전화망을 통한 자동음성번역 서비스 시스템 설계

이성주, 이영직, 양재우
한국전자통신연구원

Design of an Automatic Speech translation system on the Telephone Line

Sung-Joo Lee, Yunggik-Lee, Jea-Woo Yang
Electronics and Telecommunications Research Institute
Email: lee1862@etri.re.kr

요 약

본 논문에서는 현재 ETRI에서 개발 중인 유/무선 전화망을 통한 다국어간 대화체 음성번역서비스 시스템에 대해서 소개한다. 전화망을 통한 자동음성번역서비스 시스템은 여행대화영역을 서비스 대상영역으로 하고 있고 자동음성번역서비스를 필요로 하는 사용자들은 동일한 장소에서 대면하고 있으며 서로 다른 언어를 사용하기 때문에 서로 의사 소통에 어려움을 겪고 있다고 가정한다. 따라서 여기서 말하는 자동음성번역 시스템의 특징은 인간과 기계간의 인터페이스를 그 대상으로 하는 것이 아니라 인간과 인간사이의 인터페이스를 그 대상으로 하고 있다는 점이다. 인간과 인간사이의 인터페이스 상황에서는 인간의 이해력이 시스템 오류를 정정할 수 있는 여지를 지니고 있다. 따라서 시스템이 사용자의 말하는 의도 혹은 개념만 잘 전달할 수 있다면 서로 다른 언어를 사용하는 사용자들 사이에서도 이러한 시스템을 통한 의사소통이 가능하다. 자동음성번역서비스 시스템은 크게 음성인식모듈, 문장해석 및 번역 모듈, 음성합성모듈, 시스템통합 모듈 그리고 전화망 인터페이스 모듈로 나뉜다. 여기서는 자동음성번역 서비스 시스템의 각 모듈들의 주요 특징과 상호 인터페이스 방법에 대해서 소개한다.

1. 서론

C-STAR[1] 국제공동연구과제에서는 C-STAR III 과제를 2000년부터 시작하고 있다. C-STAR 국제공동연구과제의 목적은 자동음성번역 연구 분야에서 국제간의 협력 증진을 그 목표로 하고 있으며 1991년에 공식적으로 발족하였다. 현재 7개의 회원국으로 구성되어 있는데 그 구성은 다음과 같다.

- ATR(일본), CMU(미국), IRST(이태리), CLIPS(프랑스), ETRI(한국), NLPR(중국), UKA(독일)

2000년부터 시작된 C-STAR III 과제에서는 그 동안의 자동음성번역 분야의 연구성과를 바탕으로 유/무선 전화망을 이용한 자동음성번역 서비스 시스템을 개발하고 이를 바탕으로 실제 서비스 과정에서 음성 데이터를 수집하려고 한다. 이러한 실제 음성 데이터는 자동음성번역 시스템의 성능향상에 기여할 수 있을 뿐만 아니라 자동음성번역연구 분야의 연구를 활성화하는데 기여할 것으로 기대한다.

여기서는 자동음성번역을 필요로 하는 두 사용자가 서로 대면하고 있는 상황을 그 서비스 대상으로 하는데 이러한 상황이 해외 여행자가 서로 다른 언어장벽으로 인하여 어려움을 겪는 실제상황과 보다 유사하다고 할 수 있다. 따라서 사용자 전화망 인터페이스 시스템은 이러한 상황을 지원할 수 있도록 설계되어야 한다. 한국전자통신연구원에서는 2000년 C-STAR 회의에서 다이얼로직 CTI 보드[2]를 지원하는 전화망 인터페이스 API를 개발하여 자동음성번역 서비스 시스템과 연동, 시연한 바 있다. 그리고 일본 Advanced Telecommunication Research Institute International (ATR)에서 자동음성번역 서비스 시스템의 전체 형상을 제안하였고 이를 기반으로 하여 미국 CMU에서 전화망 인터페이스 모듈의 첫 번째 버전을 2001년 초에 배포하였다. 한국전자통신연구원에서는 2001년 10월 전화망 인터페이스 모듈을 개발 2002년 3월 중국에서 개최된 C-STAR 회의에서 중국측 공동연구 파트너인 NLPR과 자동음성번역 서비스 시스템의 국제공동시연을 보인 바 있다. 여기서는 이러한 자동음성번역서비스 시스템의 주요 특징과 설계방법 등에 대해 소개한

다.

