

자유발화음성 및 텍스트코퍼스 구축에 관한 검토

이용주*, 김종진*, 김철수*, 김경호*, 이영직**
 * 원광대학교 컴퓨터공학과, ** 한국전자통신연구소

Some considerations for construction of spontaneous speech/text corpus

Yong-Ju Lee*, Jong-Jin Kim*, Cheol-Su Kim*, Kyong-Ho Kim*, Young-Jik Lee**
 * Dept. Computer Eng., Won Kwang University, ** ETRI

요약

최근의 음성연구의 관심은 낭독음성(read speech)에서 자유발화음성(spontaneous speech)으로 옮겨가고 있다. 본고에서는 자유발화음성을 대상으로한 음성변역 및 대화시스템의 연구동향과 함께 자유발화의 음성 및 텍스트코퍼스 구축을 위한 몇몇 사항들을 살펴보고, 필자들이 현재 수집중인 코퍼스의 예를 소개한다.

1. 서론

음성에 의한 대화는 인간끼리의 가장 기본적인 의지전달의 수단이며 기계와 인간이 자유롭게 음성으로 대화하는 것은 음성정보처리의 궁극적인 목표이다. 또한 언어가 서로 다른 외국인 간의 회화를 자동으로 통역해주는 기술 또한 언어장벽을 극복하기 위한 음성언어분야의 오랜 꿈이었다. 이 음성변역도 대화음성의 이해의 바탕위에서 이루어져야 한다. 지금까지 음성인식 또는 이해 연구의 대상은 단어단위에서 문장단위로 나아가고 특히 문장의 경우는 미리 한정된 내용의 낭독음성(read speech)을 대상으로 한 연속음성이 주 대상이었다. 그러나 최근의 음성변역 및 음성대화시스템 연구의 주된 관심은 자연스럽게 발생한 회화음성 즉, 자유발화음성(Spontaneous speech)이다. 이러한 자유발화음성을 대상으로 한 연구는 선진 각국에서도 초보적인 상태로서, 아직 그 언어적, 음성학적 특성마저도 잘 밝혀져 있지 않다. 그러나 궁극적인 목표가 자유발화음성의 이해이고 이를 위한 장기적이고 조직적인 연구의 시도가 시작되고 있어 우리의 경우도 이제부터 이러한 기본적인 연구를 시작할 필요가 있다.

자유발화음성을 대상으로 한 연구의 가장 초기단계에서 필요로 하는 것이 대화 음성 및 텍스트 데이터이다. 자연스러운 자유발화음성의 확보를 위한 방법론이 아직 확립되어 있는 상태는 아니며 오히려 여러가지 시도를 통해 이를 확인하려는 단계로 볼 수 있다.

본 고에서는 자유발화음성을 대상으로한 음성변역시스템이나 대화시스템 개발의 동향과 함께 이를 위한 가장 기본적이고 우선되는 작업인 자유발화 음성 및 텍스트코퍼스 구축에 관하여 살펴본 후, 필자들이 시험적으로 수행중인 코퍼스의 예를 소개하고자 한다.

2. 자유발화음성

자유발화에 대하여 현재까지 확립된 정의는 없다. 여기서는

음어에 대한 정의에 관하여 외국에서 지금까지 논의되고 있는 내용들을[34] 잠시 소개하므로써 국내에서의 이런 논의에 계기로 삼고자 한다.

(참고: 이에 대한 우리말 용어도 아직 통일적으로 정해진 바 없고 자유발화나 자연발화 등으로 불리워 질수 있겠으나 학회에서의 용어통일을 위한 논의를 기대하며 편의상 여기서는 우선 자유발화로 부르기로 한다)

- 낭독음성(read speech)과는 달리 텍스트를 보지않고 발생하는 것, 즉, 발화와 개념의 문장화가 거의 동시에 일어나는 형태로써, 이런 경우에는 다음과 같은 현상이 일어난다.
 - 0 개념의 문장화에 따른 어러 (비문법적인 문장)
 - 0 발생속도 및 개념의 문장화에 따른 시간적인 불일치(머뭇거림, 간부시, 무의미어의 삽입)
 - 0 발화 시작후에 개념의 문장화 변경(바꿔말하기)
 - 0 발생과 개념의 문장화가 동시에 일어나므로 조음을 위한 운동기능에 신경집중이 어려움, 그리고 발화속도의 변동폭이 크다.

