

# SMS 인증 기반의 보이스포탈에서의 음성인식을 위한 CTI 모듈 구현

오세일<sup>o</sup>, 김봉현, 고진환, 박원배  
경북대학교 대학원 정보통신학과

## Voice Portal based on SMS Authentication at CTI Module Implementation by Speech Recognition

Oh Se Il <sup>o</sup>, Kim Bong Hyun, Koh Jin-Hwan, Park won bea  
Dept of Information and Communication Kyungpook National University  
E-mail : ohseil@inc.knu.ac.kr

### 요 약

전화를 통해 인터넷 정보를 들을 수 있는 보이스 포탈(Voice Portal) 서비스가 인기를 얻고 있다. Voice Portal 서비스란 알고자 하는 정보를 Speech Recognition System에 음성으로 명령하면 전화를 통해 음성으로 원하는 정보를 듣는 서비스이다. Authentication의 절차를 수행하는 SMS (Short Message Service) 서버 Module, PSTN과 Database 서버사이의 Interface를 제공하는 CTI (Computer Telephony Integration) Module, CTI 서버와 WWW (World Wide Web) 사이의 Voice XML Module, 정보를 검색하기 위한 Searching Module들이 필요하다. 본 논문은 Speech Recognition technology를 기반으로 한 CTI Module 설계를 구현하였다. 또한 인정 방식으로 Random한 일회용 password를 기반으로 한 SMS Authentication을 택하므로 더욱 더 안정된 서비스 제공을 목적으로 하였다.

#### 1. 서론

최근 인터넷 검색 정보를 phone을 통해 들을 수 있는 Voice Portal 서비스가 인터넷의 새로운 시장으로 급부상 되고있다[3]. Voice Portal 서비스란 이용자가 phone을 걸어서 Keyword를 말하게되면 Speech Recognition system이 인터넷을 검색해 관련정보를 phone로 들려주는 서비스다. PC를 사용하지 않고도 인터넷 정보검색이 가능하며, Interface가 phone이므로 유동성과 신속성이 보장된다는 점에서 더욱 더 서비스의 전망을 밝게 하고 있다. 음성으로 전해듣는 서비스로는 이미 UMS (Unified Messaging System)나 무선 인터넷 음성 이메일 서비스 등이 있지만, Voice Portal 서비스는 음성으로 명령을 내린 후 다시 음성으로 인터넷 정보를 듣는다는 점에서 전혀 차원이 다른 서비스이다. Voice Portal 서비스의 핵심 기술들은 Speech Recognition, Voice Synthesis, CTI, Voice XML등으로 거론 될 수 있다[1][2][4][6].

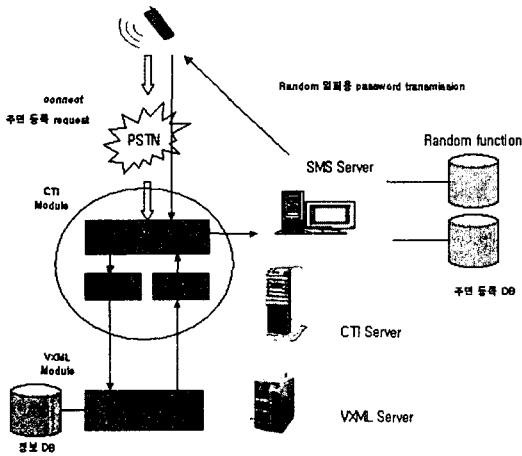
Speech Recognition과 Voice Synthesis는 IBM, 모

토롤라, L&H, Speechwork, Microsoft사 등이 세계적으로 그 기술을 인정받고 있다. 이 중 Microsoft사의 Speech Group에서 자체 개발 한 Speech Engine을 사람들에게 무료로 배포함으로써 Speech Recognition 및 Voice Synthesis에 개발 속도를 가중시키고 있다. MS사의 Speech Engine은 Win32기반의 ActiveX 컴포넌트형태로 VUI (Voice User Interface)구현이 아주 용이하다. 현재 이 엔진은 SAPI (Speech API)의 이름으로 5.0 버전까지 출시되었으며 중국어, 일본어까지 지원 가능하다. 하지만 안타깝게도 한국어는 지원하지 않는다.[7][8] CTI 핵심으로는 흔히 이용되고 있는 MS사와 Intel에서 개발한 TAPI (Telephony API)을 들 수 있다. 이 TAPI는 phone기능을 제공하는 응용프로그램 개발을 위한 Interface며, Voice Portal 서비스에서 CTI 서버가 phone망과 Interface 유지 역할을 담당한다[9]. 본 논문의 Authentication 방식은 SMS Authentication 방식을 기반으로 보다 안정된 시스템을 구현하였다. SMS Authentication

방식이란 휴대폰에 Random한 일회용 password를 SMS 전송 Authentication하는 방식을 말한다[5].