서론에 이어 2장에서는 자동음성번역 서비스 시스템의 전체 형상을 간단히 설명하고 3장에서는 이러한 자동음성번역서비스 시스템과 사용자 전화망 인터페이스를 위한 전화망 인터페이스 모듈에 대해 설명한다. 그리고 마지막 4장에서는 전화망 인터페이스 모듈의 이슈들에 대해 설명하고 결론을 맺는다.

2. 자동음성번역 서비스 시스템

C-STAR III 과제를 위한 자동음성번역 서비스 시스템의 형상은 C-STAR II 과제를 통하여 개발한 자동음성번역 서비스 시스템의 형상에 기반을 두고 있다. 하지만 C-STAR III 시스템의 경우 전화망을 통한 사용자 서비스를 대상으로 하고 있기 때문에 자동음성번역 시스템과 전화망과의 인터페이스를 위한 전화망 인터페이스 모듈이 필요하게 된다.

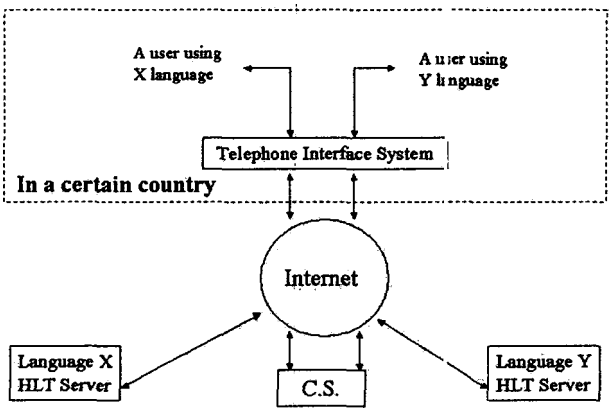

그림1 . 자동음성번역 서비스 시스템의 형상

그림 1은 다국어를 지원하는 전화망을 이용한 자동음성번역 서비스 시스템의 전체 형상을 나타낸 것이다. 여기서 HLT 서버는 Human Language Translation 서버의 약자로 HLT 서버는 음성 인식기, 기계번역기, 음성 합성기와 이들 세 가지 모듈을 통합하고 제어하는 모듈로 구성되어 있다. 그리고 Communication Switch(CS)[1]는 C-STAR II 과제를 통하여 개발된 모듈로 Interchange Format(IF)[1]의 전송을 위한 모듈이다. IF 역시 C-STAR II 과제 수행 시 개발 된 하나의 중간언어이며 현재에도 계속적인 연구가 진행 중에 있다. 위 그림에서 보는 바와 같이 X 언어를 사용하는 사용자는 유/무선 전화기를 이용하여 Y 언어를 사용하는 사용자와 대화할 수 있으며 이러한 자동음성번역 서비스 시스템간의 데이터 송수신은 인터넷 소켓을 이용하게 된다.