- 즉석에서 생각나대로 하는 발성

즉 어떤 상황에서 한정된 시간에 한정된 추론과 한정된 지식에 의해 얻어진 의미내용을 한정된 시간에 발화 하는 것. 따라서, 비록 원고가 준비되지 않았다 하더라도 어떻게 얘기할 것인가를 미리 생각하고 머리속에서 정리한 후 하는 말은 엄격한 의미에서 자유발화라 할 수 없다는 견해도 있다.

이밖에도

- 인간이 커뮤니케이션에 이용하는 음성
 - 발화자에게 제한을 주지않는 자유스런 발화
 - Unplanned utterance
- 등이 논의 되고 있으나 아직 확정된 정의는 없다.

3. 자유발화음성 및 그 응용에 관한 연구사례

3.1 음성변역연구

음성 및 언어정보처리의 종합기술이라 할 수 있는 자동통역에 관한 연구는 미국, 유럽, 일본 등에서 음성 인식, 음성합성, 기계번역 등 요소 기술 별로 각기 활발하게 연구되어 왔고 최근에는 최종목표를 자동통역연화의 구현으로 정하고 일부에서는 제한적인 음성변역 또는 자동통역의 실험시스템들을 발표하였다.[1]

음성변역을 목표로 한 최초의 실험 시스템은 영국의 British Telecom 연구소에서 1987년에 발표한 SALT(Speech Automatic

자유발화 음성 및 텍스트코퍼스 구축에 관한 검토

Language Translation)으로 음성인식과 음성합성, 그리고 정형 문 변환에 의해 호텔예약 등 1000단어정도로 구성된 400여 종류의 상용회화문을 대상으로 하였다.[2]

1960년대부터 음성언어 정보처리의 응용기술연구를 꾸준히 해 왔던 일본은 1984년 우정성이 자동통역전화시스템 개발을 위한 국제공동연구를 제안하였고.[3] 1986년에 자동통역전화 반을 전문적으로 연구하는 자동번역전화연구소를 ATR(Advanced Telecommunications Research) 산하에 설립하였다.[4] ATR에서는 1986년부터 15년간 900억엔의 예산규모로 국제회의가 가능한 일영 자동번역전화시스템 개발계획을 마련하였고[5] 그 첫단계 7년의 연구를 1993년 3월에 마치고 4월부터는 신설된 음성번역통신연구소 (Interpreting Telecommunications Research Laboratories)로 이어져 연구가 계속되고 있다.[6] 미국의 CMU에서는 일본 ATR의 연구와 연계하여 영어를 독일어와 일본어로 번역해주는 "JANUS"라는 실험시스템을 1991년에 발표하였다.[7]

또한 이들과 연계하여 독일의 Siemens AG와 Karlsruhe대학교의 연구팀들도 독일어를 중심으로 영어와 일어로 번역하는 시스템을 구성하였고, 1993년초에 이들 미국, 독일, 일본의 3국간에 통합시험을 보인바 있다.[8]

한편, 일본의 NEC는 1991년에 영어와 일본어간의 쌍방향 자동번역시스템을 발표하였다. 이 시스템은 연속음성인식, 간단한 기계번역, 그리고 음성합성기술을 결합한 쌍방향 음성통역시스템으로, 연구회의 표 예약, 관광안내와 같은 정형적인 회화를 대상으로 하고 있다.

또, ATT Bell 연구소에서는 1991년에 연속음성인식, 언어처리 및 음성합성을 결합한 쌍방향 음성번역 실험을 영어와 스페인어 간에 은행에서의 간단한 회화를 대상으로 수행한 바 있다. 이밖에도 독일의 Ministry of Research and Technology(BMET)에서는 대화시에 즉석에서 사용할 수 있는 휴대용 동역지원시스템을 2000년까지 개발하기 위한 'Verbaobil'이라는 프로젝트가 시작되었다. 여기서는 2가지의 시스템을 목표로 하고 있는데, 첫번째는 사용언어가 서로 다른 두 사람이 대화할 때 각각 상대방국어로 통역해주는 것이며, 또 하나는 두 사람 모두 모국어 가 아닌 공통 언어(공통어)로 대화 할때, 모르는 단어나 구 또는 문장을 부분적으로 통역해주는 시스템이다. 대상 텍스트로는 "상함을 위한 시간착속(시애틀링)"이 대상텍스트로 연구되고 있다.[9]

