

Implementation of Interface to Support Mobile Accessibility Using Speech I/O APIs

Seungchur Oh[†] · Young-Sun Yun^{**}

ABSTRACT

Due to the increased use of mobile devices, there is a lot of discussion on mobile accessibility. Mobile accessibility means that everyone, who includes the disabled, the elderly people, can easily use the functions of mobile devices. In this paper, we presented and implemented a mobile interface using a speech I/O APIs to improve the accessibility. The proposed interfaces are implemented on Android platforms and they used speech recognition and text-to-speech APIs supported as built-in services. In addition, to facilitate the internet access for visually impaired or blind people, we also implemented the web browsing application (web reader).

Keywords : Mobile Accessibility, Speech Information Technology, Speech Based Phone Services, Speech Based SMS Services, Mobile Web Reader

음성 입출력 API를 이용한 모바일 접근성 지원 인터페이스 구현

오 승 철[†] · 윤 영 선^{**}

요 약

최근에 모바일 기기의 확산으로 인하여 모바일 접근성에 대한 많은 논의가 이뤄지고 있다. 모바일 접근성이란 모바일 기기를 사용하여 애플리케이션을 이용하고자 하는 장애인, 고령자 등을 포함한 모든 사람들에게 모바일 기기의 활용 가능성이 제공됨을 말한다. 본 논문에서는 음성 서비스 기술을 이용하여 모바일 접근성을 향상하기 위한 인터페이스를 제안하고 구현하였다. 제안된 방법은 안드로이드 스마트폰에서 구현하였으며, 시스템에서 제공하는 음성 인식과 음성 합성 기술을 이용하였다. 또한, 전맹인과 저시력자들의 인터넷 접근을 도와주기 위하여 접근성을 고려한 웹 리더 기능을 응용 프로그램 차원에서 설계하고 구현하였다.

키워드 : 모바일 접근성, 음성 정보 기술, 음성 전화 서비스, 음성 문자 서비스, 모바일 웹 리더

1. 서 론

최근 정보통신 기기와 통신서비스의 발달로 사용자들은 시간과 장소의 제약 없이 다양한 형태의 서비스들에 접속하여 정보를 검색하거나 서비스를 제공받을 수 있다. 인터넷에 연결하여 다양한 서비스에 접근할 수 있는 스마트폰의 비율은 2012년 4월 현재 전체 휴대폰 사용자의 50%를 상회하는 것으로 집계되고 있다. 일반적으로 스마트폰으로 일컬어지는 모바일 환경은 크게 아이폰과 안드로이드폰으로 대변되는 양강체제로 형성되고 있으며, 선발 주자인 아이폰을

추격하기 위하여 안드로이드나 윈도우 진영에서는 새로운 기술들이 운영체제에 포함되는 속도가 빨라지고 있다. 새로운 기능의 추가에 중점을 두다 보니 일반 사용자들 위주의 인터페이스가 개발되고 있으며, 상대적으로 신체나 노화에 따른 불편함을 느끼는 사용자들에 대한 배려가 부족한 실정이다. 최근 들어 웹 접근성이나 모바일 접근성에 대한 관심이 증가하고 있으며, 관련 지침[1-3] 등이 정부 차원에서 제시되고 권장되고 있으나 여전히 장애인이나 노약자들에 대한 접근성 지원을 위한 시스템 구축이 더딘 형편이다. 아이폰의 경우 VoiceOver 기능을 도입하여 화면에 표시되는 거의 대부분의 텍스트를 음성 출력을 통하여 사용자에게 제공하나, 안드로이드의 경우 버튼 터치와 같은 이벤트 위주의 출력만을 제공하고 있다. 특히 안드로이드의 경우 모바일 접근성을 지원하기 위해서는 사용자가 별도의 프로그램을 설치해야하며, 프로그램(앱)의 정보를 음성으로 제공하는 경우에는 운영체제에서 지원하지 않기 때문에 별도의 인터페이스가 필요한 실정이다.

* 이 논문은 2012학년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

† 정 회 원: 삼성전자 소프트웨어센터 Convergence Solution 팀

** 중 심 회 원: 한남대학교 정보통신공학과 교수

논문접수: 2012년 7월 17일

수정일: 1차 2012년 8월 29일, 2차 2012년 9월 12일

심사완료: 2012년 9월 12일

* Corresponding Author: Young-Sun Yun(yssyun@hannam.kr)

본 논문에서는 모바일 환경 (특히, 안드로이드 환경)에서 서비스이용에 불편한 사람들이 쉽게 정보에 접근하고 단말기를 이용할 수 있도록 모바일 접근성[4-6] 향상을 위해 음성 서비스 기술을 이용한 시스템을 구현하고 관련 인터페이스를 제안한다. 제안된 방법을 통하여 프로그램 차원이 아닌 운영체제 상에서 모바일 접근성이 지원 및 확장되기를 바란다. 구현된 인터페이스는 블루투스 헤드셋을 통한 프로그램 실행, 구글 음성 인식 기반의 명령어 입력 및 터치를 통한 제어, 음성 합성을 통한 정보 제공 등이다. 이들 인터페이스를 통하여 제공되는 기능은 네트워크 기반의 음성 다이얼링 및 텍스트 입력, 문자의 수신시 발신자의 정보 및 내용을 음성 합성을 통해 사용자에게 음성으로 정보를 제공하는 것이다. 또한 기기에 설치되어 있는 어플리케이션의 이름을 통해 실행시킬 수 있는 기능과 인터넷 정보에 쉽게 접근할 수 있도록 홈페이지 내용을 음성으로 출력하고 터치나 음성 명령에 의한 웹 검색기능을 지원하는 방법 (웹 리더 앱)을 구현하여 스마트폰에서 접근성 향상을 위한 방안을

제안하였다. Table 1에 기존의 인터페이스와 본 논문에서 제안한 인터페이스를 비교하였다.

본 논문의 구성은 다음과 같다. 2장에서는 모바일 접근성을 향상하기 위한 방법으로 음성 정보 서비스의 설계 및 활용 방안을 설명하였으며, 3장에서는 음성 정보 서비스를 이용하여 전화 및 문자의 수·발신 서비스 및 명령어 실행에 대한 구현을 기술한다. 4장에서는 모바일 접근성 향상을 위한 웹 리더의 구현 및 설계, 그리고 5장에서 웹 리더를 구현하기 위해 적용한 제스처와 음성 명령에 대한 인터페이스를 소개한다. 마지막으로 구현된 내용에 대하여 평가를 하며, 결론을 맺고자 한다.

