□ 기술해설 □

대화 음성인식 및 번역

한국전자통신연구원 이영직·박 준

1. 대화체 음성인식의 필요성

사람이 다른 사람과 서로의 뜻을 주고 받는데 가장 많이 사용하는 수단은 음성이다. 따라서 사람이 평소에 자신의 뜻을 나타내는 데에는 말이 가장 많이 쓰이게 마련이다. 즉 음성은 사람에 있어 가장 자연스러운 의사전달 수단이다.

말에도 여러 종류가 있다. 자신이 혼자 중얼거리는 독백체, 미리 문장을 준비한 뒤, 이를 읽는 낭독체, 다른 사람과 말을 주고 받는 대화체 등이 그것이다. 이중 가장 많이 사용되는 것이 바로 대화체 음성이다.

대화체 음성은 다음과 같은 이유로 다른 음성과 차이가 생긴다. 우선 대화체 음성은 발성자가 생각을 해 나가면서 발성을 하게 된다. 이런 원인으로 생기는 현상이 무의미어, 반복발성, 말바꾸기, 비문법적 문장, 생략, 많은 지시대명사 등이다. 말을 해 나가던 도중 생각이 잘 진행되지 않으면, 자신이 생각하는 중임을 알리기 위해 음, 아 등의 무의미어를 발성하거나 앞에서 한 말을 다시 발성하게 된다. 그러는 사이에 생각이 진전되면 다음 말을 이어나가게 된다. 또, 생각과 동시에 말을 해야 하므로 자신이 발성한 문장이 문법에 잘 맞는지 확인할 틈이 없으므로 비문법적인 문장을 발성하게 된다. 실제로 두 사람간의 대화를 받아쓴 것을 살펴보면 심지어는 말하는 사람의 의도를 거의 알 수 없는 경우도 허다하다. 아울러 자신의 입에서 말은 나갔지만 자신의 생각이 바뀌면 앞의 말을 바꾸어 발성하게 되어 말바꾸기가 생긴다.

대화체 음성이 다른 음성과 차이가 나는 또 다른 이유는 서로 대화를 진행하면서 두 발성자가 서로의 생각을 공유하기 때문이다. 대부분의 사람들은 대화를 진행하면서 상대방의 의식의 흐름에 자신의 흐름을 맞추어 나가게 된다. 이러한 원인으로 대화체 음성에서는 그 사람, 그것 등 지시대명사가 많아지게 된다. 아울러 주어나 목적어 혹은 동작을 설명하는 말이 흔히 생략되는 현상 역시 이러한 원인에 기인한다.

현재 많은 연속 음성 인식 연구는 낭독체를 대상으로 하고 있다. 그러나 앞에서도 지적하였듯이 낭독체 음성은 사용자의 측면에서 볼 때 매우 부자연스러운 것이다. 실제 사람들이 일상생활에서 사용하는 음성은 대부분 대화체 음성이다. 대화체 음성과 낭독체 음성의 차이점을 음성 처리 핵심 기술의 측면에서 정리하면 아래의 표와 같다.

위의 표를 볼 때 대화체 음성 처리 기술은 낭독체 음성 처리 기술에 비해 그 난이도가 훨씬 높으나, 사용자의 측면에서 더욱 자연스러운 것임을 알 수 있다. 예를 들면, 문장 인식률

핵심 기술	대화체 음성	낭독체 음성
음성 DB	두명 이상의 사람이 대화함 자연스럽게 발성	한명이 주어진 문장을 발성함 문장 낭독
음성 인식	발성속도가 변함 부정확한 발음 무의미어 많음	일정한 속도 또박또박 발음함 무의미어가 없음
언어 번역	비문법적 문장 인식 오류가 많음	정형문 번역 인식 오류가 적음

90%인 낭독체 음성인식기에 대화체 문장을 입력하면 그 문장 인식률이 40%이하로 떨어진다.

