



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0129165  
(43) 공개일자 2017년11월24일

- |  |  |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)<br/>G06F 3/01 (2006.01) G06F 3/0484 (2013.01)<br/>G06F 3/16 (2006.01)</p> <p>(52) CPC특허분류<br/>G06F 3/013 (2013.01)<br/>G06F 3/04842 (2013.01)</p> <p>(21) 출원번호 10-2017-7027275</p> <p>(22) 출원일자(국제) 2016년03월15일<br/>심사청구일자 없음</p> <p>(85) 번역문제출일자 2017년09월26일</p> <p>(86) 국제출원번호 PCT/IB2016/000412</p> <p>(87) 국제공개번호 WO 2016/151396<br/>국제공개일자 2016년09월29일</p> <p>(30) 우선권주장<br/>62/135,904 2015년03월20일 미국(US)</p> | <p>(71) 출원인<br/>페이스북, 인크.<br/>미국, 캘리포니아 94025, 멘로 파크, 윌로우 로드 1601</p> <p>(72) 발명자<br/>틀 마틴 헨릭<br/>덴마크 1870 프레데릭스베르 불로스바예 24비 스트리트 티브이<br/>프리즘 요나스<br/>덴마크 2300 코펜하겐 5. 6 이슬란드 브뤼게 30비<br/>산 아구스틴 로페즈 하비에르<br/>덴마크 2200 코펜하겐 2. 티에이치 야트바에 55</p> <p>(74) 대리인<br/>방해철, 김용인</p> |
|--|--|

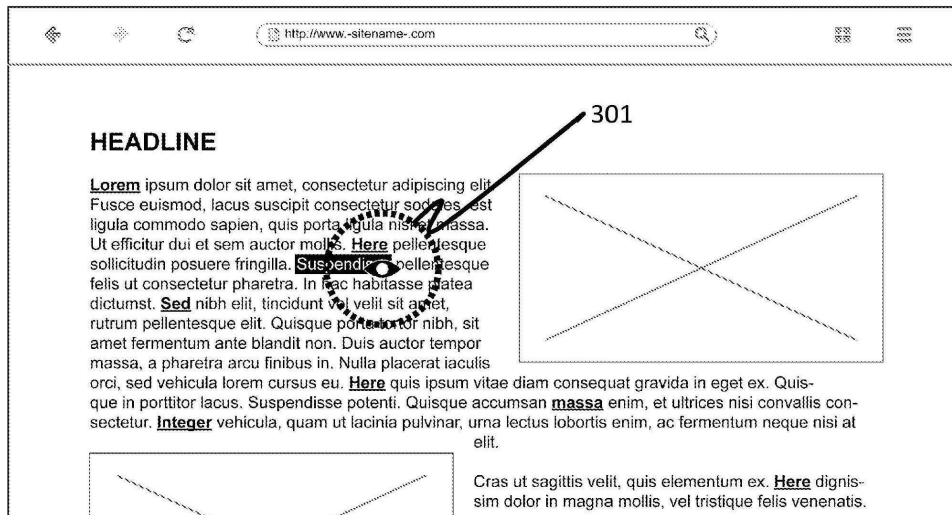
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 시선 추적과 음성 인식을 조합하여 제어를 개선하는 방법

(57) 요약

본 발명은 디스플레이 스크린상에 디스플레이되는 객체를 추후 제어 및 조작을 위해 위치시키고 선택하는 속도 및/또는 정확성을 향상시키도록 시선 추적과 음성-인식 제어 기술을 조합하는 방법이다.

대표도 - 도3



(52) CPC특허분류

*G06F 3/167* (2013.01)

*G06F 2203/0381* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

사용자가 응시하고 있는 디스플레이 스크린상의 영역을 결정하는 단계;

한 구어(spoken word) 또는 복수의 구어를 인식하는 단계;

상기 디스플레이 스크린에 디스플레이되는 객체와 상기 구어 또는 복수의 구어를 연관시키는 단계;

상기 디스플레이 스크린상에 디스플레이되는 상기 객체를 사용자가 응시하고 있는 상기 디스플레이 스크린상의 상기 영역으로 제한하는 단계; 및

상기 사용자가 응시하고 있는 디스플레이 스크린상의 상기 영역 내 상기 디스플레이 스크린상에 디스플레이되는 상기 객체를 상기 구어 또는 복수의 구어와 연관시키는 단계를 포함하는 방법.

#### 청구항 2

제 1 항에 있어서,

상기 사용자가 응시하고 있는 디스플레이 스크린상의 상기 영역 내 상기 디스플레이 스크린상에 디스플레이되는 상기 객체를 상기 구어 또는 복수의 구어와 연관시키는 것의 신뢰도를 결정하는 단계;

기결정된 신뢰도와 상기 신뢰도를 비교하는 단계; 및

상기 신뢰도가 상기 기결정된 신뢰도보다 크다면, 상기 사용자가 응시하고 있는 디스플레이 스크린상의 상기 영역 내 상기 디스플레이 스크린상에 디스플레이되는 상기 객체와 상기 구어 또는 복수의 구어의 연관을 수용하는 단계를 더 포함하는 방법.

