



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0108643  
(43) 공개일자 2009년10월15일

- |  |   |
|--|---|
| <p>(51) Int. Cl.<br/>G06Q 30/00 (2006.01) H04B 1/40 (2006.01)</p> <p>(21) 출원번호 10-2009-7017831(분할)</p> <p>(22) 출원일자 2005년03월22일<br/>심사청구일자 2009년08월26일</p> <p>(62) 원출원 특허 10-2006-7022234<br/>원출원일자 2006년10월25일<br/>심사청구일자 2006년10월25일</p> <p>(85) 번역문제출일자 2009년08월26일</p> <p>(86) 국제출원번호 PCT/IB2005/000743</p> <p>(87) 국제공개번호 WO 2005/093623<br/>국제공개일자 2005년10월06일</p> <p>(30) 우선권주장<br/>10/810,924 2004년03월26일 미국(US)</p> | <p>(71) 출원인<br/>노키아 코포레이션<br/>핀란드핀-02150 에스푸 카일알라덴티에 4</p> <p>(72) 발명자<br/>코프라 토니<br/>핀란드 핀-01450 반타아 히르비티에 15 에이 비 6<br/>미코 마키파<br/>핀란드 핀-00830 헬싱키 아이로란타 9 에이<br/>바나넨 마우리<br/>핀란드 핀-33500 탐페레 소르사푸이스토 1 에이 10</p> <p>(74) 대리인<br/>리엔목특허법인</p> |
|--|---|

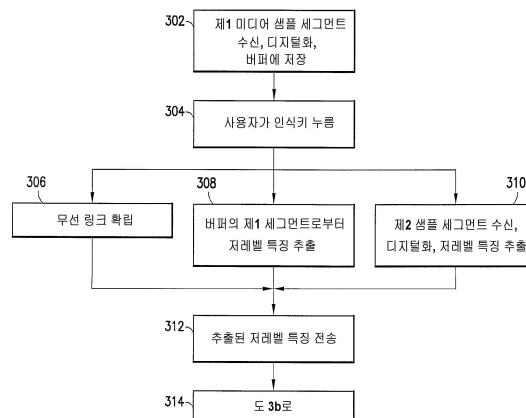
전체 청구항 수 : 총 31 항

(54) 네트워크 휴대용 장치에서의 특징 추출

(57) 요약

이동국(MS)은 미디어 샘플을 수신하기 위한 변환기 또는 다른 인터페이스, 미디어 샘플의 디지털 버전으로부터 특징을 추출하기 위한 프로세서, 이러한 추출된 특징들을 전송하기 위한 전송기 및 바람직하게는 미디어 샘플을 식별하기 위한 사용자 입력 전에 수신된 음악 부분에 대한 특징 추출을 시작하기 위한 버퍼를 가진다. 특징은 미디어 파일의 디지털 버전의 콘텐츠를 기술한다. 서버는 MS로부터 추출된 특징을 수신하고, MS-추출 특징 및/또는 MS로부터 전송된 디지털 미디어 샘플의 세그먼트로부터의 저레벨 특징으로부터 추출되는 고레벨 특징일 수 있는 추가적인 특징을 추출한다. 서버는 데이터베이스를 검색하기 위해 모든 추출된 특징을 사용하고 미디어 샘플이 취해진 근본 파일을 식별한다. MS에서의 일회 사용자 입력은 특징 추출, 링크 확립, 특징 추출을 시작하게 한다. 추가적 입력으로 MS로 다운로드된 전체 미디어 파일의 복사본을 구입한다.

대표도 - 도3a



## 특허청구의 범위

### 청구항 1

장치로서,

미디어 샘플을 수신하도록 구성된 인터페이스;

상기 미디어 샘플의 디지털 버전으로부터 제 1 세트 특징(feature)과 제 2 세트 특징을 추출(extract)하도록 구성된 프로세서;

미디어 인식 서비스와 무선 링크를 확립하고 상기 무선 링크를 통해 상기 추출된 제 1 세트 특징을 전송하고 추가 특징 요청 수신 시에는 상기 무선 통신 링크를 통해 추출된 제 2 특징을 자동으로 전송하는 전송기를 포함하되, 상기 장치는 동시에 상기 무선 링크를 확립하고 상기 제 1 및 제 2 특징을 추출하고; 및

상기 무선 링크를 통해 요청 메시지 안에서 상기 추가 특징에 대한 요청을 수신하기 위한 수신기를 포함하는 장치.

### 청구항 2

제1항에 있어서, 상기 인터페이스는 변환기를 포함하는 것을 특징으로 하는 장치.

### 청구항 3

제2항에 있어서, 상기 변환기는 마이크روف론을 포함하고 상기 미디어 샘플은 오디오 샘플을 포함하는 것을 특징으로 하는 장치.

### 청구항 4

제2항에 있어서, 상기 변환기는 카메라를 포함하고 상기 미디어 샘플은 비주얼(visual) 샘플을 포함하는 것을 특징으로 하는 장치.

### 청구항 5

제1항에 있어서, 상기 인터페이스는 케이블 및 무선 링크 중 하나를 포함하는 것을 특징으로 하는 장치.

### 청구항 6

제5항에 있어서, 상기 인터페이스가 수신하는 상기 미디어 샘플이 디지털 버전인 것을 특징으로 하는 장치.

### 청구항 7

제1항에 있어서, 상기 전송기는, 상기 하나 이상의 추출된 특징을 포함하면서 상기 미디어 샘플의 어떠한 디지털 버전 부분도 포함하지 않고 있는 메시지를 전송하도록 구성된 것을 특징으로 하는 장치.

### 청구항 8

제1항에 있어서, 상기 프로세서가 상기 미디어 샘플의 상기 디지털 버전에 기초하여 추출할 여러 가지 특징을 적응적으로 선택하도록 구성된 것을 특징으로 하는 장치.

### 청구항 9

제1항에 있어서, 상기 프로세서는 상기 미디어 샘플의 상기 디지털 버전에 기초하여 추출할 하나 이상의 타입(type)의 특징을 적응적으로 선택하도록 구성되되, 상기 프로세서는 상기 적응적으로 선택된 타입(type)의 하나 이상의 특징을 추출하고, 상기 전송기는 상기 선택된 타입의 특징 식별자를 전송하도록 구성된 것을 특징으로 하는 장치.

### 청구항 10

제1항에 있어서,

상기 전송기로 하여금 상기 제 1 세트 특징을 전송하도록 구성된 사용자 인터페이스; 및

상기 미디어 샘플의 디지털 버전의 적어도 일부를 저장하도록 구성된 버퍼를 더 포함하되, 상기 프로세서는 상기 사용자 인터페이스에서 사용자 입력 전에 제 1 세트의 적어도 일부를 추출하도록 구성된 것을 특징으로 하는 장치.

**청구항 11**

제1항에 있어서, 단일 사용자 입력에 의해 개시되는 사용자 인터페이스, 및

상기 제 1 및 제 2 특징을 추출하고 상기 원격 미디어 인식 서비스와 무선 통신 링크를 확립하도록 하는 프로세서를 더 포함하는 것을 특징으로 하는 장치.

**청구항 12**

제11항에 있어서, 상기 단일 사용자 입력은 상기 제 2 특징의 적어도 일부를 저장하는 것을 시작하도록 상기 변환기와 상기 프로세서 사이에 위치하는 버퍼 구동을 개시(incite)하는 것을 특징으로 하는 장치.

**청구항 13**

제1항에 있어서, 상기 제 1 세트 및 제 2 세트 특징이 MPEG-7 기술자(descriptor)를 포함하는 것을 특징으로 하는 장치.

**청구항 14**

제1항에 있어서, 상기 제 1 세트 특징은 상기 미디어 샘플의 디지털 버전의 재구성 불가(non-reconstructive)인 것을 특징으로 하는 장치.

**청구항 15**

제1항에 있어서, 상기 제 1 세트 및 제 2 세트 특징들은, 조합 상으로, 상기 미디어 샘플의 디지털 버전의 재구성 불가(non-reconstructive)인 것을 특징으로 하는 장치.

**청구항 16**

제1항에 있어서, 상기 요청 메시지는 적어도 타입(type)에 의해 각각의 추가적인 특징을 특정하게 식별하고, 상기 제 2 특징은 상기 식별된 타입의 특징 만을 포함하는 것을 특징으로 하는 장치.

**청구항 17**

미디어 파일에 대한 시그널링 정보를 원격 미디어 인식 서비스로 전송하기 위한 방법에 있어서,

휴대용 무선 장치에서 미디어 샘플을 수신하고,

상기 휴대용 무선 장치에서, 상기 미디어 샘플의 디지털 버전으로부터 제 1 특징과 제 2 특징을 동시에 추출하고,

상기 휴대용 무선 장치로부터 상기 추출된 제 1 특징을 포함하는 메시지를 전송하고,

상기 휴대용 무선 장치에서, 하나 이상의 추가 특징을 요청하는 요청 메시지를 수신하고,

상기 요청된 하나 이상의 추가 특징과 일치하는 추출된 제 2 특징을 포함하는 메시지를 상기 휴대용 무선 장치로부터 전송하는 것을 특징으로 하는 미디어 파일에 대한 시그널링 정보를 원격 미디어 인식 서비스로 전송하기 위한 방법.

**청구항 18**

제17항에 있어서, 상기 휴대용 무선 장치에서, 상기 제 1 및 제 2 특징 추출과 동시에 상기 원격 미디어 인식 서비스와 무선 통신 링크를 확립하는 것을 특징으로 하는 미디어 파일에 대한 시그널링 정보를 원격 미디어 인식 서비스로 전송하기 위한 방법.

**청구항 19**

장치로서,

적어도 하나의 프로세서,  
 컴퓨터 프로그램 코드를 포함하는 적어도 하나의 메모리를 포함하되,  
 상기 적어도 하나의 메모리와 컴퓨터 프로그램 코드는 상기 적어도 하나의 프로세서와 함께 적어도,  
 미디어 샘플을 수신하고,  
 상기 미디어 샘플의 디지털 버전으로부터 제 1 및 제 2 특징을 추출하고,  
 미디어 인식 서비스와 무선 통신 링크를 확립하고,  
 상기 무선 통신 링크를 통해 상기 추출된 제 1 특징을 전송하고,  
 추가 특징 요청을 수신하였을 때, 상기 무선 통신 링크를 통해 상기 제 2 특징을 자동으로 전송하되, 상기 장치는 상기 무선 통신 링크를 확립하는 것과 상기 제 1 및 제 2 특징을 추출하는 것을 동시에 하고,  
 상기 무선 통신 링크를 통해 요청 메시지에서 상기 추가 특징에 대한 요청을 수신하는 것을 실행하는 것을 특징으로 하는 장치.

**청구항 20**

컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체로서,  
 상기 사용자 입력에 응답하여 디지털 미디어 샘플로부터 제 1 특징과 제 2 특징을 추출하고, 그와 동시에 원격 미디어 인식 서비스와 무선 통신 링크를 확립하기 위한 제 1 컴퓨터 명령 세트;  
 상기 무선 통신 링크를 통해 제 1 메시지 내에 추출된 제 1 특징을 전송하기 위한 제 2 컴퓨터 명령 세트; 및  
 상기 수신된 요청 메시지에 응답하여 상기 무선 통신 링크를 통해 추출된 제 2 특징을 전송하기 위한 제 3 컴퓨터 명령 세트를 포함하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체.

**청구항 21**

제20항에 있어서, 상기 제 1 및 제 2 메시지는 상기 디지털 미디어 샘플의 어떠한 부분도 포함하지 않는 특징을 포함하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체.

**청구항 22**

제20항에 있어서, 상기 응답 메시지는 추가 특징의 수를 특정하고, 제 1 컴퓨터 명령 세트는 상기 특정화된 수를 포함하는 제 2 특징을 적응적으로 선택하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체.

**청구항 23**

제20항에 있어서, 제 1 컴퓨터 명령 세트는 상기 요청 메시지에 기초하여 추출할 한 타입(type)의 특징을 적응적으로 선택하고, 상기 적응적으로 선택된 타입의 제 1 세트 특징을 추출하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체.

**청구항 24**

제20항에 있어서, 상기 제 1 컴퓨터 명령 세트는 상기 디지털 미디어 샘플의 제 1 시간 제한(time-bounded) 세그먼트로부터 제 1 특징들을 추출하도록 구성되고,  
 상기 제 2 컴퓨터 명령 세트는 상기 제 1 특징들과 함께 상기 제 1 시간 제한 세그먼트가 아닌 제 2 시간 제한 세그먼트를 전송하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체.

