(19)대한민국특허청(KR)
(12) 공개특허공보(A)

(51) 。Int. Cl.
G10L 11/02 (2006.01)

(11) 공개번호 10-2007-0088469
(43) 공개일자 2007년08월29일

(21) 출원번호 10-2007-7002573
(22) 출원일자 2007년02월01일
심사청구일자 2007년02월01일
번역문 제출일자 2007년02월01일
(86) 국제출원번호 PCT/CA2006/000512
국제출원일자 2006년04월03일
(87) 국제공개번호 WO 2006/133537
국제공개일자 2006년12월21일

(30) 우선권주장 11/152,922 2005년06월15일 미국(US)

(71) 출원인 큐엔엑스 소프트웨어 시스템즈 (웨이브마커스) 인코포레이티드
캐나다 브이6비 2케이4 브리티쉬 콜롬비아 뱅쿠버 애보트 스트리트 #302-134

(72) 발명자 헤더링톤 필
캐나다 브리티시 콜럼비아 브이3에이치 5에이치7 포트 무디펀웨이 드라이브 23
에스코트 알렉스
캐나다 브리티시 콜럼비아 브이6비 1비7 뱅쿠버 리차드 스트리트504-1295

(74) 대리인 최정환

전체 청구항 수 : 총 39 항

(54) 음성 엔드-포인터

(57) 요약

룰 기반형 엔드-포인터는 오디오 스트림에 포함된 구두 발성을 배경 잡음 및 비음성 천이로부터 분리한다. 상기 룰 기반형 엔드-포인터는 여러 음성 특성에 기초하여 구두 발성의 시작 및/또는 끝을 결정하는 복수 개의 룰을 포함한다. 상기 룰은 이벤트, 이벤트들의 조합, 이벤트의 지속 기간, 또는 이벤트에 관한 지속 시간에 기초하여, 오디오 스트림 또는 오디오 스트림의 부분을 분석할 수 있다. 상기 룰은 오디오 스트림 그 자체의 특성, 오디오 스트림에 포함된 예상된 응답 또는 환경적 조건을 포함하는 여러 요소들에 따라 수동으로 또는 동적으로 주문 제작될 수 있다.

대표도

도 1

특허청구의 범위

청구항 1.

오디오 음성 세그먼트의 시작과 끝 중 적어도 하나를 결정하는 엔드-포인터로서,

음성 이벤트를 포함하는 오디오 스트림의 부분을 식별하는 보이스 트리거링 모듈; 및

상기 보이스 트리거링 모듈과 통신하고, 상기 오디오 스트림의 적어도 일부를 분석하여 상기 음성 이벤트에 관한 오디오 음성 세그먼트가 오디오 엔드포인트 내에 있는지 여부를 결정하는 복수의 시간 지속 기간 룰을 포함하는 룰 모듈

을 포함하는 엔드-포인터.

청구항 2.

제 1항에 있어서, 상기 보이스 트리거링 모듈은 모음을 식별하는 것인 엔드-포인터.

청구항 3.

제 1항에 있어서 상기 보이스 트리거링 모듈은 S 또는 X 사운드를 식별하는 것인 엔드-포인터.

청구항 4.

제 1항에 있어서, 상기 오디오 스트림의 부분은 프레임을 포함하는 것인 엔드-포인터.

청구항 5.

제 1항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 에너지의 부족을 분석하는 것인 엔드-포인터.

청구항 6.

제 1항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 에너지를 분석하는 것인 엔드-포인터.

청구항 7.

제 1항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 경과된 시간을 분석하는 것인 엔드-포인터.

청구항 8.

제 1항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 미리 정해진 수의 파열음을 분석하는 것인 엔드-포인터.

청구항 9.

제 1항에 있어서, 상기 룰 모듈은 상기 오디오 음성 세그먼트의 상기 시작 및 끝을 검출하는 것인 엔드-포인터.

청구항 10.

제 1항에 있어서, 에너지 검출기 모듈을 더 포함하는 엔드-포인터.

청구항 11.

제 1항에 있어서, 마이크로폰 입력부, 처리 유닛 및 메모리와 통신하는 처리 환경을 더 포함하고, 상기 룰 모듈은 상기 메모리 내부에 상주하는 것을 특징으로 하는 엔드-포인터.

청구항 12.

복수의 결정 룰을 갖는 엔드-포인터를 이용하여 오디오 음성 세그먼트의 시작과 끝 중 적어도 하나를 결정하는 방법으로서,

오디오 스트림의 부분을 수신하는 단계와;

상기 오디오 스트림의 상기 부분이 트리거링 특성을 포함하는 지를 결정하는 단계와;

적어도 하나의 시간 지속 기간 결정 룰을 상기 트리거링 특성에 관한 상기 오디오 스트림의 부분에 적용하여 상기 오디오 스트림의 상기 부분이 오디오 엔드포인트 내에 있는지를 결정하는 단계

를 포함하는 방법.

청구항 13.

제 12항에 있어서, 상기 결정 룰은 상기 트리거링 특성을 포함하는 상기 오디오 스트림의 상기 부분에 적용되는 것인 방법.

청구항 14.

제 12항에 있어서, 상기 결정 룰은 상기 트리거링 특성을 포함하는 상기 부분보다는 상기 오디오 스트림의 다른 부분에 적용되는 것인 방법.

청구항 15.

제 12항에 있어서, 상기 트리거링 특성은 모음인 것인 방법.

청구항 16.

제 12항에 있어서, 상기 트리거링 특성은 S 또는 X 사운드인 것인 방법.

청구항 17.

제 12항에 있어서, 상기 오디오 스트림의 부분은 프레임인 방법.

청구항 18.

제 12항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 에너지의 부족을 분석하는 것인 방법.

청구항 19.

제 12항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 에너지를 분석하는 것인 방법.

청구항 20.

제 12항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 경과된 시간을 분석하는 것인 방법.

청구항 21.

제 12항에 있어서, 상기 룰 모듈은 상기 오디오 스트림의 부분에서 미리 정해진 수의 파열음을 분석하는 것인 방법.

청구항 22.

제 12항에 있어서, 상기 룰 모듈은 상기 잠재적 음성 세그먼트의 시작 및 끝을 검출하는 것인 방법.

청구항 23.

오디오 스트림 중 오디오 음성 세그먼트의 시작과 끝 중 적어도 하나를 결정하는 엔드-포인터로서,

상기 오디오 스트림의 적어도 하나의 동적 양태를 분석하여 상기 오디오 음성 세그먼트가 오디오 엔드포인트 내에 있는지 여부를 결정하는 복수의 시간 지속 기간 룰을 포함하는 엔드-포인터 모듈; 및

상기 엔드-포인터 모듈과 통신하고, 상기 복수의 룰 중 하나 이상의 시간 지속 기간을 변경하는 프로화일 정보를 저장하도록 구성된 메모리

를 포함하는 엔드-포인터.

청구항 24.

제 23항에 있어서, 상기 오디오 스트림의 동적 양태는 화자의 적어도 하나의 특성을 포함하는 것인 엔드-포인터.

청구항 25.

제 24항에 있어서, 상기 화자의 특성은 상기 화자의 말하는 페이스를 포함하는 것인 엔드-포인터.

청구항 26.

제 23항에 있어서, 상기 오디오 스트림의 동적 양태는 상기 오디오 스트림 중의 배경 잡음을 포함하는 것인 엔드-포인터.

청구항 27.

제 23항에 있어서, 상기 오디오 스트림의 동적 양태는 상기 오디오 스트림 중의 예상된 사운드를 포함하는 것인 엔드-포인터.

청구항 28.

제 27항에 있어서, 상기 예상된 사운드는 화자에게 부여되는 질문에 대한 적어도 하나의 예상된 대답을 포함하는 것인 엔드-포인터.

청구항 29.

제 23항에 있어서, 마이크로폰 입력부, 처리 유닛 및 메모리와 통신하는 처리 환경을 더 포함하고, 상기 엔드-포인터 모듈은 상기 메모리 내부에 상주하는 것인 엔드-포인터.

청구항 30.

오디오 스트림 중의 오디오 음성 세그먼트의 시작과 끝 중 적어도 하나를 결정하는 엔드-포인터로서,

주기적인 오디오 신호를 포함하는 오디오 스트림의 부분을 식별하는 보이스 트리거링 모듈; 및

복수의 룰에 기초하여, 인식 장치에 입력되는 상기 오디오 스트림의 양을 변경하는 엔드-포인트 모듈

을 포함하고,

상기 복수의 룰은 상기 주기적인 오디오 신호에 관한 오디오 스트림의 부분이 오디오 엔드포인트 내에 있는지를 결정하는 시간 지속 기간 룰을 포함하는 것인 엔드-포인터.

청구항 31.

제 30항에 있어서, 상기 인식 장치는 자동 음성 인식 장치인 엔드-포인터.

청구항 32.

오디오 음성 세그먼트의 시작 및 끝 중 적어도 하나를 결정하는 명령어 세트를 담고 있는 컴퓨터 판독 가능한 저장 매체로서,

음파를 전기적 신호로 변환하고;

상기 전기적 신호의 주기성을 식별하며;

상기 식별된 주기성에 관한 상기 전기적 신호의 가변적인 부분을 분석하여 상기 전기적 신호가 오디오 엔드포인트 내에 있는지를 결정하는 것

를 포함하는 컴퓨터 판독 가능 저장 매체.

청구항 33.

제 32항에 있어서, 상기 전기적 신호의 가변 부분을 분석하는 것은 유성음 사운드 전의 시간 지속 기간을 분석하는 것을 포함하는 것인 컴퓨터 판독 가능한 저장 매체.

청구항 34.

제 32항에 있어서, 상기 전기적 신호의 가변 부분을 분석하는 것은 유성음 사운드 후의 시간 지속 기간을 분석하는 것을 포함하는 것인 컴퓨터 판독 가능한 저장 매체.

청구항 35.

제 32항에 있어서, 상기 전기적 신호의 가변 부분을 분석하는 것은 유성음 사운드 전 또는 후의 천이의 수를 분석하는 것을 포함하는 것인 컴퓨터 판독 가능한 저장 매체.

청구항 36.