#### 청구항 3

제 1 항에 있어서,

응시 좌표의 정확도, 응시 좌표의 잡음, 응시 좌표의 신뢰도, 스크린상의 객체의 위치 또는 이들의 임의의 조합에 기반하여 상기 신뢰도의 값을 결정하는 단계를 더 포함하는 방법.

#### 청구항 4

제 1 항에 있어서,

상기 사용자가 응시하고 있는 디스플레이 스크린상의 상기 영역 내 상기 디스플레이 스크린상에 디스플레이되는 상기 객체를 상기 구어 또는 복수의 구어의 인식과 연관시키는 것의 확률도를 결정하는 단계;

기결정된 확률도 값과 상기 확률도를 비교하는 단계; 및

상기 확률도가 기결정된 확률도 값보다 크다면, 상기 사용자가 응시하고 있는 디스플레이 스크린상의 상기 영역 내 상기 디스플레이 스크린상에 디스플레이되는 상기 객체와 상기 구어 또는 복수의 구어의 연관을 수용하는 단계를 더 포함하는 방법.

#### 청구항 5

제 4 항에 있어서,

음성 인식의 신뢰도, 응시 고정으로부터 각 객체 사이의 거리, 응시 고정의 지속시간, 응시 고정과 음성 명령의 발생 사이의 경과 시간 또는 이들의 임의의 조합에 기반하여 상기 확률도를 결정하는 단계를 더 포함하는 방법.

#### 청구항 6

사용자가 응시하고 있는 디스플레이 스크린상의 영역 내에 존재하는 객체를 결정하는 단계;

상기 객체에 기반하여 음성 인식 엔진의 어휘를 구축하는 단계;

상기 어휘를 사용하여 한 구어 또는 복수의 구어를 인식하는 단계; 및

상기 구어 또는 복수의 구어와 응시 영역 내에 존재하는 상기 객체를 연관시키는 단계를 포함하는 방법.

**청구항 7**

제 6 항에 있어서,

상기 사용자의 응시 고정시마다 상기 음성 인식 엔진의 상기 어휘를 업데이트하는 단계를 더 포함하는 방법.

**발명의 설명**

**기술 분야**

[0001] 가령 개인 컴퓨터, 스마트폰, 태블릿 등과 같은 컴퓨팅 장치는 그래픽 사용자 인터페이스(GUI)를 이용하여 그 사용자들에 의한 제어를 용이하게 한다. 이미지, 단어 및 영숫자의 문자들을 포함할 수 있는 객체가 스크린에 디스플레이될 수 있고; 사용자는 커서-제어 장치(예컨대, 마우스 또는 터치 패드)를 이용하여 상호작용형 스크린 요소의 선택 및 결정을 표시하도록 스위치한다.

**배경 기술**

[0002] 다른 사례로, 커서 및 스위치 이외에, 시스템은 터치 감응 스크린을 사용할 수 있고, 이로써 사용자는 손가락 또는 스타일러스로 그 스크린 위치를 터치함으로써 어떤 것을 식별하고 선택한다. 이 방식으로, 예컨대 사용자는 가령 "프린트"와 같은 제어 아이콘을 선택하거나 하이퍼링크를 선택할 수 있다. 또한, 사용자는 텍스트 편집 및/또는 복사하여 붙여넣기 상호작용을 위해 일련의 영숫자 문자 또는 단어를 선택할 수 있다. 커서 제어 및 터치-제어 패널은 사용자가 물리적으로 제어 장치를 조작하여 스크린 아이템의 위치를 파악하고 선택하도록 설계된다.

[0003] 그러나, 제어 서브시스템을 물리적으로 움직이거나 터치하지 않는 그러한 제어를 위한 대안의 수단들이 있다. 이런 대안 중 하나는 관심 있는 스크린 영역 및 상호작용적 선택을 위한 스크린 아이템을 식별하도록 스크린으로의 사용자의 응시가 이용될 수 있는 시선 추적(eye tracking)을 이용한다. 또 다른 대안은 음성 인식을 이용하고 스크린에 디스플레이되는 관련 아이템과 인식된 단어와 연관시킨다. 시선 추적이나 음성 인식 제어 모두 그 자체로는 스크린 객체의 위치를 파악하고 선택함에 있어서 가령 커서 제어 또는 터치 제어만큼 정확하지 못하다. 시선 추적의 경우, 한 지점 또는 지점들의 작은 클러스터보다는 한 스크린 영역으로 결정이 종종 제한된다. 그 스크린 영역 내에 또는 그 근처에 하나 이상의 스크린 객체가 있다면, 선택이 모호할 수 있다. 마찬가지로, 텍스트 및 객체 선택이 가득한 스크린에서 인식된 단어를 하나의 관련된 스크린 객체 또는 단어로 결정하려고 할 때 음성 인식 서브시스템도 또한 모호성으로 어려움을 겪을 수 있다. 따라서, 그 결과, 이런 제어 방법은 시선 추적 제어의 경우 스크린 객체의 수를 제한하고 객체들 사이의 거리를 증가시키도록 주밍(zooming)할 수 있거나; 정확한 제어 또는 선택 해석의 확률을 증가시키기 위해 반복되는 구두 명령을 필요로 할 수 있다.