국내에서도 능동적이고 편리한 통신서비스의 하나로써 언어의 장벽을 극복하기 위한 자동통역전화의 15년에 걸친 장기개발계획이 한국통신의 사업계획으로 마련되어 우선 한일간의 자동통역전화용 1차 목표로 하여 1991년부터 7년간 이에 필요한 요소기술을 한국전자통신연구소와 한국통신의 소프트웨어 연구소가 함께 개발하고 있으며, 그 일환으로 KRD와의 분담연구에 의해 95년도에는 호텔예약에 관한 통합시험을 계획하고 있다.[10][11]

최근 과학기술처의 지원을 기대하며 휴대형통역기개발을 위한 연구계획이 대학을 중심으로 준비되고 있다.[12]

3.2 음성대화시스템의 연구[13][14]

음성에 의한 인간과 기계의 대화를 목표로 한 새로운 관점의 음성어해시스템인 음성대화시스템은 종래의 연속음성인식 시스템과는 달리 자연언어처리와 멀티미디어기술도 포함된 복합시스템이다. 각 요소기술의 고도화와 함께 음성대화현상의 분석, 인간과 기계간의 대화모델, 이를 통합하는 방법, 통합시스템의 평가방법등 여러가지 연구과제가 복잡하게 얽혀있다. 특히 자연언어처리와 연속음성인식의 통합은 음성대화 코퍼스와 함께 중요한 과제이다. 이 연구에는 다음과 같은 항목을 주 연구대상으로 하고 있다[15].

- 자유발화(Spontaneous Speech)
- 자연스러운 발성을 대상으로 해서 불필요한 말, 포유즈, 주

변잡음도 포함한 음성인식에 도전한다. 또, 문법도 명확하지 않으므로 새로운 언어기술, 언어해석 기술이 필요하다.

대화모델

사람과 기계와의 대화모세, 기계로부터의 자연스런 응답(CRT 표시/음성응답)을 실현하기 위한 대화모델의 연구이다. 특히 이 대화모델의 연구에는 대화모델의 평가문제, 인식에러의 대처방법, 멀티미디어로서의 인터페이스 등 여러가지가 포함되어 있다. 현재는 대상을 목적지향의 텍스트로 설정하여 실제로 파이롯시스템을 작성해 가는 연구가 진행되고 있다.

대화 텍스트 DB

실제로 사람과 기계와의 대화를 모의(Wizard system)하여 대화텍스트 DB를 수집한다. 이 텍스트 DB는 대화에 나타나는 언어현상의 해석뿐만 아니라, 단어의 출현, 개념의 전달, 대화모델에서 상태의 추이 등 통계적 또는 해석적 연구의 기반이 되기도 한다.

이들 연구에서는 미국 DARPA(Defence Advanced Research Projects Agency)의 ATIS (Air Travel Information System)프로젝트가 DB의 수집, 시스템 작성에서 앞서가고 있고 유럽에서도 ESPRIT의 일환인 SUNDIAL(Speech Understanding and Dialogue)프로젝트를 중심으로 연구되고 있으며[16], 일본에서도 이분야 연구가 각 대학, 연구소 및 기업을 중심으로 활성화 되고있다.

미국의 DARPA 프로젝트하에 연구되고 있는 예는

- CMU (Carnegie Mellon)의 ATIS 시스템(Pheenix)[17]
- SRI의 ATIS 시스템[18]
- BRN의 ATIS 시스템 (HARCH)[19]
- VTI의 ATIS 시스템[20]

등이 있다 특히 이 프로젝트에서는 대화 코퍼스의 수집에도 많은 힘을 기울이고 있다.

유럽에서도 ESPRIT프로젝트의 일환으로 음성대화시스템 SUNDIAL(Speech Understanding and Dialogue)연구가 Logica, CNET, Siemens, CSELT 등을 중심으로 수행되고 있다 [16]. 이 프로젝트에서는 영어, 중국어, 독일, 이탈리아를 대상으로 하여 전화에 의한 음성대화시스템 구축을 목표로 하고 있다. 인식대상으로는 열차의 안내 및 예약, 비행기정보 안내 및 예약, 그리고 호텔예약을 다루고 있다.

일본에서는 연구소, 대학 및 기업이 개별적으로 다음과 같은 시스템을 개발하고 있다.