제 32항에 있어서, 상기 전기적 신호의 가변 부분을 분석하는 것은 유성음 사운드 전의 연속된 침묵 지속 기간을 분석하는 것을 포함하는 컴퓨터 판독 가능한 저장 매체.

청구항 37.

제 32항에 있어서, 상기 전기적 신호의 가변 부분을 분석하는 것은 유성음 사운드 후의 연속된 침묵 지속 기간을 분석하는 것을 포함하는 컴퓨터 판독 가능한 저장 매체.

청구항 38.

제 32항에 있어서, 상기 컴퓨터 판독 가능한 저장 매체는 차량 온-보드 컴퓨터에 장착되는 것인 컴퓨터 판독 가능한 저장 매체.

청구항 39.

제 32항에 있어서, 상기 컴퓨터 판독 가능한 저장 매체는 오디오 시스템과 통신하는 것인 컴퓨터 판독 가능한 저장 매체.

명세서

기술분야

본 발명은 자동 음성 인식 기술에 관한 것으로서, 보다 구체적으로는 구두 발성(spoken utterance)을 배경 잡음 및 비음성 천이(non-speech transients)로부터 분리하는 시스템에 관한 것이다.

배경기술

차량 환경 내부에서, 탑승자에게 보이스 입력에 기초한 내비게이션 지시를 제공하기 위해 자동 음성 인식(ASR; Automatic Speech Recognition) 시스템이 이용될 수 있다. 이러한 기능은 화면에 정보를 수동으로 키이 입력하거나 화면으로부터 정보를 읽으려고 시도하는 동안에 운전자의 주의가 도로에서 벗어나지 않는다는 점에서 안전에 대한 우려를 증가시킨다. 또한, ASR 시스템은 오디오 시스템, 기후 제어 또는 다른 차량 기능을 제어하는 데에 이용될 수 있다.

ASR 시스템은 사용자가 마이크로폰에 음성을 입력할 수 있도록 해주고 신호를 컴퓨터가 인식하는 명령어로 전환시켜 준다. 상기 명령을 인식하면, 컴퓨터는 소정의 애플리케이션을 실행할 수 있다. ASR 시스템에서 실행할 때에 한 가지 요소는 구두 발성을 정확히 인식하는 것이다. 이는 그 발성의 시작 및/또는 끝의 위치를 결정하는 것("엔드-포인팅")을 필요로 한다.

일부 시스템은 오디오 프레임 내의 에너지를 검색한다. 그 에너지를 검출하면, 상기 시스템은, (구두 발성의 시작 시간을 결정하기 위하여) 상기 에너지가 검출되는 포인트에서 소정의 시간 주기를 빼고, (구두 발성의 종료 시간을 결정하기 위하여) 상기 에너지가 검출되는 포인트에서 소정의 시간을 추가함으로써, 구두 발성의 엔드-포인트를 예측한다. 다음에, 이러한 선택된 오디오 스트림 부분은 구두 발성을 결정하기 위한 시도시 ASR로 보내진다.

음향 신호 내의 에너지는 많은 소스로부터 오는 것일 수 있다. 예컨대, 차량 환경 내부에서, 음향 신호 에너지는 도로의 융기부에 부딪히는 소리(road bumps), 문을 쾅 닫는 소리(door slams), 탁 하는 소리(thumps), 깨지는 소리(cracks), 엔진 잡음, 공기 이동 등과 같은 과도 잡음(transient noise)으로부터 유래할 수 있다. 에너지의 존재에 집중하는 상기 시스템은 이러한 과도 잡음을 구두 발성인 것으로 잘못 해석할 수 있고, 상기 신호의 주변 부분을 ASR 시스템에 전송하여 처리할 수도 있다. 따라서, ASR 시스템은 과도 잡음을 음성 명령인 것으로 인식하기 위해 불필요한 시도를 할 수가 있어, 폴스-포지티브(false positives)를 발생시키고 실제 명령에 대한 응답을 지연시킨다.

따라서, 과도 잡음 조건에서 구두 발성을 식별할 수 있는 지능형 엔드-포인터 시스템에 대한 요구가 있다.

발명의 상세한 설명

룰 기반형 엔드-포인터(rule-based end-pointer)는 오디도 스트림 중의 오디오 음성 세그먼트의 시작, 끝 또는 시작 및 끝을 결정하는 하나 이상의 룰을 포함한다. 상기 룰은 이벤트의 발생 또는 이벤트의 조합, 또는 음성 특성의 존재/부존재 지속 기간과 같은 여러 가지 요소에 기초할 수 있다. 또한, 상기 룰은 침묵 기간, 유성음 오디오 이벤트(voiced auido event), 무성음 오디오 이벤트 또는 이러한 이벤트의 임의의 조합; 이벤트의 지속 기간; 또는 이벤트에 관한 지속 시간을 분석하는 것을 포함할 수 있다. 적용되는 룰 또는 분석되는 오디오 스트림의 콘텐츠에 따라, 상기 룰 기반형 엔드-포인터가 전송하는 오디오 스트림의 양은 변할 수 있다.

동적 엔드-포인터는 오디오 스트림과 관련된 하나 이상의 동적 양태(dynamic aspects)를 분석할 수 있고, 그 분석된 동적 양태에 기초하여, 오디오 음성 세그먼트의 시작, 끝 또는 시작과 끝을 결정할 수 있다. 분석될 수 있는 동적 양태는 (1) 음성을 말하는 화자의 페이스, 화자의 피치(pitch) 등과 같은 오디오 스트림 그 자체, (2) 발성자에게 부과되는 질문에 대한 예상된 응답(예를 들면, "YES" 또는 "NO")과 같은, 오디오 스트림 중의 예상된 응답, 또는 (3) 배경 잡음 레벨, 에코 등과 같은 환경적 조건 등을 포함하는데, 이들에 제한되는 것은 아니다. 오디오 음성 세그먼트를 엔드-포인팅하기 위하여, 상기 룰은 상기 하나 이상의 동적 양태를 이용할 수 있다.

본 발명의 다른 시스템, 방법, 특징 및 이점은 이하의 도면 및 상세한 설명의 검토를 통해 당업자에게 명백하거나 명백해질 것이다. 이러한 모든 추가의 시스템, 방법, 특징 및 이점은 본 설명 내에 포함되고, 본 발명의 범위 내이며, 후술하는 청구범위에 의해 보호되도록 하기 위한 것이다.

본 발명은 이하의 도면 및 설명을 참고하여 더 잘 이해될 수 있다. 도면의 요소는 반드시 비례하여 나타낸 것은 아니며, 대신 본 발명의 원리를 설명할 때 강조하여 표시하였다. 또한, 도면에서, 동일한 도면 부호는 상이한 도면 전체에 걸쳐 대응 부분을 나타낸다.

실시예

룰 기반형 엔드-포인터는 트리거링 특성(triggering characteristic)에 대해 오디오 스트림의 하나 이상의 특성을 검사할 수 있다. 트리거링 특성은 유성음 또는 무성음을 포함할 수 있다. 발성 코드(vocal cord)가 진동할 때 발생되는 유성음 세

그먼트(예컨대, 모음)는 거의 주기적인 시간-도메인 신호를 발산한다. (영어에서 "f"를 말할 때와 같이) 발성 코드가 진동하지 않을 때 발생되는 무성음 사운드는 주기성이 부족하고, 잡음형 구조와 비슷한 시간-도메인 신호를 갖고 있다. 오디오 스트림 중의 트리거링 특성을 식별하고 음성 사운드의 자연적인 특성에 대해 작용하는 룰 셋트를 채용함으로써, 상기 엔드-포인터는 음성 발성의 시작 및/또는 끝을 결정하는 것을 개선할 수 있다.

별법으로서, 엔드-포인터는 오디오 스트림의 적어도 하나의 동적 양태를 분석할 수 있다. 분석될 수 있는 오디오 스트림의 동적 양태는 (1) 음성을 말하는 화자의 페이스, 화자의 피치 등과 같은 오디오 스트림 그 자체, (2) 상기 화자에 부여되는 질문에 대한 예상된 응답(예컨대,"YES" 또는 "NO")과 같은 오디오 스트림 중의 예상된 응답, 또는 (3) 배경 잡음 수준, 에코 등과 같은 환경적 조건을 포함하지만, 이들에 제한되는 것은 아니다. 상기 동적 엔드-포인터는 룰 기반형일 수 있다. 엔드-포인터의 동적 특성은 음성 세그먼트의 시작 및/또는 끝을 결정하는 것을 개선해 준다.

도 1은 보이스에 기초하여 음성 엔드-포인트를 수행하기 위한 장치(100)의 블록도이다. 엔드-포인팅 장치(100)는 하나 이상의 운영 시스템과 연계하여 하나 이상의 프로세서 상에서 구동될 수 있는 소프트웨어 또는 하드웨어를 포함할 수 있다. 엔드-포인팅 장치(100)는 컴퓨터와 같은 처리 환경(102)을 포함할 수 있다. 처리 환경(102)은 처리 유닛(104) 및 메모리(106)를 포함할 수 있다. 처리 유닛(104)은 양방향 버스를 통해 시스템 메모리(106)에 액세스함으로써 연산 동작, 로직 동작 및/또는 제어 동작을 수행할 수 있다. 메모리(106)는 입력 오디오 스트림을 저장할 수 있다. 메모리(106)는 오디오 음성 세그먼트의 시작 및/또는 끝을 검출하는 데에 사용되는 룰 모듈(108)을 포함할 수 있다. 메모리(106)는 또한 오디오 세그먼트 중의 트리거링 특성을 검출하는 데에 사용되는 보이스 분석 모듈(116) 및/또는 오디오 입력을 인식하는 데에 사용될 수 있는 ASR 유닛(118)을 포함할 수 있다. 또한, 메모리 유닛(106)은 엔드-포인터의 동작 중에 얻어지는 버퍼링된 오디오 데이터를 저장할 수 있다. 처리 유닛(104)은 입출력(I/O) 유닛(110)과 통신한다. I/O 유닛(110)은, 음파(sound waves)를 전기적 신호(114)로 변환하는 장치로부터 입력 오디오 스트림을 수신하고, 전기적 신호를 오디오 사운드(112)로 변환하는 장치로 출력 신호를 전송한다. I/O 유닛(110)은 처리 유닛(104), 전기적 신호를 오디오 사운드(112)로 변환하는 장치, 음파를 전기적 신호(114)로 변환하는 장치 사이에서 인터페이스로서 작용할 수 있다. I/O 유닛(112)은 음파를 전기적 신호(114)로 변환하는 장치를 통해 수신한 입력 오디오 스트림을 음향 파형에서 컴퓨터가 이해 가능한 포맷으로 변환한다. 유사하게, I/O 유닛(110)은 처리 환경(102)으로부터 전송된 신호를, 전기적 신호를 오디오 사운드(112)로 변환하는 장치를 통해 출력하기 위한 전기적 신호로 변환할 수 있다. 처리 유닛(104)은 도 3 및 도 4의 흐름도를 실행하도록 적절히 프로그램될 수 있다.