Authentication 후 CTI Module을 통한 Automatic Speech Recognition을 거쳐 원하는 정보를 Voice XML에 의해 처리 후 음성으로 다시 전해준다. 다시 말하면, SMS 전송 통한 Authentication을 거친 후 알고자 하는 회사의 주식종목을 말하면 음성 인식 후 해당 회사의 주식정보를 음성으로 들려주도록 하였다. Speech Engine으로 SAPI를 사용하였고 phone망과의 Interface는 TAPI를 사용하였다.

2. Voice Portal System 구조



Architecture of Voice Portal System

Voice Portal System의 Architecture는 Random 일회용 password Authentication의 절차를 수행하는 SMS Server와 PSTN과 Voice XML Server간의 Interface를 유지하는 CTI Module과 Voice XML의 파싱 (Parsing) 기능, 주식검색엔진 (Stock Searching Engine)과의 Interface를 유지하기 위한 Voice XML Module, WWW 정보를 검색하는 Searching Engine Module로 크게 나눌 수 있다.

2.1 CTI Module

사용자 Interface를 지원하는 Module이다. 사용자로부터 검색하고자 하는 Keyword를 받아서 Voice XML 프로세스에게 전달하는 역할을 한다. 이 CTI Module의 주 기능은 다음과 같다

▶Authentication : 주민등록번호와 암호를 DTMF (Dial Tone Multi Frequency)로 인식한다. SMS 서버에서 발생된 Random 일회용 password를 발송 및 검증한다.

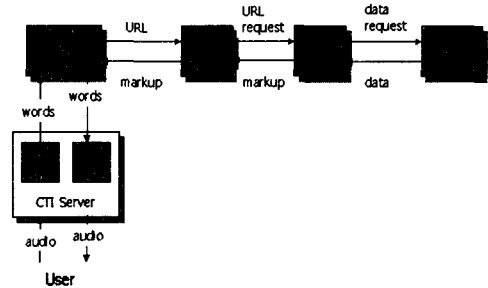
▶STT (Speech-To-Text) : 사용자 Keyword를 Speech Recognition을 통해서 Text로 변환하며 Voice XML 프로세스와 Data 교환을 주로 Text 기반으로 한다.

▶TTS (Text-To-Speech) : Voice XML에서 넘겨받은 Text를 Voice Synthesis를 통해 사용자에게 전달한다. Voice Font에 의한 다양한 음성으로 서비스가 가능하다.

▶처리과정 모니터링(Process Monitoring) : 사용자의 서버 접속부터 연결이 끊어질 때까지 사용자의 이용 프로세스를 모니터링 한다.

2.2 Voice XML Module

인터넷 기반의 콘텐츠와 정보를 음성으로 서비스할 수 있는 문서로 만들기 위한 Module로서 일종의 XML형식의 문서 처리를 한다. WWW과 연동하여 필요한 정보를 검색할 수 있도록 구현하였다.



The Architecture of Voice XML Process

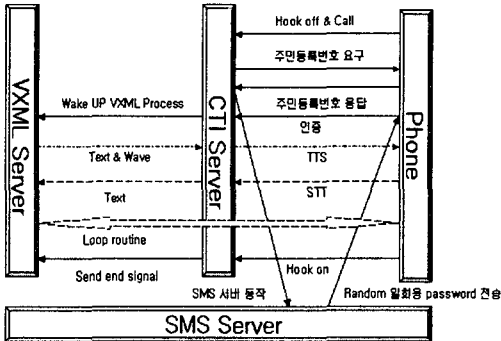
2.3 CTI Module 동작 절차

- 1) CTI 서버에 접속되면 먼저 주민등록번호를 묻는다. 주민등록번호가 일치한다면 그 번호에 부합하는 폰 번호로 SMS 전송을 하게 된다.
- 2) SMS로 전달 된 Random한 일회용 password를 검증한다.
- 3) 검증이 끝난 후 PSTN망을 통해 들어오는 Voice를 먼저 Microsoft사의 Speech Engine을 통해 Speech Recognition 과정을 거쳐 Text로 변환된다.
- 4) Control Socket을 통하여 먼저 Voice XML Process에게 Control Signal을 보내 Voice XML Process를 깨운다.
- 5) Data Socket을 통하여 Keyword에 해당하는 Text를 Voice XML Process에게 전송한다
- 6) Keyword에 대응하는 정보를 DB를 통해서 처리한다.
- 7) Text형식으로 CTI 서버에 전해진다.
- 8) CTI 서버에서는 TTS Engine을 통하여 Voice Synthesis 시킨 후 TAPI(Telephony Application Protocol Interface)를 통해 해당 사용자에게 서비스한다.