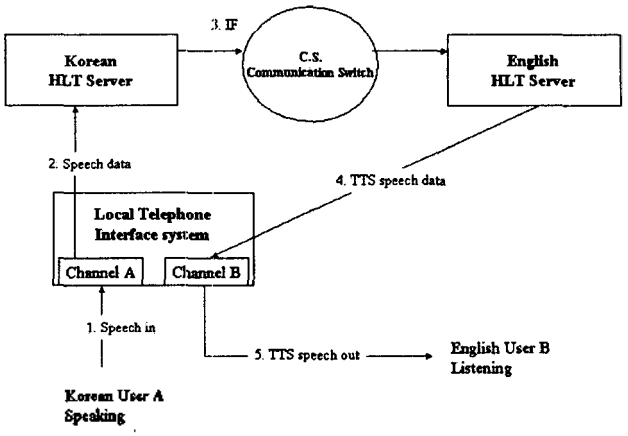

그림 2. 자동음성번역서비스 시스템의 데이터 흐름 예

그림 2는 한국어 사용자와 영어 사용자간의 자동음성번역 서비스의 예를 나타낸 것이다. 한국어 사용자가 유/무선 전화기를 이용하여 대화를 시작하면 사용자의 음성데이터는 전화망 인터페이스 모듈을 통하여 한국어 HLT 서버로 전송된다. 한국어 HLT 서버에서는 음성인식과 기계번역의 과정을 거쳐 음성데이터를 IF로 변환하고 이러한 IF는 CS를 통하여 영어 HLT 서버로 전송되어진다. 그러면 영어 HLT 서버에서는 IF 번역과 음성합성과정을 거쳐 합성된 음성데이터를 전화망 인터페이스 모듈로 전송하고 전화망 인터페이스 모듈은 이를 영어를 사용하는 사용자에게 들려줌으로써 영어로 된 번역 음성을 영어사용자가 들을 수 있게 되는 것이다.

3. 자동음성번역시스템을 위한 전화망 인터페이스 모듈설계

앞에서 언급한 바와 같이 전화망을 이용한 자동음성번역 서비스 시스템은 통역서비스를 필요로 하는 두 사용자가 서로 대면하고 있으며 각각 전화를 사용할 수 있는 상황을 그 서비스 대상으로 하고 있다. 따라서 통역서비스를 위해서는 두 개의 전화채널이 필요하며 서비스 시스템에 전화를 거는 사용자는 대화를 원하는 상대방의 전화번호를 입력할 필요가 있다.

3.1. 언어 할당 방법

C-STAR III 과제에 참여하는 회원국은 모두 7개 나라로 전화망 인터페이스 모듈은 7개 국어를 지원하여야 한다. 전화망 인터페이스 모듈의 입장에서 보면 하나의 통역서비스를 제공하기 위해서 전화를 받는 채널과 전화를 거는 채널의 두 전화채널이 필요하고 각국의 언어를 사용자가 선택하게 하는 방법으로는 크게 전화번호와 전화버튼(DTMF)의 두 가지 방법이 있다. 여기

서 전화를 받는 채널을 소스(source)채널이라 하고 전화를 거는 채널을 데스티네이션(destination)채널이라고 하자.

1) 전화번호 방법: 특정 언어에 특정 전화번호를 할당하는 방법
2) 전화버튼 방법: 특정 언어에 특정 전화버튼을 할당하는 방법

따라서 소스채널과 데스티네이션채널에 특정 언어를 할당하는 방법으로 다음과 같이 크게 3가지 방법을 들 수 있다.

1) 소스(전화번호), 데스티네이션(전화번호)
- 장점: 전화번호만으로 통하여 두 사용자의 사용언어를 모두 알 수 있다.
- 단점: 소스채널에만 7*6=42개의 채널을 할당하여야 하므로 많은 비용이 든다.

2) 소스(전화버튼), 데스티네이션(전화버튼)
- 장점: 전화채널 두 개로 통역서비스가 가능하므로 비용이 저렴하다.
- 단점: 통역 서비스 시 어떤 언어로 된 안내메시지로 서비스를 진행하여야 할 지 알 수 없고 사용자가 안내메시지를 이해하지 못 할 경우 서비스 진행이 불가능하다.

3) 소스(전화번호), 데스티네이션(전화버튼)
- 소스채널 7개가 필요하고 사용자의 언어를 전화번호에 할당하였으므로 서비스 진행이 용이하다. 따라서 비교적 적은 채널(적은 비용)로 서비스가 가능하다.

따라서 여기서는 비교적 적은 수의 전화채널로 통역서비스가 가능한 세 번째 방식을 전화망 인터페이스 모듈에 적용하기로 하였다.

3.2. 서비스 흐름

앞장에서 설명한 언어할당 방식을 적용한 전화망 인터페이스 모듈의 서비스 흐름은 다음과 같다.