도 2는 차량(200)에 합체된 엔드-포인터 장치(100)를 나타낸다. 차량(200)은 운전자 좌석(202), 탑승자 좌석(204) 및 뒷좌석(206)을 포함할 수 있다. 또한, 차량(200)은 엔드-포인터 장치(100)를 포함할 수 있다. 처리 환경(102)은 전자 제어 유닛, 전자 제어 모듈, 바디 제어 모듈과 같은 차량(200)의 온-보드 컴퓨터에 합체될 수 있으며, 또는 하나 이상의 허용 가능한 프로토콜을 이용하여 차량(200)의 기존 회로와 통신할 수 있는 별도의 후공장 유닛(after-factory unit)일 수 있다. 일부 프로토콜은 J1850VPW, J1850PWM, ISO, ISO9141-2, ISO14230, CAN, High Speed CAN, MOST, LIN, IDB-1394, IDB-C, D2B, Bluetooth, TTCAN, TTP 또는 FlexRay라는 상표명으로 판매되는 프로토콜을 포함할 수 있다. 전기적 신호를 오디오 사운드(112)로 변환하는 하나 이상의 장치는 전방의 탑승자 공간과 같이,차량(200)의 탑승자 공간에 배치될 수 있다. 이러한 구성에 제한되는 것은 아니지만, 음파를 전기적 신호(114)로 변환하는 장치는 입력 오디오 스트림을 수신하는 I/O 유닛(110)에 연결될 수 있다. 별법으로서, 또는 추가적으로, 전기적 신호를 오디오 사운드(212)로 변환하는 추가의 장치 및 음파를 전기적 신호(214)로 변환하는 장치는 뒷좌석의 탑승자로부터 오디오 스트림을 수신하여 그 탑승자에 정보를 출력하기 위하여 차량(200)의 뒷좌석 공간에 배치될 수 있다.

도 3은 음성 엔드-포인터 시스템의 흐름도이다. 상기 시스템은 입력 오디오 스트림을 프레임과 같은 여러 이산 구역(discrete sections)으로 분할하여, 그 입력 오디오 스트림이 프레임-바이-프레임(frame-by-frame)에 기초하여 분석될 수 있도록 동작할 수 있다. 각 프레임은 전체 입력 오디오 스트림의 약 10 ms 내지 약 100 ms 범위의 임의의 곳을 포함할 수 있다. 상기 시스템은 입력 오디오 데이터를 처리하기 시작하기 전에, 입력 오디오 데이터의 약 350 ms 내지 약 500 ms와 같이 미리 정해진 크기의 데이터를 버퍼링할 수 있다. 블록(302)으로 나타낸 바와 같이, 에너지 검출기는 잡음과는 별개로 에너지가 존재하는지 여부를 결정하는 데에 이용될 수 있다. 상기 에너지 검출기는 존재하는 에너지의 크기와 관련하여, 프레임과 같은 오디오 스트림의 일부를 검사하고, 그 크기를 잡음 에너지의 추정치와 비교한다. 잡음 에너지의 추정치는 일정하거나 동적으로 결정될 수 있다. 그 차이(dB) 또는 파워의 비는 순간적인 신호 대 잡음비(SNR)일 수 있다. 분석 전에, 프레임은 비음성인 것으로 추정될 수 있어, 상기 에너지 검출기가 프레임 내에 에너지가 존재하는 것으로 결정하면, 그 프레임은 블록(304)으로 나타낸 것과 같이, 비음성인 것으로 표시된다. 에너지가 검출된 후에, 프레임$_n$으로서 나타낸 현재 프레임의 보이스 분석은 블록(306)으로 표시한 것과 같이 일어날 수 있다. 보이스 분석은 2005년 5월 17일에 출원된 미국 출원 번호 제11/131,150호에 설명된 것과 같이 일어날 수 있으며, 그 명세서 내용은 본 명세서에 참고로 합체된다. 상기 보이스 분석은 프레임$_n$ 내에 존재할 수 있는 임의의 트리거링 특성을 체크할 수 있다. 상기 보이스 분석은 오디오 "S" 또는

"X"가 프레임$_n$ 내에 존재하는지 여부를 체크할 수 있다. 별법으로서, 상기 보이스 분석은 모음의 존재를 체크할 수 있다. 제한하려는 것이 아닌 설명의 목적을 위해, 도 3의 나머지는 보이스 분석의 트리거링 특성으로서 모음을 사용하는 것으로서 설명한다.

프레임 내의 모음의 존재를 식별할 수 있는 다양한 방식의 보이스 분석이 있다. 한 가지 방식은 피치 추정기(pitch estimator)를 사용하는 것이다. 피치 추정기는 모음이 존재할 수 있다는 것을 나타내는 프레임 내의 주기적 신호를 검색할 수 있다. 또는, 피치 추정기는 모음의 존재를 나타낼 수 있는 미리 정해진 수준의 특정 주파수에 대하여 프레임을 검색할 수 있다.

상기 보이스 분석에 의해 프레임$_n$에 모음이 존재하는 것으로 결정되면, 프레임$_n$은 블록(310)으로 나타낸 것과 같이, 음성으로서 표시된다. 다음에, 상기 시스템은 하나 이상의 이전의 프레임을 검사할 수 있다. 상기 시스템은 블록(312)으로서 나타낸 바와 같이, 바로 직전의 프레임(프레임$_{n-1}$)을 검사할 수 있다. 상기 시스템은 이전의 프레임이, 블록(314)으로 나타낸 바와 같이, 음성을 포함하고 있는 것으로 이전에 표시되었는지를 결정할 수 있다. 이전의 프레임이 이미 음성으로서 표시되었다면(즉, 블록(314)에 대한 대답이 "YES"), 상기 시스템은 음성이 프레임 내에 포함되어 있다고 이미 결정하였고, 블록(304)으로 표시한 것과 같이, 새로운 오디오 프레임을 분석하기 위하여 이동한다. 이전의 프레임이 음성으로서 표시되어 있지 않다면(즉, 블록(314)에 대한 대답이 "NO"), 상기 시스템은 그 프레임이 음성으로 표시되어야 하는지를 결정하기 위하여 하나 이상의 룰을 이용할 수 있다.

도 3에 나타낸 바와 같이, 결정 블록 "엔드포인트 외부"로서 표시한 블록(316)은 상기 프레임이 음성으로 표시되어야 하는지를 결정하기 위하여 하나 이상의 룰을 이용하는 루틴(routine)을 이용할 수 있다. 하나 이상의 룰은, 프레임 또는 프레임 그룹과 같이, 오디오 스트림의 임의의 부분에 적용될 수 있다. 상기 룰은 검사 하의 현재 프레임이 음성을 담고 있는지 여부를 결정할 수 있다. 상기 룰은 음성이 프레임 또는 프레임 그룹에 존재하거나 존재하지 않는지를 나타낼 수 있다. 음성이 존재한다면, 그 프레임은 엔드-포인트 내부에 있는 것으로서 표시될 수 있다.

음성이 존재하지 않는다고 상기 룰이 나타내면, 그 프레임은 엔드-포인트 외부에 있는 것으로서 표시될 수 있다. 결정 블록(316)이 프레임$_{n-1}$이 엔드-포인트 외부에 있다고 나타내면(즉, 어떠한 음성도 존재하지 않는다), 새로운 오디오 프레임, 즉 프레임$_{n+1}$이 시스템에 입력되고, 블록(304)에서 나타낸 것과 같이, 비음성으로서 표시된다. 결정 블록(316)이 프레임$_{n-1}$이 엔드-포인트 내부에 있다고 나타내면(즉, 음성이 존재한다), 프레임$_{n-1}$은 블록(318)에서 나타낸 것과 같이, 음성으로서 표시된다. 이전의 오디오 스트림은, 블록(320)에서 표시한 바와 같이, 메모리 내의 마지막 프레임이 분석될 때까지, 프레임-바이-프레임 방식으로 분석될 수 있다.

도 4는 도 3에 나타낸 블록(316)에 대한 보다 상세한 흐름도이다. 상기한 바와 같이, 블록(316)은 하나 이상의 룰을 포함할 수 있다. 그 룰은 음성의 존재 및/또는 부존재와 관련한 임의의 양태와 관련 있을 수 있다. 이러한 방식으로, 상기 룰은 구두 발성의 시작 및/또는 끝을 결정하는 데에 이용될 수 있다.

상기 룰은 이벤트(예컨대, 유성음 에너지, 무성음 에너지, 침묵의 부존재/존재 등) 또는 이벤트들의 임의의 조합(예컨대, 무성음 에너지에 침묵이 후속하고, 이 침묵에 유성음 에너지가 후속하는 경우, 유성음 에너지에 침묵이 후속하고, 이 침묵에 무성음 에너지가 후속하는 경우, 침묵에 무성음 에너지가 후속하고, 이 무성음 에너지에 침묵이 후속하는 경우 등)을 분석하는 것에 기초할 수 있다. 구체적으로, 상기 룰은 침묵 주기로부터 에너지 이벤트로의 천이 또는 침묵 주기로부터 에너지 이벤트로의 천이를 검사할 수 있다. 어떤 룰은, 음성이 무성음 이벤트 또는 모음 앞의 침묵으로부터의 단지 하나의 천이만을 포함할 수 있다는 룰을 이용하여, 모음 앞의 천이의 수를 분석할 수 있다. 또는, 어떤 룰은, 음성이 무성음 이벤트 또는 모음 후의 침묵으로부터의 단지 2개의 천이만을 포함할 수 있다는 룰을 이용하여 모음 후의 천이의 수를 분석할 수 있다.

