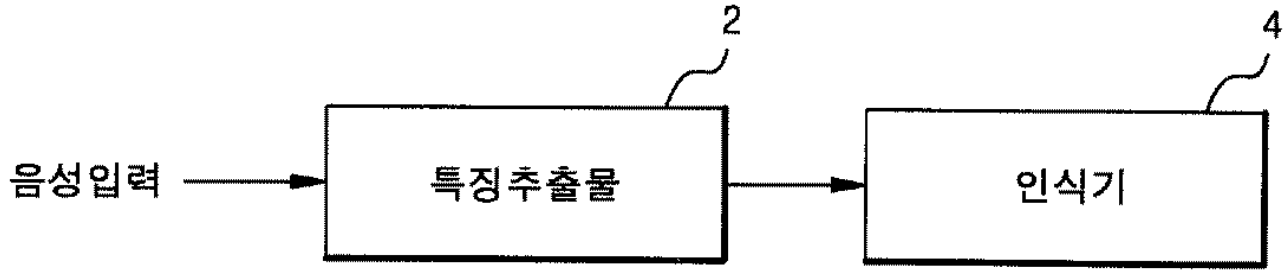
청구항 1에 있어서,

비동기식 SRAM을 더 포함하고,

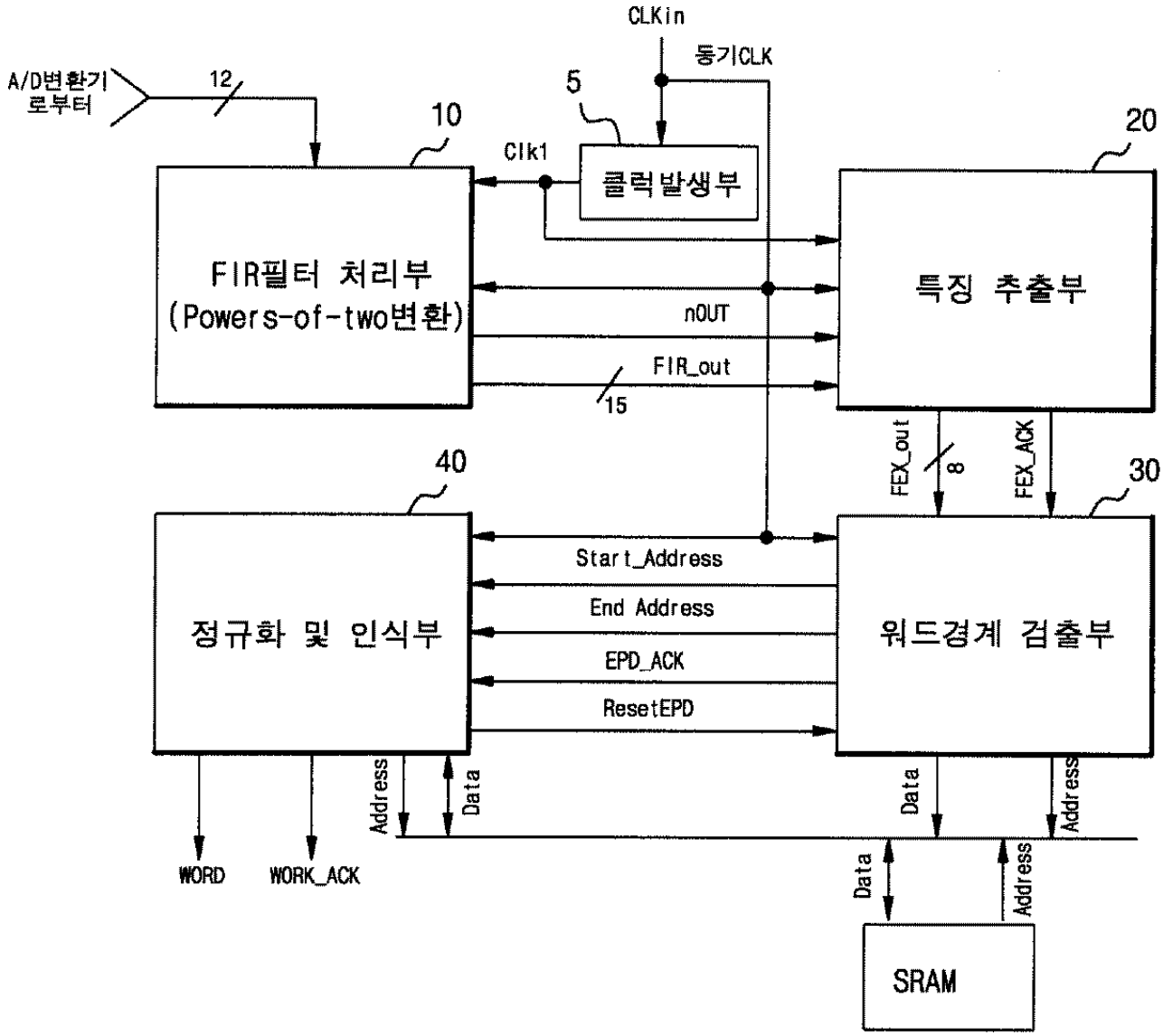
상기 SRAM은 상기 특징추출부에서의 특징을 저장하고 있다가 상기 정규화 및 인식부에서 읽어들이는 동작을 하는 것을 특징으로 하는 음성인식장치.