- ATR 자동번역전화연구소 및 음성번역통신연구소의 음성번역을 위한 대화어해시스템[8]
- 전자기술종합연구소의 교통안내를 위한 음성대화시스템[23]
- NTT의 초연구소의 음성언어연구를 테스트베드으로써 '음성언어 시스템'의 연구[24]
- 오사카대학의 음성대화시스템[25]
- 동요하시 기술과학대학의 UNIX에 대한 질의응답을 테스트로 한 대화음성어해시스템[26]
- 교토 공업성유대학 관광출판을 작성하는 것을 테스트로 한 음성대화시스템 [27]
- NEC의 일영쌍방향 자동통역실험시스템 INTERTALKER를 위한 음성어해연구[28]
- HITACHI의 인터리어 상담용 음성대화시스템[29]
- TOSHIBA의 멀티모달 음성대화시스템(TOSBURG)[30] 등이 있다.

국내에서의 대화음성어해를 목표로한 시스템의 구체적인 개발에는 아직 없다.

최근의 자유발화 대화어해 연구는 아직까지 많은 제한을 안고 있는 자유발화만이 아니라 다른 모드의 정보(예를 들어, 제스처, 터치스크린, 마우스 등)를 같이 이용하려는 움직임이 있

다. 즉, 음성번역을 포함한 대화성인해의 연구는 인간의 고도의 언어지식을 이용하여야 하므로 그 표현형태에 따른 많은 애매성을 포함하고 있어서, 타 미디어정보를 동시에 활용하므로써 이를 해결 하고자하는 시도들이 점차 활발해 지고 있으며 이른바 음성언어정보의 멀티모달 인터페이스라 할 수 있다 [31]. MIT가 음성과 계층유어를 이용하여 휴먼인터페이스를 대폭적으로 향상 시켜 개발한 "Put that There"[32] 나, 다양한 미디어정보의 상호이용에 의해 언어해석의 애매성을 보완한 연구들도 그 좋은 예가 될 것이다. 이밖에도 음성미디어와 시각 미디어에 의한 독순(lip reading)의 통합연구가 있다. 이러한 연구는 입모양에 의한 자모음인식, 세그먼트인식 성능의 향상, 잡음환경에서의 음성인식에 응용이 기대되고 있다. 또한 눈의 움직임(시선)과 음성인식의 통합연구도 시도되고 있으며, 대화 음성인해의 좋은 응용예인 자동통역 시스템으로서 멀티모달 음성번역 통신이 국내외에서 제안되고 있다[21][22][33].

3.3 대화 음성 및 텍스트 코퍼스의 수집

음성대화시스템의 음성/텍스트 코퍼스의 수집은 단순히 대화시스템의 평가만이 아니라 대화시스템연구의 장래를 위해서도 매우 중요하다. 또 다량의 대화 텍스트코퍼스를 수집하는 것은 연속음성인식에서 좋은 성능을 보이고 있는 통계적 언어모델 수집을 대화시스템의 언어처리에도 도입하는데 불가결하다. 여기서는 먼저 NIST(National Institute of Standard and Technology)를 중심으로 DARPA프로젝트에서 수행되고 있는 음성대화 코퍼스의 수집에 대해 TI(Texas Instrument)에서의 최근동향을 소개한다. [1][14]