하나 이상의 룰은 여러 가지 지속 기간 주기를 검사할 수 있다. 구체적으로, 상기 룰은 어떤 이벤트(예컨대, 유성음 에너지, 무성음 에너지, 침묵의 부존재/존재 등)에 대한 지속 시간을 검사할 수 있다. 어떤 룰은, 음성이 약 300 ms 내지 400 ms 범위 내의 모음 앞의 지속 시간을 포함할 수 있고 약 350 ms일 수 있다는 룰을 이용하여, 모음 앞의 지속 시간을 분석할 수 있다. 또는, 어떤 룰은 음성이 약 400 ms 내지 약 800 ms의 범위 내의 모음 후의 지속 시간을 포함할 수 있고 약 600 ms일 수 있다는 룰을 이용하여 모음 후의 지속 시간을 분석할 수 있다.

하나 이상의 룰은 이벤트의 지속 시간을 검사할 수 있다. 구체적으로, 상기 룰은 소정 타입의 에너지 지속 시간 또는 에너지 부족을 검사할 수 있다. 무성음 에너지는 분석될 수 있는 에너지의 한 가지 종류이다. 어떤 룰은, 음성이 약 150 ms 내지 약 300 ms 범위 내의 연속한 무성음 에너지의 지속 시간을 포함할 수 있고 약 200 ms일 수 있다는 룰을 이용하여, 연속

한 무성음 에너지의 지속 시간을 분석할 수 있다. 별법으로서, 연속한 침묵은 에너지의 부족으로서 분석될 수 있다. 어떤 룰은, 음성이 약 50 ms 내지 약 80 ms 범위 내의 모음 앞의 연속한 침묵의 지속 시간을 포함할 수 있고, 약 70 ms일 수 있다는 룰을 이용하여 모음 앞의 연속한 침묵의 지속 시간을 분석할 수 있다. 또는, 어떤 룰은, 음성이 약 200 ms 내지 약 300 ms 범위 내의 모음 후의 연속 침묵의 지속 시간을 포함할 수 있고 약 250 ms일 수 있다는 룰을 이용하여, 모음 후의 연속한 침묵의 지속 시간을 분석할 수 있다.

블록(402)에서, 분석되는 프레임 또는 프레임 그룹이 배경 잡음 레벨보다 높은 에너지를 갖고 있는지 여부를 결정하기 위한 체크가 수행된다. 배경 잡음 레벨보다 높은 에너지를 갖고 있는 프레임 또는 프레임 그룹은 소정 타입의 에너지의 지속 기간 또는 이벤트에 관한 지속 기간에 기초하여 추가로 분석될 수 있다. 분석되는 프레임 또는 프레임 그룹이 배경 잡은 레벨보다 높은 에너지를 갖고 있지 않다면, 그 프레임 또는 프레임 그룹은 연속한 침묵의 지속 기간, 침묵 주기로부터 에너지 이벤트로의 천이, 또는 침묵 주기로부터 에너지 이벤트로의 천이에 기초하여 추가로 분석될 수 있다.

분석되는 프레임 또는 프레임 그룹에 에너지가 존재한다면, "에너지" 카운터는 블록(404)에서 증가된다. "에너지" 카운터는 시간의 양을 카운트한다. 그 카운터는 프레임 길이만큼 증가한다. 프레임 크기가 약 32 ms라면, 블록(404)은 "에너지" 카운터를 약 32 ms만큼 증가시킨다. 결정 블록(406)에서, 상기 "에너지" 카운터의 값이 시간 문턱값(time threshold)을 초과하는지 여부를 확인하기 위하여 체크가 수행된다. 결정 블록(406)에서 평가된 문턱값은 음성의 존재 및/또는 부존재를 결정하는 데에 사용될 수 있는 연속한 무성음 에너지 룰에 대응한다. 결정 블록(406)에서, 연속한 무성음 에너지의 최대 지속 기간에 대한 문턱값은 평가될 수 있다. 결정 블록(406)이 "에너지" 카운터의 값이 문턱값 설정치를 초과한다고 결정하면, 분석되는 프레임 또는 프레임 그룹은 블록(408)에서 엔드-포인트 외부에 있는 것으로 지정된다(즉, 어떠한 음성도 존재하지 않는다). 그 결과, 도 3을 다시 참조하면, 상기 시스템은, 새로운 프레임, 즉 프레임$_{n+1}$이 시스템에 입력되어 비음성으로서 표시되는 블록(304)으로 점핑한다. 별법으로서, 블록(406)에서 복수의 문턱값이 평가될 수 있다.

블록(406)에서 "에너지" 카운터의 값이 어떠한 시간 문턱값도 초과하지 않는다면, "노에너지(noEnergy)" 카운터가 분리 문턱값(isolation threshold)을 초과하는지 여부를 결정하기 위하여 결정 블록(410)에서 체크가 수행된다. "에너지" 카운터(404)와 유사하게, "노에너지" 카운터(418)는 시간을 카운트하고, 분석되는 프레임 또는 프레임 그룹이 잡음 레벨보다 높은 에너지를 갖고 있을 때 프레임 길이만큼 증가된다. 상기 분리 문턱값은 2개의 파열음 이벤트(plosive event) 사이의 시간의 양을 규정하는 시간 문턱값이다. 파열음은 축어적으로, 화자의 입으로부터 폭발하는 자음(consonant)이다. 공기가 잠시 차단되어 압력을 증가시켜 파열음을 방출한다. 파열음은 "P", "T", "B", "D" 및 "K" 사운드를 포함할 수 있다. 이 문턱값은 약 10 ms 내지 약 50 ms의 범위 내에 있을 수 있고, 약 25 ms일 수 있다. 분리된 무성음 에너지 이벤트가 상기 분리 문턱값을 초과한다면, 침묵에 의해 둘러싸인 파열음(에컨대, STOP의 P)은 식별되었고, "분리된이벤트(isolatedEvent)" 카운터(412)가 증가된다. "분리된이벤트" 카운터(412)는 정수값으로 증가된다. "분리된이벤트" 카운터(412)를 증가시킨 후에, "노에너지" 카운터(418)는 블록(414)에서 리셋된다. 이 카운터는 리셋되는데, 왜냐하면 분석되는 프레임 또는 프레임 그룹 내에서 에너지가 발견되었기 때문이다. "노에너지" 카운터(418)가 상기 분리 문턱값을 초과하지 않는다면, "노에너지" 카운터(418)는 "분리된이벤트" 카운터(412)를 증가시키는 일이 없이 블록(414)에서 리셋된다. 다시, "노에너지" 카운터(418)가 리셋되는데, 왜냐하면 분석되는 프레임 또는 프레임 그룹 내에서 에너지가 발견되었기 때문이다. "노에너지" 카운터(418)를 리셋한 후에, 외부 엔드-포인트 분석은 블록(416)에서 "NO" 값을 반송함으로써, 분석되는 프레임 또는 프레임 그룹이 엔드-포인트 내부에 있는 것으로서 지정한다(예컨대, 음성이 존재한다). 그 결과, 다시 도 3을 참조하면, 상기 시스템은 318 또는 322에서 상기 분석된 프레임을 음성으로서 표시한다.

별법으로서, 결정 블록(402)이 잡음 레벨 보다 높은 에너지가 없다고 결정하면, 분석되는 프레임 또는 프레임 그룹은 침묵 또는 배경 잡음을 포함하고 있다. 이러한 경우에, "노에너지" 카운터(418)는 증가된다. 결정 블록(420)에서, "노에너지" 카운터의 값이 시간 문턱값을 초과하는지 여부를 확인하기 위한 체크가 수행된다. 결정 블록(420)에서 평가된 문턱값은 음성의 존재 및/또는 부존재를 결정하는 데 이용될 수 있는 연속한 무성음 에너지 룰 문턱값에 대응한다. 결정 블록(420)에서, 연속한 침묵의 지속 시간에 대한 문턱값이 평가될 수 있다. 결정 블록(420)이 "노에너지" 카운터의 값이 문턱값 설정치를 초과한다고 결정하면, 분석되는 프레임 또는 프레임 그룹은 블록(408)에서 엔드-포인트 외부에 있는 것으로서 지정된다(예컨대, 어떠한 음성도 존재하지 않는다). 그 결과, 다시 도 3을 참조하면, 상기 시스템은 새로운 프레임, 즉 프레임$_{n+1}$이 시스템에 입력되어 비음성으로서 표시되는 블록(304)으로 점핑한다. 별법으로서, 블록(406)에서 복수의 문턱값이 평가될 수 있다.

"노에너지" 카운터(418)의 값이 어떠한 시간 문턱값도 초과하지 않는다면, 최대 수의 허용된 분리된 이벤트가 일어났는지 여부를 결정하기 위하여, 결정 블록(422)에서 체크가 수행된다. "분리된이벤트" 카운터는 이 체크에 대답하기 위하여 필요한 정보를 제공한다. 허용된 분리된 이벤트의 최대 수는 구성 가능한 패러미터이다. 소정의 문법이 예상된다면(예컨대, "YES" 또는 "NO" 대답), 허용된 분리된 이벤트의 최대 수는 엔드-포인터의 결과를 "엄밀하게(tighten)" 하도록 설정될 수

있다. 허용된 분리된 이벤트의 최대 수가 초과되었다면, 분석되는 프레임은 블록(408)에서 엔드-포인트의 외부에 있는 것으로서 지정될 수 있다(예컨대, 어떠한 음성도 존재하지 않는다). 그 결과, 다시 도 3을 참조하면, 상기 시스템은 새로운 프레임, 즉 프레임$_{n+1}$이 시스템에 입력되어 비음성으로서 표시되는 블록(304)으로 점핑한다.