**발명의 내용**

**해결하려는 과제**

[0004] 시선 추적 제어와 음성 인식 제어를 조합함으로써, 사용자는 위치 및 선택의 정확성을 효과적으로 증가시킬 수 있고, 이로써 하나 이상의 제어 기술을 사용할 때 현재 요구되는 주밍 또는 구두 명령의 반복을 줄일 수 있다.

**과제의 해결 수단**

[0005] 본 명세서에서 개시되고 청구되는 방법을 통해, 독립적으로 구현되는 시선 추적 및 음성 인식 제어는 전체 제어가 더 신속하고/하거나 더 정확해지도록 협력할 수 있다.

[0006] 본 명세서에서 개시되고 청구되는 방법은 음성 인식 제어와 시선 추적을 결합하는 통합형 제어 시스템으로 이용될 수 있다.

[0007] 본 명세서에서 개시되고 청구되는 방법은 애플리케이션을 실행하거나, 클라이언트 사용자 시스템을 사용하여 서버-기반 HTML 페이지 집합과 상호작용(예컨대, 인터넷을 통해 웹사이트와 상호작용)하기 위한 준비로서 시스템

을 부팅함으로써 발생할 수 있는 스크린 객체의 위치판단 및 선택에 적용될 수 있다. 실제로, 시선 추적 및 음성 인식 제어 서브시스템과 연계하여 본 방법은 기반 플랫폼 사양에 관계없이 스크린에 디스플레이된 객체의 상호작용에 대한 향상된 제어를 제공할 수 있다.

[0008] 본 명세서에서 개시되고 청구되는 방법은 시선 추적의 속성을 사용하여 음성 인식 제어의 모호성을 감소시키며; 음성을 사용하여 시선 추적 제어의 모호성을 감소시킨다. 그 결과 제어에 시너지 효과가 나타난다; 즉, 시선 추적 또는 음성 인식을 각각 제어하는 것에 비하여 제어 속도 및 정확성이 월등하다.

**발명의 효과**

[0009] 본 발명의 내용 중에 포함되어 있다.

**도면의 간단한 설명**

[0010] 도 1은 비-텍스트 객체와 텍스트 객체를 디스플레이하는 디스플레이 스크린을 도시한다. 예컨대, 스크린은 가령 컴퓨터 모니터, 스마트폰 스크린, 태블릿 스크린 등과 같은 임의의 시스템 디스플레이 및 제어 스크린일 수 있다.

도 2는 사용자의 응시가 실질적으로 비-텍스트 객체에 있다고 시선 추적 제어가 결정하는 도 1의 스크린을 도시한다.

도 3은 사용자의 응시가 실질적으로 텍스트 객체를 포함하는 스크린 영역에 있다고 시선 추적 제어가 결정하는 도 1의 스크린을 도시한다.

도 4는 시선 추적과 음성 인식의 조합이 어떻게 위치와 선택을 결정하는 신뢰도를 증가시켜서 그 정확도를 증가시키는지를 도시하는 예시적인 흐름도를 도시한다.

도 5는 시선 추적과 음성 인식의 조합이 어떻게 위치와 선택을 결정하는 확률도를 증가시켜서 그 정확도를 증가시키는지를 도시하는 예시적인 흐름도를 도시한다.

도 6은 시선 추적과 음성 인식의 조합이 사용자의 응시 스크린 영역으로 결정되는 더 작은 스크린 영역 내 발생과 해석된 단어를 연관시킴으로써 어떻게 단어 그룹 내에서 선택된 단어를 결정하는 확률도를 증가시키는지를 도시하는 예시적인 흐름도를 도시한다.

**발명을 실시하기 위한 구체적인 내용**

[0011] 모든 종류의 상호작용형 컴퓨팅 시스템이 발달해오면서, GUI는 시스템과 사용자 사이의 주요한 상호작용 메커니즘이 되었다. 이미지, 영숫자 문자, 텍스트, 아이콘 등일 수 있는 스크린상에 디스플레이되는 객체를 가지고, 사용자는 GUI의 일부를 이용하며 이를 통해 사용자는 스크린 객체를 위치판단하고 선택할 수 있다. 주요한 2개의 흔한 GUI 서브시스템은 커서 제어 장치(예컨대, 마우스 또는 터치 패드) 및 선택 스위치를 사용하여 스크린 객체를 위치판단하고 선택한다. 스크린 객체는 프린트 버튼과 같은 제어 아이콘일 수 있고, 따라서 그 위치를 판단하고 선택함으로써 디스플레이된 문서 파일이 프린트될 수 있다. 스크린 객체가 글자, 단어 또는 하이라이팅된 텍스트 부분이라면, 선택하여 편집, 삭제, 복사하고 붙여넣기 또는 유사한 작업에 이용할 수 있다. 오늘날, 많은 장치는 손가락이나 스타일러스 터치로 스크린 객체의 위치를 판단하거나 선택할 수 있는 터치 패널 스크린을 사용한다. 2가지 경우 모두, 그 제어에는 스크린 객체의 위치를 판단하고 선택하기 위해 사용자가 제어 장치에 물리적으로 관여할 필요가 있다.