1) 통역서비스를 필요로 하는 사용자가 전화망 인터페이스 모듈의 소스채널로 전화를 건다.
2) 소스채널은 사용 가능한 데스티네이션 채널이 있는 경우 사용자에게 대화상대방의 언어 전화버튼 입력을 유도하기 위한 안내메시지를 들려준다. 그렇지 않은 경우에는 서비스를 진행할 수 없음을 사용자에게 알리고 전화를 끊는다.
3) 사용자가 언어버튼을 입력하면 이 번호가 유효한 버튼인지를 확인하고 사용자가 버튼을 잘 못 입력한 경우 다시 언어버튼을 입력하도록 유도하는 안내 메시지를 들려준다.
4) 사용자가 언어버튼을 입력하게 되면 전화망 인터페이스 모듈은 서비스에 필요한 HLT 서버들의 정보를 얻게 되고 이를 바탕으로 두 HLT 서버에 접속을 시도한다. 만약 접속이 실패한 경우 사용자에게 접속에 실패하였다는 정보를 알려주고 서비스를 종료한다.
5) 사용자에게 대화를 원하는 상대방의 전화번호를 입력하도록 유도하는 안내메시지를 들여주고 사용자의 전화입력을 기다린다.
6) 사용자가 전화번호를 입력하면 데스티네이션 채널을 열고 상대편으로 전화를 건다. 전화접속에 실패한 경우 서비스를 종료하도록 유도하는 메시지를 사용자에게 들려준다.
7) 전화가 성공적으로 연결된 경우 사용자 모두에게 별표버튼을 누른 후 대화를 시작하라는 메시지를 들려주어 서로 대화를 유도한다.
8) 자동통역 서비스를 계속 진행한다.
9) 전화연결이 끊어지면 서비스를 종료한다.

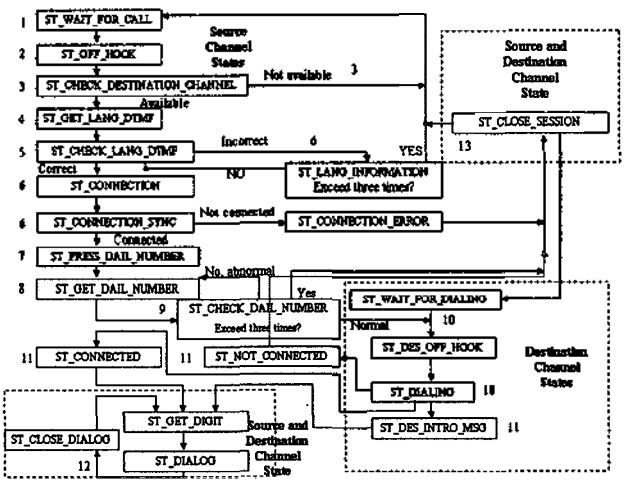

그림 3. 전화망 인터페이스 모듈의 상태도

그림 3은 앞에서 설명한 서비스 흐름에 따른 전화망 인터페이스 모듈의 상태도이다. 이러한 상태도는 다이얼로직 CTI 보드를 이용하여 Asynchronous 방식[2]으로 전화망 인터페이스 모듈 설계 시 반드시 필요한 요소이다. 여기서는 소스채널과 데스티네이션채널이 상태를 따로 사용하거나 공유하게 되는데 이때 상태를 공유하는 부분은 사용자의 음성입력을 바탕으로 자동음성번역서비스가 진행되는 부분과 서비스가 종료되는 부분이며 그 외의 상태는 서로 공유하지 않고 있다.

4. HLT 서버

Human Language Translation (HLT) 서버는 크게 네 가 부분으로 나눌 수 있으며 다음과 같다.

1) 음성 인식기
2) 음성 합성기
3) 문장 해석기 및 문장 생성기 (기계번역모듈)
4) 제어 및 통합 모듈

4.1. 음성 인식기

음성 인식기는 대화체연속음성을 그 인식 대상으로 하고 있다. 무제한 대화체연속음성인식의 경우 그 성능이 현저히 낮기 때문에 여기서는 여행대화영역만을 그 인식 대상으로 하여 그 성능을 보장하려 하였다. 하지만 여행대화영역 또한 그 영역이 광범위하여 음성인식 성능을 보장할 수 없으므로 그 영역을 세분화하는 방법이 필요하다. 따라서 여행대화영역을 9가지 영역으로 세분화하였는데 그 내용은 다음과 같다.