**청구항 25**

제20항에 있어서, 상기 적어도 제 1 및 제 2 특징은 타임포인트(timepoint)를 정의하고, 상기 제 1 컴퓨터 명령 세트는 상기 디지털 미디어 샘플로부터 하나 이상의 타임포인트를 추출하고, 상기 제 1 및 제 2 메시지 중의 하나는 타임포인트, 상기 디지털 미디어 샘플의 스펙트럴 슬라이스 및 상기 스펙트럴 슬라이스를 상기 타임포인트로 연계시키는 식별자를 포함하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독

매체.

**청구항 26**

제20항에 있어서, 상기 제 1 특징들은 상기 디지털 미디어 샘플을 재구성 불가능한 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 이동국 내의 컴퓨터 가독 매체.

**청구항 27**

다수의 추출된 특징을 데이터베이스에 저장된 특징 세트와 정확하게 일치시키기 위한 컴퓨터 프로그램을 탑재한 컴퓨터 가독 매체로서,

네트워크를 통해 제 1 특징을 포함하는 제 1 메시지를 수신하기 위한 제 1 컴퓨터 명령 세트;

수신된 특징과 일치하는 모든 일치 세트를 찾기 위해 특징 세트 데이터베이스를 검색하고, 각각의 일치 세트 중에서 구별되는 하나 이상의 추가적인 특징을 판별하기 위한 제 2 컴퓨터 명령 세트;

상기 하나 이상의 추가적인 특징을 명기한 요청 메시지를 상기 네트워크를 통해 전송하기 위한 제 3 컴퓨터 명령 세트;

상기 네트워크를 통해 상기 하나 이상의 추가적인 특징과 일치하는 제 2 특징을 포함하는 제 2 메시지를 수신하기 위한 제 4 컴퓨터 명령 세트; 및

상기 수신된 제 2 특징을 사용하여 매칭 세트로부터 정확하게 하나의 특징을 식별하기 위한 제 5 컴퓨터 명령 세트를 포함하는 컴퓨터 프로그램을 탑재한 컴퓨터 가독 매체.

**청구항 28**

제27항에 있어서, 각각의 특징 세트는 미디어 파일 타이틀과 관련되어 있고, 상기 컴퓨터 프로그램은 네트워크를 통해 상기 메시지의 전송자에게 상기 미디어 파일 타이틀을 포함하는 응답 메시지를 전송하는 제 5 컴퓨터 명령 세트를 더 포함하는 컴퓨터 프로그램을 탑재한 컴퓨터 가독 매체.

**청구항 29**

제27항에 있어서, 상기 요청 메시지는 상기 추가 특징의 넘버(number)와 상기 하나 이상의 추가적 특징의 타입(type) 중 적어도 하나를 포함하는 것을 특징으로 하는 컴퓨터 프로그램을 탑재한 컴퓨터 가독 매체.

**청구항 30**

이동국으로서,

미디어 샘플을 수신하기 위한 수단;

원격 미디어 인식 서비스와 무선 통신 링크를 확립과 상기 미디어 샘플의 디지털 버전으로부터 제 1 특징으로 추출을 동시에 하기 위한 프로세싱 수단을 포함하되, 상기 프로세싱 수단은 상기 제 1 특징을 추출하기 위해 사용자 입력에 응답하고, 식별된 추가적인 특징과 일치하는 제 2 세트 특징을 자동으로 추출하기 위해 추가적인 특징을 식별하는 요청 메시지에 응답하고,

무선 통신 링크를 통해 독립된 메시지로 상기 제 1 및 제 2 세트 특징을 전송하기 위한 수단; 및

상기 요청 메시지를 수신하기 위한 수단을 포함하는 것을 특징으로 하는 이동국.

**청구항 31**

제30항에 있어서, 상기 수신하기 위한 수단은 변환기를 포함하고, 상기 추출하기 위한 수단은 디지털 프로세서를 포함하는 것을 특징으로 하는 이동국.

**명세서**

**발명의 상세한 설명**

**기술분야**

<1> 본 발명은 디지털 미디어 파일 인식과 이동 전자제품 분야와 관련된다. 구체적으로 넓게는 본 발명은 이동 무선 전화기에서와 같은 이동국 내에서 오디오 샘플의 적어도 일부 인식에 관련된다.

### 배경 기술

<2> 대중 음악 소유자들은 최근에서야 인터넷과 같은 전자통신망을 통해 그들의 저작물의 디지털 복사본을 다운로드 하는 것을 받아들였다. 이렇게 하는 잘 알려진 웹 사이트가 <http://www.apple.com/itunes/>인데, 여기서 사용자는 PC 등을 통해 웹사이트를 방문하고, 직접 곡명을 수동으로 선택하고, 웹 사이트를 방문하기 위해 그들이 사용한 PC로 선택한 곡들의 디지털 버전(예를 들면 MP3, AAC)을 다운로드한다. 사용자는 그러고 난 후 PC 상에서 음악을 실행하거나 iPod®나 다른 전용 디지털 뮤직 플레이어같은 이동 장치로 전송한다.

<3> 최근 음악 다운로드 기술의 제2세대가 시장에 등장했는데, 사용자는 웹 페이지 상에서 곡명에 의해 노래를 손으로 직접 선택할 필요가 없게 되었다. 그 대신, 노래는 서버나 혹은 웹 페이지에 연관된 다른 컴퓨터에 의해 '칭취되고', 디지털 음악의 데이터베이스로부터 그 특정 노래를 식별하도록 디지털적으로 해석된다. 예를 들면, [http://shazam.com/uk/do/help\\_faqs\\_shazam#4](http://shazam.com/uk/do/help_faqs_shazam#4) 웹 사이트는 다음과 같은 동작을 설명한다. 사용자가 술집이나 자동차 같은 곳에서 음악을 듣고 음악이 연주되는 동안 자신의 이동국(MS)에 코드를 입력한다. 상기 음악은 마치 보이스(voice) 입력이 보내지는 것 같이 표준 MS 링크를 통해 shazam.com과 같은 호스팅 웹 사이트로 보내진다. 호스팅 웹 사이트는 (변환된) 생음악 입력을 수신하고, 그것을 해석하고, 자신의 데이터베이스 내의 음악 중 하나와 일치시킨다. Shazam.com 기술은 이동국을 단지 아날로그 입력을 디지털로 변환하고 변환된 신호를 패킷화하는 통로로만 사용하여 서버에서 음악을 식별하고 상기 이동국으로 결과를 되돌려 주는 데 한정된 것처럼 보인다.

<4> 대부분의 음악 다운로드 서비스는 개인 컴퓨터(PC)를 통해 액세스되지만, 이동 무선(Over-The-Air) 다운로드 서비스 역시 소개되었다. 이동 장치는 작은 스크린 크기, 제한된 파워 서플라이, 신뢰성 및 데이터 접속 속도에 관한 제한을 가지고 있는데, 이러한 것들은 음악 탐색과 음악 서비스로부터 이동 단말기로의 음악 전송을 힘들게 한다. 보통 실제 음악 서비스는 최소 500,000 곡의 컬렉션을 가진다. 순차적인 프레젠테이션이 스크린 콘텐츠와 부합해야만 하는, 이동 장치 사용자 인터페이스를 사용하는 경우에 있어서는 음악 카탈로그를 탐색하는 것이 힘든 일이다. 이는 계층 분류가 복잡해지고 음악 탐색을 힘들게 만든다.

<5> 사람들이 술집이나 콘서트 홀 혹은 자동차 같은 공공장소에서 음악을 듣는 것은 일상적인 일이 되었다. 전통적으로, 사람들은 음악의 어느 특정 부분에 관심을 갖게 되나 정작 음악의 제목은 모른다. 그러면 친구에게 묻거나, 레코드 가게 점원, 혹은 라디오 방송으로부터 노래 제목을 우연히 듣게 되고 노래를 구별할 수 있도록 가수와 노래 제목을 알게 된다. 몇몇 개인들은 관련된 서비스 번호를 호출하고 라이브 '방송'으로 그들의 이동 단말기를 통해 음악을 듣는 것에 의해 음악 제목을 알 수 있도록 Shazam.com 사이트를 이용할 수 있다. 그리고 나서 이 사람은 레코드 가게로 가서 Shazam.com이나 유사한 식별 서비스에 의해 되돌려받은 식별에 기초하여 인터넷 서비스로부터 상기 노래를 구입한다. 하지만, 항상 다음과 같은 단계가 따른다. 1) 음악 샘플 칭취; 2) 음악 식별; 3) (식별에 기초하여) 구입을 위해 음악 찾기; 4) 노래 복사본 구입; 5) 구입된 복사본의 전달 및 개인 음악 컬렉션에 추가.

<6> 1)부터 3)까지의 단계는 전통적인 방법에서는 수동이다. 인터넷 기반 음악 구입은 4), 5) 단계를 자동화한다. Shazam.com 같은 서비스는 단계 1)에서 전자적 '칭취'에 기반하여 단계 2)를 자동화한다. 하지만, 상기 단계를 수행하기 위해 발명자에게 알려진 모든 선행기술 접근 방식은 노래 복사본을 식별하고 구입하기 원하는 개인에 의한 수동 단계를 필요로 한다. 본 발명은 능률을 추구하고 위의 프로세스를 좀더 자동화한다.

<7> 막대한 양의 디지털 파일들이 생성되고 저장되고 있기 때문에, 디지털 파일들이 식별되는 방식을 표준화할 필요가 발생하였다. MPEG-7은 콘텐츠 자체가 아닌, 멀티미디어 파일의 콘텐츠를 식별하는 정보 표시를 표준화하기 위해 Moving Pictures Expert Group(MPEG)에 의해 개발되고 있다. MPEG-7이 표준화된 접근 방식이 되건 그렇지 않건, 멀티미디어 파일을 효과적으로 분류, 탐색 및 검색하는 발전된 도구 세트를 기술한다. 이러한 도구는 본 발명이 어떻게 식별 및 음악 혹은 다른 미디어 구입을 자동화하는지 아래에서 설명하는 데 있어 예시적 방식으로 사용된다.

### 발명의 내용

#### 해결 하고자하는 과제

<8> 사용자가 우연히 음악과 같은 미디어 파일이 플레이 되는 시점에서 미디어 파일의 일부 특징으로 미디어 파일 전체를 획득할 수 있는 방법이나 장치가 요구된다.

**과제 해결수단**

<9> 상기의 과제를 해결하기 위한 장치로서, 미디어 샘플을 수신하도록 구성된 인터페이스; 상기 미디어 샘플의 디지털 버전으로부터 제 1 세트 특징(feature)과 제 2 세트 특징을 추출(extract)하도록 구성된 프로세서; 미디어 인식 서비스와 무선 링크를 확립하고 상기 무선 링크를 통해 상기 추출된 제 1 세트 특징을 전송하고 추가 특징 요청 수신 시에는 상기 무선 통신 링크를 통해 추출된 제 2 특징을 자동으로 전송하는 전송기를 포함하되, 상기 장치는 동시에 상기 무선 링크를 확립하고 상기 제 1 및 제 2 특징을 추출하고; 및 상기 무선 링크를 통해 요청 메시지 안에서 상기 추가 특징에 대한 요청을 수신하기 위한 수신기를 포함하는 장치가 제공된다.

**효과**

<10> 상기의 과제 해결 수단에 의하여 본 발명에서는 사용자가 예를 들어 카페에서 호감이 가는 어떤 음악을 들었지만 가수나 노래 제목을 모르는 경우에, 사용에서의 해결 방안을 제공한다.

<11> 또한 본 발명은 중단 사용자가 음악을 알아내고 장치로 직접 그 특정 음악의 복사본을 구입하는 기회를 자동적으로 얻을 수 있는 간단한 방법을 제공한다. 비록 음악과 오디오 파일에 주로 참조되어 기술되고 있으나, 본 발명은 또한 비디오 파일과 비디오 샘플, 이미지 인식과 조합된 오디오/비디오 파일 및 샘플에 적용될 수도 있다.