[0012] 사용자는 보통 커서 제어로 정확히 스크린 객체의 위치를 판단하고 선택할 수 있다. 때때로 사용자는 의도된 스크린 객체의 위치를 정확히 판단하고 선택하기 위해 객체들을 더 크게 만들고 객체들을 서로 더 멀리 떨어지게 하도록 스크린의 일부를 확대해야 한다. 이런 주밍 기능은 손가락-터치 제어에서 더 일반적인데, 이때 몇몇 작은 스크린 객체를 갖는 영역에서의 손가락 터치는 주밍이 적용될 때까지 부정확하다.

[0013] 또한, GUI는 물리적 관여 없이 스크린 객체의 위치판단 및 선택을 가능하게 하는 역할을 할 수 있다. 예컨대, 시선 추적 제어를 이용하는 GUI는 사용자가 스크린상의 어디를 응시하고 있는지(예컨대, 위치)를 결정하고 선택 제어를 위한 몇몇 방법(예컨대, 응시 유지 시간)을 사용할 수 있다. 이는 마우스를 사용하여 커서를 스크린 객체로 이동시킨 후 버튼을 눌러 선택 의도를 나타내는 것과 유사할 수 있다.

[0014] 또한, 음성 인식 기반 제어도 물리적 관여가 필요하지 않는 제어 기술의 역할을 할 수 있다. 객체의 스크린은

객체와 관련된 구어(spoken words)의 어휘를 가질 수 있고, 사용자가 단어 또는 문구를 말할 때, 제어 시스템은 그 단어를 인식하고 특정 스크린 객체와 연관시킨다. 예컨대, 사용자가 "원 A"라고 말하여 GUI 시스템이 그 객체를 하이라이팅하게 한 후, "선택하라"를 말하여 GUI 시스템이 그 객체를 선택하고 그 하이라이팅을 제거함으로써, 중앙에 문자 A가 있는 원인 객체를 갖는 스크린이 위치판단되고 선택될 수 있다. 명백히, 스크린상에 많은 객체가 있다면, "원"이라고 말할 때, 다양한 크기와 색을 갖는 5개의 원이 있는 경우, 동일한 설명을 갖는 몇몇은 모호해질 수 있다. 더 높은 신뢰도 또는 더 높은 확률 추정을 가지도록 하기 위해, 시스템은 사용자가 추가적인 서술을 하도록 유도한다.

