

한국디지털미디어고등학교 "S.R.S SE"

거동이 불편한 장애인을 위한 음성인식 컴퓨터 제어 시스템

한국디지털미디어고등학교 HCI Lab(대표 김대욱)에서 개발한 S.R.S(Speech Recognition Secretary) SE는 거동이 불편한 장애인을 위한 음성인식 컴퓨터 제어 시스템으로서 거동이 불편해 키보드와 마우스를 이용할 수 없는 장애인과 발음이 부정확한 장애인들 일지라도, 음성을 이용하여 사용자의 요구에 따라 컴퓨터 및 가전 기기를 제어할 수 있으며, 이러한 기술을 모바일 환경에서도 사용이 가능하도록 개발된 소프트웨어입니다.



SRS는 일상생활에서 사용하는 자연어를 이용하여 컴퓨터가 수행해야 하는 일을 일련의 스크립트로 작성 한 뒤 음성과 1:1매칭 시켜서 그 명령을 수행 합니다. 발음이 부정확한 장애인의 경우 기존 음성데이터 베이스를 활용할 경우 인식률이 저조 하며, 몸이 불편한 장애우들에게 많은 량의 음성 데이터베이스 훈련을 요구하는 것 또한 무리가 있습니다. 따라서 간단한 훈련으로 높은 인식률을 나타 낼 수 있는 Dynamic Time Warping 알고리즘을 사용했습니다.

SRS는 크게 Secretary Server와 Mobile Client로 구성되어 있으며, Secretary Server는 사용자의 PC에 직접 설치되어 음성을 입력받아 명령을 수행하는 기능을 하고, Mobile Client는 원격지에서 컴퓨터에게 명령을 내릴 수 있습니다.

SRS를 이용 할 경우 우리 사회의 장애인과 비장애인간의 정보, 문화적 격차를 해소할 수 있으며, 장애인들이 사회 구성원으로 많은 기여를 하고 경제적으로도 독립함으로써 수동적 복지가 아닌 능동적 복지를 실현 할 수 있는 사회가 될 것이다.

Speech Recognition Secretary SE

1. 작품명 : Speech Recognition Secretary SE

2. 제작자 : 한국디지털미디어고등학교 (HCI Lab)

대표자 : 김 대욱

개발참여자 : 김 대욱

주소 : (425-030) 경기도 안산시 단원구 와동 산 27-1

전화 : 031) 439 - 5578

팩스 : 031) 402 - 8363

Email : kdw234@naver.com

3. S/W 요약설명

거동이 불편해 키보드와 마우스를 이용할 수 없는 장애인과 발음이 부정확한 장애인들 일지라도, 음성을 이용하여 사용자의 요구에 따라 컴퓨터 및 가전 기기를 제어 할 수 있으며, 이러한 기술을 모바일 환경에서도 사용이 가능하도록 개발된 소프트웨어

3.1 개발 배경

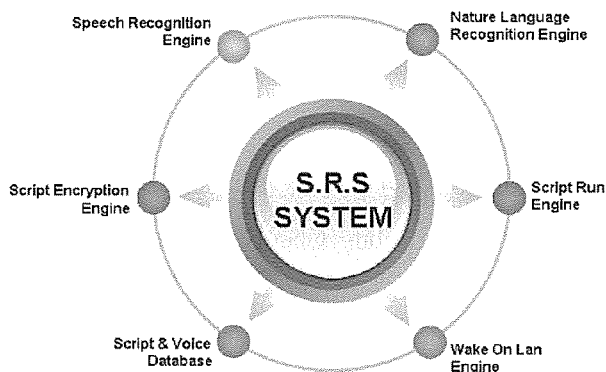
보건복지부 산하 한국 보건사회 연구원에서 조사한 장애인 실태조사 결과로는 2004년 6월 기준으로 국내의 장애인 수는 약 140만 명에 달하고 있습니다. 그러나 사회적 편견을 이유로 장애인들이 등록을 기피하는 현실에 비추어 본다면 실질적인 장애인의 수는 등록 장애인보다 3배 이상 많을 것으로 보건복지부 및 관련기관들은 추산하고 있습니다.