2.4 서버간의 통신

자바로 구현되어진 Voice XML 프로세스와 Visual Basic으로 구현되어진 CTI 프로세스간의 Data 교환

은 Socket을 이용하도록 구현하였다. 포트 1944번은 Control Socket으로 제어를 위한 Control Signal을 보낼 때 사용되며, 주로 Voice XML Browser를 깨울 때, 사용자의 접속해제, Phrase Parameters의 통신유무를 알릴 때 사용한다. 포트 1943번은 Data Socket으로 Keyword를 비롯한 data 교환을 위해 만들어 놓은 Socket이다.



Communication for Processes

### 3. Voice Portal System 구현

#### 3.1 구현 환경

본 논문에서 구현한 CTI Module은 Visual Basic을 이용하여 하나의 어플리케이션으로 구현되어있다. 구현환경은 다음과 같다.

▶ Software

- 서버 System/운영 체제 : WindowNT 4.0
- 사용 언어 : Visual Basic 6.0
- Automatic Speech Recognition엔진 : Microsoft SAPI(Speech API)
- 데이터베이스 : Microsoft Access
- Development Tool Kits : Parity Software CallSuite

▶ Hardware

- CTI board : Dialogic D/41ESC

#### 3.2 SAPI를 이용 Speech Recognition 구현

처음 폼이 Load 시킬 때 사용할 Speech Engine도 같이 초기화 및 Load 시킨다. 다음의 소스는 폼이 로드 될 때의 소스이다.

```
Private Sub Form_Load()
Dim WaveID As Integer
Dim IType As Long
If Command <> "" Then
    VBocx1.PhoneLine = Val(Command)
Else
    VBocx1.PhoneLine = 1 'Use Phone Line
End If
'Automatic Speech Recognition Engine initialization
IType = GRAMMAR_CFG
```

```
MBocx1.EngineMode = "English for telephone
(Microsoft Command & Control)"
Call
MBocx1.AssignDialogicDevice(VBocx1.PhoneLine)
Call
MBocx1.LoadGrammar(App.Path & "\factory.grm"
, IType)
MBocx1.ActivateRule (App.Path & "\factory.grm")
VBocx1.Log = LOG_Detailed
Call VBocx1.OnHook
Call Trace("Application started")
'Process monitoring
End Sub
```

#### 3.3 SAPI를 이용 Voice Synthesis 구현

Voice Synthesis는 Voice XML Module로부터 Data Socket을 통해 Data를 받으면 처리한다. 먼저 CTI Board에 출력라인을 할당하고 그 출력라인으로 Voice XML로부터 받은 Data를 Voice Synthesis 하여 보낸다. Voice Synthesis는 TAPI에서 정의된 메소드 Speak 메소드를 이용한다.

```
Private Sub datask_DataArrival(ByVal bytesTotal
As Long)
Dim Receive_Data As String
Dim stt As String
datask.GetData Receive_Data
Trace ("Receive Data=" & Receive_Data)
'Output line allocation
Call
CBocx1.AssignDialogicDevice
(0, VBocx1.PhoneLine)
'Output line, From the Voice XML the data voice
synthesis which it receives
Call CBocx1.Speak(Receive_Data, 0, -True)
stt = Recognition()
'Automatic Speech Recognition
End Sub
```

#### 3.4 에러 처리

phone을 끊을 때 또는 갑작스런 에러가 발생 시 처리하는 이벤트가 아래의 코드처럼 Hangup 이벤트와 ErrorTrap 이벤트이다. CTI 서버에게 강제로 phone망과의 단절을 만들고 대기상태로 Return된다.