본 고에서는 대화체 음성인식 및 번역 기술에 대해 논한다. 현재 대화체 음성인식은 주로 자동통역 기술의 개발에 사용되고 있다. 따라서 제2절에서는 자동통역 기술의 필요성 및 그 연구 내용을 기술하고, 제3절에서는 자동통역 기술의 국내외 연구동향을 기술한다. 제4절에서 자동통역의 핵심기술별로 현재의 기술 현황을 살펴보고, 제5절에 향후 연구 방향을 제시한다.

2. 자동통역 기술의 필요성

음성은 의사 전달의 가장 기본적인 수단으로, 음성에 대한 연구가 통신 분야 연구를 선도해 왔다. 예를 들어 전화의 발명, 교환기의 발명, 이동 통신 기술의 발달 등이 바로 음성을 효율적으로 전달하고자 하는 요구에서 발전된 것이다. 그런데, 이러한 음성을 통한 의사 전달도 같은 언어를 사용하는 사람들 사이에서만 그 기능을 발휘한다. 외국인과 만나더라도 그 나라의 말을 모르거나 상대방이 한국어를 모른다면 전혀 의사 전달이 되지 않는다. 만약 한국말을 인식하고 이해한 뒤 이를 다른 나라 말로 번역할 수 있다면, 이는 언어 장벽을 해소하여 지구촌을 한 가족으로 만드는 궁극적인 의사 전달 도구가 될 것이다.

21세기에는 세계화가 가속되어 대화체 자동통역의 요구가 국제간 화상 회의뿐만 아니라 산업계 전반에 걸쳐 필요하게 될 것이다. 예를 들어, 국외 여행자의 수 증가는 놀랄 만하다. 이러한 상황에서 보면, 여행 관련 분야의 자동통역이 가능하다면 많은 사람이 이를 사용할 것으로 본다. 차세대 고부가가치 통신의 중요한 서비스 중의 하나인 대화체 자동통역 연구는 기술적 측면이나 수요 측면에서 볼 때 지금 기술을 확보해야 할 분야이다.

다가올 2000년대는 반도체 및 컴퓨터 성능이 비약적으로 향상이 되고, 이를 토대로 자동통역 기술이 실상용화 될 것이다. 그러나, 아직은 시장이 형성되어 있지 않고 그 형상의 변화 가능성이 크므로, 수익성을 따져야 하는 산업체가 주도적으로 이 기술에 장기적인 투자를 하기 힘들다. 따라서, 국책연구소의 주도하에 자동통역에 대한 기초 기반연구를 추진하여 기술을 축적하므로써, 다가오는 2000년대를 대비하여야 한다.

본 고의 끝부분에 그려진 그림 1에 나타난 바와 같이, 자동통역 기술[1]은 음성인식, 언어번역, 음성합성의 세 가지 핵심 기술이 결합된 복합 기술로서, 외국인과 자국어를 이용하여 자유롭게 의사를 주고 받을 수 있도록 해주는 기술이다.

이 중 음성인식 기술은 사람의 음성을 받아들여 이를 문자열로 바꾸는 기술이며, 언어번역 기술은 주어진 문장의 의미를 파악하여 이를 외국어로 바꾸어 주는 기술이고, 음성합성 기술은 임의의 문자열을 음성으로 바꾸는 기술이다.