- [0015] 따라서, 시선 추적 또는 음성 인식 제어를 사용하는 것의 균형은 포인팅/선택 장치 또는 스크린과의 물리적 관여의 필요성을 제거하지만 덜 정확한 위치판단 및 선택의 결정을 수용하고 있다. 흔히, 덜 확고한 결과로서, 시스템이 가령 커서, 터치 패드 또는 터치 스크린과 같은 더 확고한 제어에 상응하는 확률을 가진 객체의 위치 및 선택을 결정할 수 있기 전에 수행되는 더 많은 단계들이 있을 수 있다.
- [0016] 통상, 유형-선택 커서(type-selecting cursor)는 독립해있거나 단어에 포함되어 있는 영숫자 문자보다 더 작다. 따라서, 사용자가 오타를 바로잡고 있다면, 사용자는 하나의 글자를 선택하고 그 글자를 삭제하거나 변경할 수 있다. 터치 제어를 사용하면, 손가락이나 스타일러스 터치의 영역은 통상 커서 포인터보다 더 크다. 유사한 오타 수정을 위해 단어에 포함된 글자를 선택하는 것이 어려울 수 있다. 사용자는 터치 지점이 하나의 의도된 글자 타겟으로 결정될 수 있기 위해 정확한 글자를 선택하거나 단어를 더 큰 비율로 확장(즉, 주밍)시키도록 몇몇의 포인팅 시도를 해야 할 수 있다.
- [0017] 사용자가 어떤 GUI 위치판단 및 선택 기술을 사용하는지에 관계없이, 폰트 크기 및 비-텍스트 객체 치수는 제어 결정에 영향을 줄 수 있지만, 일반적으로 물리적 관여가 필요 없는 기술은 반복하는 주밍 단계 없이 작은 문자들을 갖는 밀집된 텍스트 및 작은 치수를 갖는 비-텍스트 객체를 수용할 수 없다.
- [0018] 본 명세서에서 개시되고 청구되는 방법은 실제로 스스로 제어 기술 중 어느 것을 사용하여 스크린 객체를 위치판단하고 선택하는 정확성을 향상시키도록 시선 추적 및 음성 인식 제어 기술을 함께 사용한다. 본 방법은 디스플레이되는 객체를 갖는 임의의 시스템에 적용되며, 이로써 사용자는 스크린 객체를 위치판단하고 선택하여 시스템이 하나 또는 복수의 스크린 객체 상에서 임의의 작업 또는 작업들을 수행하게 함으로써 상기 시스템과 상호작용한다. 이런 시스템은 디스플레이되는 객체의 디스플레이, 위치판단, 선택 및 작동을 협력하여 지원하는 하드웨어, 펌웨어 및 소프트웨어의 조합을 포함할 수 있다. 본 방법은 시선 추적과 음성 인식 제어를 통합하는 통합형 제어 서브시스템의 일부로서; 또는 별도의 시선 추적과 음성 인식 제어 서브시스템이 상호작용할 수 있는 시스템의 일부로서 시스템 하드웨어 및/또는 소프트웨어와 상호작용하는 것을 포함할 수 있다. 따라서, 본 명세서에서 개시되고 청구되는 방법 발명은 임의의 특정 시스템 아키텍처 또는 하드웨어와 소프트웨어의 파싱으로 범위가 제한되지 않아야 한다.
- [0019] 시선 추적 기술 또는 서브시스템은 아키텍처 또는 구현에 관계없이 임의의 이런 기술 또는 서브시스템을 일컬으며, 사용자의 시선 또는 시선들이 디스플레이 스크린의 임의의 영역 어디를 대략 응시하고 있는지를 결정할 수 있다. 또한, 시선 추적 기술 또는 서브시스템은 사용자가 응시된 영역에서 하나 이상의 객체를 선택했고 위치판단했다고 결정할 수 있다. 객체는 선택한다면 동작을 개시하는 아이콘 또는 링크일 수 있다.
- [0020] 음성 인식 기술 또는 서브시스템은 아키텍처 또는 구현에 관계없이 임의의 이런 기술 또는 서브시스템을 일컬으며, 사용자의 구어 또는 단어들의 문구를 인식하고 디스플레이된 객체 및/또는 작업 명령과 인식된 단어 또는 문구를 연관시킬 수 있다.
- [0021] 도 1은 스크린상의 객체의 디스플레이를 도시한다. 객체는 가령 영숫자 문자, 단어, 문장 및 단락과 같은 텍스트 객체; 및 이미지, 라인 아트, 아이콘 등을 포함하는 비-텍스트 객체로 구성된다. 이 도면은 예로서 스크린상의 객체의 레이아웃 및 콘텐츠를 제한하는 것으로 이해되어서는 안된다.
- [0022] 시선 추적 제어 기술로 사용자는 사용자의 시선 또는 시선들이 도 1의 스크린에서 응시하고 있는 영역을 결정할 수 있다. 예컨대, 도 2에서, 시선 추적 제어 서브시스템은 사용자의 시선이 비-텍스트 객체의 부분을 응시하고 있고 응시 영역은 201의 원형 영역으로 정의된다고 결정했다.
- [0023] 도 3은 사용자의 시선이 301의 원형 영역인 텍스트 객체의 일부를 응시하고 있다고 시선 추적 제어 서브시스템이 결정한 도 1의 스크린을 도시한다.
- [0024] 도 2에서, 비-텍스트 객체가 201보다 작았고 하나 이상의 이런 영역(201)에 위치해있었다면, 시선 추적 서브시스템은 그때 영역(201) 내에 어느 객체가 사용자의 관심 객체인지를 결정할 수 없었을 것이다. 후속 단계와 연

계함으로써, 단 하나의 객체만이 영역(210)에 위치하도록 스크린 객체는 확대될 수 있다. 그러나, 정확성을 위한 후속 단계는 시간을 증가시킨다. 또한, 첫번째 주밍 시도에도 여전히 영역(201) 내에 2 이상의 객체가 남아 있는 경우도 있을 수 있다. 따라서, 관심 객체를 결정하기 위해 두번째 주밍 동작이 이루어져야 할 수 있다. 이때 다시 더 많은 시간이 사용된다.

[0025] 도 3에서, 응시 영역(301)은 복수의 영숫자 문자 및 단어를 포함한다. 이때 다시, 시선 추적 제어 서브시스템은 어느 문자 또는 단어가 관심 객체인지를 상세하게 결정할 수 없다. 다시 한번, 어느 글자 또는 단어가 관심 객체인지를 결정하기 위해 반복되는 주밍 동작이 이루어져야 할 수 있다. 비-텍스트 객체의 경우와 같이, 주밍 동작이 적용될 때마다 더 많은 시간이 요구된다.

[0026] 도 1과 관련하여 음성-인식 기술을 사용하면, 가시적인 전체 스크린 및 그 임의의 객체들이 사용자의 선택 객체일 수 있다. 예컨대, 사용자가 "단어 'here'를 삭제하라"고 말하면, 음성 인식 서브시스템은 먼저 단어 "here"를 인식한 후 그 단어를 스크린 객체 중에서 그 단어의 임의의 인스턴스와 연관시켜야 한다. 도 1에 도시된 바와 같이, 단어 "here"에 대한 3개의 인스턴스가 있다. 따라서, 음성 인식 서브시스템은 단 하나의 객체 선택에 대한 명령을 결정할 수 없다. 결국, 예컨대 사용자가 "예"라고 말할 때까지 "here"의 각 인스턴스를 하이라이팅하는 반복적인 과정이 있어야 할 수 있다. 이는 더 많은 시간이 걸릴 수 있다.