이런 장애인의 정보화 수준은 정부에서 추진한 무료 정보화 교육 등

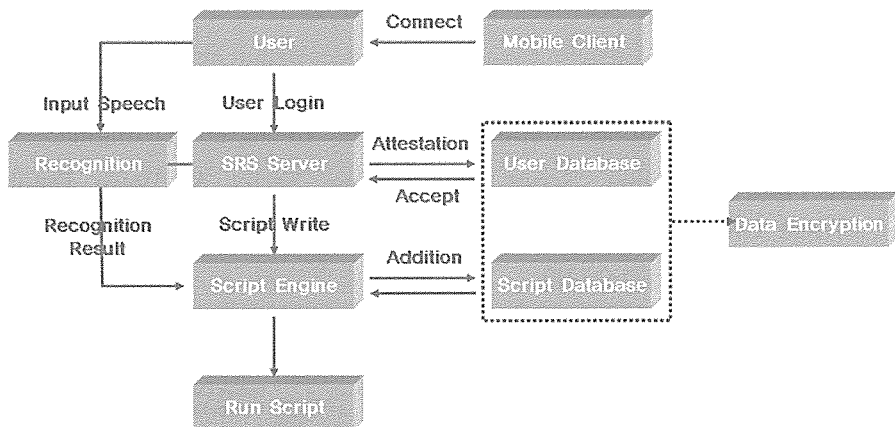
정보격차 해소 정책으로 2000년에 비해 대폭 상승하였으나, 아직도 전체 국민 수준에 비해서는 매우 낮은 실정입니다. 2003년 기준으로 국민 전체의 인터넷 이용률이 65.5%인데 반해, 장애인의 경우에는 27.6%만이 인터넷을 활용하고 있어 37.9% 포인트의 격차가 나타났습니다. 통계수치로 볼 때 장애인의 정보화 수준은 일반 국민과 비교하여 매우 낮은 실정으로 나타났습니다. 미국이나 영국의 경우 각각 일반국민과 장애인의 인터넷 이용률의 격차가 이 14.8% 포인트와 21% 포인트로 우리나라가 정보화 강국으로 부상하였음에도 불구하고 아직까지 장애인에게는 정보화가 다소 거리가 있는 실정입니다. 따라서 장애인 복지를 위해 IT 기술을 적용하는 첫걸음으로 거동이 불편한 장애인들이라도 SRS를 사용하여 원하는 정보를 쉽게 얻고, 장애인, 비장애인들이 차별 없이 모두에게 편리한 컴퓨터 제어 인터페이스를 제공하고자 Speech Recognition Secretary SE를 개발하게 되었습니다.

3.2 시스템 개요

Speech Recognition Secretary SE는 아래 그림과 같이 Speech Recognition Engine, Nature Language Recognition Engine, Script Encryption Engine, Script Run Engine, Script & Voice Database, Wake On Lan Engine으로 구성 되어져 있으며 각 구성 요소 마다 하는 일은 다음과 같습니다.



- Speech Recognition Engine : 입력한 음성을 기존에 저장된 음성과 비교 하여 가장 유사한 결과를 반환한다.
- Nature Language Recognition Engine : 자연어의 뜻을 인식하고 컴퓨터가 처리 할 수 있는 형태의 명령어로 변환한다.
- Script Run Engine : 저장된 스크립트 명령을 순차적으로 수행 한다.
- Script Encryption Engine : 스크립트 데이터베이스의 내용을 암호화 하여 각 스크립트의 내용을 암호화 한다.
- Wake On Lan Engine : 사용자가 원격에서 컴퓨터를 켜고 끌 수 있도록 매직패킷을 생성하여 전송한다.
- Script & Voice Database : 사용자의 음성과 스크립트를 1:1 매칭 시켜 저장



사용자는 Secretary Server에서 컴퓨터 비서 시스템이 수행해야 할 스크립트를 작성하고 스크립트 호출 시 사용될 음성과 1:1 매칭 시켜 Database에 저장한다. Mobile Client는 Server에 접속하여 기존에 작성된 스크립트의 목록을 전송 받고 음성 입력 시 음성 Database에서 가장 유사한 음성을 찾아 그 음성에 할당된 스크립트를 수행한다.

3.3 시스템 특징

Speech Recognition Secretary SE의 특징은 다음과 같습니다.

1) 효율성

- 사용자가 요구에 맞게 기능을 추가 할 수 있어서 효율적인 작업 수행이 가능합니다.
- 일상생활에서 사용하는 자연어를 이용한 스크립트 프로그래밍을 지원하므로 어려운 스크립트 명령어를 외울 필요 없이 누구나 쉽게 기능을 추가 할 수 있습니다.