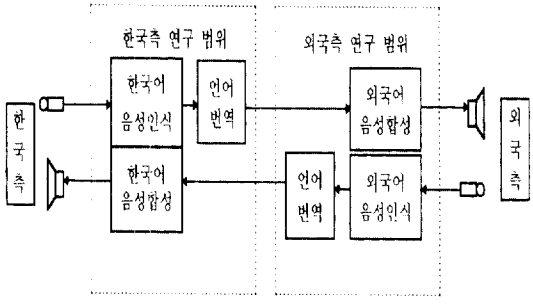


그림 1 자동통역 시스템의 기본 구성도 및 연구범위 분담

자동통역 시스템은 서로 다른 두 개 이상의 언어를 대상으로 동작되는 시스템이므로, 한 언어에 대한 연구만으로는 시스템을 꾸밀 수 없다. 즉 한영 자동통역 시스템의 경우, 한국어 인식, 한국어 합성, 한영 번역 기술만으로는 전체 시스템을 꾸밀 수 없으며, 영어 인식, 영한 번역, 영어 합성 기술이 추가되어야만 하나의 시스템이 구성된다. 따라서 이 분야의 연구는 국제공동연구를 통해 추진하는 것이 가장 효율적이다.

3. 국내외 연구동향

이러한 자동통역 연구를 수행하기 위해 모인

연구그룹 중 가장 큰 그룹이 바로 국제 자동통역 연구 컨소시움(C-TAR : Consortium for Speech Translation Advanced Research)이다. C-STAR는 대화체 자동통역을 연구하는 세계 각지의 연구 그룹이 자발적으로 형성하였다. 이 컨소시움은 핵심 그룹과 참여그룹으로 구성된다. 이 중 핵심그룹은 자국어에 대하여 음성인식, 음성합성 및 언어번역 기술을 연구하여 1999년에 국제간 실험을 하기로 약속한 기관들이며, 참여 그룹은 음성인식/합성 번역 기술의 일부를 연구하여 이를 C-STAR 회의에 발표하는 기관들이다. C-STAR 핵심그룹은 이 목표를 달성하기 위해 1996년에 중간 실험, 1999년에 국제간 실험을 하기로 결정하였다.

현재 C-STAR에는 미국의 CMU, 일본의 ATR, 독일의 Karlsruhe 대학 및 Siemens, 한국의 ETRI, 이태리의 IRST, 프랑스의 CLIPS가 핵심 그룹으로 참여하고 있으며, 미국의 MIT, AT&T, Lincoln Labs., 영국의 SRI Cambridge, 프랑스의 LIMSI, 독일의 DFKI, 스웨덴의 Telia Research AB, 스위스의 Geneva 대학, Laussanne 대학, IDIAP, 호주의 Ottago 대학이 Affiliate 그룹으로 참여하고 있다. 1996년 9월에 중국의 Chinese Academy of Science, Harbin 대학이 Affiliate 그룹에 참여한 점은 특기할 만하다.

C-STAR 컨소시움은 그 추진 전략이 매우 잘 짜여져 있다. 이 컨소시움은 그 실험 도메인, 성능 평가 방안, 다국간 번역 중간언어 연구 등의 분야에서 working group을 형성하여 공동 연구를 추진하고 있다. 아울러 매년 1~2회의 국제 회의를 개최하여 기술교류를 하고 있으며, WWW이나 전자메일을 이용하여 데이타를 주고 받는 등 활발한 공동연구를 하고 있다. 특히 1996년 9월에 개최된 중간실험 결과를 보면 1999년의 국제간 실험이 점차 가시화되고 있음을 알 수 있다. 1996년 9월에는 일본 교토에서 각 기관의 중간 실험 결과를 발표하는 워크샵을 개최하였다. 자동통역 기술은 아직 기술이 완성되지 않은 기술 분야로, 모든 영역의 대화를 수용하지 못하고 있다. C-STAR에서는 여행계획 대화영역을 대상으로 다국간 자동통역을 실험하고 있다.

C-STAR의 전신으로 미국 CMU, 일본 ATR, 독일 Siemens는 1987년부터 1993년까지 국제학회 등록 영역을 대상으로 자동통역 전화를 시도한 바 있다. 이 연구의 결과로 1993년 1월에 일본 ATR과 미국의 CMU간 자동통역 전화 시범이 있었는데, 번역에 약 10초~20초가 소요되었으며, 인식 대상 어휘 수는 약 1,500단어였다. 아울러 인식 대상 음성은 주어진 문장을 읽는 낭독체 음성이었다.