2. 음성 정보 서비스 설계 및 활용

본 장에서는 안드로이드 플랫폼을 통하여 음성인식과 음성합성 기능을 통하여 휴대폰의 기본기능인 전화, 문자, 프로그램의 실행을 제어하는 기능의 설계 및 구현 과정을 설명한다. 또한, 터치 제스처에 따라 인터넷 접근 시 다양한 제어를 통하여 웹 접근성 향상을 지원하도록 하였다.

구현된 서비스는 시각 장애인의 스마트폰 접근성 향상을 위하여 기본 기능인 전화와 문자의 수신 확인에 대한 불편을 해소하고, 편리한 인터넷 정보 접근을 위하여 음성인식 기술을 통하여 사용자에게 블루투스 헤드셋을 통한 음성 입력 서비스를 제공하고 음성합성 기술인 TTS를 활용하여 정보들을 음성으로 제공하는 것이 목적이다. 일반적으로 접근성의 지원은 운영체제에서 지원되어야 하나, 현재까지는 운영체제에서의 지원이 미흡하기 때문에, 개발된 서비스를 장애인이나 노약자등과 같은 사용자들에게 앱(어플리케이션)의 형태로 제공하여 안드로이드에서 미흡한 접근성의 기능을 보완하도록 하였다.

Table 1. Interface comparison between typical system and proposed system

		typical(screen)	proposed (speech,touch)
telephone	send (call)	telephone number or contact name input	speech command
	receive	slide	speech command
	information	sender information on screen	speech synthesis
tet	send	screen keyboard input	speech command
	receive	text message output	speech synthesis
internet	control	movement by touch	speech command and touch
	information	text output	speech synthesis
App. Execution		touch	speech command and touch

2.1 시스템 구조

안드로이드 시스템에서 전화와 문자, 웹, 음성인식, 음성합성 기능을 활용하기 위하여 Fig. 1과 같은 구조의 API를 활용한다. 전화 기능(Call Service)을 제어하기 위하여



Fig. 1. System APIs of Android OS

Telephony API를 통하여 전화의 수·발신을 제어하였으며, 문자 서비스(SMS : Short Message Service) 제어를 위하여 SMSManager API를 활용 하였다. 위의 2가지 기능은 응용 프로그램 내에서 동작하는 기능이 아닌, OS 상에서 백그라운드 작업으로 실행하여 해당 이벤트가 발생하였을 경우에만 동작되기 때문에 서비스(백그라운드 작업으로 동작하며 운영체제의 시작 시 자동으로 실행되는 프로그램)로 등록하는 과정이 필요하다.

구현 서비스는 수신과 발신으로 구분하여 수신시에는 발신자 정보 서비스, 문자 읽기, 웹 리더를 제공하며, 발신 시에는 음성 전화 서비스, 문자 작성 서비스를 제공하도록 하였다.

인터넷을 이용할 경우 기존의 방식과 동일하게 화면을 통하여 웹 브라우저의 내용을 표시하고, 브라우저에서 표현되는 내용을 분석하여 음성 출력을 사용자에게 제공하도록 하였다. 또한 브라우저 제어를 음성 인식 또는 터치 제스처로 활용하도록 하였다. 명령 실행의 경우 블루투스 헤드셋의 볼륨 키를 눌렀을 때 음성 입력을 받아 음성 인식 과정을 거친 후 명령이 실행되도록 구현하였다. 위의 기능들은 비장애인에게도 자신이 원하는 서비스를 선택적으로 제공받을 수 있도록 설정 환경을 별도로 제공하였다.

2.2 음성 인식 및 합성의 활용

안드로이드에서는 음성인식 및 음성합성 기능을 지원하는 Speech API와 TTS(Text to Speech) API를 제공한다. 해당 API를 이용하면 사용자는 스마트 폰에서 음성 인식 및 합성을 지원하는 응용 프로그램을 개발할 수 있으며, API를 통해 음성을 통한 서비스를 구현할 수 있다.

1) 음성 인식 기능의 활용

기존의 음성 인식 서비스는 단순히 음성을 통한 인터넷 검색으로 활용되어 왔으나, 본 연구에서는 명령어 실행과 같이 활용도를 높였다 (최근의 안드로이드폰에는 명령어 실행을 위한 별도의 음성 인식기가 내장되었다). 명령 실행을 위한 음성 명령어는 Table 2와 같이 구현하였다.

Table 2. The context forms of implemented speech commands

Command := telephone text execution
Target := contact_name application_name
Execution := Command Target

스마트폰에서 입력 장치(내장 마이크)를 통해 수집된 음성 데이터는 구글의 음성 인식 서버에 전송되어 음성 인식이 실행된 후, 인식 결과가 다시 폰에 전달된다. 따라서 음성 명령을 실행하기 위해서는 인식 결과를 이용하여 명령을 실행하도록 구현해야 한다(내장 음성 인식기의 경우 사용자 또는 프로그램 개발자의 사용자화가 어려우나, 구글 인식기

를 사용하는 경우 공개된 API를 이용하여 사용자화가 쉽다). 전화 발신은 “전화 OOO”로 발성을 하게 되면 명령어 인식 후 주소록에서 이름을 검색하여 저장된 번호로 전화를 발신한다. 비슷한 방식으로 문자 발신은 해당 전화번호가 존재하면 발신하고자 하는 메시지를 음성으로 입력하고 문자 메시지를 전달하게 된다. 마지막으로 프로그램의 실행은 “실행 카메라”와 같이 발성하면, 설치된 프로그램을 검색하여 해당 프로그램을 동작시킴으로써 사용자에게 서비스를 제공한다.