허용된 분리된 이벤트의 최대 수가 도달되지 않았다면, "에너지" 카운터(404)는 블록(424)에서 리셋된다. "에너지" 카운터(404)는 에너지가 없는 프레임이 식별되었을 때 리셋될 수 있다. "에너지" 카운터(404)를 리셋한 후에, 외부엔드-포인트 분석은, 블록(416)에서 "NO" 값을 반송함으로써, 분석되는 프레임이 엔드-포인트 내부에 있는 것으로서 지정한다(예컨대, 음성이 존재한다). 그 결과, 다시 도 3을 참조하면, 상기 시스템은 318 또는 322에서 상기 분석된 프레임을 음성으로서 표시한다.

도 5 내지 도 9는 시뮬레이션한 오디오 스트림의 일부 미가공 시계열(raw time series), 이들 신호의 여러 특성 플롯, 대응하는 미가공 신호의 분광 사진(spectrograph)을 보여준다. 도 5에서, 블록(502)은 시뮬레이션한 오디오 스트림의 미가공 시계열을 나타낸다. 상기 시뮬레이션한 오디오 스트림은 구두 발성 "NO"(504), "YES"(506), "NO"(504), "YES"(506), "NO"(504), "YESSSSS"(508), "NO"(504), 수 많은 "클리킹(clicking)" 사운드(510)를 포함한다. 이들 클리킹 사운드는 차량의 회전 신호가 관여될 때 발생되는 사운드를 나타낼 수 있다. 블록(512)은 상기 미가공 시계열 오디오 스트림에 대한 여러 특성 플롯을 나타낸다. 블록(512)은 x-축을 따라 샘플의 수를 표시한다. 플롯(514)은 엔드-포인터의 분석의 한 가지 대표도이다. 플롯(514)이 제로 레벨에 있을 경우, 엔드-포인터는 구두 발성의 존재를 결정하지 않는다. 플롯(514)이 비-제로 레벨에 있을 경우, 엔드-포인터는 구두 발성의 시작 및/또는 끝의 경계를 정한다. 플롯(516)은 배경 에너지 레벨보다 높은 에너지를 나타낸다. 플롯(518)은 시간-도메인 내의 구두 발성을 나타낸다. 블록(520)은 블록(502)에서 식별된 대응 오디오 스트림의 스펙트럼 대표도이다.

블록(512)은 엔드-포인터가 입력 오디오 스트림에 어떻게 응답하는지를 나타낸다. 도 5에 도시한 바와 같이, 엔드-포인터 플롯(514)은 "NO" 신호(504) 및 "YES"(506) 신호를 정확하게 캡처한다. "YESSSSS"(508)이 분석되는 경우, 엔드-포인터 플롯(514)은 잠시 후미의 "S"를 캡처하지만, 모음 후의 최대 기간 또는 연속한 무성음 에너지의 최대 지속 기간이 초과되었다는 것을 발견하면, 엔드-포인터는 컷오프된다. 상기 룰 기반형 엔드-포인터는 엔드-포인터 플롯(514)에 의해 정해진 오디오 스트림 부분을 ASR에 전송한다. 블록(512) 및 도 6 내지 도 9에서 도시한 바와 같이, ASR에 전송된 오디오 스트림 부분은 어느 룰이 적용되는지에 따라서 변한다. "클릭"(510)은 에너지를 갖고 있는 것으로서 검출되었다. 이는 블록(512)의 가장 우측부에서 배경 에너지 플롯(516)으로 나타내어진다. 그러나, "클릭"(510)에서 어떠한 모음도 검출되지 않았기 때문에, 엔드-포인터는 이러한 오디오 사운드를 배제한다.

도 6은 엔드-포인팅된 "NO"(504)의 상세도이다. 구두 발성 플롯(518)은 시간 스미어링(time smearing)으로 인해 하나의 프레임 또는 두 개만큼 지체된다. 상기 플롯(518)은, 상기 에너지 플롯(516)으로 나타내어지는, 에너지가 검출되는 기간 전체에 걸쳐 연속된다. 구두 발성 플롯(518)이 상승된 후에, 그 플롯은 평평하게 되고 배경 에너지 플롯(516)을 따라간다. 엔드-포인터 플롯(514)은 음성 에너지가 검출될 때 시작한다. 플롯(518)에 의해 나타내어지는 기간 동안, 엔드-포인터 룰 중 어느 것도 위반되지 않으며, 오디오 스트림은 구두 발성인 것으로 인식된다. 엔드-포인터는 모음 룰 후 연속 침묵의 최대 지속 기간 또는 모음 룰 후 최대 시간이 위반되었을 경우에 최우측에서 컷오프된다. 도시한 바와 같이, ASR로 보내지는 오디오 스트림 부분은 대략 3150 샘플들을 포함한다.

도 7은 엔드-포인팅된 "YES"(506)의 상세도이다. 구두 발성 플롯(518)은 다시, 시간 스미어링으로 인해 하나의 프레임 또는 두 개만큼 지체된다. 엔드-포인터 플롯(514)은 에너지가 검출될 때 시작한다. 엔드-포인터 플롯(514)은 에너지가 잡음으로 떨어질 때, 즉 모음 룰 후 최대 시간 또는 연속한 무성음 에너지 룰의 최대 지속 시간이 위반되었을 때까지 계속된다. 나타낸 바와 같이, ASR로 보내지는 오디오 스트림 부분은 대략 5550 샘플들을 포함한다. 도 6 및 도 7에서 ASR로 보내진 오디오 스트림의 양의 차이는 상이한 룰을 적용하는 엔드-포인터에서 비롯되는 결과이다.

도 8은 엔드-포인트된 "YESSSSS"(508)의 상세도이다. 엔드-포인터는 합리적인 시간 크기 동안만, 가능한 자음으로서 모음후 에너지(post-vowel energy)를 받아들인다. 합리적인 시간 기간 후에, 어느 모음 룰 후 최대 시간 또는 연속한 무성음 에너지 룰의 최대 지속 기간이 위반되었을 수도 있고, 엔드-포인터는 떨어져 ASR로 건네지는 데이터를 제한한다. 나타낸 바와 같이, ASR로 보내지는 오디오 스트림 부분은 대략 5750 샘플들을 포함한다. 구두 발성이 추가의 약 6500 샘플들에 대해서 계속되지만, 엔드-포인터는 합리적인 시간 후에 컷오프되므로, ASR로 보내진 오디오 스트림의 양은 도 6 및 도 7에서 보내진 것과는 상이하게 된다.

도 9는 몇몇 "클릭"(510)이 후속하는 엔드-포인트된 "NO"(504)의 상세도이다. 도 6 내지 도 8에서와 같이, 발성 구두 플롯(518)은 시간 스미어링 때문에 하나의 프레임 또는 두 개만큼 지체된다. 엔드-포인터(514)는 에너지가 검출될 때 시작

한다. 제1 클릭음은 엔드-포인트 플롯(514)에 포함되어 있는데, 왜냐하면 배경 잡음 에너지 레벨보다 높은 에너지가 있고 이 에너지는 자음, 즉 후미의 "T"일 수 있기 때문이다. 그러나, 제1 클릭음과 다음 클릭음 사이에 약 300 ms의 침묵이 있다. 이 예에서 사용되는 문턱값에 따르면, 이 침묵 기간은 모음 룰 후 연속한 침묵의 엔드-포인터의 최대 지속 기간을 위반한다. 따라서, 엔드-포인터는 그 제1 클릭음 후의 에너지를 배제하였다.

엔드-포인터는 오디오 스트림의 적어도 하나의 동적 양태를 분석함으로써 오디오 음성 세그먼트의 시작 및/또는 끝을 결정하도록 구성될 수도 있다. 도 10은 오디오 스트림의 적어도 하나의 동적 양태를 분석하는 엔드-포인터 시스템의 부분 흐름도이다. 글로벌 양태의 초기화는 단계(1002)에서 수행될 수 있다. 글로벌 양태는 오디오 스트림 자체의 특성을 포함할 수 있다. 제한하기 위한 것이 아닌 설명의 목적을 위해, 이들 글로벌 양태는 음성을 말하는 화자의 페이스 또는 화자의 피치를 포함할 수 있다. 단계(1004)에서, 로컬 양태의 초기화가 수행될 수 있다. 제한하기 위한 것이 아닌 설명의 목적을 위해, 이들 로컬 양태는 예상된 화자의 응답(예컨대, "YES" 또는 "NO" 응답), 환경적 조건(예를 들면, 시스템 내의 에코 또는 피드백의 존재에 영향을 미치는 개방 또는 폐쇄된 환경) 또는 배경 잡음의 추정을 포함할 수 있다.

상기 글로벌 및 로컬 초기화는 시스템의 동작 중 전체에 걸쳐 여러 시간에서 일어날 수 있다. 배경 잡음의 추정(로컬 양태 초기화)은 시스템에 먼저 전력이 공급될 때마다, 및/또는 소정의 시간 후에 실행될 수 있다. 음성을 말하는 화자의 페이스 또는 피치의 결정(글로벌 초기화)은 더 작은 비율로 분석되고 초기화된다. 유사하게, 어떤 응답이 예상되는 로컬 양태는 더 작은 비율로 초기화될 수 있다. 이 초기화는 ASR이 어던 응답이 예상되는 엔드 포인터와 통신할 때 일어날 수 있다. 환경 조건에 대한 로컬 양태는 파워 사이클 당 단 한번 초기화하도록 구성될 수 있다.

초기화 기간(1002, 1004) 동안, 엔드-포인터는 도 3 및 도 4와 관련하여 상기한 바와 같이, 그 디폴트 문턱값 설정치에서 동작할 수 있다. 임의의 초기화에 문턱값 설정치 또는 타이머의 변화가 요구된다면, 상기 시스템은 적절한 문턱값을 동적으로 변경할 수 있다. 별법으로서, 초기화 값에 기초하여, 상기 시스템은 시스템의 메모리에 미리 저장되어 있는 특정 또는 일반적인 사용자 프로화일을 재호출(recall)할 수 있다. 이 프로화일은 모든 또는 특정의 문턱값 설정치 및 타이머를 변경할 수 있다. 초기화 프로세스 동안 상기 시스템이, 사용자가 빠른 페이스로 말을 한다고 결정하면, 특정 룰의 최대 지속 기간은 상기 프로화일에 저장된 레벨로 감소될 수 있다. 또한, 나중에 사용할 사용자 프로화일을 생성 및 저장하기 위하여, 상기 시스템이 상기 초기화를 실행하도록 상기 시스템을 트레이닝 모드에서 동작시킬 수 있다. 나중에 사용할 목적으로 하나 이상의 프로화일이 시스템의 메모리 내에 저장될 수 있다.