**발명의 실시를 위한 구체적인 내용**

<12> 본 발명은 사용자가 예를 들어 카페에서 호감이 가는 어떤 음악을 들었지만 가수나 노래 제목을 모르는 경우에, 사용에서의 해결 방안을 제공한다. 본 발명은 중단 사용자가 음악을 알아내고 이동국으로 직접 그 특정 음악의 복사본을 구입하는 기회를 자동적으로 얻을 수 있는 간단한 방법을 제공한다. 비록 음악과 오디오 파일에 주로 참조되어 기술되고 있으나, 본 발명은 또한 비디오 파일과 비디오 샘플, 이미지 인식과 조합된 오디오/비디오 파일 및 샘플에 적용될 수도 있다.

<13> 어떤 실시예에서, 본 발명은 미디어 샘플을 수신하기 위한 변환기와 같은 인터페이스, 수신된 미디어 샘플의 디지털 버전으로부터 적어도 하나의 특징을 추출하기 위한 프로세서, 및 무선 통신 링크를 통해 적어도 하나의 추출된 특징을 전송하기 위한 전송기를 포함하는 이동국(MS)이다. 변환기는 입력 신호를 다른 형태의 출력 신호를 바꾸는 (예를 들어, 마이크로폰, 디지털 카메라, 광전지 같은) 어떠한 요소도 가능하다. 수신된 미디어 파일은 선형기술에서 알려진 바와 같이 MS 내에서 디지털화되는데, 디지털 버전이라는 용어는 MS 변환기에 입력이 될 수 있는, 사람이 식별할 수 있는 미디어 샘플과 구별하기 위해 단지 사용될 뿐이다. 디지털 버전은 아날로그 미디어 샘플을 재생하는 데 사용될 수 있으나, 어떤 실시예에서는 추출된 특징과 함께 시간 종속적인(time-dependent) 특징과 관련된 하나 이상의 스펙트럴 슬라이스와 같은 디지털 버전의 세그먼트를 전송할 수 있다.

<14> 앞에서의 MS 인터페이스는 변환기일 필요는 없다. MS 인터페이스는 MS가 미디어 샘플을 받을 수 있게 하는 케이블 혹은 무선 (이동, 블루투스, 광, 라디오 방송) 링크를 포함할 수 있다. 어떤 경우에, MS는 디지털 샘플을 수신하고 그러한 수신된 미디어 샘플은 (비록 몇몇 경우에 MS가 좀더 효과적인 처리를 위해 수신된 디지털 미디어 샘플 입력을 변경할 수 있음에도 불구하고) 특징이 추출되는 미디어 샘플의 디지털 버전이다. MS가 다른 전자 장치(예를 들어, 케이블로 연결된 MP3 플레이어, 전자우편을 거쳐 연결된 다른 MS 등) 또는 MS로부터 기능적으로 별개인 통합 장치(예를 들어 통합 MP3 플레이어, 통합 디지털 라디오)로부터 미디어 샘플을 받는 경우 이러한 경우가 있을 수 있다. 또 다른 경우에, MS는 변환기를 통하지 않는 아날로그 미디어 샘플을 받고 (예를 들어 통합 아날로그 FM 라디오, 아날로그 입력 케이블 컨넥터) 이 샘플을 특징이 추출될 수 있는 디지털 버전으로 변환한다.

<15> 다른 실시예에서는, 본 발명은 이동국 내에 미디어 샘플을 처리하기 위해 컴퓨터 가독 매체 상에 구현된 컴퓨터 프로그램을 포함한다. 컴퓨터 프로그램은 기능(function)에 의해, 그리고 스페이스(space)에 의해서일 필요는 없는, 파싱(parse) 가능한 두 세트의 명령을 포함한다. 컴퓨터 명령의 제1 세트는 디지털 미디어 샘플로부터 적어도 하나의 특징을 추출할 것이다. 컴퓨터 명령의 제2 세트는 무선 통신 링크를 통해 적어도 하나의 추출된 특징을 전송할 것이다. 이동국에 관해서는, 바람직하게는 디지털 미디어 샘플의 어떠한 부분도 전송되지 않지만, 어떤 실시예에서는, 컴퓨터 명령의 제1 세트에 의해 이미 추출된 특징과 더불어 디지털 미디어 샘플의 세그먼트가 전송될 수 있다.

- <16> 다른 실시예에서는, 본 발명은 다수의 추출된 특징이 데이터베이스에 저장된 특징 세트와 정확하게 매치시키는 컴퓨터 가독 매체 상에 구현된 컴퓨터 프로그램을 포함한다. 이 컴퓨터 프로그램은 역시나 기능(function)에 의해 파싱(parse)될 수 있는 세가지 세트의 명령을 포함한다. 컴퓨터 명령의 제1 세트는 네트워크를 통해, 바로 앞선 실시예에 따라 동작하는 이동국에 의해 보내지는 메시지와 같은 수신된 특징을 포함하는 메시지를 수신한다. 컴퓨터 명령의 제2 세트는 메시지에서부터 추가적인 특징을 추출할 것이다. 이러한 추가적인 특징은 수신된 특징으로부터 추출될 수도 있고(예를 들면 디지털 미디어 샘플 그 자체로부터 직접 온 것이 아닌 고레벨 추출), 스펙트럴 슬라이스와 같은 수신된 미디어 샘플의 세그먼트와 메시지 내에서 다른 특징과 슬라이스의 연관을 구별하는 특징 혹은 양자 모두로부터 추출될 수도 있다. 컴퓨터 명령의 제3 세트는 추출된 추가적인 특징과 결합된 다수의 수신된 특징들이 데이터베이스의 단 하나의 특징과 정확히 일치할 때까지 특징 세트의 데이터베이스를 탐색할 것이다.
- <17> 본 발명의 실시예에서의 다른 측면과 이익이, 첨부된 도면과 연결된 이하의 설명을 참조하여 더욱 명확해 질 것이다. 하지만 도면은 예시적인 것으로 주어질 뿐 본 발명의 제한을 정의로서가 아님을 인식하여야 한다.
- <18> 여기서 사용된 용어 설명은 이하 상세한 설명을 더욱 명확히 할 것이다. 미디어 샘플은 해석이 이루어지는 어떤 크기라도 되는 음성, 영상 혹은 동영상 신호의 일부이다. 미디어 샘플이 마이크로폰이나 카메라를 거쳐 수신되는 어떤 경우에, 미디어 샘플은 사람이 식별할 수 있다. 아래의 설명은 예를 들어 전통적인 FM 라디오를 통해 일반적으로 흘러나오는 노래의 연속적이고 일관적인 10초 부분과 같은 노래의 시간 제한 부분 상황에서 이루어진다. 그 샘플의 특징, 혹은 바람직하게는 그것의 디지털 버전은, 샘플의 디지털 해석으로부터 수집되거나 추출될 수 있는 샘플 콘텐츠의 디지털 마커(marker), 기술자(descriptor), 혹은 다른 식별자(identifier)이다. 타임포인트(timepoint)는 샘플 내의 시간적 일순간이다. 특징은 종종 타임포인트와 관련된다. 예를 들면, 샘플에서 최대치는 오직 한 타임포인트에서만 일어나거나 반복적인 주파수 패턴은 동등한 간격을 사이에 둔 타임포인트에서 시작할 것이다. 특징적인 음향 시퀀스는 식별가능한 베이스 시퀀스의 종단으로부터 고정된 오프셋 타임(offset time)에서만 발생할 것이다. 좀더 상세한 특징 설명이 아래에 제시된다. 샘플의 다양한 특성들을 정량화하기 위해 샘플들로부터 특징들이 추출된다. 음악을 예로 들면, 음악 데이터베이스를 탐색하고 수천 혹은 수백만 노래 가운데 샘플된 곡명(혹은 버전, 가수 등)을 찾아내기 위해 노래 샘플의 특징이 추출된다. 데이터베이스가 전체 노래들로부터 사전에 추출되어진 특징을 포함하는 경우에는, 현재 샘플로부터 추출된 특징과 사전에 추출된 특징 간에 일대일 비교가 이루어질 것이다. 이러한 방식으로, 비교적 전체 중 적은 샘플만으로 노래가 찾아질 수 있다. 특징이 미디어 샘플 혹은 그 샘플의 디지털 버전으로부터 추출될 수 있는 반면에, 추가적인 특징들은 이미 추출된 특징에서 추출될 수 있다. 예를 들어, 미디어 샘플 혹은 그 샘플의 디지털 버전으로부터 추출된 제1 특징 세트는 그들간에 일정한 자동 상호연관(autocorrelation)을 보여줄 것이다. 그러한 자동 상호연관은 오직 추출된 특징의 제1 세트의 해석에 의해서만 결정된다.
- <19> 본 발명이 동작되는 통신 시스템(20)이 도 1에 나타난다. FM 라디오 같은 미디어 소스(22)는 노래의 일부와 같은 미디어 샘플(24)을 제공한다. 무선 통신 능력을 가진, 특별히 무선 전화 통신 능력을 가진 이동국 MS(26) 혹은 유사한 이동 장치는 디지털화, 처리 및 전송을 위해 미디어 샘플(24)을 수신한다. 비록 MS(26)가 노래 전체를 수신할 수 있음에도 불구하고, 미디어 샘플 혹은 그것의 디지털화된 버전은 본 발명에 따라 MS(26)가 동작하게 되는 노래의 일부이다. MS(26)는 미디어 샘플(24)의 적어도 일부의 디지털 버전을 생성하고, 그것으로부터 하나 이상의 특징을 추출하고 무선 링크(28)를 통해 통신 서버(30B)와 통신 중인 기지국(base station)(30A)을 포함하는 통신 서비스(30)로 상기 특징들을 전송한다.
- <20> 통신 서비스(30)는 기지국(30A)에서 MS(26)로부터의 호출을 수신하고 그 호출을 추적하고 MS(26)로부터의 전송을 상업 음악 인식 서비스(34)로 방향을 전환하게 한다. 통신 서비스는 MS(26)로부터 수신된 바로 그 신호를 사용되는 통신 방식에 따라 (디코드, 압출 풀기, 에러를 찾아내고 정정하는 등과 같이) 변경할 수도, 하지 않을 수도 있으나, 근본적으로 중요한 데이터, 전송된 특징을 변경하지는 않는다. 더불어, 통신 서비스(30)는 아래에 설명되는 바와 같이 MS(26)가 받을 수 있도록 단문 응답 메시지를 작성하는 것을 수행한다.
- <21> 상업 음악 인식 서비스(34)는 통신 서비스(30)를 통해 MS(26) 전송을 수신한다. 이 전송은 바람직하게는 상업 음악 인식 서비스(34)에서 좀더 상세히 해석되는 미디어 샘플(24)의 특징을 포함한다. 선택적으로, MS-추출 특징은 샘플의 다른 부분으로부터 상업 음악 인식 서비스(34)가 추가적인 특징을 추출하는 그러한 샘플의 다른 부분과 함께 전송된다. MS 내의 프로세싱 파워와 배터리 효율성이 증가함에 따라, MS(26)로부터의 전송은 상업 음악 인식 서비스(34)에 의한 더 이상의 특징 추출이 필요하지 않을 정도로 충분히 종합적인 특징 세트를 포함할 수 있다. 어떠한 경우에도 상기 전송이 MS(26)를 통해 단지 전화상으로 전체 미디어 샘플을 라이브 전송하는 것은 아니다. 그와 상관없이, 상업 음악 인식 서비스(34)는 상기 전송을 수신하고, MS(26)로부터 수신된 메시지로

부터 더 많은 관련 특징을 추출할 것이며, 이들을 상기 메시지에서 수신된 MS-추출 특징과 결합시킨다.