2) 음성 합성 기능의 활용

음성 합성 기능은 안드로이드에서 제공하는 TTS를 활용하였으며, 해당 서비스는 통신망의 이벤트(전화·문자 수신)가 발생한 상황을 인지하여 서비스를 제공한다. 서비스의 발생 상황을 감지하기 위하여 TTS 서비스를 백그라운드 프로세스로 동작시켰으며, 이벤트 발생 시 바로 사용자에게 음성 서비스를 제공할 수 있도록 하였다. 전화 수신인 경우 통신망에서 전송하는 발신자의 정보 중에서 발신번호를 추출하여 해당 번호를 주소록에서 검색하고 이름을 출력한다. 해당 전화번호가 주소록에 존재하지 않을 경우 번호를 음성으로 출력한다. 문자 수신인 경우 전화 수신 경우와 동일하게 발신자 정보를 출력하며, 추가로 수신된 메시지를 사용자에게 음성으로 전달한다.

3) 웹 브라우저 기능의 지원

현재 안드로이드 폰에서 제공되는 접근성 기능은 사용자의 인터페이스 동작을 검출하여 버튼이나 메시지에 대한 정보를 음성 출력으로 제공하고 있다. 그러나 운영체제의 반응이 아닌 프로그램의 반응인 경우에는 음성 지원이 서비스되지 않는다(웹 접속의 경우에는 안드로이드 브라우저를 사용하기 때문에 프로그램을 통한 반응에 해당한다). 웹 정보를 음성으로 지원하기 위해서는 웹 문서를 분석하고 Tag의 분리 및 XML문서의 해석을 통해 텍스트와 링크 정보를 추출한 후 음성으로 출력한다. 사용자는 웹 문서 정보를 음성 출력으로 제공 받고, 화면의 터치 동작이나 음성 명령에 의하여 해당 문서의 링크로 이동하여 웹 문서를 탐색 하게 된다.

3. 음성 인터페이스를 이용한 정보 서비스 응용

3.1 전화 서비스

시각 장애인이 스마트폰을 이용하여 전화를 이용하기엔 불편한 점이 많다. 비장애인들은 화면으로 주소록 정보를 확인하고, 터치로 상대방에게 전화를 발신하거나 직접 숫자패드에서 전화번호를 입력하지만, 시각 장애인들은 터치 사용이 불편하기 때문에 음성을 통한 제어 및 음성 합성을 이용하여 사용자가 전화 서비스를 이용할 수 있도록 하였다 (Fig. 2 참조).

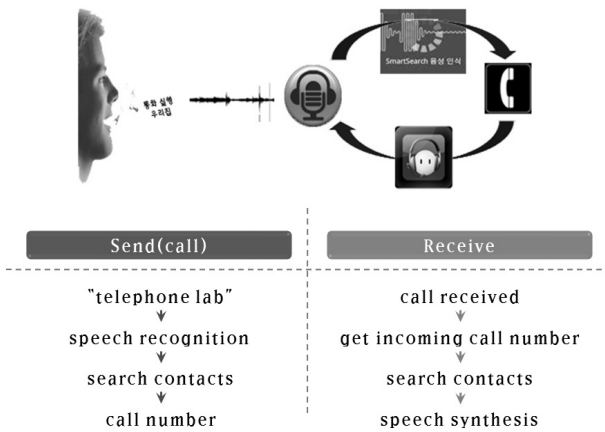


Fig. 2. Diagram of telephone service

1) 전화 수신

a) Telephony - 전화 통신

스마트폰을 위한 플랫폼의 가장 기본적인 구성 요소는 바로 전화 기능이다. 전화를 걸거나 받기, 문자 메시지 및 멀티미디어 메시지(MMS)들을 보내고 받는 것이 가능할 뿐만 아니라 전화와 관련된 모든 기능을 지원한다. 이들 중 수신된 전화 확인을 위하여 TelephonyManager를 이용하여 해당 음성 네트워크의 상태, 기기에 할당된 전화번호, SIM (Subscriber Identity Module)카드 정보 등과 같은 데이터를 검색하는데 사용된다. Telephony는 전화망을 통한 전자적인 음성 통신을 나타내는 일반적인 용어라고 할 수 있다.

TelephonyManager를 이용하여 PhoneStateListener의 형태로 이벤트 리스너를 폰에 적용함으로써 응용 프로그램에서 전화의 사용가능 여부 및 언제 전화가 오는지, 통화중인 상태인지, 통화가 끝나는지 등에 대한 정보를 알 수 있다.

Fig. 3은 수신된 전화의 정보를 얻는 경우에 사용되는 Telephony 시스템 구조를 보여 주고 있다. Telephony 시스템 구조는 크게 휴대폰 제조업체에서 제공하는 부분과 안드로이드에서 제공하는 두 부분으로 구성된다. 휴대폰 제조업체에서는 전화의 발·수신이나 SMS 전달 등의 기능을 제공하기 위하여 RIL(Radio Interface Layer)를 라이브러리 형태로 제공하며, solicited call과 unsolicited call로 구성된다. solicited call의 경우 응용 프로그램에서 RIL 데몬에게 전화의 발신 등을 요청하고 전화가 연결되면 다시 응용 프로그램에게 제어를 넘기는 것을 말하며, unsolicited call은 전화 상태나 SMS의 수신등에 의하여 RIL 데몬이 응용 프로그램을 호출하는 것을 의미한다. 이 구조를 통하여 TelephonyManager 클래스에서는 Call Tracker의 정보에 수신되어 있는 Incoming Call의 정보를 얻는다. Fig. 4는 unsolicited call 과정에서 전화가 수신되는 과정을 보인다.

b) Contacts - 주소록

안드로이드 SDK에는 휴대폰에 저장된 전화번호 내용을 PhoneNumberUtils 클래스를 이용하여 제공한다. 이 클

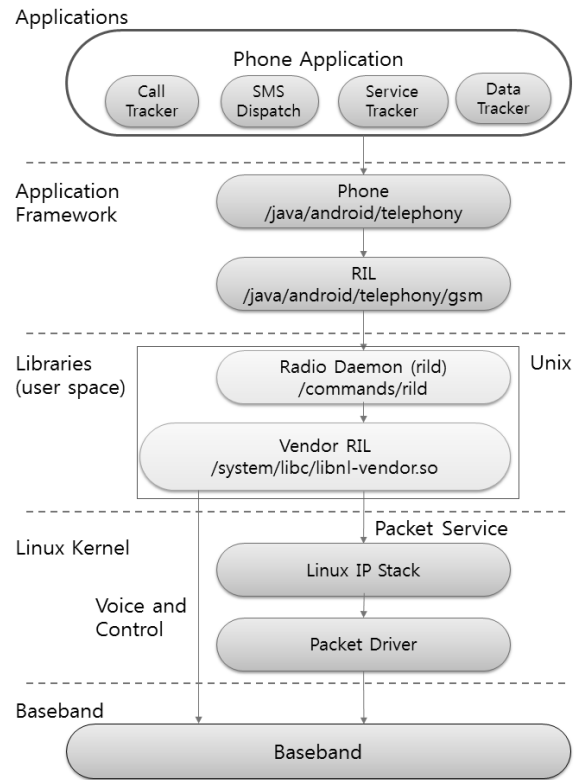


Fig. 3. Structure of telephony system[7]