도 1에서 설명한 엔드-포인터와 유사한 동적 엔드-포인터를 구성할 수 있다. 또한, 동적 엔드-포인터는 처리 환경과 ASR 사이에 양방향 버스를 포함할 수 있다. 상기 양방향 버스는 처리 환경과 ASR 사이에서 데이터 및 제어 정보를 전송할 수 있다. ASR로부터 처리 환경으로 보내진 정보는, 화자에게 부여되는 질문에 응답하여 소정의 응답이 예상된다는 것을 나타내는 데이터를 포함할 수 있다. ASR로부터 처리 환경으로 보내진 정보는 오디오 스트림의 양태를 동적으로 분석하는 데에 사용될 수 있다.

동적 엔드-포인터의 동작은, "엔드포인트 외부" 루틴, 즉 블록(316)의 하나 이상의 룰 중 하나 이상의 문턱값이 동적으로 구성될 수 있다는 것을 제외하고는 도 3 및 도 4를 참조하여 설명한 엔드-포인터와 유사하다. 다량의 배경 잡음이 있다면, 결정 블록(402)에서 잡음보다 큰 에너지에 대한 문턱값은 이러한 조건을 책임지기 위하여 동적으로 상승될 수 있다. 이러한 재구성을 수행하면, 상기 동적 엔드-포인터는 더 많은 천이 사운드 및 비음성 사운드를 거절할 수 있어, 폴스 포지티브의 수를 감소시킬 수 있다. 동적으로 구성 가능한 문턱값은 배경 잡음 레벨에 한정되지 않는다. 동적 엔드-포인터에 의해 이용되는 임의의 문턱값은 동적으로 구성될 수 있다.

도 3, 도 4 및 도 10에 나타낸 방법은 신호 담지 매체, 컴퓨터 판독 가능한 매체(예컨대, 메모리)에 인코딩되거나, 하나 이상의 집적 회로와 같은 소자 내부에 프로그램되거나 또는 컨트롤러 또는 컴퓨터에 의해 처리될 수 있다. 상기 방법이 소프트웨어에 의해 수행된다면, 그 소프트웨어는, 룰 모듈(10)에 상주하거나 그 모듈과 인터페이스를 이루는 메모리 또는 임의의 통신 인터페이스에 상주할 수 있다. 상기 메모리는 논리 함수(logical function)를 실행하기 위한 실행 가능한 명령어들의 순서 리스트를 포함할 수 있다. 논리 함수는 디지털 회로, 소스 코드, 아날로그 회로, 또는 전기적, 오디오 또는 비디오 신호를 통하는 것과 같은 아날로그 소스를 통해 실행될 수 있다. 상기 소프트웨어는 명령 실행 가능한 시스템, 장치 또는 디바이스에 의해 또는 이들과 연계하여 사용하기 위하여, 임의의 컴퓨터 판독 가능한 매체 또는 신호 담지 매체에 내장될 수 있다. 이러한 시스템은 컴퓨터 기반 시스템, 프로세서 포함 시스템, 또는 명령 실행 가능한 시스템, 장치, 또는 명령을 실행할 수 있는 디바이스로부터 명령을 선택적으로 페치할 수 있는 다른 시스템을 포함할 수 있다.

"컴퓨터 판독 가능한 매체", "기계 판독 가능한 매체", "전파 신호(propagated-signal)" 매체 및/또는 "신호 담지 매체"는 명령 실행 가능한 시스템, 장치 또는 디바이스에 의해 또는 그 시스템, 장치 또는 디바이스와 연계하여 사용하기 위한 소프트웨어를 포함하고, 저장하고, 통신하며, 전파 또는 운송하는 임의의 수단을 포함할 수 있다. 기계 판독 가능한 매체는 선

택적으로, 전자, 자기, 광, 전자기, 적외선 또는 반도체 시스템, 장치, 디바이스 또는 전파 매체일 수 있지만, 이들에 제한되는 것은 아니다. 기계 판독 가능한 매체의 비제한적인 예로서 다음과 같은 것이 있다. 즉, 하나 이상의 와이어를 구비하는 전기적 접속 "전자 장치", 휴대형 자기 또는 광 디스크, "RAM"(전자 장치)과 같은 휘발성 메모리, "ROM"(전자 장치), 소거 가능하고 프로그램 가능한 ROM(EPROM 또는 플래시 메모리)(전자 장치), 또는 광 섬유(광). 기계 판독 가능한 매체는 또한 유형 매체를 포함할 수 있는데, 이 매체에는, 소프트웨어가 전자적으로 이미지 또는 다른 포맷으로 저장됨에 따라(에컨대, 광 스캔을 통해), 소프트웨어가 프린트되어지고 그 후 컴파일링되고 및/또는 해석되거나 그렇지 않으면 처리된다. 다음에, 상기 처리된 매체는 컴퓨터 및/또는 기계 메모리에 저장될 수 있다.

본 발명의 다양한 실시예를 설명하였지만, 당업자는 다른 많은 실시예 및 변형이 본 발명의 범위 내에서 가능하다는 것을 이해할 것이다. 따라서 본 발명의 범위는 오직 첨부된 청구범위와 그 등가물에 의해서만 제한된다.

도면의 간단한 설명

도 1은 음성 엔드-포인팅 시스템의 블록도이다.

도 2는 차량에 합체되는 음성 엔드-포인팅 시스템의 일부를 보여주는 도면이다.

도 3은 음성 엔드-포인터의 흐름도이다.

도 4는 도 3의 일부에 대한 보다 상세한 흐름도이다.

도 5는 시뮬레이션한 음성 사운드의 엔드-포인팅을 나타낸다.

도 6은 도 5의 시뮬레이션한 음성 사운드의 일부에 대한 상세한 엔드-포인팅을 나타낸다.

도 7은 도 5의 시뮬레이션한 음성 사운드의 일부에 대한 제2의 상세한 엔드-포인팅을 나타낸다.

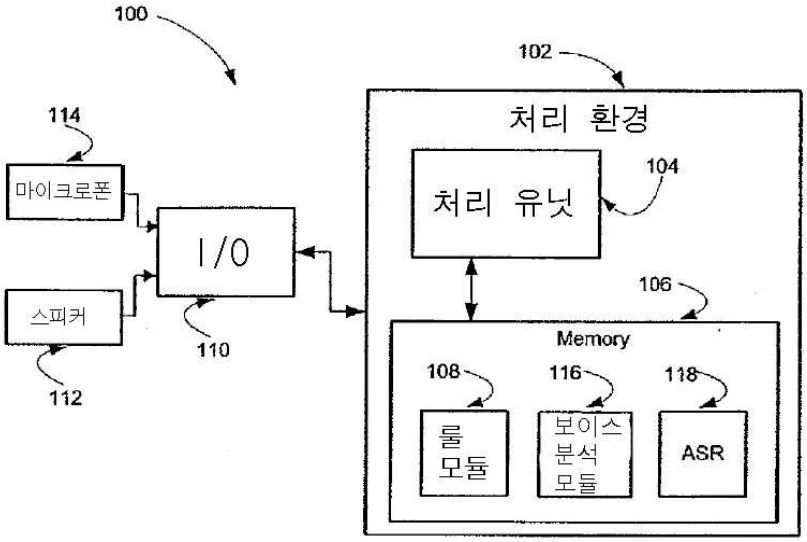
도 8은 도 5의 시뮬레이션한 음성 사운드의 일부에 대한 제3의 상세한 엔드-포인팅을 나타낸다.

도 9는 도 5의 시뮬레이션한 음성 사운드의 일부에 대한 제4의 상세한 엔드-포인팅을 나타낸다.

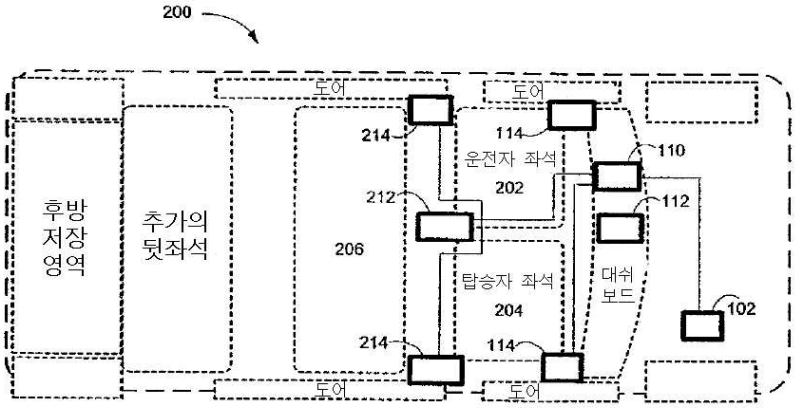
도 10은 음성에 기초한 동적 음성 엔드-포인팅 시스템의 부분 흐름도이다.