TI의 대화데이터 수집의 경우, 테스트로는 ATIS(Air Travel Information System)을 대상으로하며 기본적으로는 인간과 기계간의 대화의 텍스트코퍼스 수집이 목표이다. 수집시스템에는 소위 Wizard of Oz방법을 사용하고 있다. 즉, 두사람의 인간(Wizard)이 숨어 있어서, 한사람은 유저의 발화를 단어열로 변환하고 또 한사람의 Wizard는 단어열을 보고 OAG(Official Airline Guide)의 검색언어인 SQL을 시스템에 입력한다. 검색 결과는 응답문과 함께 유저의 CRT상에 표시한다. 대부분의 유저는 기계와 대화를 하고있다고 믿고 대화를 계속한다. 그래서 이와 같은 시스템을 Wizard 시스템이라고 부르며 사람과 기계간의 대화데이터 수집에 유효한 방법이다. Wizard시스템으로부터 음성 및 CRT출력으로 ATIS시스템이 소개된다. 거기에는 ATIS가 음성에 의한 정보검색실현시스템이라는 것과 음성입력의 조작방법, OAG의 개요가 설명된다. 그리고 OAG를 이용하여 주어진 여행시나리오에 따라 여행계획을 새우도록 요청한다. 유저가 시스템에 음성입력으로 검색을 시작하면 유저의 발화, 그리고 시스템의 Wizard가 변환한 단어열(transcription)과 SQL, 시스템의 응답이 기록(session log)되어 수집된다. 이와 같이 해서 유저가 시나리오의 여행계획을 마칠때까지 대화가 계속되며 데이터의 수집도 계속된다. 나중에 음성 및 단어열의 Transcription(ML_input)을 참고하여 숫자 등의 발성도 단어열로 한 prompting-text와 불필요어도 써 넣은 SR_output을 작성한다. 이렇게 해서 수집된 데이터는 NIST에서 일괄 관리하고 ATIS의 연구기관들에 배포된다. 1991년 5월에는 이러한 ATIS 테스트의 대화코퍼스의 수집을 가속화 하기위해 MADCOB(Multi-site ATIS Data Collection Working group) 이라는 워킹그룹이 구성되었다. 이 워킹그룹에는 AT&T, BBN, CMR, MIT, SRI, Paramax(옛 Unisys)가 참가하고 있다. 이 6개기관은 ATIS데이터 수집과 이를 이용한 시스템의 작성 및 평가를 주요 임무로 하고 있다. NIST는 모아진 데이터의 체크, 배포, 평가를 위한 테스트데이터의 선정 및 평가를 임무로 하고 있다.

각 연구기관에서의 대화데이터 수집방법은 다소 다르다. TI와 같이 Wizard시스템이 원칙이나, BBN, SRI와 같이 일부 음성인식시스템을 이용하는 기관도 있고 AT&T가 같이 음성응답만으로 대화를 하는 기관도 있다.

일본에서도 음성번역 및 대화시스템 연구의 일환으로 음성번역 Simulation환경(멀티모달 인터페이스의 경우는 ATR의 BAMI)이나 NTT, KDD, ETL 및 그밖의 기업 및 대학 연구소에서 Wizard of Oz법 및 인식시스템을 이용한 대화음성의 수집이 활발하다. 또한 ATR에서는 음성대화과 유사한 특성을 나타내면서 수집효율을 높일 수 있는 "키보드를 이용한 모의대화"도 제안하였다 [35].

현재까지 국내에서 음성번역 또는 대화시스템 구축을 위한 대화음성 및 관련 텍스트데이터의 수집 예는 아직 발표된 바 없고 호텔예약 등 대화예문을 가상으로 상정한 것만이 사용되고 있다. 필자들은 현재 실제상황을 Simulate하여 모의대화음성을 수집하고 이를 이용한 대화 텍스트코퍼스의 작성을 시작하였다.

4. 모의대화에 의한 음성 및 텍스트코퍼스의 수집

여기서는 현재 필자들이 수집하고 있는 모의대화에 의한 음성 및 텍스트 데이터 수집과정을 소개한다.

4.1 테스트의 설계 및 대화음성의 수집방법 검토

4.1.1 테스트

음성대화 및 번역시스템의 개발을 위한 음성 및 텍스트코퍼스는 현재의 기술수준으로는 목표저항의 한정된 분야를 설정하고 단계적으로 기술을 확보해야 한다. 일반적으로 대상이 되고 있는 것은 관광안내, 여행안내, 회사나 학교안내 등과 같은 안내테스크, 호텔이나 열차 예약, 국제회의 개최 등 다양하다. 여기서는 음성번역에 이용할 수 있는 것으로 한정하여 두사람간의 약속시간을 정하는 스케줄링 테스트의 코퍼스를 수집하였다. 이는 미국의 Carnegie Mellon 대학에서의 코퍼스 수집과 동일한 테스트코퍼에 공통이 되는 부분은 그들의 결과를 참조하였다. 따라서 그들이 만든 개인 일정표를 바탕으로 하여 이를 한국어 상황에 맞게 변형하여 사용하였다.

4.1.2 수집방법

(1) 기본 상황

발성시 기본 상황은 선행화자가 약속요청자, 후행화자는 그 대상자의 임장이라 가정하고 선행화자가 먼저 약속을 요청하는 것으로 하였다. 발생시에는 발생유도를 위하여 무엇보다에 만나려 하는가 안을 약속 요청자에게 알려주고 그밖의 어떠한 선행조건도 주지 않으며 발생자가 일정표를 보고 즉흥적으로 발생하도록 하였다.