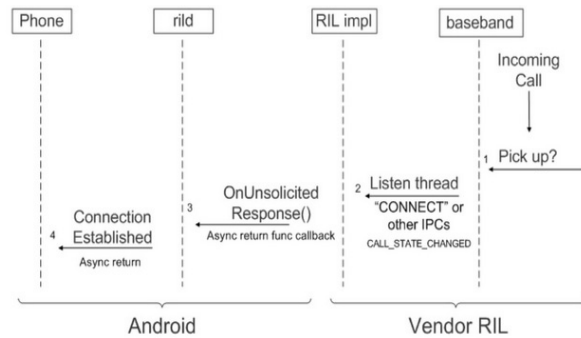


Fig. 4. Process of incoming call[7]

래스는 문자열 데이터를 전화번호 형식으로 파싱하고, 알파벳 키패드를 숫자로 바꿔 주며, 전화번호의 속성을 판별(시내·외 전화 및 국제전화 판단)한다. PhoneNumberUtils 클래스는 전화번호를 파싱하기 위한 많은 정적 Helper 메소드들을 가지고 있으며, Contacts의 정보를 읽어 해당 정보 중 PhoneNumber에 대한 정보만을 수신 전화 정보와 비교하여 발신자의 정보를 제공하며, 주소록에 없을 경우 해당 전화번호를 알려 준다.

c) Intercept calls - 전화번호 해석

전화 수신시 발신자의 전화번호와 주소록을 비교하여 주소록에 있는 경우 발신자 이름을 알려주는 일을 수행한다. 이 기능을 위해서 Broadcast Receiver를 이용하여 수신되는

전화를 가로채기한다. Broadcast Receiver를 사용하기 위하여 onReceiver 메서드를 구현한 후 android.intent.action.NEW_OUTGOING_CALL을 필터링하여 전화번호 키에서 인텐트(intent) 데이터를 얻는다.

d) Text to Speech - 음성 출력

전화번호 해석(Intercept calls)을 통하여 전화번호를 발신자의 이름으로 변환 한 후 해당 Data를 음성 합성 API를 이용하여 스피커로 재생한다.

2) 전화 발신

안드로이드의 전화 발신은 API를 직접 사용하지 않고 내장 인텐트를 사용해서 전화를 건다. Intent.ACTION_CALL 액션과 Tel:Uri만 있으면 내장된 다이얼 응용 프로그램을 호출하여 전화를 걸 수 있다. 비장애인들의 경우 시각을 통한 터치 행태로 전화 발신이 이루어지나, 시각 장애인은 시각에 의존한 환경을 사용할 수 없기 때문에 블루투스를 통하여 통화기능 사용 시 바로 해당 API를 호출 하고, 음성을 통하여 연락처의 전화번호를 가져와 발신 서비스를 제공하도록 하였다.

3.2 문자 서비스

일반적인 문자 서비스는 사용자가 문자 전송 프로그램을 실행하여 자신이 원하는 상대방의 번호를 확인 후 키패드를 통하여 메시지를 작성한다. 그러나 시각 장애인의 경우 터치 키보드로 입력하면서 입력 문자를 소리로 확인하기 때문에 많은 불편함이 존재한다. 또한 문자의 수신시에도 시각으로만 정보를 제공해 주고 있어 정보 접근의 한계성을 준다. 이러한 문제로 음성을 통한 프로그램의 제어 및 문자의 수신시에 음성 TTS를 활용하여 전달 정보를 제공하도록 구현하였다(Fig. 5 참조).

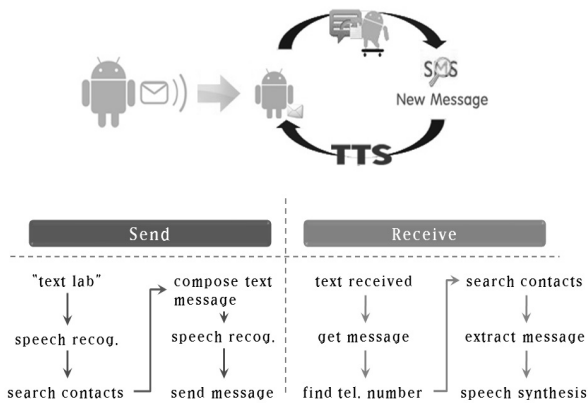


Fig. 5. Diagram of text service

1) 문자 수신

SMS는 휴대폰에서 가장 많이 사용되는 통신 수단의 하나로 짧은 문자메시지와 같이 적은 양의 데이터를 보내는데 사용된다. 안드로이드에서는 내장된 메시지 프로그램을

사용하여 수신한 SMS 메시지를 전송하거나 화면에 표시한다. SDK에서 제공하는 API를 이용하여 사용자 화면의 처리, 메시지의 전송 및 수신이 가능하다. 수신된 SMS를 확인하여 메시지를 읽고 음성으로 제공하기 위해서는 SMS의 Broadcast Receiver객체를 이용해서 적절한 인텐트를 등록해야만 한다. SMS가 수신되면 onReceiver() 메소드가 호출된다. 이 메소드는 수신된 자료를 담은 Bundle 객체에서 정보를 얻는다. Bundle 객체에는 PDU(무선 메시지 프로토콜에서 쓰이는 자료 형식) 자료를 담은 여러 개의 바이트 배열들이 존재 한다.