- <22> 음악 인식 서비스(34)와 통합되었을 수도, 아닐 수도 있는 음악 시그내처 데이터베이스(song signature database)(36)는, (MS(26)와 상업 음악 인식 서비스(34) 모두에 의해) 미디어 샘플(34)로부터 추출된 특징의 전체 세트와 비교되는 데이터베이스를 제공한다. 이 데이터베이스는 바람직하게는 각각의 수많은 파일을 위해, 기초가 되는 미디어 파일이나 그것의 샘플로부터 추출된 다수의 특징을 저장한다. 이러한 저장된 특징 세트는 음악 시그내처 데이터베이스(36)에 있는 모든 다른 파일 중에서 미디어 샘플(24)이 꺼내진 기초가 되는 파일을 고유하게 식별하기 위해 사용될 수 있다. 선택적으로, 음악 시그내처 데이터베이스(36)는 각각의 파일 혹은 곡들을 위해 여러 개의 추출된 고유하지 않은 특징과 (기준 샘플과의 조합이 기초가 되는 파일이나 노래를 고유하게 식별하는 그러한) 기준 샘플, 혹은 다른 노래나 파일로부터 데이터베이스에 있는 노래나 파일을 고유하게 식별할 수 있는 어떠한 다른 방법을 저장할 수 있다.
- <23> 음악 인식 서비스(34)는 원래의 샘플과 음악 시그내처 데이터베이스(36)에 저장된 샘플을 비교하고 음악 시그내처 데이터베이스(36)에서 MS(26)에서 전송된 특징과 일치하는 것만을 찾기 위해 탐색한다. MS(26)와 인식 서비스(36) 각각이 특징을 추출하는 데 있어서, 음악 인식 서비스(34)는 두 가지 별개의 기능을 수행한다. 추출된 특징을 사용하여 데이터베이스를 탐색하는 것과 전송된 미디어 샘플의 일부로부터 더 많은 특징을 추출하는 것이다. 바람직하게는, 인식 서비스(34)는 이 두 기능을 병렬적으로 수행한다. 음악 인식 서비스(34)가 초기에 MS(26)에 의해 사전에 추출된 특징만을 사용하여 시그내처 데이터베이스(36)를 탐색할 때, 특정 노래가 찾아지지 않을 수 있지만 이러한 초기 탐색에 사용된 특징의 수에 따라 시그내처 데이터베이스(36)에서 잠재적으로 일치하는 것들은 상당히 작은 서브세트(subset)로 줄어들 수 있다. 동시에, 인식 서비스(34)는 MS(26)에 의해 보내진 미디어 샘플(24)의 부분으로부터 추가적인 특징을 추출한다. 각각의 추가적인 특징이 추출됨에 따라, 잠재적으로 일치하는 서브세트는 오직 한 개의 일치만이 찾아질 때까지 줄어든다. 그 시점에서, 인식 서비스(34)는 추가적 추출을 종료하고, 시그내처 데이터베이스(36)의 미디어 파일로 미디어 샘플(24)을 일치시키기 위해 필요한 전체 프로세스가 최소화된다.
- <24> 선택적으로, 인식 서비스(34)는 대신 일단 MS(26)로부터 전송된 메시지에서 추가적인 특징을 추출하고, MS(26)로부터 수신된 메시지와 추출된 특징을 컴파일하고, 특정 일치를 찾기 위해 전체 시그내처 데이터베이스(36)의 탐색을 오직 한번 수행할 수도 있다. 선택은 사용자에게 응답성(데이터베이스 탐색 시간)과 많은 MS로부터의 수많은 동시다발적 메시지에서 추가적인 특징을 추출하기 위해 인식 서비스에서 이용가능한 프로세싱 파워 간의 시장 주도적 밸런스에 달려있다.
- <25> 일단 정확한 일치가 발견되면, 인식 서비스(34)는 샘플 식별 메시지(만일 일치된 것이 발견되지 않으면 식별된 것이 없음 메시지)를 통신 서비스(30)를 통해 MS(26)로 보낸다. MS(26)는 사용자에게 노래 이름 혹은 다른 미디어 파일을 알려주기 위해 디스플레이 사용자 인터페이스 상에 곡명과 가수를 표시한다.
- <26> 바람직하게는, 본 발명은 단지 노래나 미디어 파일을 식별하는 것을 넘어서 MS(26)에서 고객에게 다운로드되어 식별된 미디어 파일의 복사본을 사용자가 구입하도록 하는 링크도 자동적으로 제공한다. 선택적으로, 본 발명은 MS(26)에게 제 1 복사본을 제공하고, 개인 컴퓨터(44) 같은 다른 장치에 제 2 복사본을 제공하도록 구성될 수 있는데, 각각의 복사본은 다운로드 상 이용될 수 있는 링크 28, 46 형식에 맞추어진다. 이는 제1 복사본이 MS(26)에서 (대역폭을 고려하는 것과 함께) 좀더 제한된 저장 및 사운드 재생 능력을 위해 최적화하는 코덱으로 압축된 식별된 미디어 파일일 수 있고, 동일한 기초가 되는 미디어 파일 제 2 복사본이 PC(44)로의 링크(46)에서 사용가능한 더 큰 대역폭을 위해 최적화하는 제 2 코덱으로 압축될 수 있다는 면에서 이점이 있다. 고객은 그리고 난 후 제2 복사본을 높은 재생 충실도의 이동 음악을 위한 Rio® 혹은 iPod® 같은 전용 이동 음악 장치로 업로드할 수 있다. 이러한 두 복사본 다운로드 대안에 대한 상세한 기술은 공통으로 소유되고 2004년 3월 2일 출원된 미국 특허출원번호 10/792,547의 주제이기도 하고 본원에 참조로 결합되어 있다.
- <27> 사용자에게 식별된 미디어 파일 복사본을 제공하는 것은 다운로드되어야 하는 실제 파일이나 노래를 저장하는 파일 혹은 음악 데이터베이스(40)와의 협력을 필요로 한다. 시그내처 데이터베이스(36)이 (일치하는 것이 찾아지면 MS(26)로 보내져야 되는 곡명/작곡가와 같은) 특징 및 파일/노래 식별을 저장하는데 있어, 파일/노래 저장 데이터베이스(40)는 시그내처 데이터베이스(36)에서 일치된 특징과 일치하는 실제의 파일/노래를 저장한다. 두 데이터베이스(36, 40)는 비록 꼭 그럴 필요는 없으나 하나로 합쳐질 수도 있다. 시그내처 데이터베이스(36)로부터의 식별은 구별된 파일/음악 데이터베이스(40)로부터 특정 파일/노래를 손쉽게 선택하는데 사용될 수 있다. 후자의 데이터베이스(40)으로부터의 파일/노래는 본 발명이 앞서서처럼 바람직하게는 파일 압축 이후 MS(26)의 사용자에게 노래의 복사본과 그 식별을 제공하도록 확장된 때에 사용자에게 다운로드되는 것이다.

- <28> 노래가 그러한 식으로 다운로드 된 때에, 바람직하게는 네트워크(32)에서 음악 데이터베이스(40)에 연결된 음악 서비스(38)는 MS(26)로부터 요청을 수신하고 요청된 파일(노래, 이미지, 텍스트 등) 및 파일을 기술된 형식(스타일, 템플릿, 스크립트 등)으로 제시하기 위한 메타데이터로 응답한다. 음악 서비스(38)는 음악 저장 데이터베이스(40)로부터 데이터를 문의한다. MS(26)가 통신 서비스(30)를 통해 네트워크(32)에 연결되어 있으므로, MS(26)로 가는 도중에 그곳을 거쳐가게 된다.
- <29> 다운로드 서비스(42)는 음악 서비스(38)를 통해 MS(26)에서 시작된 모든 개인적인 다운로드 트랜잭션을 관리하기 위해 포함될 수 있다. 다운로드 서비스(42)는 저작권 제한과 부합되고 금액청구를 하기 위해 저장 데이터베이스(40)로부터 파일 각각의 다운로드를 추적한다. 본 발명은 버튼이나 소프트 키를 한 번 누름과 같이 MS(26)의 사용자 인터페이스 UI에서 한 번의 입력으로 자동적으로 실행되는 종단간(end-to-end) 트랜잭션을 피한다. 이러한 종단간 실시예에서, 단일 입력은 미디어 샘플의 포착, 특징 추출 및 MS(26)에서 호출 확립 그리고 MS(26)로부터 추출된 특징과 미디어 샘플 부분의 전송을 개시한다. 노래는 (어떤 것은 MS(26)에 의해 추출되고, 어떤 것은 인식 서비스(34)에 의해 추출된) 추출된 특징에 의해 시그내처 데이터베이스(36)를 사용하여 식별되고, 음악 서비스(38) 혹은 인식 서비스(34) 중 하나는 MS(26)로 MS 사용자에게 노래를 식별하는(예를 들어 곡명 및 가수) 메시지를 보낸다.
- <30> 메시지는 또한 음악 서비스(38)로의 링크를 제공하여 사용자가 바람직하게는 MS(26)에서 한번 더 단일 입력으로 MS(26)로 노래를 다운로드 하도록 요청할 수 있다. 음악 서비스는 (음악 데이터베이스(40)로부터 노래를 디지털적으로 선택하도록 사용되는 바와 같은 곡명/가수일 수도 혹은 아닐 수도 있는)노래 식별자를 저장하거나 그 식별자를 다운로드 서비스(42)로 전송하여 다운로드 서비스(42)가 그것을 저장한다. MS(26)가 노래 다운로드를 요청하는 때에, 음악 서비스(38)는 음악 데이터베이스(40)로부터 일치하는 노래를 선택하고, 통신 서비스 서버(30B)를 통해 MS(26)의 사용자에게 금액청구를 준비하는 다운로드 서비스(42)에게 통보하고, 음악 서비스(38)는 MS(26)에 다운로드 되도록 음악 데이터베이스(40)로부터 식별된 노래를 제공한다. 사용자는 식별 메시지에서 제공된 링크를 이용하여 MS(26)로부터 직접 음악 서비스(38)를 액세스할 수 있거나 혹은 현재 실행 상 무선 이동 전화 링크(28)같이 대역폭에서 일반적으로 그렇게 제한적이지 않은 광대역 폭 혹은 다른 PC 링크를 통해 노래가 다운로드될 것이기 때문에 사용자는 더 높은 재생 충실도 버전(다른 압축 코덱)의 음악을 얻을 수 있도록 PC(44)로부터 링크를 사용할 수 있다.
- <31> 도 2a 내지 2b는 통신 시스템(20)의 분배된 성분 가운데서 미디어 샘플의 해석을 광범위하게 설명한다. 도 2a는 블럭도에서 MS(26) 내의 동작을 보여준다. 미디어 샘플(24)은 MS(26)로의 다른 어떤 입력처럼 디지털화되고, MS(26)는 음악 인식 서비스(34)에 의해 동작되는 것과 같은 서버로의 (예를 들면 무선 링크(28) 및 네트워크(32)를 통해서와 같이) 데이터 접속을 확립한다. MS(26)는 디지털화된 오디오 입력 신호(201) 전처리를 시작하고 그 신호로부터 특징을 추출하는데, 상기 입력 신호는 추후에 데이터 패킷으로 서버에 보내질 것이다. 선택적으로 MS(26)는 전처리(202)가 프로세스를 시작하기 위한 인식키 누름 혹은 다른 사용자 입력에 앞선 시점에서 입력이 되는 신호에서 시작할 수 있도록 지속적으로 오디오 입력을 버퍼링해온다. 대부분의 경우 사용자는 인식키 누름 이전에 어느 정도 노래를 이미 듣고 있다.
- <32> 인식을 위한 특징은 예를 들면 스펙트럴 편평성(spectral flatness), 스펙트럴 센트로이드, 리듬 및/또는 음정일 수 있다. 특징들은 파일을 분류하고 식별할 수 있도록 미디어 파일이나 샘플의 내용을 특징지우지만, 이들은 MS(26)로의 입력이 되는 (아날로그) 미디어 샘플(24)의 단순한 디지털 재형상화는 아니다. 비록 MS(26)가 코덱더를 사용함에 의해서와 같이 입력 미디어 샘플(24)을 디지털화하지만 그것만으로는 특징 추출이 아니다. 바람직하게는, 특징들은 재구성할 수 있는 것이 아니다. 즉, 원래의 미디어 샘플은 단지 재처리 혹은 샘플로부터 추출된 특징들을 숨씨있게 조작해서 재구성될 수 있는 것은 아니다. 재구성 불가 특징들은 미디어 파일 콘텐츠를 기술하고 식별하지만 콘텐츠를 재생성하지는 못한다. 모든 특징들이 재구성 불가일 필요는 없지만, MS(26)에서 재구성 불가 특징을 추출하는 것은 항상 그렇지는 않다고 하더라도 대부분의 경우 유선 링크에 넘겨주기 위해 재구성 특징보다 더 작은 패킷화된 메시지를 만들어내게 된다.
- <33> 적절한 특징들이 예를 들면 참조로 결합되어 있는 MPEG-7 표준(ISO/IEC 15938, 여기에서 오디오 관련 실시예를 위해 특별히 ISO/IEC 15938-4, INFORMATION TECHNOLOGY - MULTIMEDIA CONTENT DESCRIPTION INTERFACE - PART 4: AUDIO)에 설명되었다. MPEG-7에 관련된 문헌들은 본 발명에 따라 MS(26)에 의해 수행되는 바와 같이 미디어 샘플로부터 직접 추출되는 특징을 표시하기 위해 용어 기술자(term descriptor)를 사용한다. MPEG-7 문헌은 기술자(descriptor)로부터 추출된 특징을 기술하기 위해 기술 서술 체계(description scheme, DS) 혹은 기술틀(description tool) 용어를 사용한다. 기술자(descriptor)가 미디어 샘플 디지털 버전의 저레벨 해석인 경우, 서술 체계(description scheme)는 (즉, 샘플로부터 직접 나온 것이 아닌 기술자(descriptor) 간의 상호 관계로

부터 나온) 고레벨 해석이다. MPEG-7는 현재 개선 중에 있고 이하는 현재 상기 표준에서 계획된 특정 기술자(descriptor) 및 기술 체계의 요약이다.