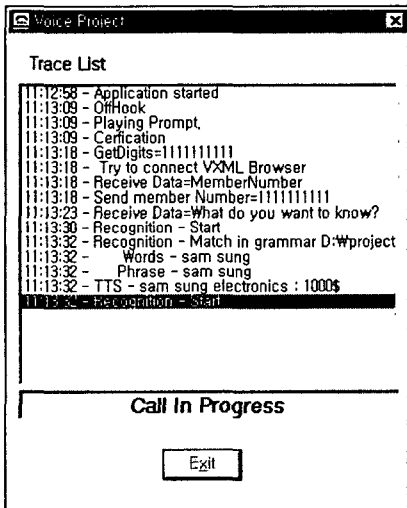
```
Hangup:
'Set Text to indicate Waiting for Call
StatusBar.Caption = "Waiting for Call"
'Set phoneline on-hook
VBocx1.OnHangupErr = False
VBocx1.OnHook
VBocx1.OnHangupErr = True
Call Trace("OnHook")
'Enable the exit button now that the call is
```

```

complete.
ExitButton.Enabled = True
Exit Sub
ErrorTrap:
'If it is a hangup, exit normally
If VBocx1.HangupDetected Then
'Set Text to indicate Waiting for Call
StatusBar.Caption = "Waiting for Call"
Call Trace("Caller hung up.")
VBocx1.OnHangupErr = False
VBocx1.OnHook
VBocx1.OnHangupErr = True
'Enable the exit button now that the call
is complete.
ExitButton.Enabled = True
Exit Sub
End If
'Otherwise, log the error
Call Trace("Error #" & Err.Number & ": " &
Err.Description)
Beep
GoTo Hangup
End Sub
    
```

4. 구현 결과 및 검토

CTI 서버에 동작하는 어플리케이션은 Trace List 를 통해 현재 접속해 있는 사용자의 시간 때별로 모니터링 하여 출력하도록 Interface를 꾸몄다. 사용자가 접속한 상태에서는 이 프로세스를 강제로 종료를 시킬 수 없도록 하였다.



The executive scene of CTI modul

Interface를 통해 사용자가 정보를 검색하고, 검색된 데이터를 Voice XML 프로세스로부터 얻어 사용자에게 Voice Synthesis를 통해 서비스 되도록 한다. 프로세스간의 통신으로 인한 데이터 검색 지연시간을 줄이기 위해 Text 기반의 통신을 택하였다. Speech Engine으로는 MS사의 Speech Group에서 무료로 배포하는 SAPI를 사용하여 Recognition 및 Synthesis를 구현하였다. SAPI는 화자 독립방식 Speech Engine이며, 그 인식률은 Native일 경우 약 95%정도에 이른다. 구현 된 System은 다음의 개선 사항이 있다. Speech Engine인 SAPI는 중국어, 일본어까지 지원하지만 아직 한국어를 지원하지 않는 단점이 있다. 따라서 본 논문에 구현된 Voice Portal은 영어를 기반으로 구현되었다. 차후 1~2년 안에 한국어를 지원하리라 예상되며, 본 System도 MS사의 한,영 버전으로 Update 해야 할 것이다. 또한 DB의 확장성을 키워서 접속해 있는 사용자 정보를 System 관리자가 쉽게 알 수 있는 Interface와 Phone Line의 확장으로 인한 멀티프로세스에 의한 CTI Module 을 구현 할 것이다.

참고 문헌

- [1] 김희동 “컴퓨터텔레포니결합(CTI) 기술”, 전자공학회지 제24권, pp30-39, 1997.7.
- [2] C.R.Strathmeyer, "An Introduction to Computer Telephony", *IEEE Comm Magazine*, pp.106-111.,1996. May
- [3] “음성인식의 현주소” 마이크로소프트웨어 9월호 pp254-279
- [4] “음성서비스 작성언어 VoiceXML” 마이크로소프트웨어 9월호 pp268-280
- [5] 오세일, 박원배 “Login Establish Forms through Wireless Systems” CICE2000 pp91-94 2000.11
- [6] 구명완 “음성인식기술의 현황과 전망”, 전자공학회지 제20권, pp41-50, 1993.5
- [7] SAPI(Speech API) SDK 5.0 <http://www.microsoft.com/it/>
- [8] SAPI(Speech API) online document <http://microsoft.com/IIT/projects/sapisdk.htm>
- [9] “Introduction to TAPI 3.0” [http://msdn.microsoft.com/library/backgrnd/html/msdn\\_tapi\\_30.htm](http://msdn.microsoft.com/library/backgrnd/html/msdn_tapi_30.htm)
- [10] <http://www.parity.com/>
- [11] <http://www.newgen.co.kr/solution/ums.html>
- [12] <http://lead2000.sicc.co.kr/contents/theme/theme9708.htm>
- [13] “Special Issue on Computer Telephony”, *IEEE Comm Magazine* 1996. April
- [14] 전자신문 “CTI업계, 음성인식 사업 강화”, 1999.11.11

본 논문에 구현 된 Speech Recognition based on SMS Authentication을 이용한 CTI Module은 보다 안정 된 Authentication을 거쳐 phone이라는 단일한