안드로이드 SDK가 제공하는 SmsMessage 클래스의 정적 Method createFromPdu()를 호출해서, PDU 자료로부터 단문 메시지 객체를 만들어 낸다. 그리고 그 객체에 대한 getDisplayMessageBody()를 호출해서 실제 문자 메시지를 얻는다. 상대방이 보내는 메시지의 길이가 SMS 서비스의 최대 문자 길이보다 긴 경우, 메시지를 보내는 쪽의 안드로이드는 해당 메시지를 분할하여 다중(Multipart) 메시지로 만든 후 전송한다. onReceive()의 반복문은 다중 메시지를 처리하기 위한 것이다. 이러한 onReceive() 호출 메소드를 등록한 후에는, 응용프로그램이 SMS를 받을 수 있도록 응용 프로그램을 하나의 Broadcast Receiver로 시스템에 등록하는 과정이 필요하다.

SMS 수신 시 해당 호출 메소드를 통하여 수신된 새로운 메시지의 Text를 읽어와 Voice Service로 제공한다. 이때 제공되는 부분은 수신된 메시지의 발신자 정보를 분리하여 자신의 주소록에 발신자 정보와 매칭이 되는 번호를 검색하여 주소록에 존재할 경우 해당 주소록에 저장된 이름을 반환받아 누구의 메시지인지 알려주며, 메시지의 본문을 분리하여 본문만을 사용자에게 음성으로 제공한다. 또한 주소록에 존재하지 않을 경우에는 수신 전화번호를 알려주도록 하였다.

2) 문자 송신

문자를 송신하기 위해서는 SMS Message 관련 작업 중 먼저 SmsManager의 인스턴스를 얻는 과정(객체 생성)이 필요하다. 이 과정은 정적 getDefault 메소드에서 완료 되며, SmsManager는 PendingIntent(PendingIntent)를 생성한다. PendingIntent는 나중에 사용할 인텐트의 명세서라고 할 수 있다. 기본으로 사용되는 인텐트를 다른 응용 프로그램에 전달하고 그 응용 프로그램에서 동일한 허가권을 갖는 경우, 인텐트 호출과 무관하게 문자를 송신할 수 있다.

안드로이드 플랫폼에서는 Activity 라이프 사이클과 함께 분리된 프로세스 로직을 사용하고 있으며 응용 프로그램 라이프 사이클이 종료된 후(Beyond the grave)에도 특정 인텐트를 사용할 수 있도록 한다. 인텐트를 생성한 응용프로그램이 종료되었다 할지라도 한번 생성된 인텐트는 나중에 사용할 수 있다. PendingIntent는 필요한 Activity, Broadcast, 혹은 서비스를 지정할 수 있다. 응용 프로그램에서는 인텐트를 얻기 위하여 getActivity메서드를 사용하며, 컨텍스트

(Context), 요청코드(request code), 인텐트, 그리고 플래그들(Flags)을 지정한다. 플래그는 해당 인스턴스가 존재하지 않을 경우 참조하는 Activity(또는 브로드 캐스트나 서비스)의 새로운 인스턴스를 생성해야 하는지에 나타내며, 펜딩인텐트를 생성하면 전송하는 메시지의 목적지 주소에 대한 유효성을 검사한 뒤 (PhoneNumberUtils의 또 다른 메서드를 사용) SMS메니저의 sendTextMessage 메소드를 사용하여 메시지를 전송한다. 이 과정에서 메시지의 내용을 작성하는 시작 부분과, 작성 부분, 전송 부분을 구글 음성 인식을 통하여 해결 하였다.

3.3 명령어 실행 서비스

앞에서 설명한 서비스들은 시각장애인이 모바일 기기를 쉽고 편리하게 사용할 수 있도록 모바일 접근성 기능을 제공하는 것에 목적이 있다. 명령어 실행 서비스는 개별적인 제공 서비스를 모두 포함한다고 할 수 있다. 명령어 실행 서비스에서 제공하는 것은 안드로이드에서 접근하는 모든 응용 프로그램의 실행을 의미한다. 사용자는 본 연구에서 구현된 프로그램을 실행하여 자신이 원하는 작업을 선택 한다. 이 작업은 기본적으로 전화, 문자, 프로그램으로 구분되며, 블루투스 헤드셋의 볼륨 키에 변화가 생길 경우에 구현된 프로그램을 실행하도록 하였다.

1) 음성 전화 서비스

사용자가 지정한 단어를 인식하며, “전화” 단어를 인식 후 자신의 주소록에 있는 이름을 검색하여 해당되는 전화번호를 얻어 전화 프로그램을 실행한다. 해당 주소록에 있는 전화번호가 아닌, 전화번호를 통한 입력을 원할 경우에는 이름이 아닌 숫자 인식기를 통하여 해당 텍스트를 인식하여 URL로 변환한다.

2) 문자 서비스

문자 서비스에는 “문자”라는 단어를 인식하지만, 세부 기능으로 2가지로 나뉜다.

- 문자 읽기: 3.2의 1)에서 구현한 문자 수신 기능이 메시지

수신 시 제공되는 서비스라 한다면 문자 읽기는 기존의 저장된 메시지를 가져와 텍스트 메시지를 음성으로 출력한다.

- 문자 전송: 안드로이드 운영체제에서는 문서 작성 시 음성 입력을 제공한다. 본 연구에서는 휴대기기를 꺼내지 않고 블루투스를 통한 제어 및 실행을 통하여, 해당 서비스를 실행한 후, 음성으로 입력된 텍스트를 문자로 보낼 수 있다.

3) 프로그램 실행

안드로이드 플랫폼에서 제공하는 모든 프로그램에는 자신의 앱 이름을 할당 한다. 실행하고자 하는 앱 이름을 음성으로 입력 받아, 지정된 프로그램을 실행 한다.

4. 모바일 접근성 지원을 위한 웹 리더(Web Reader) 구현

본 장에서는 웹 접근성 향상을 위한 방법으로, 기존에 시각으로만 전달되던 정보들을, 시각 장애자나 저시력을 가진 사용자들이 음성 입·출력과 제스처를 통하여 웹 정보를 효율적으로 이용할 수 있도록 표준 인터페이스를 제안하고, 전맹인과 저시력자의 2가지 모드로 분류하여 앱으로 구현하였다. 현재 많은 사이트에서 자체의 모바일 홈페이지를 구축하고 있지만, 여전히 웹 접근성 지침을 따르지 않고 있기 때문에 모든 홈페이지에 적용 가능한 웹 리더를 구축하기는 어렵다. 최근에 애플의 Safari에서 텍스트와 이미지만을 추출하여 사용자가 본문을 읽기 쉽도록 하는 Reader(읽기) 기능을 지원하고 있지만 여전히 동적 페이지 위주의 포털 등의 경우에는 적용이 어렵다. 따라서 본 연구에서는 국내 포털 점유율(네이버 63.8%, 다음 15.9%, 구글 14% 순)이 가장 높은 네이버(http://www.naver.com)를 주 대상으로 선정하였다[8]. 기본적으로 모바일 네이버 페이지의 구성은 Fig. 6과 같다.