- <34> 기술자(descriptor)는 신호 크기의 통계학적 모델, 신호의 기본 주파수, 신호에 존재하는 소스 수의 계산, 스펙트럼 틸트(tilt), 감성적 콘텐츠, 외부 음향 효과 모델 그리고 모든 경우의 구체적인 혹은 추상적인 특징과 같은 저레벨 특징, 동영상 콘텐츠의 기본적인 품질을 나타낸다. 본 발명의 바람직한 실시예에서는 MS(26)가 기술자(descriptor)를 추출한다.
- <35> 서술 체계(DS, description scheme)는 기술자(descriptor)의 구조적 조합이다. 이 구조는 문서의 구조를 직접적으로 표현하기 위해 문서에 주석을 달거나 혹은 고레벨 개념의 더욱 풍성한 표현을 형성하는 특징의 조합을 생성하기 위해 사용된다. 예를 들면, 클래식 음악 DS는 소나타 형태의 음악 구조를 (예외를 허용하면서) 암호화한다. 다양한 스펙트럼 및 시간 기술자(descriptor)는 음색이나 짧은 음향 효과를 기술하기 위해 적절한 DS를 형성하도록 결합된다. 본 발명의 바람직한 실시예에서, 네트워크(32) 상의 서버는 DS를 추출한다.
- <36> 기술자(descriptor)는 미디어 샘플 콘텐츠의 특정한 특성 표시의 구성과 의미를 정의하는 것으로 여겨질 수 있다. 예를 들면, 이미지 파일의 색은 하나의 특성(characteristic)이다. 기술자(descriptor)가 둘 중에 좀더 기본이고 특정한 특성이 여러 개의 기술자(descriptor)에 의해서만 완벽하게 설명될 수 있을 때에, 특징은 기술자(descriptor) 혹은 특성일 것이다. '컬러' 특성과 연관된 가능한 기술자(descriptor)는 컬러 히스토그램, RGB 벡터 혹은 스트링을 포함한다. MS(26)는 주어진 값 세트에 대한 기술자(descriptor)의 표시, 특정 기술자(descriptor)에 대한 값을 추출한다. 예를 들면, RGB = (255, 255, 255), 컬러 스트링 = "red"이다. 어떤 기술자(descriptor)는 어떤 종류의 미디어 파일과 관련되거나 유효하지만 다른 것은 그렇지 않다. 예를 들면, 시각적 미디어 샘플에 대한 기술자(descriptor)는 기본적 구조적 특성 내에서는 격자형 레이아웃과 히스토그램; 컬러 특성 내에서는 컬러 스페이스, 주요 컬러, 컬러 히스토그램과 컬러 양자화; 질감 특성 내에서는 공간적 이미지 밀도 분포와 균질 질감; 형상 특성 내에서는 물체 경계 박스, 영역-기반 형상, 윤곽-기반 형상 및 3D형상 기술자; 동작 특성 내에서는 카메라 모션, 오브젝트 모션 궤도, 파라미터적 오브젝트 모션, 모션 액티비티 및 모션 궤도 양상(속도, 방향, 가속도)을 포함할 수 있을 것이다. 오디오 샘플에 관련된 기술자(descriptor)는 음색 주해 특성에서는 단어와 음운의 격자와 메타데이터; 음질 특징에서는 우수(even) 하모닉에 대한 기수(odd) 하모닉 비율 및 하모닉 어택 간섭성(harmonic attack coherence); 멜로디 특성에서는 콘투어(contour) 및 리듬을 포함할 수 있다.
- <37> 현재 MPEG-7에는 오디오 샘플을 기술하기 위한 17가지 시간적(temporal) 그리고 공간적(spatial) 기술자(descriptor)가 있다. 이들은 기본(basic), 기본 스펙트럴(basic spectral), 신호 파라미터(signal parameters), 음색 시간(timbral temporal), 음색 스펙트럴(timbral spectral), 스펙트럴 기저(spectral basis)들이다. 더불어, 휴지(休止) 기술자(silence descriptor)는 오디오 샘플의 콘텐츠를 식별하는 데 있어 상당히 유용한 것으로 증명되었다. 두 가지 기본 오디오 기술자(descriptor)는 모든 종류의 오디오 샘플에 적용가능한, 일반적 사용을 위해 시간적으로 샘플링된 스칼라 값들이다. 파형 기술자(descriptor)는 전형적으로는 디스플레이 목적 상 (최소 및 최대) 오디오 파형 엔벨로프(envelope)를 기술한다. 파워 기술자는 시간적으로 평탄화된 순간 파워인데, 신호 및 샘플의 핵심 개요로서 유용하고, 파워 스펙트럼과 연관되어 있다.
- <38> 4가지 기본적 스펙트럼 오디오 기술자는 모두가 오디오 신호 또는 샘플의 단일 시간-주파수 해석으로부터 유도되는 기본적인 공통 기반을 갖는다. 이들 모두는 제 1 기술자인 오디오 스펙트럼 엔벨로프 기술자(AudioSpectrumEnvelope Descriptor)에 의해 제공되는데, 제 1 기술자는 로그-주파수 스펙트럼이고, 이차 제수 혹은 다중 옥타브(octave)에 의해 구분된다. 이 오디오 스펙트럼 엔벨로프는 오디오 샘플의 단기간 파워 스펙트럼을 기술하는 벡터이다. 이것은 스펙트로그램(분광 사진)을 디스플레이한다든가, 데이터의 정제되지 않은 청각화(auralization)를 합성하는데 사용되거나, 탐색과 비교를 위한 일반적 목적의 기술자로서 사용된다. 다른 스펙트럼 오디오 기술자는 로그-주파수 파워 스펙트럼의 중심(중심(centroid) 혹은 형상), 중심(centroid) 주변 스펙트럼 분포, 여러 가지 주파수 밴드나 빈(bin)에 대한 스펙트럼 편평도를 표시한다.
- <39> 두가지 신호 파라미터 기술자는 주로 주기적 혹은 반주기적 신호에 적용된다. 이들은 오디오 샘플의 기본 주파수(신뢰도(confidence measure))와 (예를 들면 음악 톤과 목소리 연설, 금속성음과 벨소리, 'f' 음과 같은 마찰음과 악기들의 조밀한 혼합들을 구별하기 위한) 하모닉을 기술한다.
- <40> 두가지 음색 시간 기술자는 사운드 세그먼트의 시간적 특성을 기술하고, 특별히 음악적 음색, 피치(pitch) 및 소리 세기와 독립된 톤의 질(質)에 유용하다. 한 기술자는 사운드의 "어택(attack)" 특성(신호가 휴지(休止) 상태에서 최대치로 상승하는데 걸리는 시간)을 기술화하고, 다른 기술자는 신호가 어느 정도 시간 안에 어디로 신

호가 집중되는지를 나타내는 신호 엔벨로프 혹은 센트로이드 특성을 기술화한다. 예를 들면, 후자의 기술자는 두 음조의 길이와 어택이 동등할 때 감쇠하는 피아노 음조와 유지되는 오르간 음조를 구별할 수 있다.

- <41> 다섯 가지 스펙트럴 기술자는 특별히 음악 음색 인지에 적용가능한 선형 주파수 공간에서의 스펙트럴 특징이다. 하나는 선형 파워 스펙트럼에서 빈의 주파수 지수 가중 평균(power-weighted average)이고, 이는 앞에서 명시된 센트로이드 기술자와 유사하지만 다른 악기에 대한 사운드의 "높은 가락"을 구별한다. 나머지 음색 스펙트럴 기술자는 신호나 샘플의 하모닉 일정 간격 컴포넌트 상에서 동작한다. 이러한 이유로, 기술자(descriptor)는 선형 주파수 공간에서 계산되고, 스펙트럴 센트로이드, 스펙트럴 편차 및 스펙트럴 확산 기술자를 포함한다.
- <42> 두가지 스펙트럴 기저 기술자는 간결성과 인식을 돕기 위해 고차원 스펙트럴 공간의 저차원 투사(投射)를 나타낸다. 그러한 한가지 기술자는 정규화된 파워 스펙트럼의 특이값 분해(singular value decomposition)로부터 유도되는 일련의 (잠재적으로 시변(time-varying) 및/또는 통계적으로 독립된) 기본 함수이다. 다른 하나는 (앞에 것과 결합된) 감소된 랭크 기저(rank basis)로의 투사 이후 스펙트럼의 저차원 특징을 나타낸다. 휴지(休止) 세그먼트는 "휴지(silence)"의 단순 의미를 오디오 세그먼트에 단순히 덧붙인다. 비록 이것이 아주 단순하기는 하지만, 매우 효과적인 기술자이다. 이것은 오디오 스트림의 추가적 분할을 돕는데 사용되거나 분할 프로세스를 하지 않도록 하는 힌트로 사용될 수 있다.
- <43> 더불어, 기술자들은 스펙트로그램의 독립적인 서브스페이스를 관찰하고 간결하게 표시하는데 사용될 수 있다. 종종 이러한 독립 서브스페이스(혹은 그것의 그룹들)는 다른 사운드 소스와 강하게 상호연관된다. 그래서 더 적은 스페이스를 사용하면서도 스펙트로그램으로부터 좀더 돌출된 특징과 구조를 얻는다.
- <44> 서술 체계(Description Schemes : DS)는 다수개의 주파수 빈(bins) 혹은 스펙트럴 슬라이스(slice)에 대해 동일 종류의 기술자 중에 혹은 기술자들 간에 상호연관성을 반영한다. 그러한 방법으로, DS는 기술(記述)적 풍족함을 위해 어느 정도의 보편성을 희생한다. 일단 MS(26)에서 샘플이나 샘플의 세그먼트로부터 충분한 수의 기술자가 추출되면, 디지털화된 오디오 샘플을 전송할 필요가 없다. 인식 서비스(34)는 단지 그것으로부터 추가적인 특징을 추출할 수 있는 추출된 특징만 있으면 된다. 본 발명의 어떤 실시예에서 이러한 추가적인 특징들은 MPEG-7의 DSs와 유사하다. 예를 들면, 오디오 시그내처 DS는 스펙트럴 편평도 기술자를 통계적으로 간략화한다. 이는 오디오 샘플의 강건성(robust) 자동 식별 목적을 위한 고유한 콘텐츠 식별자를 제공할 수 있다.
- <45> 음색 DSs는 악기 사운드의 지각적 특징을 기술하는데 목적을 두고 있다. 음색은 동일한 피치와 소리 세기 사운드를 가진 두 사운드가 다르게 만드는 지각적 특징이다. 음색 DSs는, 사운드의 "어택(attack)," "선명도(brightness)," "풍요로움(richness)" 과 같은 관념, 특정 악기 혹은 악기 그룹(타악기) 등에 대한 하모닉, 간섭적인, 지속적인 사운드 및 비지속적 사운드와 관련될 수 있는 감소된 기술자 세트로 이러한 지각적 특징을 기술한다. 다른 음색 DS는 음색 시간 기술자를 거리 측량을 이용하여 스펙트럴 센트로이드와 비교한다.
- <46> 멜로디 DSs는 효율적이고, 강건(robust)하며, 표현적인 멜로디 유사성 매칭(matching)을 구현하기 위한 단(單)선용 멜로디 정보를 위해 풍부한 표현을 포함하고, 극도의 간결함에 대해 멜로디 콘투어(contour) DS(5-단계 콘투어는 인터벌이 대(大) 혹은 소(小) 인터벌, 위, 아래 혹은 동일로 양자화된 경우에 인접한 음(音)들 간의 인터벌 차이를 나타낸다), 효율적인 멜로디 콘투어 표시 및 좀더 장황하고, 완전하며, 표현적인 멜로디 표시를 위한 멜로디 시퀀스 DS(확장된 기술자 세트이고 고정밀 피치 인터벌 인코딩)를 포함한다. 어느 하나 혹은 양자는 멜로디에 대한 보완 정보를 포함하도록 확장될 수 있다. 애플리케이션에 의해 원하는 바대로 사용되도록 가사, 키, 박자 및 개시음과 같은 일련의 선택적 보완 기술자들이 이러한 핵심 기술자 주변에 배열된다.
- <47> 몇몇 "인식(recognition)" DSs는 근원적인 미디어 파일을 탐색하고 색인하는데 있어 보탬이 되고, 히든 마르코프(hidden Markov) 혹은 가우시안 혼합 모델과 같이 그것으로부터 통계학적 모델이 형성된 토대(foundation)로서 저레벨 스펙트럴 기본 기술자를 사용한다. 결과적인 개연적(probabilistic) 분류자는 연설과 음악과 같은 폭넓은 사운드 클래스를 인식할 수 있거나, 남성, 여성, 트럼펫 혹은 바이올린과 같은 좀더 좁은 카테고리들을 식별하도록 훈련될 수 있을 것이다. 다른 애플리케이션은 장르 분류와 목소리 인식을 포함한다. 추가적인 DSs는 오디오 스트림 내에서 발성(spoken)된 콘텐츠를 상술한다.
- <48> 추출될 수 있는 특징에 관한 더 자세한 것은 본 원에 참조로 결합된 2002년 6월 27일에 공개된 미국 특허출원 번호 US 2002/0083060에서 찾을 수 있다. MS(26)에서의 전처리(202)의 양은 그 범위가 전혀 없는 것에서부터 특징 벡터의 완전한 추출까지 변동한다. 각각의 특징에 대한 전체 특징 추출은 바람직하게는 여러 단계로 나누어 지는데, 그 중 몇몇은 비(非)추출 신호(예를 들면 MS(26)에서 수신된 '라이브' 음악 샘플(24))에 비교하여 나중에 보내진 데이터의 양을 줄여준다. 예를 들면, 스펙트럴 특징에 대한 첫번째 단계는 순간 자기상관