도면

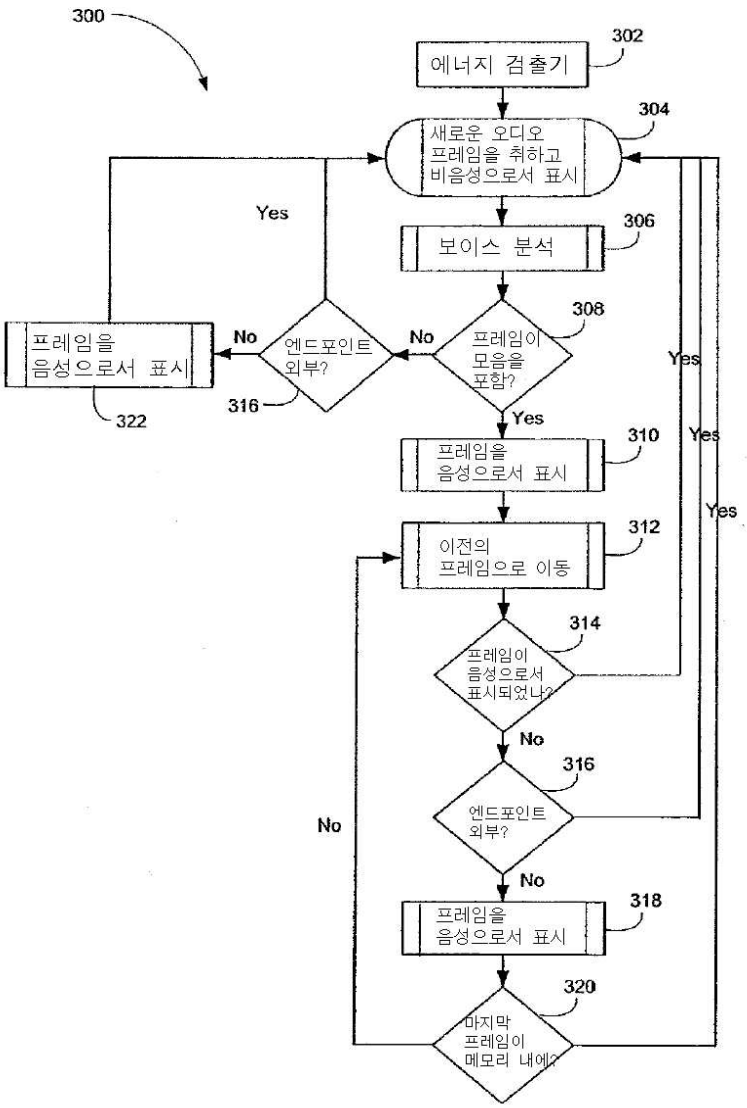
도면1


100
102
처리 환경
114
마이크로폰
I/O
104
처리 유닛
스피커
106
Memory
110
112
108
116
118
룸
모듈
보이스
분석
모듈
ASR

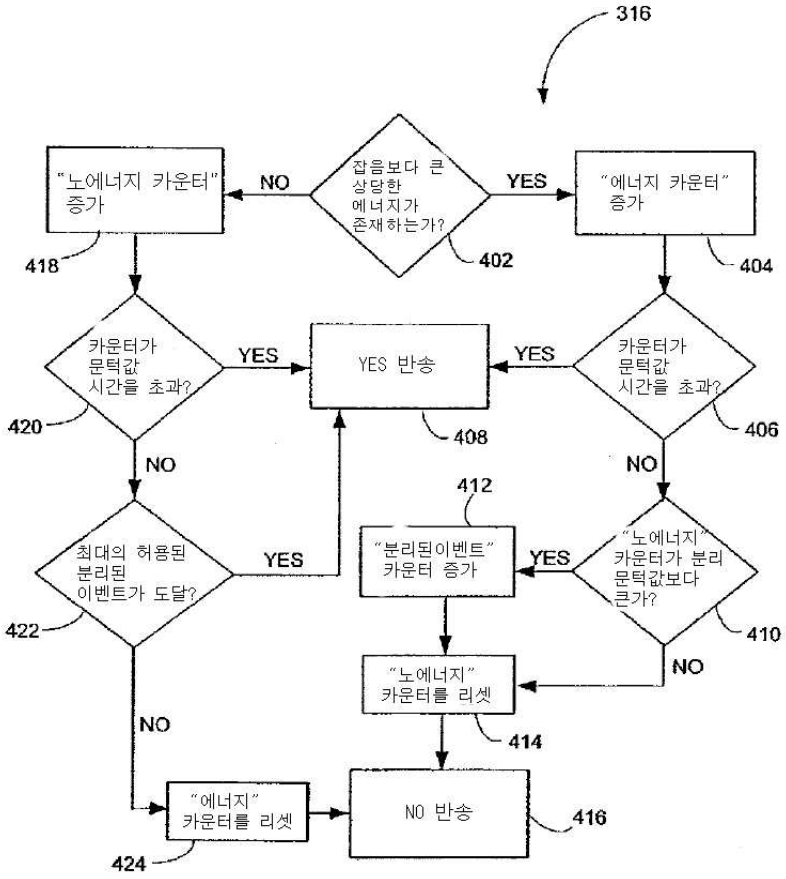
도면2


200
도어
도어
214
114
운전자 좌석
202
110
212
112
후방
저장
영역
추가의
뒷좌석
206
탑승자 좌석
204
대쉬
보드
102
214
114
도어
도어

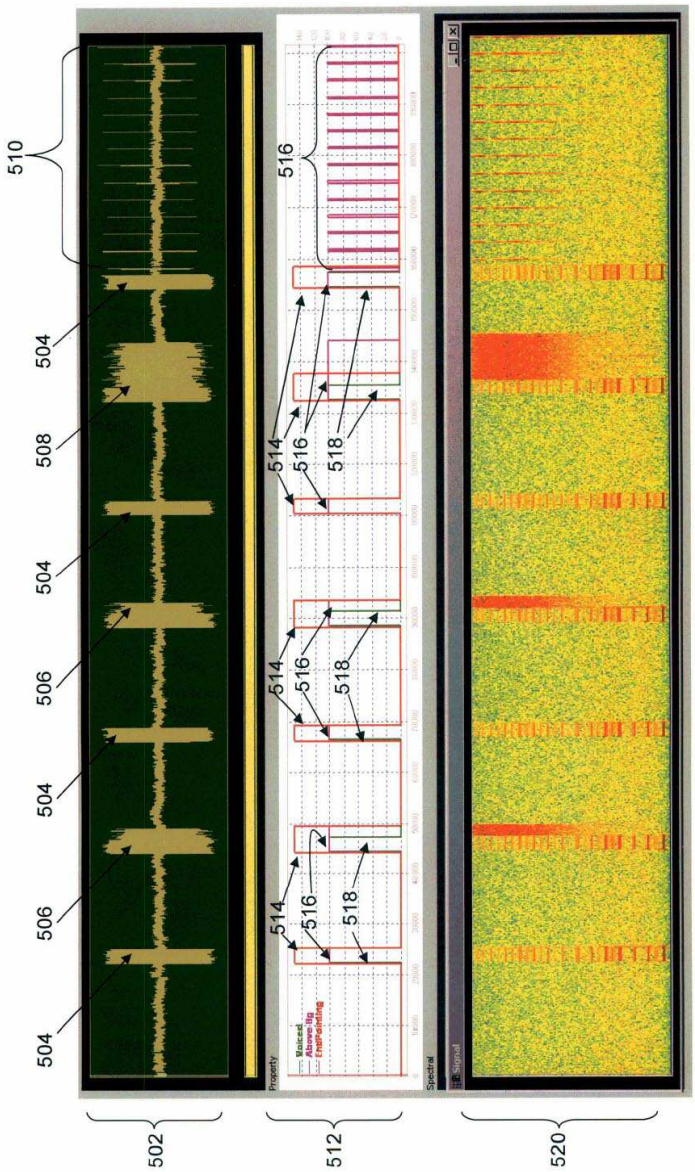
도면3


300
에너지 검출기
302
새로운 오디오 프레임을 취하고 비음성으로서 표시
304
보이스 분석
306
프레임이 모음을 포함?
308
엔드포인트 외부?
316
프레임을 음성으로서 표시
322
프레임을 음성으로서 표시
310
이전의 프레임으로 이동
312
프레임이 음성으로서 표시되었나?
314
엔드포인트 외부?
316
프레임을 음성으로서 표시
318
마지막 프레임이 메모리 내에?
320
Yes
No

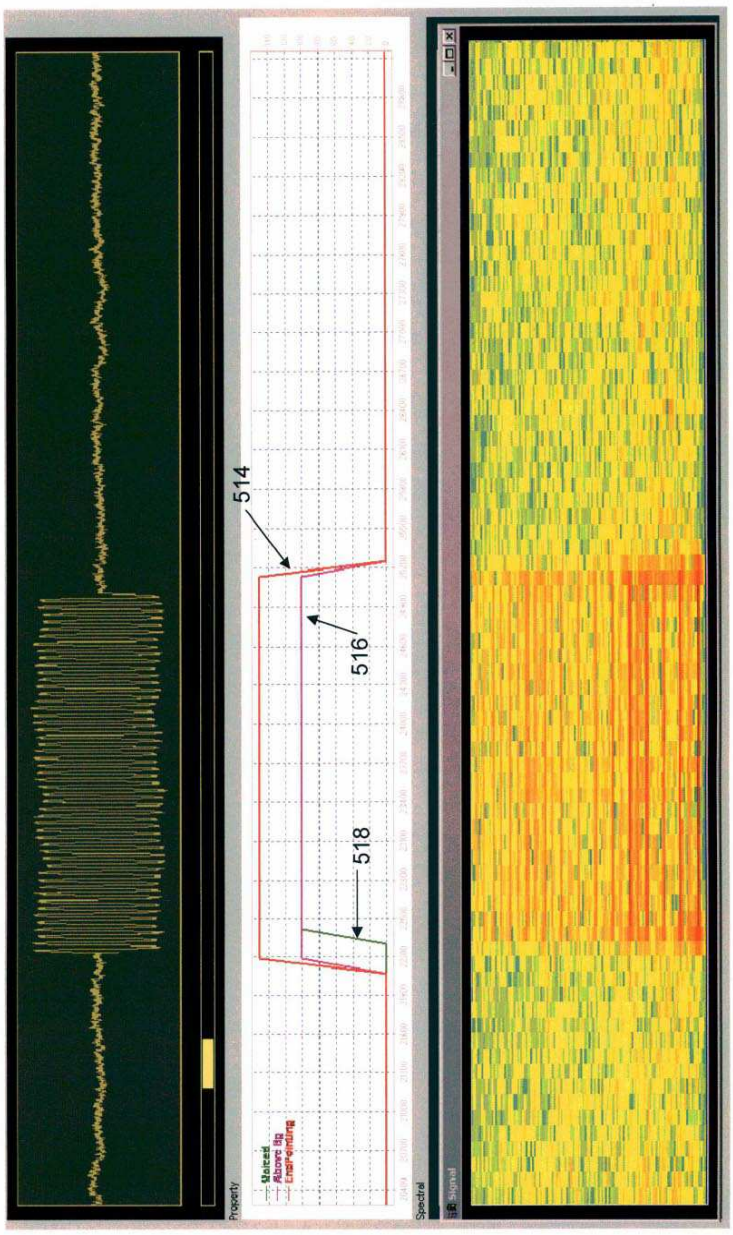
도면4


316
"노에너지 카운터" 증가
NO
잡음보다 큰 상당한 에너지가 존재하는가?
YES
"에너지 카운터" 증가
418
402
404
카운터가 문턱값 시간을 초과?
YES
YES 반송
YES
카운터가 문턱값 시간을 초과?
420
408
406
NO
NO
412
최대의 허용된 분리된 이벤트가 도달?
YES
"분리된이벤트" 카운터 증가
YES
"노에너지" 카운터가 분리 문턱값보다 큰가?
422
410
NO
"노에너지" 카운터를 리셋
NO
414
"에너지" 카운터를 리셋
NO 반송
416
424