한편, 독일에서는 BMBF의 지원으로 만남약속 분야를 대상으로 대화체 음성인식 및 자연어처리 연구를 수행하고 있다. 이 프로젝트의 이름은 Verbmobil인데, 1991년부터 1996년까지 매년 90억원을 투자하여 휴대용 통역기의 개발을 추진하고 있다. 이 시스템은 영어를 약간 할 줄 아는 독일인과 일본인이 영어로 말을 하다가 잘 표현이 안 되는 부분에서 버튼을 누르고 자국어로 말을 하면, 이를 인식하여 영어로 번역한 뒤, 이를 영어 합성음으로 들려 주는 시스템이다. 이 과제에는 Siemens, Daimler-Benz, Debis Systemhouse, Philips, Alcatel, IBM 등 7개의 기업체 및 21개의 대학이 참여하고 있다. 이 과제의 초기 단계로 1991년부터 3년간 사전 조사를 수행하였으며, 1996년에 Phase 1을 끝내고, Phase 2 과제 제안을 하는 중이다. 이 시스템의 인식 대상 어휘는 1,000내지 3,000 어휘이다.

4. 핵심기술 현황

앞에서 말한 바와 같이 자동통역 기술은 대화체 음성인식 기술, 대화체 언어번역 기술, 및 음성합성 기술이 통합된 것이다. 본 절에서는 대화체 음성인식 기술과 대화체 언어번역 기술에 한정하여 그 현황을 살펴 본다.

현재 연구되고 있는 자동통역 시스템은 여행계획 혹은 만남약속 분야의 대화를 대상으로 연구가 수행되고 있다. 그 이유는 아직 대화체 음성인식 기술이 모든 대화영역을 처리할 수 없기 때문이다. 이 경우 대화체 음성인식기의 인식 대상 어휘는 1,000 내지 3,000개가 되며, 언어에 따라 5,000어휘까지 되기도 한다. 음성인식 성능은 인식 대상 어휘에 영향을 많이 받

는다. 현재 세계 최고 수준은 2,000어휘에서 84%, 5,000어휘에서 80% 정도이며, 후자가 한국전자통신연구원의 결과이다. 대화체 음성인식의 출력은 n-best 문장 출력에서 단어 격자 출력으로 바뀌어 가고 있다.

음성인식의 성능 향상을 위해 화자적응, 미세한 음소모델링, 언어모델 보강 등의 방법을 사용하고 있다. 화자 적응 기술은 음성인식 결과가 발성자에 따라 20~30%의 변화를 보이는 점에 착안하여 개발된 기술이다. 대부분의 경우 한 화자의 음성 특징과 표준 음성 특징간의 사상법을 사용하고 있다.

대화체 음성은 낭독체 등 다른 음성에 비해 불확실한 발성 및 좌우의 음운환경에 따른 변이음이 많이 발생한다. 이를 처리하기 위해 음소 대신 triphone, 더 나아가서는 polyphone [2] 단위로 음소를 모델함으로써 인식률을 증가시키고 있다. 또 언어 모델도 bigram에서 trigram으로 확장함으로써 성능 향상으로 도모하고 있다.

대화체 언어번역 분야의 연구는 언어 번역의 방식에서 크게 직접 번역 방식과 중간언어 방식의 두 줄기로 나누어 볼 수 있다. 직접 번역 방식은 자국어 문장을 그대로 상대국 문장으로 바꾸는 방식으로, 번역의 질을 쉽게 높일 수 있다는 장점을 가진다. 그러나 N개의 국가간에 다국어 번역을 하는 경우, N*(N-1)/2개의 번역 모듈이 필요하다는 단점이 있다. 중간언어 방식은 번역을 위한 하나의 중간언어를 정하고, 각국은 자국어에서 중간언어로, 그리고 중간언어에서 자국어로 바꾸는 모듈을 개발하는 방식이다.