(2) 발생 화자

(가) 선정 기준

- 1) 20 세 이상의 남성화자로서 표준어 사용자로서 우선으로 하나 방언 억양이 두드러지지 않은 화자의 발생도 허용한다.
- 2) 1 명의 화자가 대화에 참석하는 횟수는 5회로 제한한다. 발생시마다 서로 다른 상대방과 서로 다른 상황(캘린더)에서 발생하도록 설계한다.
- 3) 소요 발생 인력

1 차목표 : 300 set

대화형태이므로 1 세트의 대화에 2명의 발생자 필요
 $[300(\text{set}) \times 2(\text{명})] / 5(\text{set}) = 120 \text{ 명 소요}$

(4) 발생자의 연령 및 지역분포

발성자의 연령 및 지역은 고루 분포되는 것이 바람직하나 금번의 데이터 수집에서는 20 세 이상의 남성화자로 하며 연령 및 지역의 분포는 고려하지 않는다. 그러나 추후 데이터 확장시에는 이를 감안하여 전체 인구비에 맞도록 연령 및 지역 분포를 고려하여 보완한다.

자유발화 음성 및 텍스트로피스 구축에 관한 검토

(나) 발성자의 정보

각각의 발성자에 대해서 학력, 출생지, 성장지, 방언 정도 등 화자정보와 및 발성상황, 소음정도 등 녹음환경정보를 기록 유지한다.

여 "바라과 헛님" 의 일부 문장을 번갈아 가면서 발성하도록한다. 이때 녹음통제자는 각각의 발성자가 낭독하는 동안에 발성자의 성량, 음질, 갑음레벨등을 검사하며 준비 상태를 체크한다

(3) 녹음

(가) 녹음환경

자유 발화 음성 데이터의 수집 환경은 일상적인 분위기를 만들어낼 수 있는 사무실에서 수행하는 것으로 한다. 녹음 환경의 최우선의 목표는 발성자들이 가장 자연스러운 대화를 할 수 있는 분위기를 제공하는데 있으므로 특별한 장비나 또는 특별한 장소 등으로 발성자에게 심리적인 영향을 미치지 않도록 한다. 대화전에 간단한 발성연습으로 발성자의 긴장감을 완화 시킨다

4) 발성 헤더 정보 녹음

발성 헤더 정보는 녹음 통제자가 발성자들의 발성내용을 녹음하기 이전에 상황 번호와 발성자들에 대한 간단한 정보를 먼저 녹음하고 2~3 초 후 선행화자가 발성을 시작 하도록한다.

5) 발성자의 발성

선행화자는 PD 의 녹음 헤더 정보가 발성된후 2~3 초 후부터 발성을 시작한다. 발성은 최대한 자연스럽게 한다. 자연스런 대화에서는 대화도중에 두사람이 동시에발성하는 경우도 있을 수 있으나 여기서는 중간에 통역자(시스템)을 두고 대화하는 경우로 한정하므로 상대방의 발성이 끝난후 발성하도록 한다.

(나) 녹음순서

1) 태스크 설명

발성자들에게 자신들이 발성할 일경표를 배포하고 발성방법및 상황을 설명한다.

6) 녹음 종료

녹음의 종료는 발성자들의 발성 상황이 결정된후 2~3 초후 종료한다

0 사나리오의 기본 상황

- 본 시나리오는 상담을 위한 일정을 약속하는 태스크이다.

7) 녹음된 음성자료의 정리

녹음된 자료를 2 대의 DAT 를 이용하여 불필요 부분을 제거하고 디지털 다방에 의해 최종적인 Master Tape 를 작성한다

0 선행화자

- 자신의 녹음위치에 있는 발성 제어 표시등에 불이 들어와 있는 화자가 먼저 발성한다. 즉 선행화자는 약속의 요청자이며 후행화자는 그 상대자이다.

(4) 트랜스크립션

가) 1 차 트랜스크립션

1 차 트랜스크립션은 녹음된 발성자들의 대화 내용을 단순히 텍스트로 변환시키는 작업이다. 이때에는 트랜스크립션에 대한 전문 교육을 받지 않은 일반인을 사 용해서 트랜스크립션 가능하다.