(autocorrelation) 벡터를 계산하고 일정 주기 동안 이들을 평균함으로써 디지털화된 입력 신호(201)의 자기상관(autocorrelation) 추정치가 될 것이다. 다음 단계는 푸리에 변환 및 추가적인 프로세스를 수행할 수 있다. 전체적인 특징 추출 프로세스는 MS(26)와 서버(34) 간에 나누어진다. MS(26)는 수행하는 단계의 갯수를 선택하고, 전처리의 결과(MS 추출 특징)와 함께, 사이드 정보로서 수행된 단계의 수 M을 패킷화(203)한다. 바람직하게는, MS(26)는 미디어 샘플(오디오, 비디오, 이미지 등) 및/또는 채널 특징을 통해 추출된 특징을 전송할, 채널 특징의 타입(type)에 기초하는 추출 특징의 종류 및/또는 상기 수(number M)를 순응적으로 선택한다. MS(26)에 의해 아무런 특징도 추출되지 않은 미디어 샘플(24)의 세그먼트 역시 인식 서비스(34)에서 더 상세한 분석을 위해 위에서와 같이 MS(26)로부터 전송될 것이다. 패킷은 전형적인 무선 프로토콜(컨스텔레이션 맵핑, 에러 정정 등)에 의거해 암호화된다. 액세스 네트워크(32) 및 MS(26)의 성능에 기초하여 프로세싱 파워와 전송 대역간의 적절한 밸런스가 선택된다. 암호화된 패킷은 하나 이상의 채널을 거쳐 무선 링크(28)를 통해 전송된다.

<49> 도 2b에서 보는 바와 같이, 패킷이 수신되고 복호화된다(205). 무선 링크(28)에 특화된 복호화는 통신 서버(30B)에 의해 이루어지고 개별 패킷의 오프닝(206)은 단 대 단(end-to-end) 인터넷전화 합의에서 음악 인식 서비스(34)에 의해 이루어질 수 있다. 음악 인식 서비스 서버(34)는 노래 식별을 위해 요구되는 추가적인 파라미터를 추출하는데(207), 이러한 추가적인 파라미터는 MS(26)로부터 수신된 특징 및/또는 MS(26)로부터 수신된 디지털화된 미디어 샘플의 세그먼트로부터 추출될 수 있다. 도 2a 내지 2b의 예에서, 총  $N > M$  특징이, 미디어 샘플의 궁극적인 식별이 가능하도록 추출된다. M이 더 낮은 레벨의 특징을 우선적으로 나타내고 N-M이 미디어 샘플의 디지털화된 버전으로부터가 아닌 더 낮은 레벨의 특징으로부터 추출되는 더 높은 레벨의 특징을 나타내는 경우에, 많은 M 특징이 MS(26)에 의해 추출되고, 남은 N-M특징이 서버(34)에 의해 추출된다. 음악 인식 서비스 서버(34)는, 각각 근원적인 미디어 파일을 대표하고 음악 시그내처 데이터베이스(36)에 저장된 특징 세트와 추출된 특징 간에 비교(208)를 행한다. 이상적으로는, 한 특징과 데이터베이스 내의 한 특징 세트만이 추출된 특징 그룹과 일치할 것이고 일치하는 특징 세트는 미디어 샘플이 취해진 파일을 대표할 것이다. 추출된 특징에 기초하여, 미디어 샘플(24)은 음악 시그내처 데이터베이스(36)에 의해 제공된 근원 파일에 대한 특징 벡터 혹은 특징 세트에 대한 N개의 추출 특징을 비교함으로써 인식된다.

<50> 도 3a 내지 3d는 본 발명의 실시예에 따라 흐름도 형태로 다양한 단계 혹은 바람직한 방법 실시를 예시하는데, 각각의 흐름도는 보이는 것과 같이 다른 곳에 연결되어 있다. 도 3a는 미디어 샘플(24)이 초기에 수신되고 처리될 때에 MS(26) 내에서의 방법을 기술한다. 이 실시예에서, 미디어 샘플은 두 세그먼트로 구문분석(parse)되고, 버퍼에 저장된다. 이 저장은 어떤 사용자 입력이 MS에게 가르침에 따라 해석을 수행하던지 혹은 입력을 처리하라고 알려주지 않는 한 계속된다. 사용자가 인식키(304)를 누르면 MS(26)로 하여금 바람직하게는 동시에 다음 기능들을 수행하도록 트리거시킨다. MS(26)는 기지국(30A)로의 무선 링크(28)를 확립하고, MS(26)는 버퍼링된 제1 미디어 샘플(24)의 세그먼트로부터 낮은 레벨의 특징을 추출(308)하며 MS(26)는 미디어 샘플의 제 2 세그먼트를 수신하고, 디지털화 하며 그것으로부터 낮은 레벨 특징을 추출(310)한다. 제 2 세그먼트는 잠시간 저장될 것이다. MS(26)는 추출된 특징(및 앞에서 명시한 바와 같이 만일 적용 가능하다면 특정 특징을 추출하는데 있어 MS(26)에 의해 수행된 단계의 수)을 306 단계에서 확립된 무선 링크(28)를 통해 전송(312)한다. 선택적으로, 특징들은 제 1 세그먼트, 혹은 제 2 세그먼트로부터 추출되고, 추출된 특징들은 제 2 세그먼트와 함께 전송되는데, 상기 제 2 세그먼트는 단지 디지털화된 미디어 샘플의 스펙트럴 슬라이스 혹은 일련의 스펙트럴 슬라이스일 것이다. 방법 단계는 도 3b에서 계속된다(314).

<51> 도 3b는 시스템(20)의 네트워크(32) 측면에서의 단계를 상술한다. 서버(34)는 MS(26)로부터 도 3a의 단계 312에서 전송된 메시지를 수신(316)한다. 통신 서버(30B)는 패킷화된 메시지의 무선 특정 부분을 복호화할 것이고, 중요 데이터를 운반하는 개방되지 않은 패킷을 네트워크(32)를 통해 단지 우선적으로 재전송할 것이다. 음악 인식 서비스(34)는 패킷을 개방하고, 이미 추출된 저레벨 특징을 수신하고, 정확한 일치(unique match)(320)를 위해 MS(26)로부터 받은 저레벨 특징과 일치하는 특징 세트를 가지는 파일을 찾기 위해 시그내처 데이터베이스(36)를 검색(318)한다. 그 검색 결과는 일시적으로 저장된다. 오직 MS-추출 특징을 사용하여서는 정확한 일치가 찾아지지 않는다면, (블럭 316에서 일시적으로 저장된)인덱스 K가 최대값에 견주어 비교되고, 만일 넘지 않았다면, 하나 이상의 상위 레벨 특징이 블럭 326에서 수신된 하위 레벨 특징으로부터 추출된다. 블럭 328에서는, (블럭 318에서 일시적으로 저장되고) MS-추출 특징과 일치하는, 데이터베이스로부터 나온 특징 세트가 K번째 서버-추출 특징을 사용하여 검색되며, 그 결과 전체 데이터베이스(36)는 현재의 검색 수행 반복에서 다시 검색될 필요가 없게 된다. K번째 검색의 결과는 블럭 322에 일시적으로 저장되고, K는 블럭 330에서 인덱스되며, 만일 320 블럭에서 정확한 일치가 또다시 찾아지지 않는다면, K가 최대값을 넘어서 "일치하는 것 없음" 메시지가 MS로 보내질(334) 때까지 추가적인 추출 특징을 위해 루프를 반복한다. 일단 정확한 일치가 블럭 320에서 발견되면, 음악(노래)의 제목(혹은 다른 식별자), 혹은 추출된 특징과 일치하는 데이터베이스 특징 세트와 유일무이하

게 일치하는 파일을 담고 있는 응답 메시지(336)가 MS(26)로 보내진다. 응답 메시지(336)는 또한 일치되는 파일로의 직접 연결되는 링크를 우선적으로 운반하는데, 일치되는 파일은 시그내처 데이터베이스(36)와 함께 있을 수도 있고 그렇지 않을 수도 있는 음악 또는 파일 저장 데이터베이스(40)에 있게 된다.

<52> MS(26)와 음악 인식 서비스(34)간의 통신 링크는 바람직하게는 MS(26)가 일단 그것의 추출된 특징을 전송하는 때와 인식 서비스(34)가 고유 파일 식별자를 담은 응답 메시지(336)를 전송하는 때 사이에 개방된 채로 남아 있게 된다. 한 실시예에서는, MS(26)는 추출된 특징의 제 1 세트와 함께 제 1 메시지를 전송하고, 인식 서비스(34)는 그것의 데이터베이스를 검색한다(추가적인 특징을 추출할 수도 그렇지 않을 수도 있다). 어떤 정확한 일치도 발견되지 않는 때에는, 인식 서비스(34)는 MS(26)에 요청 메시지를 보내게 되는데, 요청 메시지는 MS(26)가 추출하도록 요청하는 추가적인 특징의 수 및/또는 타입(type)을 상술할 수 있다(MS(26)는 버퍼에 저장된 미디어 샘플의 디지털 버전으로부터 추출할 수 있다). 예를 들면, 인식 서비스(34)가 데이터베이스로부터 4 개의 일치 결과를 되돌려준다고 가정해 보자. 인식 서비스는 그것이 신호 엔벨로프 혹은 센트로이드 같은 저레벨 특징이든, 음색 시간 센트로이드 간의 음색 시간 차이 같은 고레벨 특징이든 간에 4개의 특징이 특정 특징에 의해 고유하게 구별될 수 있는 그러한 하나 이상의 특징한 특징을 결정할 수 있다. 전자에서, 인식 서비스(34)는 4가지 가운데서 직접적으로 구별할 특정 스펙트럴 슬라이스에 대한 센트로이드 특징을 특별히 요청할 것이다. 후자에서는, 인식 서비스가 4가지를 구별할 (예를 들어 센트로이드 간의 거리) 추가적인 특징을 추출할 수 있도록 필요한 음색 시간 센트로이드를 요청할 것이다. MS(26)는 제 2 추출에서 추출된 제 2 특징 세트와 함께 다른 메시지를 전송하고, 인식 서비스(34)는 데이터베이스를 다시금 검색한다. 이 데이터베이스의 제 2 검색은 추출된 특징의 제2 세트를 사용할 수 있고, 제2 세트로부터 추출된 추가 특징들의 제 2 세트 혹은 이들의 조합을 사용할 수 있을 것이다. 이러한 방법으로, 데이터베이스 검색이 정확한 일치를 되돌려주는데 실패할 때마다 인식 서비스(34)가 추가적인 특징을 요청하는 MS(26)와 인식 서비스(34) 간의 '대화'는 계속될 것이다. 바람직하게는, 이들 간의 통신 링크는 개방된 채로 유지된다. 왜냐하면, 패킷 전환 네트워크는 패킷이 전송되지 않고 있기 때문에 인식 서비스가 데이터베이스를 검색하는 동안 상당한 대역폭을 쓸 필요가 없다.