이렇게 할 경우 N개의 국가간에 다국어 번역을 한다면 2*N개의 번역모듈만 있으면 된다. 따라서 다국어 확장이 쉬운 장점이 있다. 그러나 여러 나라 언어를 동시에 수용하는 중간언어의 설계가 상대적으로 어려우며, 이 경우 국제간 공동연구 없이는 불가능하다는 문제가 생긴다. 현재 일본의 ATR이 직접번역 방식을, 미국의 카네기멜런 대학이 중간언어 방식을 채택하고 있으며, 한국전자통신연구원은 한일간 직접번역, 한영간 중간언어 방식을 채택하고 있다. 현재 번역률은 여행 계획 분야에서 한영의 경우 76%, 한일의 경우 94%를 나타내고 있다.

자동통역 시스템에서 언어번역의 입력은 대화체 음성을 인식한 결과이다. 앞에서도 보았듯이 음성인식 결과는 20% 정도의 오류를 포함하고 있다. 따라서 하나의 문장을 대상으로 문장을 이해하기보다는 구 단위 등 문장보다 작은 단위로 이해를 수행하는 것이 좋다. 아울러 단어 격자를 이해하는 모듈의 개발도 추진되고 있다.

자동통역 기술의 개발은 단순히 음성언어 처리 기술의 기술 수준 향상뿐만 아니라, Human Factor, 즉 사용의 편의성을 증대시키는 데에도 많은 영향을 미친다. 우선 대화 상황의 편의성을 증대시키기 위해 마이크 어레이 기술을 이용하여 사용자가 마이크를 가까이 하지 않고서도 시스템을 사용할 수 있도록 원격음성 입력 기술을 개발하고 있다. 아울러 이 시스템의 사용자가 사람이므로 대화에 필요한 주제 인식이나 지시대명사의 의미 파악을 대화자의 지능에 의존하는 방식으로 시스템을 개발하는 추세이다.

5. 향후 연구방향

자동통역 기술은 그 필요성이 지대함에도 불구하고 아직 기술이 완성된 분야가 아니다. 따라서 현재의 연구 결과가 그대로 상용화되기보다는 적용 역역의 확대, 음성인식률 증가 등의 방향으로 연구가 계속될 것이다. 아울러 연구의 중간 과정에서 개발되는 기술 중 사용자의 편이성 증가에 기여하는 부분이 계속 실용화 기술로 완성되어 활용될 것이다.

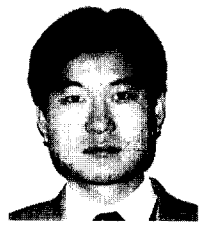
감사의 글

본 연구는 정보통신부 출연 다중매체 환경하에서의 대화체 음성번역 통신기술 개발 과제로 수행되었습니다.

참고문헌

[1] J.-W. Yang, et al., Multimedia spoken

language translation, *IEICE Transactions on Informatics&Systems*, June 1996.
[2] K.-W. Hwang, Vocabulary optimization based on perplexity, Proc. ICASSP '97, Munchen, pp. 1,419~1,422. 1997. 4.



이 영 직

1979 서울대학교 전자공학과(학사)
1981 한국과학기술원 산업전자공학과(석사)
1981~1984 삼성전자주식회사 컴퓨터개발실
1989 Polytechnic University 전기 및 전산과(박사)
1989~현재 한국전자통신연구원 음성언어연구실장, 책임연구원
관심분야 : 음성인식, 음성합성, 자동통역, 신경회로망, 패턴인식

박 준

1981 서울대학교 전자공학과(학사)
1983 서울대학교 전자공학과(석사)
1983~현재 한국전자통신연구원 음성언어연구실 선임연구원
1994 Univ. of Southern California 전기과(박사)
관심분야 : 음성인식, 음성합성, 신경회로망