0 발성 제어 표시등의 사용법

- 자신의 발성 표시등에 불이 켜져 있는 동안에만 발성을 할 수 있다

- 자신의 발성이 종료되면 상대방 화자가 발성하도록 자기측의 발성 제어 표시등 옆의 푸쉬버튼을 누른다. 자신의 불이 꺼지면 발성할 수 없다

나) 2 차 트랜스크립션

1 차 트랜스크립션된 텍스트에 대해서 녹음 대화 내용을 반복적으로 청취에 가면서 각종 트랜스크립션 기호를 삽입하는 과정이다. 이때에는 트랜스크립션에 대한 전문 교육을 마치고 이해한 사람이 2 차 트랜스크립션을 수행한다. 이때 1 차 트랜스크립션한 내용의 고정도 동시에 수행한다.

0 대화의 진행

- 발성자들은 전화기를 통해서 상대방과 이야기 하는 것처럼 대화 한다

다) 3 차 트랜스크립션

발성된 녹음 데이터에 대해서 모두 2 차 트랜스크립션을 수행한 후 2 차 트랜스크립션에 참석한 모든 사람이 모인 가운데 녹음된 내용을 다시 반복 청취하면서 2 차 트랜스크립션 된 내용과 비교하면서 누락된 트랜스크립션이나 오 기입된 트랜스크립션 기호를 교정 한 후 트랜스크립션을 완료한다

0 대화의 도입부분

- 초기 대화의 도입부분에서는 서로간의 소개나 이에 따른 간단한 종료된 상태를 기준으로 대화한다.

0 발성의 종료 조건

- 발성의 종료는 약속을 위한 구체적인 날짜와 시간이 결정되면 종료하는 것을 기본으로 한다. 그러나 상호간의 일정에서 구체적인 날짜와 시간이 결정되어질 수 없는 경우에는 다음에 다시 연락하는것으로하고 상황을 종료할 수 있다

4.2 수집시스템의 구성

음성 채널만을 이용한 모의 대화의 음성수집 시스템을 다음과 같이 구성하였다.

여기서 대화자 '갑'과 '을'의 음성은 DAT의 L채널과 R채널에 각각 녹음된다.

2) 장비 준비 상태 확인

녹음시 매번 장비의 준비 상태를 확인한다. 주로 검사해야할 준비 상태는 다음과 같다

- 발성자의 입술과 마이크로폰과의 거리
- 발성자 갑과 을의 출력 레벨
- 주변 갑음 레벨
- 발성자의 일정보

4.3 Transcription 기준 작성및 시험 데이터의 수집

본 연구에서 사용하는 트랜스크립션 기준은 공통부분은 키 네기벨런 대학의 트랜스크립션 규칙을 기준으로 하고 부족한 부분은 SSE 의 트랜스크립션 규칙 등을 참조하여 작성하였다

3) 발성 연습

발성자들의 성량, 음질, 갑음 레벨을 테스트하기 위하

4.3.1 우리말의 트랜스크립션 기준(안)

음성 데이터 헤더 정보	
녹음일자	
상황번호	
화자-A	
화자-B	

화자에 의한 잡음의 표현

<화자에 의한 잡음 표현> ::= <화자 잡음 표시 기호>
<잡음내용>
<화자 잡음 표시 기호>

<화자 잡음 표시 기호> ::= '/'
<잡음내용> ::= ah : eh : er : hm : mm : oh :
uh : um : cg : h# : lg : ls

ah	::=	아 아 아
eh	::=	에 에 에
er	::=	어 어 어
hm	::=	흠 흠 흠
mm	::=	음 음 음
oh	::=	오 오 오
uh	::=	우 우 우
cg	::=	기침소리
h#	::=	숨쉬는 소리
lg	::=	웃음소리
ls	::=	입술소리

주변잡음의 표현

<주변잡음의 표현> ::= <주변잡음표시기호><주변잡음>
<주변잡음표시기호>

<주변잡음표시기호> ::= '#'
<주변잡음> ::= C | D : H : B : E

C ::= 키보드같은 것을 두드리는 소리
(가볍고 경쾌한)

H ::= 책상을 두드리는 것과 같은 소리
(무겁고 두터운)