2) 독창성

- 개선된 Dynamic Time Warping 알고리즘을 사용하여 기존 알고리즘보다 음성 인식이 뛰어납니다.
- 음성 인식 기술을 이용하여 일반인뿐만 아니라 거동이 불편하거나 발음이 부정확한 사용자등 모두에게 편리한 인터페이스를 제공합니다.

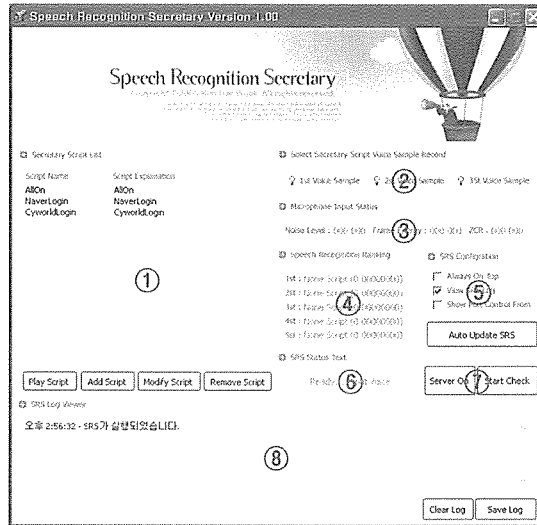
3) 기타 제품과 차별성

- 사용자가 직접 음성데이터 베이스를 훈련하여 이용하기 때문에 미리 구축된 음성 데이터베이스를 이용하는 제품보다 뛰어난 인식을 나타낸다.
- PC에서 사용 할 수 있는 모든 기능을 PDA와 같은 Mobile 단말기를 이용해서 가능하도록 구현 되어서 언제 어디서나 이용할 수 있습니다.

3.4 프로그램구성 및 주요기능

Speech Recognition Secretary SE는 Secretary Server와 Mobile Client로 구성 되어 있으며 다음과 같은 일을 수행합니다.

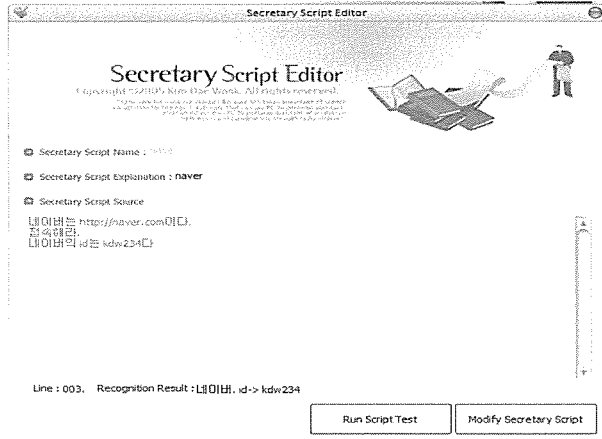
■ Secretary Server - Main Form



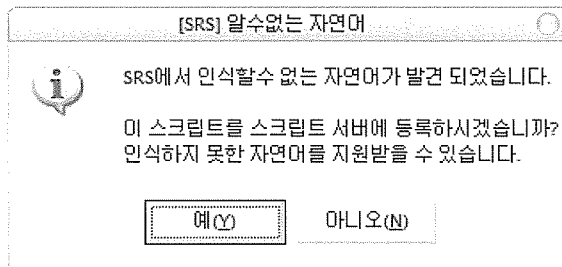
Speech Recognition Secretary SE의 메인 화면은 각 영역마다 다음과 같은 기능을 수행합니다.

- ① 현재 사용자의 컴퓨터에 설치된 스크립트의 목록을 나타냅니다.
- ② 음성인식을 위해 필요한 기준 패턴을 만들기 위해 샘플 음성을 녹음하는 부분입니다.
- ③ 현재 마이크에서 입력되고 있는 상태를 나타내는 부분입니다.
- ④ 음성인식 결과의 순위가 출력되는 부분입니다.
- ⑤ SRS SE의 기본적인 설정을 할수 있습니다.
- ⑥ 현재 마이크입력상태와 기타 여러 상태를 출력합니다.
- ⑦ 음성인식 모드 & 서버 모드를 선택할 수 있습니다.
- ⑧ SRS SE에서 사용된 스크립트 호출 기록 음성녹음 기록 등이 저장되는 프로그램 로그입니다.