웹 페이지에는 화면 구성을 위한 다양한 태그 명령어들이 존재하나, 시각 효과를 제공하는 스크립트나 동적 페이지들은 불필요한 정보들이기에 분석 단계에서 제거하였다. 정리

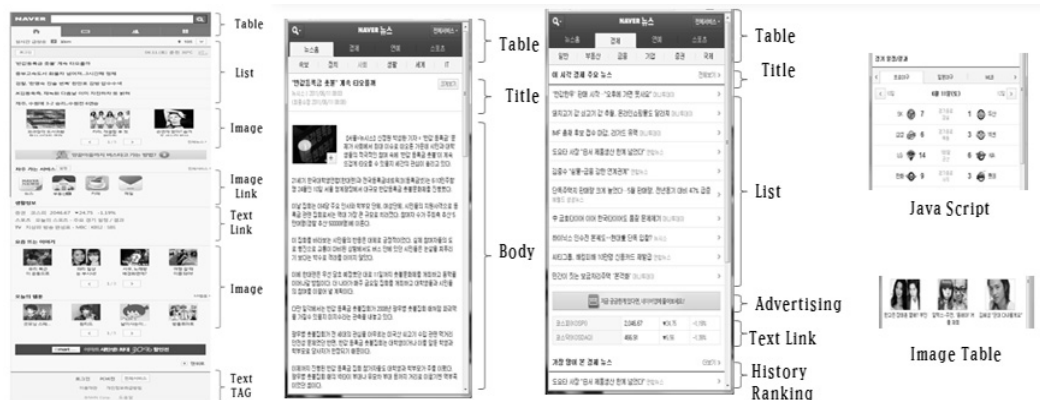


Fig. 6. Structure of Naver mobile homepage

Table 3. Types and functions along to homepage structure

Types	Functions
Table	- define HTML table - represent the various information together
Title	- define a title for current HTML document
Body	- HTML document - in this area, articles are placed
List	- support one ordered and one unordered HTML list
Advertising	- advertising information such as image or flash etc.
Link	- links to the other pages, label exists
History ranking	- on page loading, the page is placed along to ranking order
Javascript	- link to javascript file
Image table	- types of table including images and texts - images are shown in thumbnail type

한 페이지의 구성 요소들 중 텍스트와 중요하다고 판단되는 정보들을 분리하여 리스트(list)의 형태로 차례로 저장한다.

일반적인 HTML 홈페이지 포함된 정보들은 Table 3과 같다.

네이버에서 제공되는 페이지 중 상대 경로를 가지는 URL은 문서를 분석하여 절대 주소로 변환하여 직접 접근이 가능하도록 수정하였다. 즉, *main.nhn* / *list.nhn* / *rankingRead.nhn* 처럼 상대 주소로 되어 있는 경로들을 찾아 절대 주소 체계로 변환하였다. 문서 분석결과 onclick 등과 같이 사용자의 입력을 받은 경우 절대 주소로 저장되어 있기 때문에, 상대주소가 발견된 경우 절대 주소로 변환하였다. 간혹 onclick으로 반환 되는 링크 부분이 동일한 문서에 존재하지 않으면, 이전과 다음의 정보를 검색하여 절대 주소로 변환하였다.

4.1 전맹인을 위한 서비스 구축

전맹인은 시각적인 이벤트나 화면을 직접 이용할 수 없기 때문에, 터치된 화면의 이벤트를 음성으로 확인한 후 서비스를 이용하고 있다. 따라서 전맹인의 경우 웹 페이지 접근을 요청하면 해당 페이지에서 텍스트 정보들만 추출 작업을 거친다. 텍스트를 판단하는 기준으로 웹 페이지를 구성하는 가장 기본적인 요소인 링크와 텍스트 부분을 분할하여 링크로 구성된 주소와 화면에 표시된 텍스트 부분을 리스트의 형태로 저장한다. 웹페이지들은 링크와 앵커를 통하여 페이지 구성이 이루어지는 점을 착안하여, 링크나 앵커가 존재 하는 텍스트는 정보 분할의 대상이 된다. 링크와 앵커가 없는 단순한 텍스트의 경우 페이지를 구성하는 본문으로 분류하였다. 분할 과정을 반복하여 정보를 수집하면, 정보 획득의 순서(화면의 상단에서 하단까지)에 따라 TTS를 활용하여 음성으로 웹페이지를 읽어 주고, 제스처를 통하여 정보를 선택하고 원하는 페이지의 이동을 할 수 있게 설계하였다.

4.2 저시력자를 위한 서비스 활용

저시력자들은 전맹인과 동일한 방식으로 서비스를 지원 받을 수 있지만, 좀 더 직관적인 방법으로 자신이 원하는 서비스를 접근 할 수 있도록 하였다. 사용자가 터치를 통해서 원하는 영역의 링크를 호출 하면, 관련된 텍스트를 검출한다. 사용자는 검출된 텍스트를 합성된 음성으로 듣고, 음성 인식이나 제스처 입력을 통하여 원하는 페이지로 이동할 수 있다.

5. 접근성 향상을 위한 인터페이스 제안

스마트폰에서 가장 일반적인 입력장치는 터치스크린과 마이크를 통한 음성이다. 본 장에서는 웹 리더를 구현하는 과정에서 사용자의 화면 터치를 위하여 정의된 제스처와 프로그램의 제어 및 종료 실행 할 수 있도록 정의된 음성 인터페이스를 기술한다.

5.1 터치를 통한 제스처

터치는 스마트폰에서 가장 유용한 명령어 전달 방법이다. 비장애인에게도 스마트폰 사용에 있어서 터치가 없으면 활용 할 수 없을 정도로 직관적인 인터페이스라 할 수 있다. 그러나 전맹이나 저시력자들과 같이 시력이 불편한 사람이나 정확한 터치가 어려운 경우, 단말기마다 다른 터치인터페이스는 스마트폰을 이용하는데 장애가 된다. 따라서 본 논문에서는 효율적으로 터치 인터페이스를 이용하여 스마트폰을 제어할 수 있도록 제어 방법을 제안한다.