도면

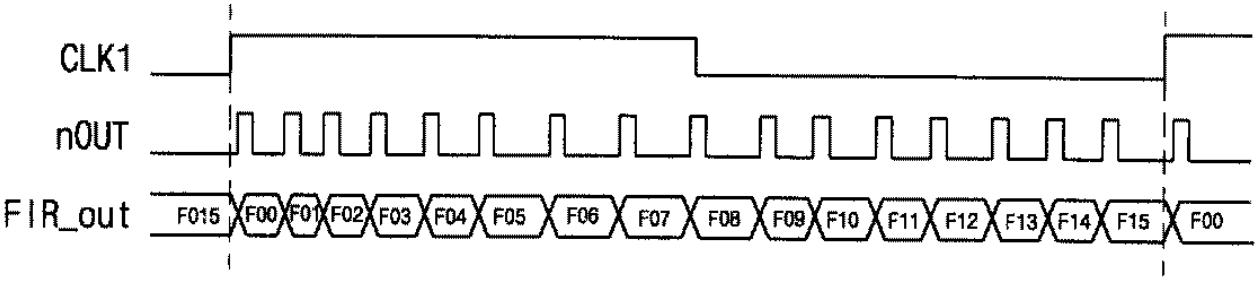
도면 1


2
4
음성입력
특징추출물
인식기

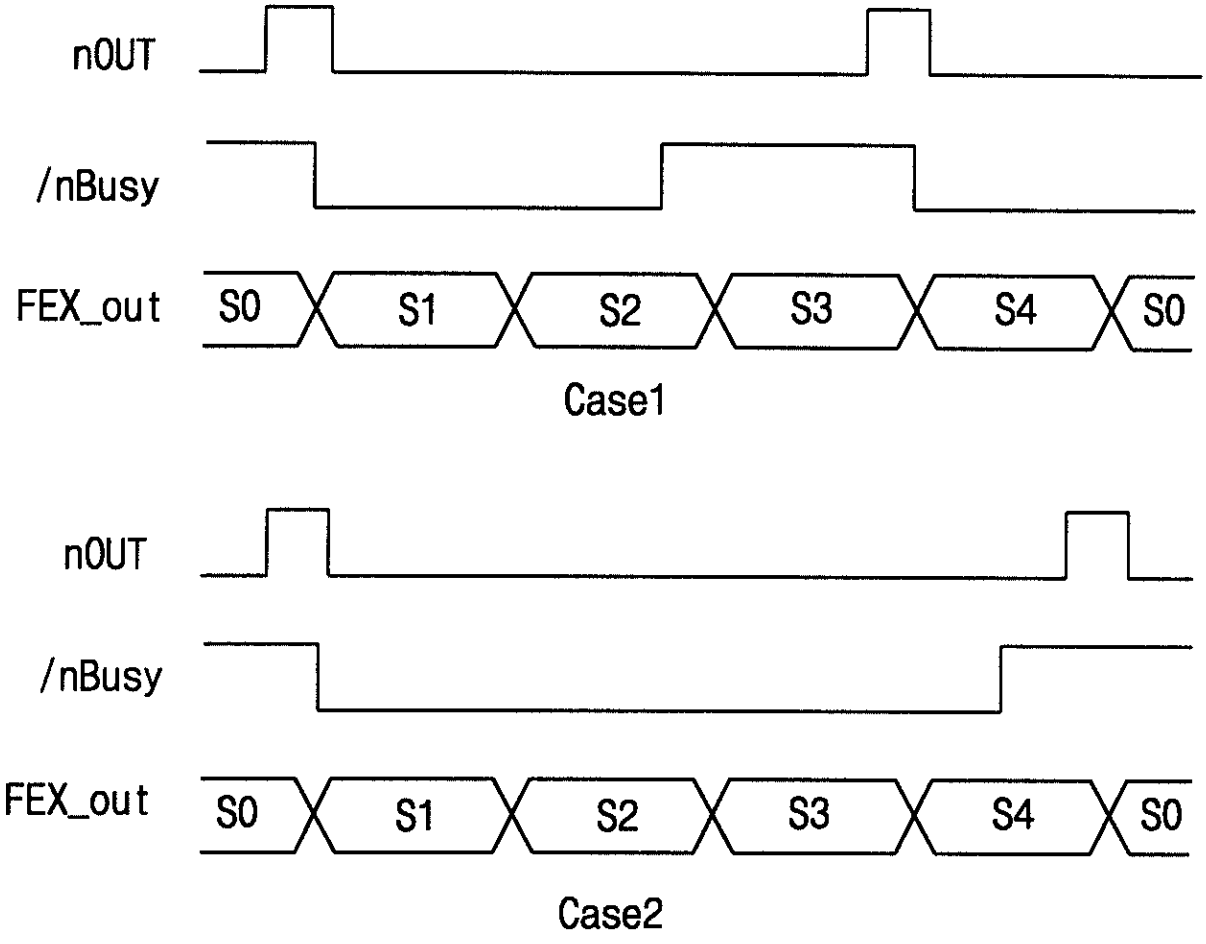
도면 2


CLKin
동기CLK
A/D변환기
로부터
12
5
10
20
Clk1
클럭발생부
FIR필터 처리부
(Powers-of-two변환)
특징 추출부
nOUT
FIR_out
15
FEX_out
8
FEX_ACK
40
30
Start_Address
End Address
EPD_ACK
ResetEPD
정규화 및 인식부
워드경계 검출부
Address
Data
Data
Address
WORD
WORK_ACK
Data
Address
SRAM