■ Secretary Server - Script Editor Form

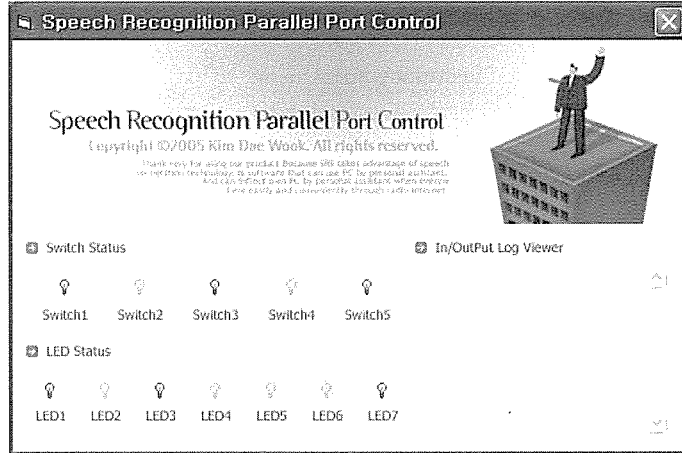


일상생활에서 사용하는 자연어를 이용해 스크립트 프로그래밍이 가능한 Script Editor를 제공합니다. SRS는 이 스크립트 에디터에서 작성한 자연어 스크립트를 이용해서 사용자가 원하는 기능을 제작할 수 있습니다. 간단한 파일 입출력부터 Parallel Port 컨트롤까지 다양한 기능을 제공하여 사용자가 원하는 기능을 제작할 수 있습니다.



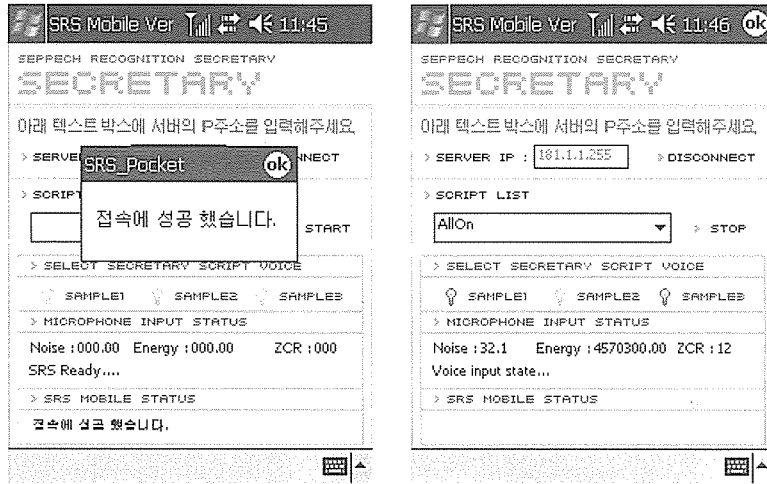
그리고 스크립트 저장 시 사용자가 입력한 자연어 스크립트를 SRS가 인식 하지 못한 경우 위의 그림과 같은 메시지를 출력합니다. SRS는 인식하지 못한 스크립트를 스크립트 서버에 저장해두고 다음 업데이트 시 사용자가 자주 사용했으나, 인식하지 못했던 자연어 스크립트를 업데이트 한다.

■ Secretary Server - Parallel Port Control



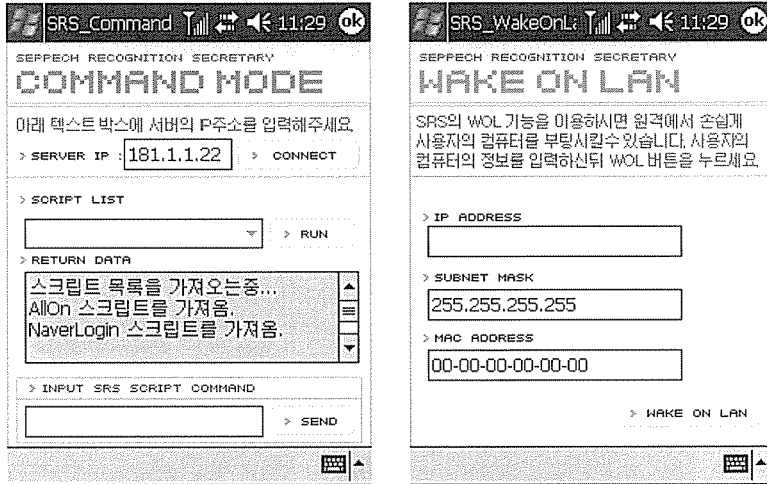
그림은 SRS Server폼에서 Show Print Port Control 버튼을 클릭하면 보이는 화면입니다 SRS Parallel Port Control에서는 유비쿼터스 시대에 가장 널리 사용되고 있는 기능인 홈 네트워크 기능을 SRS를 통해서 구현할 수 있도록 도와줍니다. PC와 연결된 Parallel Port의 신호를 입력받아 Parallel Port에 연결된 기기들의 상태를 제어 할 수 있습니다. Parallel Port Control의 기능을 스크립트로 작성해놓을 경우 음성으로 또한 제어가 가능합니다.