<53> 바람직하게는, 음악 혹은 파일 식별과 상응하는 음악 서비스 URI는 통신 서비스 30A-B로 되돌려지는데, 통신 서비스는 MS(26)로의 응답 메시지를 작성하고 MS(26)로 SMS/MMS 혹은 유사한 형태의 메시지를 보낸다. 인식 서비스 서버(34)는 또한 (음악 서버가 인식 서비스 서버로부터 독립적인 경우) 음악 서버(38)에게 MS(26)의 고유 식별자와 일치 파일로의 링크와 함께 메시지(338)를 보낸다.

<54> 블록 342는 도 3c 음악 서버(38)로 이끈다. MS(26)가 즉각 일치되는 파일 혹은 노래의 다운로드를 요청하지 않을 것이므로, 음악 서버(38)는 도 3b의 블록 338로부터 메시지를 수신(344)하고 다운로드 관리자(346)에서 정보를 저장(346)한다. 일단 음악 서버(38)가 MS(26)로부터 요청을 수신하면(도 3d를 참조), 바람직하게는 MS(26)가 인식 서비스 서버(34)로부터 응답 메시지(336)를 수신한 바로 다음에, 음악 서버(38)는 MS(26)의 고유 식별자를 다운로드 관리자(346)에 저장된 MS ID로 일치(350)시키고 일치된 파일로의 관련 링크를 꺼내온다. 지불 정보가 (바람직하게는 MS(26)로부터 첫번째 구입 이후에 파일에 보유되어 있고 MS ID와 일치한다) 승인되고 일치된 파일 혹은 음악은 MS(26)로 다운로드된다. 구매 인보이스(invoice)는 MS(26)의 통신 서버(30B)로 전송(354)되고, 방법의 네트워크 측면에 있어서 프로세스가 종료(356)된다.

<55> 도 3b의 블록 340는 MS(26) 관점으로부터 더 진행하는 도 3d로 진행된다. MS(26)는 블록 336(도 3b)로부터 응답 메시지와 링크를 수신하고 링크와 함께 음악 혹은 파일 제목 혹은 다른 식별자를 디스플레이(360)한다. 사용자는 링크를 선택한 후에, 동시에 디스플레이되도록 사용자가 여러 개를 저장해 놓은, 그리고 도 3c의 블록 348에서 표시된 파일 혹은 음악에 대한 요청을 전송(364)하고 더불어 미리 개시된 계좌에서 인출 혹은 신용 거래에 대한 인증을 구성하는, 다운로드 키를 누른다(362). 사용자는 MS(26)에서 다운로드된 파일 혹은 음악을 수신(366)하고, 트랜잭션 측면에서 MS(26)에 대한 방법이 종료된다.

<56> MS(26)의 특정 컴포넌트는 도 4a 내지 도 4b를 참조하여 상술될 것이다. 아날로그 오디오 입력(앞에서 언급된 미디어 샘플(24))을 수신하기 위한 제1 변환기 혹은 마이크로폰(48)은 샘플을 디지털화하고 컴퓨터 가독 주기억 장치(54)에 저장된 특징을 추출을 개시하는 프로세서(50)에 연결된다. 특징 추출을 위해 사용되는 특정 알고리즘은 컴퓨터 가독 명령 형태로 기억장치(54)에 저장되어 있을 것이고, 주기억 장치(54)는 단일한 인접 저장 매체일 필요는 없고 여러 개의 다른 저장 컴포넌트 가운데 분포될 수 있다. 미디어 샘플은 프로세스 되기 전에 샘플의 한정된 데이터 볼륨을 지속적으로 저장하는 버퍼 저장 장치(52)를 통과한다. 버퍼 저장 장치(52)는 마이크로폰(48)에서 가장 최근에 수신된 신호를 저장하고 그 결과, 미디어 샘플이 사용자 명령에 추출된 특징을 가져야 할때, 프로세서(50)가 현재 버퍼 저장 장치(52) 내에 있는 미디어 샘플의 일부로부터 즉시 특징을 추출하기 시작한다. 이러한 방법으로, 음악이나 파일의 식별 혹은 링크의 확립을 위한 사용자 입력 명령에 앞서 MS로의

입력인 샘플의 일부 혹은 세그먼트로부터 특징이 추출될 수 있을 것이다. 사용자 명령이 프로세스를 요구하지 않는 한, 그리고 요구할 때까지, 추가적인 버퍼 메모리에 저장된 신호에 어떠한 프로세싱도 이루어지지 않으므로, 최소한의 전력만이 소비된다.

<57> MS(26)에서 프로세서(50)와 모든 컴포넌트는 재충전 가능한 배터리(56) 혹은 착탈식 연료 전지와 같은 휴대용 파워 소스에 의해 전력을 공급받는다. 사용자 인터페이스(58)는 프로세서의 출력으로부터 사용자가 인식할 수 있는 텍스트 메시지나 이미지로 변화하기 위한 디스플레이 인터페이스(58A), 그리고 (다수의 버튼, 디스플레이 스크린(58A) 내의 터치 혹은 압력 센서, 혹은 주 기억 장치(54)에 저장된 음성 인식 소프트웨어를 가진 마이크로폰(48)과 같은) 사용자가 프로세서(50)로 명령/요청을 입력할 수 있게 해 주는 사용자 입력 메커니즘(58B)을 포함한다. 프로세서(50)는 전송과 수신이 동시에 일어나지 않도록 스위치(64)를 통해 전송기(60)와 수신기(62)에 연결된다. 각각의 전송기(60)와 수신기(62)는 안테나(66)에 연결되어 있는데, 안테나는 프로세서를 담고 있는 본체에 내장될 수도 외장될 수도 있다. 프로세서는 제2 변화기 혹은 스피커(68)에서 사용자에게 출력을 제공할 수도 있다. 비디오 수신을 위해, MS(26)는 프로세서(50)에 비주얼 미디어 샘플을 제공하기 위한 카메라(70) 혹은 다른 이미지-캡처 장치를 포함할 수 있다. 비록 도시되지는 않았으나, 카메라(70)는 또한 마이크로폰(48)과 연관된 버퍼 메모리(52)와 유사하게 동작하는 비디오 버퍼 메모리(도시되지 않음)로 입력을 제공할 수 있다.

<58> 도 5는 MS(26)의 입력 버퍼링에서 얻을 수 있는 이득을 도시하는 타이밍 다이어그램이다. 미디어 샘플(501)은 502 시점에서 시작하는 MS(26)에서 수신된다. 사용자는 시작점 502에서 청취를 시작하고, 사용자가 복사복을 갖고 싶은 그 노래 혹은 파일이라고 결정이 되면, 사용자는 예를 들어 요청 시간에 전용 인식 버튼(86)을 누른다(504). MS(26)가 전처리를 수행하기 위해서 최소 (시간 제한) 미디어 샘플  $\Delta T$ 가 필요하다고 가정해보자. 이는 처리 시간이 아니라 전체 곡이나 파일을 식별하기 위한 충분한 특징을 추출하기 위해 필요한 (예를 들면 10초) 음악 분량이라고 할 수 있을 것이다. 버퍼(52)가 채워진 경우,  $\Delta T$ 는 사용자 요청 시간 504에 앞서서 506 시점에서 시작하고 필요한 최소 샘플 타임을 버퍼링 샘플 엔드 타임(buffered sample end time) 508까지로 늘린다. 버퍼가 없을 때는 미디어 샘플이 캡처되는 최초의 시간이 요청 타임 504가 된다. 필요한 최소 샘플 타임은 버퍼링 샘플 엔드 타임(508)보다 늦은 시점에 끝나야만 한다(510). MS(26)가 미디어 샘플(24)의 버퍼링된 세그먼트를 처리하는데 걸리는 시간의 양에 따라서, 508과 510 두 엔드 타임 간의 차이(512)는 미디어 샘플 시작 타임 506, 504 간의 차이만큼 될 것이다. MS(26)가 바람직하게는 전처리된 특징의 전송으로부터 계속적인 무선 링크(28)를 유지하기 때문에, 이 차이(512)는 지연을 줄임으로써 사용자 경험을 강화할 수 있을 뿐이다.

<59> 요약하면, 본 발명은 원활한 음악 인식과 구매 시스템 및 방법을 포함하는 음악 파일에 적용된 것으로서, 앞선 음악 인식과 구매 시스템 및 방법은 흥미를 끄는 음악을 듣는데서부터 무선으로 특정 음악 한 편을 구입하는 모든 단계를 포괄하는 한결같은 사용 체인을 구현한다. 미디어 샘플의 식별이 MS(26)의 단일 버튼으로 이루어지는데 중점을 두는 새로운 기능이 있다. 버튼을 누르면 MS(26)에서 특징 추출과 전송을 일으키고, 응답으로 궁극적으로는 미디어 샘플이 취해진 근간을 이루는 음악을 식별할 뿐만 아니라 음악 서비스에서 트랙에 대한 직접 연결 링크를 제공하는 메시지를 받게 된다. 링크를 누르게 되면 사용자를 음악 서비스로 인도하는데 이곳에서 특정 작품이 무선으로 즉시 구매되고 다운로드될 수 있다. 음악 인식 프로세스의 특징 추출 부분은 서버로 보내지는 데이터 양이 줄어들도록 제 1 단계에서 MS가 수행하도록 MS와 서버로 나누어진다. 단계의 수는 MS에 의해 선택되어질 수 있고, MS가 특징 추출을 어디에서 할 수 있을지를 알기 위해 사이트 정보로서 MS에 보내진다. 이렇게 함으로써, 특징 추출에서 미디어 샘플을 형성하기 위한 계산 부담이 MS와 서버간에 나누어진다(혹은 MS가 특정 순간에 모든 특징 추출을 수행하고 서버는 단지 이미 추출된 특징을 데이터베이스와 매치시키기 때문에 적어도 서버에서는 부담이 감소된다). 바람직하게는, 이러한 특징 추출 분배는 서버로부터 받은 피드백, 네트워크에 의 링크에서 사용가능한 채널 용량의 측정을 사용함에 의해 및/또는 식별되어야 하는 미디어 샘플의 타임(type)이나 복잡도에 대한 MS의 결정에 기초하여 MS에서 적응적으로 구현된다.

<60> MS(26)의 최종 사용자는 본인이 빠르게 끊임없이 듣는 음악을 알아내고 구매할 수 있다. 사용자는 인식된 음악 정보와 실제 음악이 구매될 수 있는 시점 간의 연결관계를 느슨하게 하지 않는다. MS에서 신호 분석(특징 추출) 부분을 수행하는 것은 서버로 보내지는 데이터 양(OTA)을 줄여주고, 그렇게 함으로써 관련된 전송 비용도 줄여준다. 현재 청구되는 발명의 바람직한 실시예가 무엇인지 예시되고 설명되었지만, 당업자에게 다양한 변경과 수정이 가해질 수 있음이 인식되어야 한다. 덧붙여진 청구항에서는 본 발명의 사상과 범위에 속하는 그러한 모든 변경과 수정을 포함하는 것으로 의도되었다.