[0027] 본 명세서에서 개시되고 청구되는 발명의 일실시예로, 도 4는 예시적인 작업 흐름을 도시한다. 도 4에 도시된 흐름은 제한되는 것으로 이해되어서는 안된다. 흐름은 401에서 시작하는데, 이때 시스템은 스크린 객체를 포함하는 구성요소들을 로딩하고 파싱한다. 비록 흐름도에 도시되지 않지만, 이 동작은 반복하여 행해질 수 있다. 단계 402에서, 시선 추적 서브시스템은 반복되는 스크린 응시 좌표를 계산하고 시스템으로 전달한다. 402로부터, 응시 영역(G)이 결정된다(403). 404 및 405에서, 영역(G)이 결정되면, 시스템은 영역(G) 내 발견된 링크들에 대하여 링크(D) 및 어휘(V)의 사전을 구축한다. 컴퓨팅 장치 및/또는 음성 인식 서브시스템의 능력에 따라, 어휘(V)는 모든 응시 좌표에 대하여, 모든 고정점에 대하여, 모든 N개의 응시 좌표에 대하여, 매 T 밀리초 등에 대하여 업데이트될 수 있다. 단계 402 내지 405는 음성 명령이 수신(406)될 때까지 계속 리프레시(refresh)된다. 이후, 시스템은 어휘(V)에 기반하여 음성 명령을 인식(407)하고, 정확성의 신뢰도(C)와 함께 링크(L)를 결정(408)한다. 음성 인식에서, 음성 명령과 결합되는 이질적인 소리도 또한 인식 정확성을 감소시킬 수 있는 오디오 아티팩트(audio artifacts)를 포함할 수 있다. 이질적인 소리로 인한 부정확한 선택을 방지하기 위해, 신뢰도(C)는 임계값(th)과 비교될 수 있고, 신뢰도가 임계값보다 크다면(409), 시스템은 링크(L)를 활성화시키며(401), 그렇지 않으면 동작 402로 되돌아간다. 임계값(th)은 고정된 값을 취할 수 있거나, 예컨대 응시 좌표 내 잡음, 시선 추적 시스템에 의해 보고되는 온-스크린 정확성, 응시 좌표의 신뢰도, 스크린상의 링크(L)의 위치 또는 이들의 임의의 조합과 같은 서로 다른 인자에 따라 사례별로 계산될 수 있다. 본 명세서에서는 시선 추적 기술이 가능한 객체들의 전체 스크린을 응시 영역(G) 내 객체들만으로 감소시키는데 사용되는 경우이다. 반복되는 주밍 단계를 반복하지 않고, 델리네이터(delineator)로서 시선 추적 응시 영역(G)을 사용함으로써, 시스템은 더 적은 단계와 적은 시간을 사용하여 충분한 신뢰도로 링크(L)를 활성화시킬 수 있다.

[0028] 또 다른 실시예로, 도 5는 예시적인 작업 흐름을 보여준다. 도 5의 흐름은 제한되는 것으로 이해되어서는 안된다. 흐름은 501에서 시작하는데, 이때 시스템은 스크린 객체를 포함하는 구성요소를 로딩하고 파싱한다. 비록 흐름도에 도시되지 않지만, 이 동작은 반복하여 행해질 수 있다. 시선 제어 서브시스템은 응시 영역 좌표를 반복하여 리프레시하고, 그 데이터를 시스템에 공급한다(502). 음성 명령이 수신(503)될 때, 응시 영역(G)은 명령이 수신되는 시간으로부터 그 이전 몇몇의 기결정된 초(seconds) 수까지의 범위일 수 있는 시간 윈도우 동안 수신되는 시선 추적 좌표에 의해 결정된다(504). 영역(G)에 있는 링크의 사전(D)이 구축되고(505), 영역(G) 내 링크의 어휘(V)가 구축된다(506). 음성 명령은 확률(P)을 갖는 V에 기반하여 인식된다(507). 다수의 링크가 인식되는 경우, 각 링크에 대한 정확성 확률(P)이 예컨대 음성 인식의 신뢰도(C), 응시 지점 또는 고정 지점에서 링크까지의 거리, 상기 고정점의 지속시간, 응시되는 링크와 음성 명령 발생 사이의 경과 시간 등과 같은 서로 다른 인자에 기반하여 계산될 수 있고(508); 가장 높은 확률(P)을 갖는 링크가 선택될 수 있다. P가 임계값(th)보다 크다면(509), 링크(L)는 활성화되며(510), 그렇지 않으면 시스템은 동작 502로 되돌아가서 새로운 음성 명령을 기다린다. 임계값(th)은 고정된 값을 취할 수 있거나, 동작 409에서 상술한 바와 같이 사례별로 계산될 수 있다. 도 4 및 도 5 모두에서 링크가 활성화됨을 주목하자. 실제로, 이런 동작들은 링크에 국한되는 것이 아니며, 임의의 상호작용형 스크린 객체에 적용될 수 있다.