■ Mobile Client - Secretary Mode



서버와 통신을 하며 사용자의 음성을 인식하여 서버로 작업을 수행하도록 명령을 합니다. Server IP에 접속하고자하는 Server 컴퓨터의 IP Address를 입력한 뒤 Connect 버튼을 누르면 Server 컴퓨터에 접속을 시도하고, 접속에 성공하면 Script List를 서버에서 받아 옵니다. Script List에서 스크립트를 선택하게 되면 Select Secretary Script Voice 란에 해당 스크립트의 음성파일의 유무가 나타납니다. 노란색 불이 켜지면 해당 샘플 파일이 있다는 뜻입니다. 불이 꺼진 아이콘을 선택하면 음성 샘플 녹음을 할 수 있습니다. 음성 샘플도 준비가 다 된 상태면 Start 버튼을 클릭하여 음성인식 모드를 시작할 수 있습니다. 처음 1~2초정도는 DC bias 나 잡음의 평균을 추출해내기 위한 작업을 수행하기 때문에 처음 1~2초정도를 입력하지 않으면 좀 더 정확한 음성인식 효과를 얻을 수 있습니다.

■ Mobile Client - Command Mode & Wake On Lan



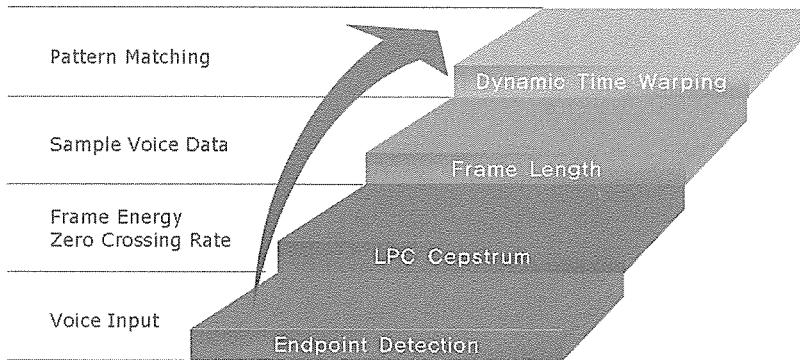
음성을 이용해서 서버에 명령을 내릴 수도 있지만 때로는 음성이 아닌 직접 명령을 내려서 결과를 받아야 할 때가 있을 수 있습니다. 이렇게 Command 모드에서는 스크립트를 바로 선택해서 실행하는 기능과 스크립트 명령어를 직접 타이핑해서 서버에 명령을 내릴 수 있습니다. Mobile Main Client처럼 접속할 서버의 IP Address를 입력한 뒤 Connect 버튼을 눌러서 서버에 접속 합니다. 그러면 서버는 스크립트 리스트를 보내주게 됩니다. 스크립트를 수행하려면 Script List에서 스크립트를 선택한 뒤 Run 버튼을 누르면 서버에서 해당 스크립트가 수행 합니다. 그리고 Input SRS Script Command란에 명령하고자 하는 내용을 직접 입력한 뒤 Send 버튼을 누르면 서버에 입력한 명령어가 서버에 전송되어 서버에 해당 명령이 수행됩니다.

WOL(Wake On Lan) 기술을 이용하여 원격지에서 꺼져 있는 사용자의 컴퓨터를 부팅시킬 수 있습니다. IP Address, Subnet Mask, Mac Address를 입력하신 뒤 Wake On Lan 버튼을 누르면 사용자의 컴퓨터에 컴퓨터를 켤 수 있는 매직패킷을 전송하여 컴퓨터를 부팅시킬 수 있습니다.