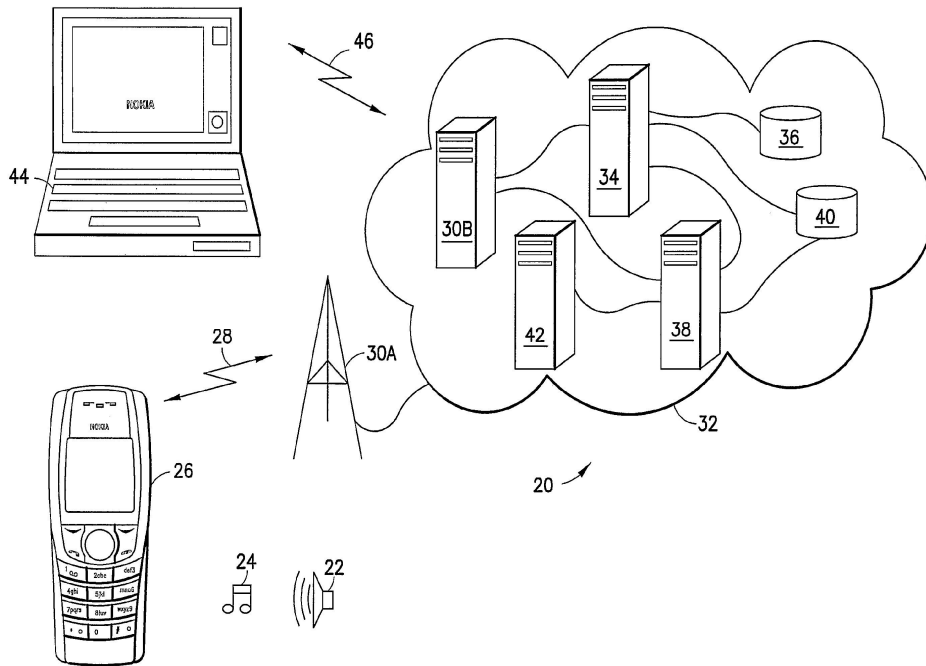
**도면의 간단한 설명**

<61> 도 1은 본 발명이 동작하는 통신 시스템의 개략도이다.

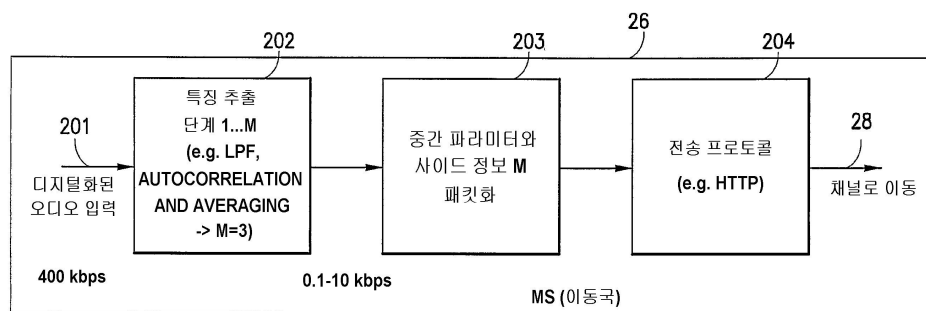
- <62> 도 2a 내지 2b는 이동국과 서버 각각에 분배된 미디어 샘플의 특징 추출을 나타내는 블록도이다.
- <63> 도 3a 내지 3d는 도 1의 통신 시스템에서 취해지는 상세한 동작을 나타내는 연속 흐름도의 일부이다. 도 3a는 이동국 내에서의 단계를 나타내고, 도 3b에서는 서버 내에서의 단계를 나타내며, 도 3c는 구매를 위한 미디어 파일 다운로드를 제공하는 서버 내에서의 단계를 나타내고, 도 3d는 서버에 응답하여 이동국에서의 추가적인 단계를 나타낸다.
- <64> 도 4는 본 발명의 가르침에 따라 특별히 구성된 이동국의 블록도이다.
- <65> 도 5는 이동국에서 입력을 연속 버퍼링하는 이익을 보여주는 시계열도이다.

도면

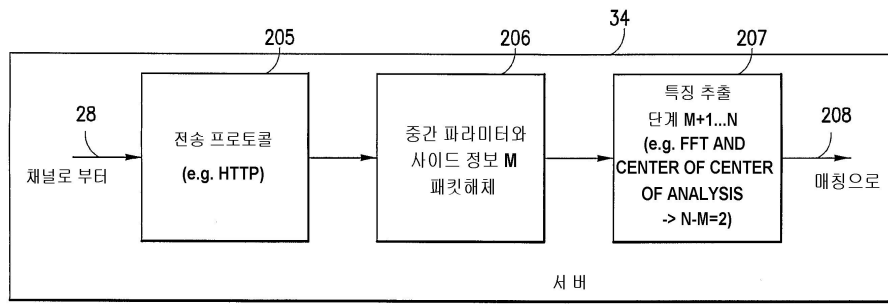
도면1



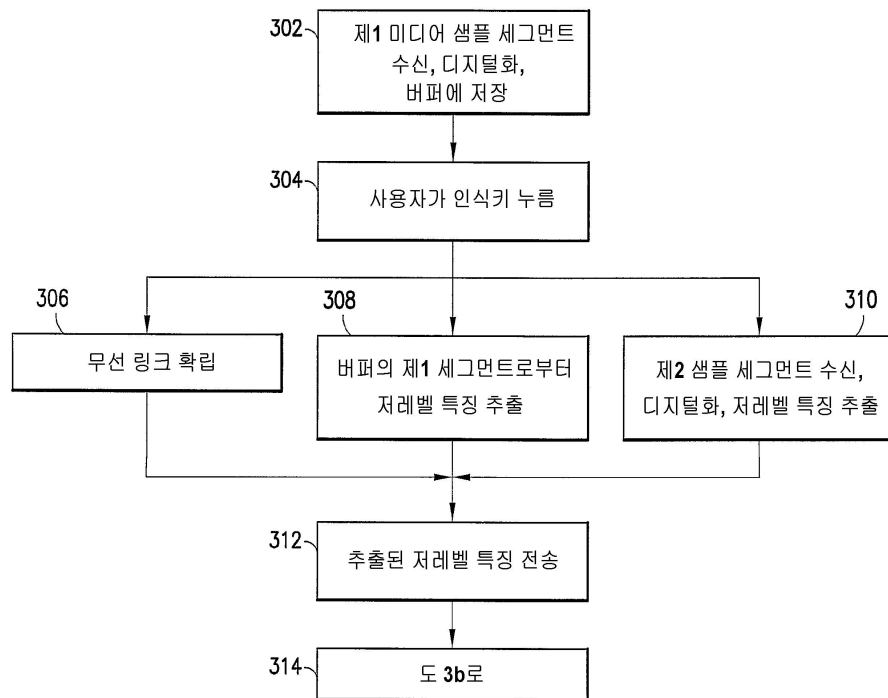
도면2a



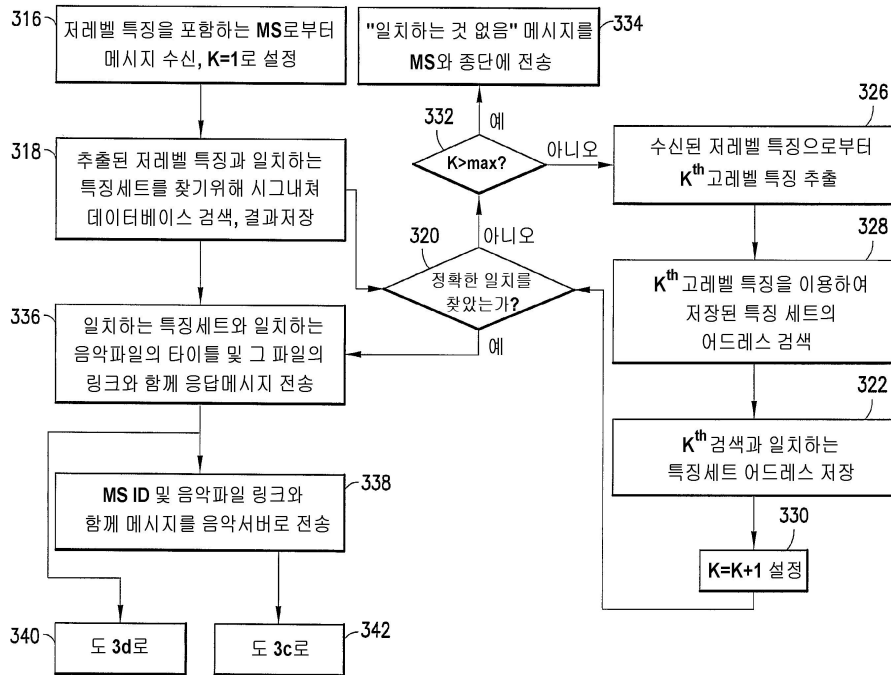
도면2b



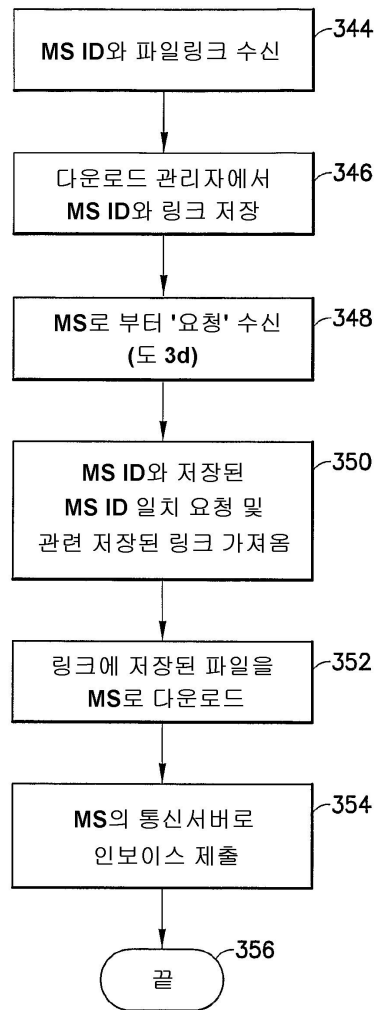
도면3a



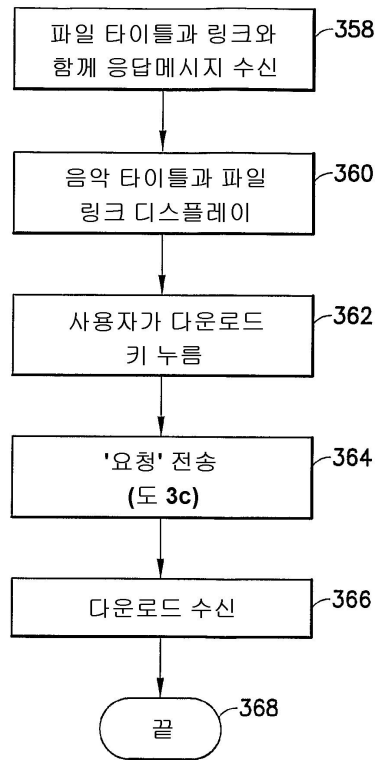
도면3b



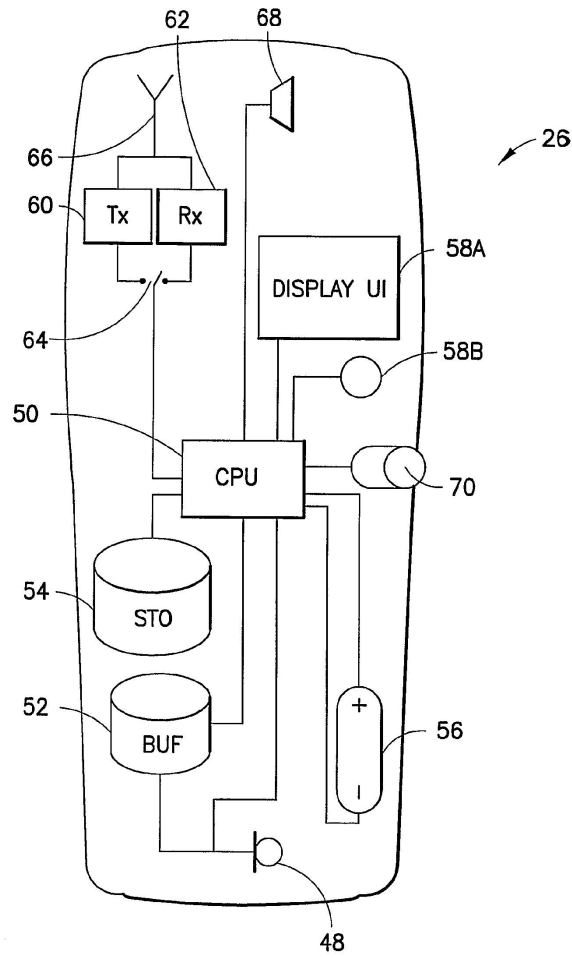
도면3c



도면3d



도면4



도면5