[0029] 또 다른 실시예로, 도 6은 예시적인 작업 흐름을 도시한다. 도 6의 흐름은 제한되는 것으로 이해되어서는 안된다. 흐름은 시스템이 스크린 객체를 포함하는 구성요소를 로딩하고 파싱하는 것으로 시작한다. 비록 흐름도에 도시되지 않지만, 이 동작은 반복하여 행해질 수 있다. 이후, 시스템은 음성 명령을 기다린다. 본 명세서에서,

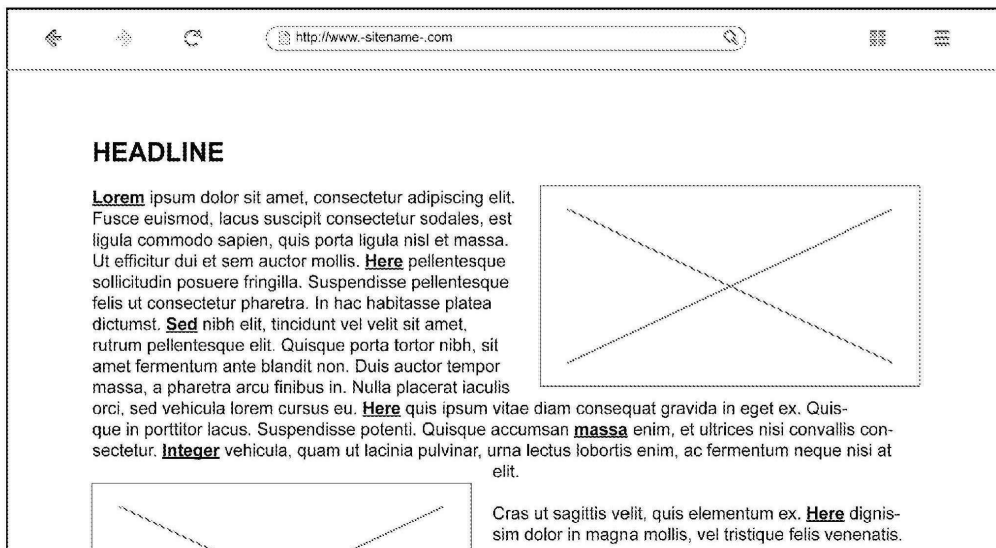
예컨대 명령은 "선택하라(select)"이다(603). 응시 영역(G)은 명령이 수신되는 시간으로부터 그 이전 몇몇의 기 결정된 초 수까지의 범위일 수 있는 시간 윈도우 동안 수신되는 시선 추적 좌표를 사용하여 결정된다(604). 본 명세서에서, 응시 영역은 도 3과 같이 텍스트 객체 위에 있다. 따라서, 텍스트(T)는 영역(G)에서 파싱되고, 어휘(V)가 구축된다(605). 어휘(V)에 기반하여, 음성 명령의 텍스트 객체가 인식된다(606). 단어(W)는 확률(P)로 평가되고(607), 임계값(th)과 비교된다(608). P가 th를 초과하면, 단어(W)는 선택된다(609). 확률(P) 및 임계값(th)은 상술한 바와 같이 계산될 수 있다.

[0030]

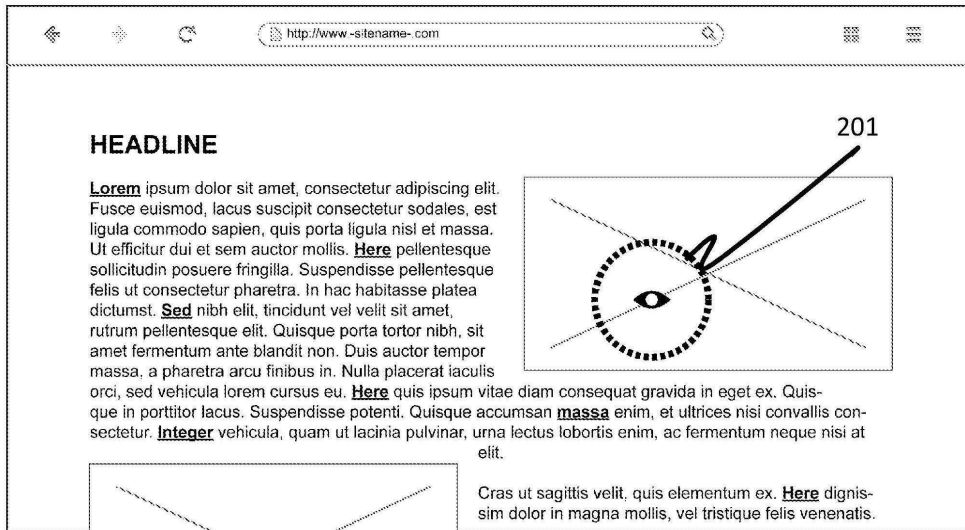
도 4, 도 5 및 도 6에 도시된 흐름은 예시이다. 각 예시에서, 전체 스크린의 객체들이 응시 영역 내 객체들로 축소되어 주밍 동작에 의하지 않고 신뢰도 또는 확률도를 높인다. 물론 응시 영역이 여전히 계속하여 몇몇의 모호한 관심 객체를 가질 수 있지만 그 개연성은 단지 음성 인식 제어만을 사용하는 것보다 훨씬 더 낮아질 수 있다. 많은 경우, 응시 영역과 조합된 구어는 임의의 주밍 동작 없이도 관심 객체를 결정하는데 충분할 것이다. 명백히, 시선 추적 및 음성 인식 기술의 조합은 시선 추적이나 음성 인식 제어 중 하나가 배타적으로 적용되는 것에 비하여 더 신속하게 관심 객체를 결정할 것이다.