3.5 시스템의 특징

■ 개선된 음성인식 알고리즘을 이용한 음성 인식률 향상

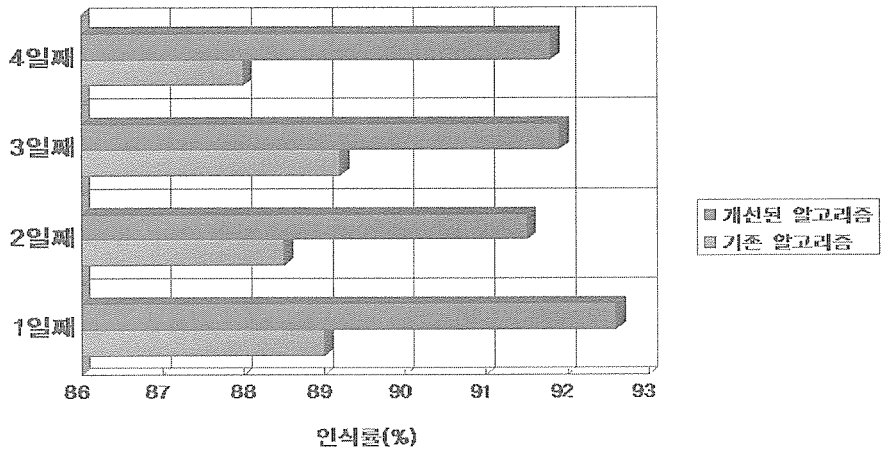
기존 공개된 음성 데이터베이스는 일반인의 음성을 기준으로 제작되어 장애인들이 이를 이용하려면 많은 음성 데이터베이스 훈련이 필요합니다. 따라서 훈련량이 적고 소규모 음성인식에 적합한 Dynamic Time Warping 알고리즘을 이용하여 음성을 인식합니다.



그리고 음성인식률 향상을 이용하여 음성 인식률을 향상 시켰습니다. 이렇게 기존 알고리즘을 개선한 결과 그 성능을 입증하기 위해 다음과 같은 실험을 했습니다.

- 실험 방법 -

- (가) 연령별 10명의 화자가 S.R.S를 이용해 인식률을 측정
- (나) 20개의 단어를 3회씩 발음한 음성을 기준으로 사용하였다.
- (다) 4일간 임의의 시간대를 선정하여 음성 인식을 테스트하였다.
- (라) 기존알고리즘과 개선된 알고리즘의 성능을 비교 분석 하였다.



위와 같은 실험을 한 결과 다음과 같이 기존 알고리즘보다 높은 음성 인식률을 나타 냈습니다.

■ 일상생활에서 사용하는 자연어를 이용한 스크립트 프로그래밍

기존 스크립트 기반 컴퓨터 제어 시스템의 경우 복잡하고 어려운 스크립트 프로그래밍 언어를 새로 익혀야만 이용 가능 했었던 점을 보완하기 위해서 일상생활에서 사용하는 자연어를 이용한 스크립트 프로그래밍을 제공합니다.

- 자연어 스크립트 예제 -

(1) 인터넷 접속 및 정보 검색

네이버는 <http://naver.com>입니다. 접속해주세요.

네이버의 id는 kdw234이고, pw는 1234입니다.

3초간 기다려주세요 네이버의 뉴스에 접속해주세요.

네이버의 Query는 장애인입니다. Search를 전송해주세요.

(3) 홈 네트워크 스위치 제어

3번 스위치를 켜주세요.

2번 스위치를 꺼주세요.

만약에 5번 스위치가 켜져 있다면 3번 스위치를 꺼주세요.

4. 개발단계별 기간 및 투입인원수

구분	개발 일정												인원	
	2003년			2004년										
	10	11	12	1	2	3	4	5	6	7	8	9		
SRS Planing														1
Interface Design														
Speech Recognition														
Script language														
Parallel Port Control														
Network														
SRS Mobile														
Wake On Lan														
Debug & Test														
계	개발기간 : 12개월 , 개발인원 1명													

5. 사용 또는 개발언어, TOOL

PC Application : Microsoft Visual C++ 6.0, Microsoft Visual Basic 6.0

Mobile Application : Microsoft Embedded Visual C++

6. 사용시스템

구분	PC Application	Mobile Application
사용 OS	Windows 98이상 OS	PPC 2003 이상 OS
CPU	펜티엄 4 1GHz 이상	PXA 270 이상
메모리	64MB 이상	32MB 이상
HDD	10 MB 이상	10 MB 이상
Sound Card	내장 사운드카드 이상	
Microphone	필수	