전맹인의 경우 화면 터치에 의하여 출력 음성을 확인하고, 이중 터치(double touch)를 통해 해당 페이지로 이동을 하게 설계되었다. 또한 싱글 드래그를 통해 현재 페이지의 출력중인 음성의 이전/다음 음성을 재생 할 수 있도록 하였으며, 멀티 드래그로 페이지의 이동 및 음성 명령을 할 수 있는 인터페이스를 구현하였다. 저시력자가 알고 싶은 정보의 영역을 터치하면 해당 영역의 정보를 음성 출력하고, 싱글 드래그의 경우 화면 이동으로 장애의 종류에 따라 기능을 구현하였다(Fig. 7).

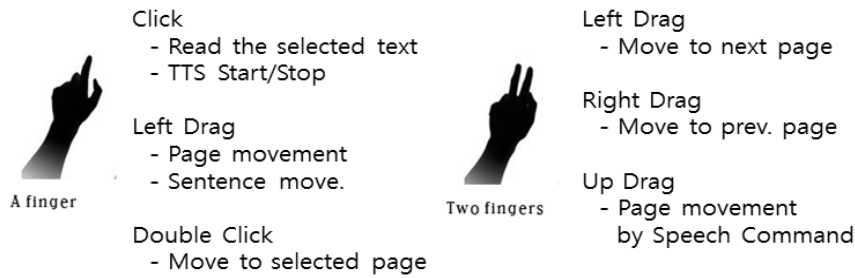


Fig. 7. Types of actions depending on finger gestures

Table 4. Implementation of speech interface for Web Reader.

		blind	low vision
Touch		TTS Stop / Start	TTS Start
Double touch		Page movement	
DRAG	single	Left	Next sentence TTS
		Right	Previous sentence TTS
	multi	Left	Next page
		Right	Previous page
		Up/Down	Speech command
Difference		TTS for each sentence	TTS for selected sentence

5.2 음성 명령

스마트폰에서 음성 인식 기능을 이용하기 위해서는 장치의 입력 장치(내장 마이크)를 통해 수집된 음성 데이터를 구글의 음성 인식 서버에 전송한 후, 인식 결과를 전달받아 목적에 맞는 동작을 구현해야 한다. 웹 접근성에 대한 인터페이스로 음성 명령 부분은 사용자의 편의를 위한 부분이라 할 수 있다. 음성 명령을 통해 사용자는 페이지의 이동과 추출된 단어의 검색, 종료 등의 명령을 제어 할 수 있다. 음성 명령을 실행하기 위한 제스처는 멀티 드래그로 구분하여 업 또는 다운하였을 경우에 음성 출력을 멈춰 스피커의 출력이 음성 인식에 영향을 받지 않도록 하였다. 제안된 인터페이스는 Table 4에 정리하였다.

6. 구현 및 평가

시각 장애인을 위한 음성정보 서비스의 적용은 안드로이드 버전 2.3(Gingerbread)로 Galaxy A, S, S2, Tap의 4가지 종류의 스마트폰으로 실험하였다.

음성 정보 서비스를 스마트폰에 적용하기 위한 기초 작업으로 음성 인식 및 합성의 기본 성능 및 음질에 대하여 평가를 하였다. 안드로이드 버전의 경우 인식은 구글 인식을 사용하며, 합성기의 경우 국내에서 개발한 합성기가 사용되기 때문에 본 연구에서 개발한 시스템과 관련이 적으나 제안한 시스템의 기초 자료로서 간단한 평가를 실시하였다.

첫째로 구글의 Speech API를 이용한 인식 성능을 평가하

였다. 남자 38명, 여자 12명, 총 50명의 화자가 유명인사 이름, 지명, 전화번호부 이름, 앱 이름, 명령어 등에 대하여 발성한 후 그 인식결과를 확인하였다. 확인된 결과 유명인사와 연락처의 이름인식에서는 평균 55% 정도의 저조한 인식률을 보였으나, 명령어나 앱(프로그램) 등의 이름에서는 약 85%의 인식률을 보였다. 일반적으로 사람이름의 경우에는 음성 인식기의 기본 통계자료로 사용되는 언어 정보와 다르기 때문에 이름 인식의 경우에는 저조한 인식률을 보인다.

둘째로 음성 합성 기능의 활용 분야에서는 단일 자음으로만 구성된 신조어를(ㅋ,ㅎ 등) 제외한 음성 합성에서는 명료한 서비스를 제공 받을 수 있었다. 최근의 문자 메시지의

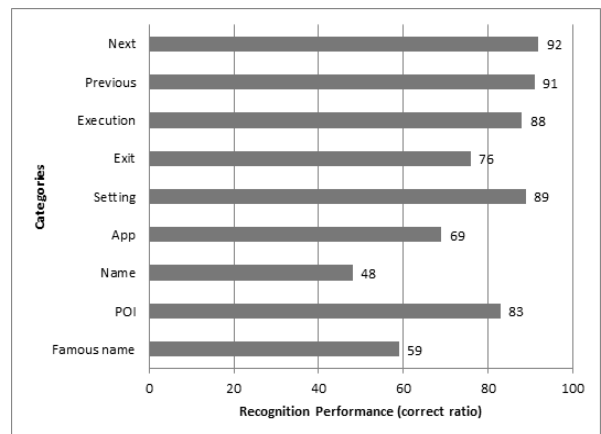


Fig. 8. Google word recognition performance (correct ratio)

경향이 신조어나 언어 파괴를 가져오는 인터넷 언어가 많이 사용되는 경향으로 자음이나 특수 문자의 결합인 경우에는 음성 합성에 의한 의사 전달에 어려움이 있었다. 그러나 일상생활에서 사용하는 단어나 문장의 경우에는 정확하게 의사 전달이 이뤄졌다.

웹 접근성 향상을 위한 웹 리더의 부분에서 실험에 적용한 웹 페이지는 모바일 네이버 페이지를 대상으로 하였고, 그에 따른 해석기를 설계하였다.