**도면**

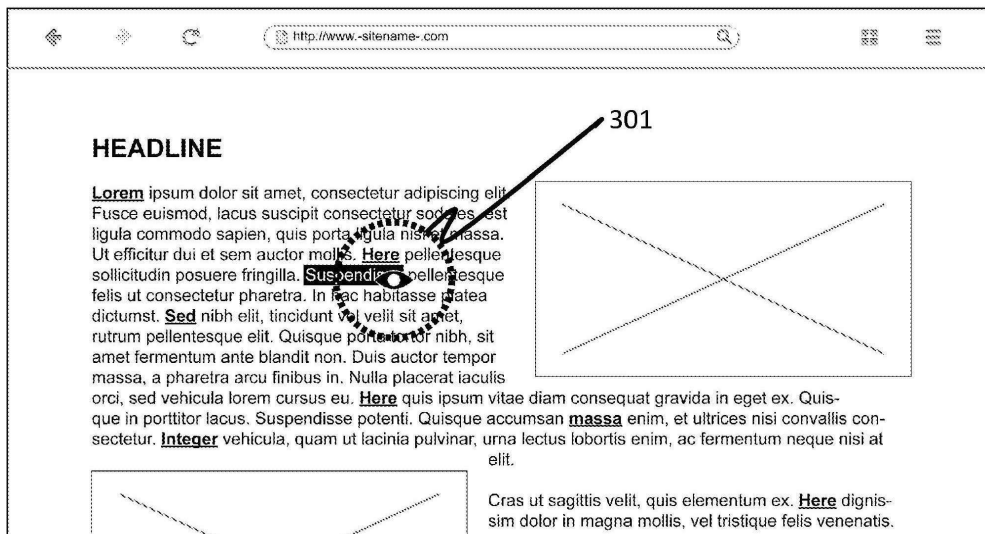
**도면1**



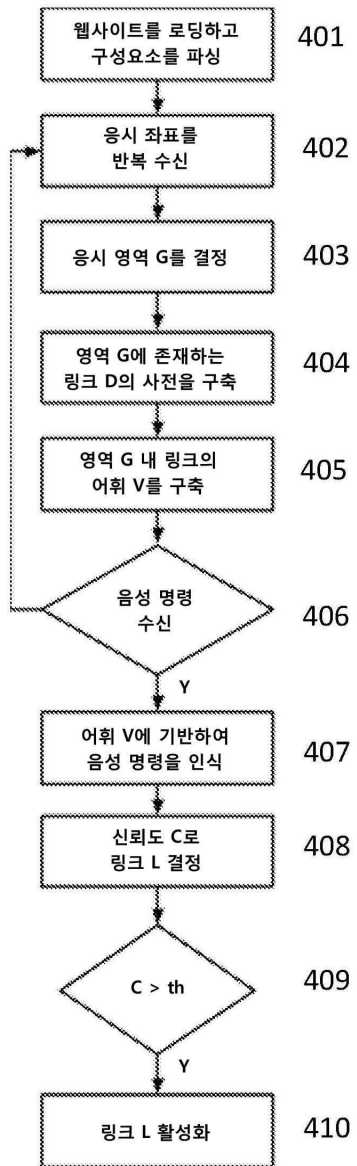
도면2



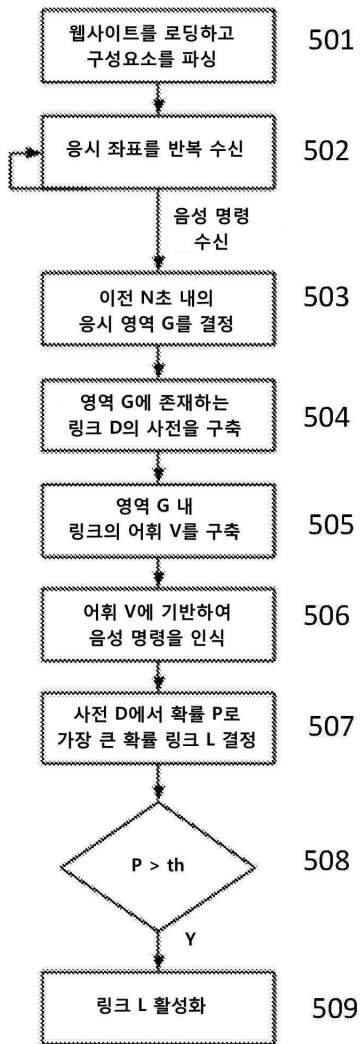
도면3



도면4



도면5



도면6