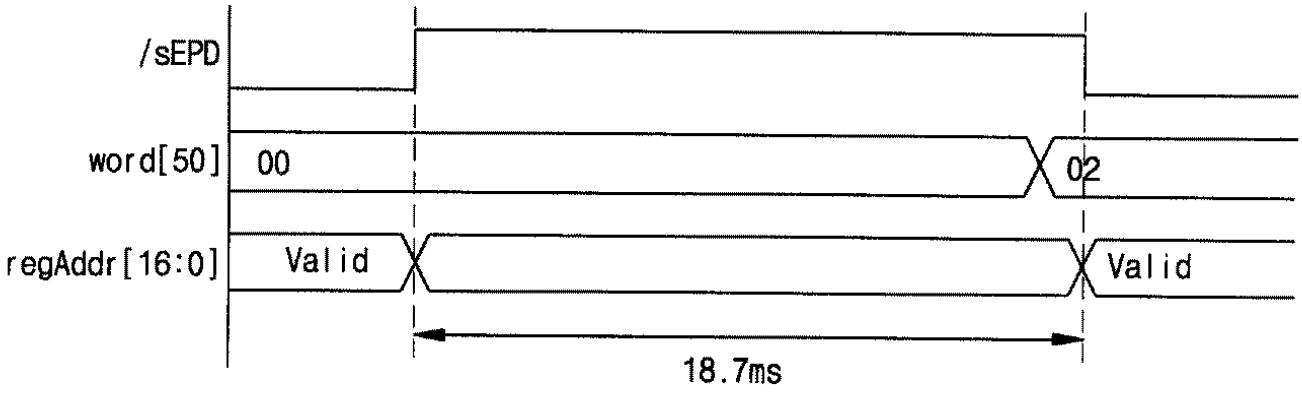
도면 3a


CLK1
nOUT
FIR_out
F015
F00
F01
F02
F03
F04
F05
F06
F07
F08
F09
F10
F11
F12
F13
F14
F15
F00

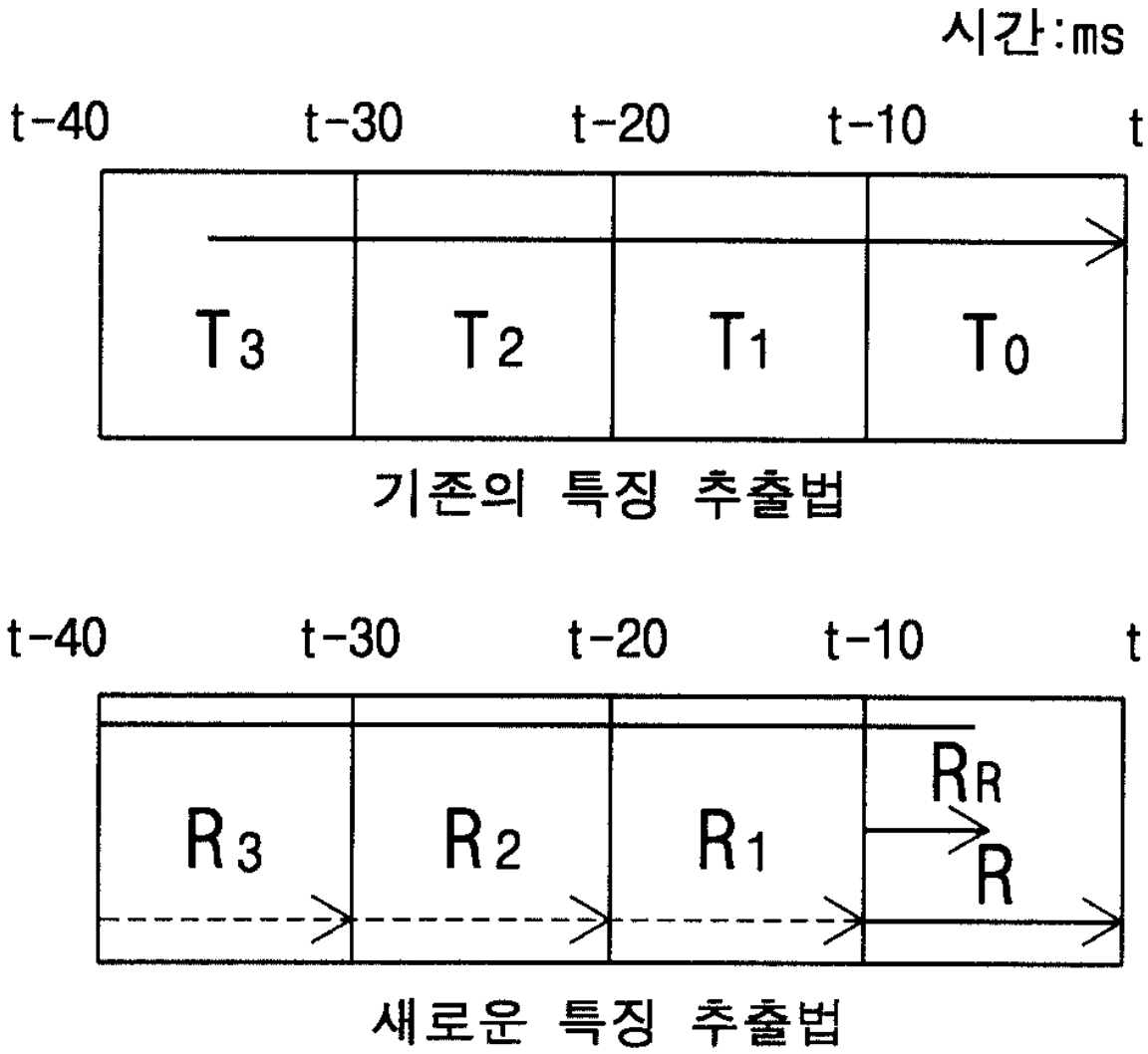
도면 3b


nOUT
/nBusy
FEX_out
S0
S1
S2
S3
S4
S0
Case1
nOUT
/nBusy
FEX_out
S0
S1
S2
S3
S4
S0
Case2

도면 3c


/sEPD
word[50]
00
02
regAddr[16:0]
Valid
Valid
18.7ms

도면 4


시간:ms
t-40
t-30
t-20
t-10
t
T3
T2
T1
T0
기존의 특징 추출법
t-40
t-30
t-20
t-10
t
R3
R2
R1
RR
R
새로운 특징 추출법