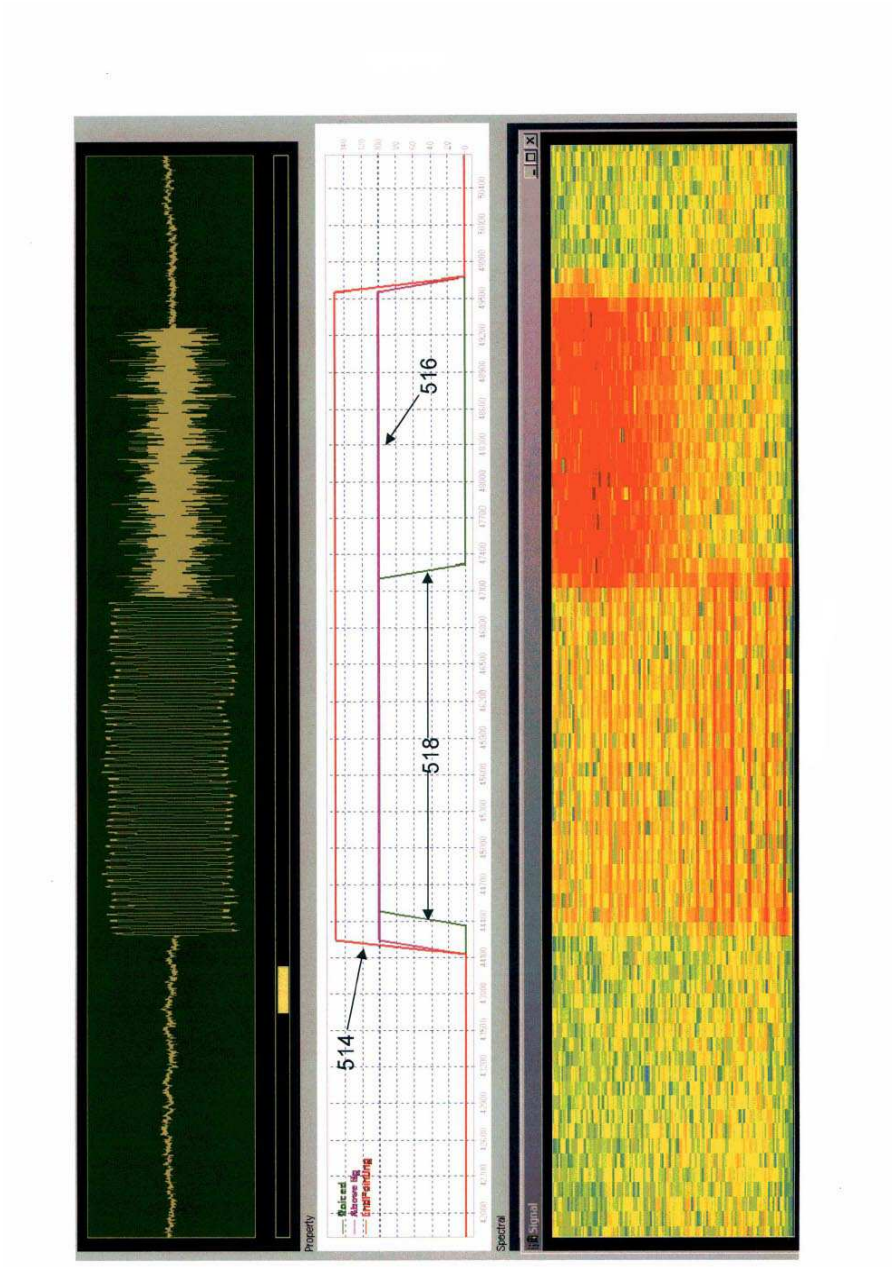
도면5


510
504
508
504
506
504
506
504
516
514
516
518
514
516
518
514
516
518
502
512
520

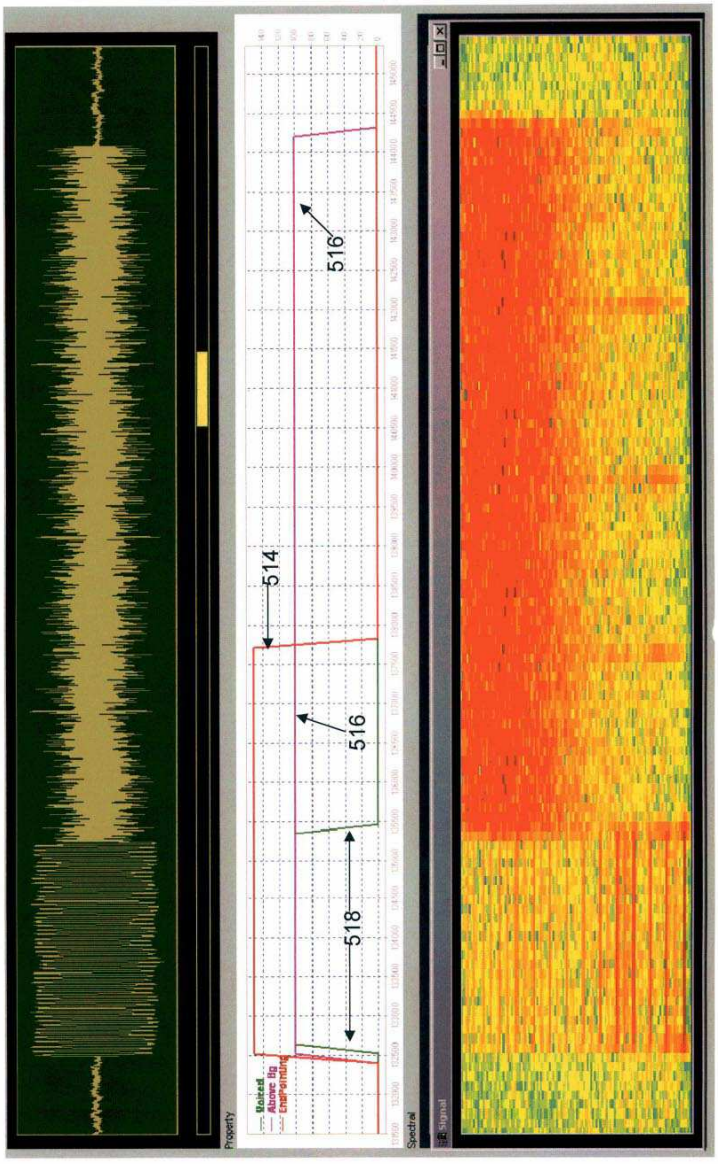
도면6



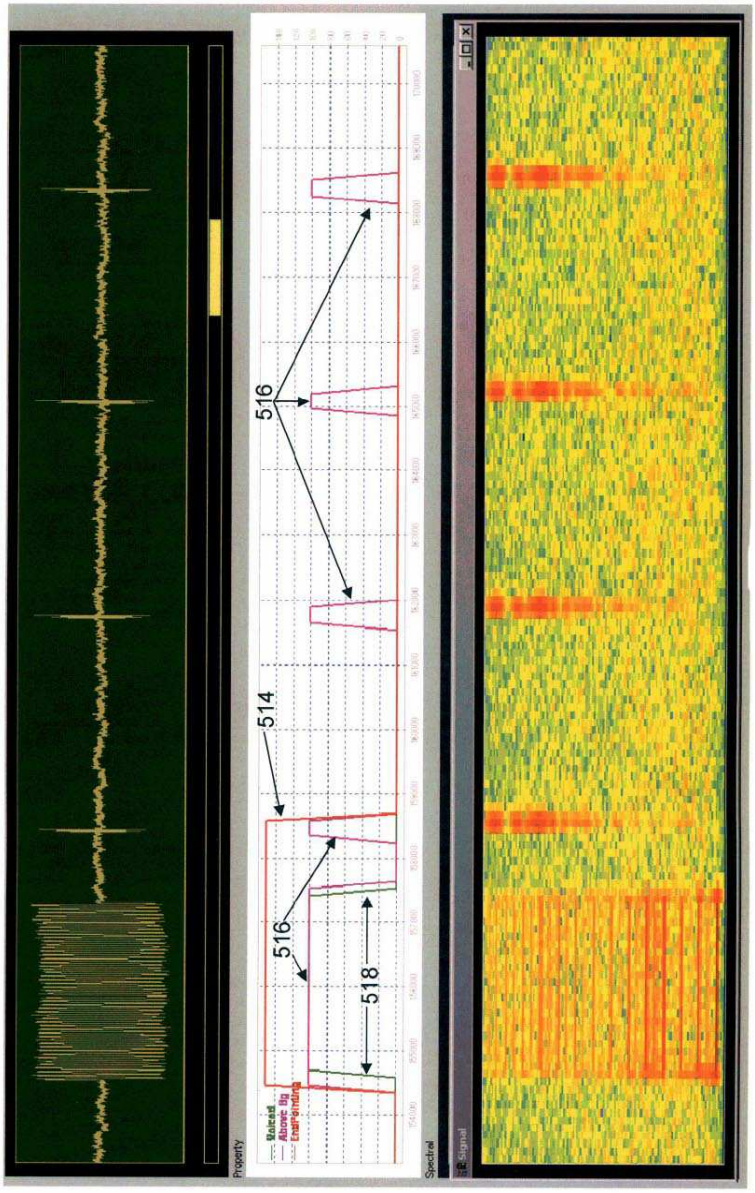
도면7



도면8



도면9



도면10