첫째로 전맹을 위한 서비스는 적절한 태그 분할이 중점이 되었으며, 동적 페이지(Flash, Script 포함)로 연결된 정보를 제외한 텍스트 정보들을 모두 제공 하였다. 그러나 상단 메뉴의 공통적인 정보(페이지 이름, 기본 메뉴) 제공 후에 사용자가 원하는 부분의 정보를 접근하기 때문에 정보 접근의 측면에서는 순차적인 음성 출력의 제약으로 인하여 시간 지연이 발생하였다.

둘째로 저시력자를 위한 서비스는 전체 문서에서 사용자의 터치로 넘어온 링크의 주소와 비교하여 해당 텍스트로 판단되는 부분을 찾기 때문에, 전맹의 사용자보다는 좀 더 쉽고, 명확하게 정보를 파악할 수 있으며, 모든 정보를 음성으로 출력하지 않고 선택된 음성만을 출력하기 때문에 원하는 정보에 쉽게 접근할 수 있다는 장점이 있다.

셋째로 본 연구에서 제안한 제스처에 따른 인터페이스는 음성 서비스의 제어 및 페이지의 이동, 음성 명령의 시작 등의 기본적인 기능들로만 구현되어 있어서 웹 접근성에 대해서 장애를 가진 이들에게 손쉬운 접근을 제공 할 수 있을 것으로 판단한다. 그러나 안드로이드에서 제공하는 터치 API 중 단일 터치(one finger)와 이중 터치(two fingers)가 명확히 구분되지 않아 단일 터치 인식 후 이중 터치 인식으로 넘어가는 단계에서 이중 터치를 별도로 처리하는 시간적 오차는 보완해야 할 문제점으로 지적되었다. 또한 이중 터치 영역에서 터치 경계선이 모호할 경우 원하지 않은 정보의 접근이 이루어져 손가락의 터치 영역에 대한 명확한 처리, 터치 영역의 경계 부분 처리에 대한 명확한 알고리즘 적용이 이루어져야 하겠다.

7. 결 론

본 논문에서는 안드로이드 기반의 스마트폰에서 구글이 제공하는 음성 인식 기능과 음성 합성 기능을 이용하여 시력이 저하되었거나 불편한 사용자의 편의성과 접근성을 향상시키기 위한 방안을 고려했다. 음성 인식의 결과로 구글 음성 인식 시스템은 자주 사용되는 단어들의 경우에 높은 인식성능을 확인할 수 있었지만, 빈도수가 낮은 일반 고유명사나 사람 이름의 경우에 낮은 인식성능을 보여주었다. 음성 합성의 경우 발신자의 정보, 메시지의 정보를 원활히 제공해 주는 것으로 확인 되었다.

모바일 접근성 향상을 위하여 안드로이드 스마트폰에서 응용 프로그램의 차원에서 웹 페이지를 손쉽게 접근할 수 있도록 음성 인식 기능과 음성 합성 기능을 이용하여 웹 서

비스 인터페이스를 설계 및 구현하였다. 웹 리더를 통하여 시각으로 제공되었던 웹 정보들은 스마트폰을 통하여 음성으로 제공될 수 있으며, 이 방법은 전체 운영체제의 변화 없이 시각 장애를 가진 이들에게 웹 접근성을 높일 수 있는 하나의 방안이라고 생각한다. 그러나 보완할 점으로는 범용의 구글 음성 인식을 그대로 사용하지 않고, 업무에 맞는 언어 모델을 사용하거나 독자적인 음성 인식을 이용한다면 한국어의 인식 성능이 높아질 수 있을 것으로 보인다.

본 논문에서 제안한 방법은 범용의 콘텐츠에 적용하기에는 어려움이 따른다. 이를 보완하기 위하여 한국정보통신 표준에서 제안하는 한국형 웹 콘텐츠 접근성 지침 2.0을 통하여, 웹 콘텐츠 개발 시 해당 지침을 적절히 적용한다면, 장애에 구애됨 없이 모든 사람들이 손쉽게 정보를 공유할 수 있을 것이다. 접근성 향상에 따른 수혜자가 장애인인 것은 분명하지만, 장애인들이 편리하게 이용 할 수 있으면, 비장애인 역시 편리하게 이용 할 수 있다는 것을 의미한다. 따라서 웹 접근성 향상은 “장애인을 위한” 것이 아닌 “모두를 위한” 것임에는 분명하다.

참 고 문 헌

- [1] Korean Web Contents Accessibility Guideline 2.0, Korea Communications Commission, 2010. 12. 31.
- [2] The guidelines for promoting the information accessibility and the use convenience for the elderly and disabled (Ministry of Public Administration and Security Notice No.2011-32), 2011. 7. 14
- [3] Mobile Applications Accessibility Guidelines (Ministry of Public Administration and Security Notice No.2011-38), 2011. 9. 22
- [4] Accessibility, Wikipedia, <http://en.wikipedia.org/wiki/Accessibility>
- [5] Web Accessibility Initiative, <http://www.w3.org/WAI/>
- [6] Joon-Ho Hyun, National Information Society Agency, Mobile Accessibility, 2010. 08
- [7] Radio Interface Layer, <http://www.kandroid.org/online-pdk/guide/telephony.html>
- [8] Ji-seong Lee, Naver dominates Korean mobile search, Seoul Economy 2012. 4. 17, <http://economy.hankooki.com/lpage/it/201204/e20120417180315117700.htm>

[부록]

구현된 시스템의 동영상 : <http://youtu.be/ccAsszPsbeM>



오 승 철

e-mail : nonstop232@nate.com

2010년 한남대학교 정보통신공학과(학사)

2012년 한남대학교 정보통신공학과(석사)

2012년~현 재 삼성전자 소프트웨어센터

Convergence Solution 팀

관심분야: 모바일 접근성, 모바일 프로
그래밍



윤 영 선

e-mail : ysyun@hannam.kr

1990년 한국과학기술원 전산학과(학사)

1992년 한국과학기술원 전산학과(석사)

1992년~1995년 ㈜핸디소프트 주임연구원

2001년 한국과학기술원 전산학과(박사)

2001년~현 재 한남대학교 정보통신
공학과 교수

2006년~2007년 한국전자통신연구원 초빙 연구원

2012년~현 재 University of Washington 방문학자

관심분야: 접근성, 모바일 접근성, 음성인식, 음성정보처리 등