1) 길안내 영역
2) 호텔 영역
3) 기차역(전철역) 영역
4) 상점 영역
5) 고속버스터미널 영역
6) 식당 영역
7) 택시 안 영역
8) 공항 영역
9) 긴급상황 영역

현재 구현된 음성 인식기는 여행계획 영역을 그 대상으로 하고 있는 연속음성 인식기를 위의 9가지 영역으로 확장하기 위하여 노력하고 있다.

4.2. 음성 합성기

음성 합성기는 음성데이터 코퍼스 기반의 음성합성 알고리즘을 기반으로 하고 있다[4]. 그리고 이를 기반으로 하여 보다 자연스러운 사용자 인터페이스를 위하여 대화체 음성 합성기 개발이 진행 중이며 남녀의 성별에 따른 대화체 음성 합성기를 현재 개발하고 있다.

4.3. 기계 번역 모듈

현재 기계 번역 모듈은 개념 기반 번역에 그 초점을 두고 있다[1]. 앞에서 설명한 바와 같이 자동음성번역 시스템은 기계와 인간 사이의 인터페이스를 대상으로 하는 시스템이 아니라 인간과 인간 사이의 인터페이스를 그 대상으로 하고 있는 시스템이므로 인간의 이해력으로 어느 정도의 에러 정정이 가능하다고 생각하고 있다. 따라서 말한 사람의 의도나 개념만을 잘 전달할 수 있다면 서로 언어가 통하지 않는 두 사람이 대화하는데 많은 도움을 줄 수 있고 계속적인 대화가 가능할 것으로 생각하고 있다. 따라서 기계 번역 모듈은 개념 기반 알고리즘에 그 기반을 두고 있으며 현재 개념 기반의 기계번역에 대한 연구가 기계번역을 담당하는 C-STAR Working Group에 의해 현재 계속 진행 중에 있다.

4.4. 제어 및 통합 모듈

제어 및 통합 모듈은 음성 인식기, 음성 합성기 그리고 기계번역모듈의 세가지 구성 요소를 통합, 제어하는 기능을 수행하며 Communication Switch(CS)와 접속, IF 및 데이터 전송의 기능을 수행하는 모듈이다[3]. 현재 다중 사용자 접속 시 이를 지원할 수 있는 형태의 제어 및 통합 모듈 형태로 개발이 진행 중에 있다.

5. 결론

C-STAR III 국제공동연구과제에서는 유/무선 전화망을 이용한 자동음성번역 서비스 시스템을 구현하기 위하여 노력하고 있으며 이러한 시스템을 통하여 실제 자동음성번역 시스템을 통하여 사용자들의 음성데이터를 수집하려고 한다. 이러한 실제 음성 데이터들은 다국어 자동음성번역의 성능을 향상 시킬 수 있을 뿐만 아니라 자동음성번역연구 분야를 연구 활동을 보다 활성화 시킬 수 있을 것으로 기대하고 있다. 따라서 본 연구원에서는 해외 여행자들을 위한 자동음성번역 시스템을 개발 중에 있으며 이러한 시스템을 통하여 서로 다른 언어장벽으로 인하여 어려움을 겪는 해외 여행자들에게 도움을 줄 수 있기를 바라고 실제 서비스를 통하여 얻어진 음성 데이터들이 자동음성번역 연구 분야의 연구를 더욱 활성화하는 역할을 할 수 있기를 바라고 있다.

참고 문헌

[1] Jun Park, Kyuwoong Hwang, Un-Cheon Choi, Junko Hosaka, Siong Hun Yi and Jae-Woo Yang: "Spoken language Translation System: Development and Demonstration", ICSP'99, Volume 2 of 2, pp.535-538, Korea, 1999.

[2] http://www.intel.com/network/csp/trans/dialogic.htm

[3] SiongHun Yi and Jun Park, "Development of communication interface in spoken language translation system," ICSP99, Volume 2 of 2, pp.553-556, Korea, 1999

[4] Sanghun Kim, Dong-Gyu Kang, and Jun Park, "Experiment for improving stability of sound and downsizing synthesis database," ICSP99, Volume 1 of 2, pp.209-212, Korea, 1999