D ::= 문여는 것과 같은 소리

B ::= 일수 없는 잡음의 시작

E ::= 일수 없는 잡음의 끝

간투사의 표현

<간투사의 표현> ::= <간투사시작기호>
<간투사>
<간투사시작기호>

<간투사시작기호> ::= '['
<간투사끝 기호> ::= ']'
<간투사> ::= 간투사

침묵의 표현

<침묵의 표현> ::= <침묵 표시 기호>
<침묵기호>
<침묵 표시 기호>

<침묵 표시 기호> ::= '#'
<침묵기호> ::= P | S

P ::= 인지할 수 있는 짧은 침묵

S ::= 인지할 수 있는 긴 침묵

반복발성의 표현

<반복발성표현> ::= <반복발성시작기호>
<반복발성>
<반복발성끝기호>

<반복발성시작기호> ::= '<'
<반복발성끝 기호> ::= '>'
<반복발성> ::= 반복 발생된 어절이나 단어

잘못 발생된 내용의 표현

<잘못발성된 표현> ::= <잘못발성된 내용 시작 표시자>
<잘못발성된내용>
{ ('<' 고정된 발성') }
<잘못 발생된 내용 끝 표시자>

<잘못발성된 내용 시작 표시자> ::= '<<'
<잘못발성된 내용 끝 표시자> ::= '>>'
<잘못발성된 내용> ::= 발생자가 잘못
발성된 내용

<고정된 발성> ::= 잘못발성된 발성된
내용에 대한
올바른 발성

편집주석의 표현

<편집기호> ::= <편집기호 시작 표시자>
<편집기호>
<편집기호 끝 표시자>

<편집기호 시작 표시자> ::= '['
<편집기호 끝 표시자> ::= ']'
<편집기호> ::= P | C | Q | E

P ::= 마침표
C ::= 콤마
Q ::= 물음표
E ::= 느낌표

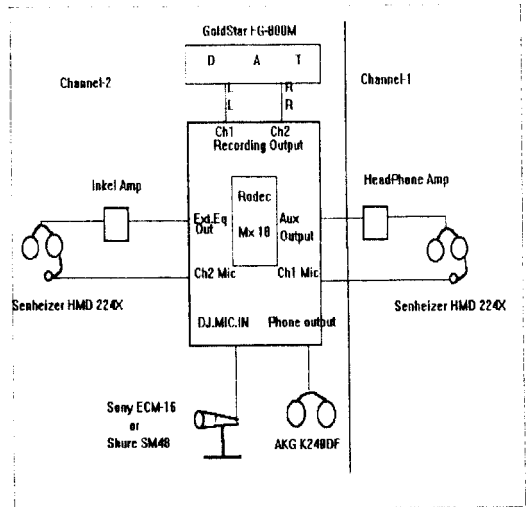


그림 1. 수집 시스템의 구성

사투리의 표현	
<사투리표현>	::= <사투리> <사투리표시자> <표준어> <사투리표시자>
<사투리>	::= 한국어의 사투리 방언
<표준어>	::= 한국어의 표준어
<사투리표시자>	::= '/'
숫자의 표현	
<숫자표현>	::= <연속숫자> <고립숫자>
<연속숫자>	::= [<한글숫자> <수단위>] <숫자표시자> [<숫자>] <숫자표시자>
<고립숫자>	::= [<한글숫자>] <숫자표시자> [<숫자> <공백>] <숫자표시자>
<공백>	::= ' ' B
<숫자표시자>	::= x
<숫자>	::= 0 : 1 : 2 : 3 : 4 : 5 : 6 : 7 : 8 : 9 : 0
<수단위>	::= 조 : 억 : 만 : 천 : 백 : 십
<한글숫자>	::= 공 : 영 일 이 삼 사 오 육 칠 팔 구 : 십
축약어의 표현	
<축약어의 표현>	::= <축약어> <축약어표시기호> <원어> <축약어표시기호>
<축약어>	::= 축약어
<축약어표시기호>	::= ~
<원어>	::= 원래의 단어
외국어의 표현	
<외국어의 표현>	::= <외국어의한국어표현> ['<외국어>']
<외국어의 한국어표현>	::= 외국어를 한국어식으로 발성한 내용
<외국어>	::= 원래의 외국어 형태 (영어, 일어...)
장음의 표현	
<장음의 표현>	::= <장음표시자>
<장음표시자>	::= '·' '·'
-	::= 길게 끌어 늘어지게 발성된 소리
·	::= 짧게 늘어지는 소리
성조의 표현	
<성조의 표현>	::= <성조시작표시기호> <성조> <성조끝표시기호>
<성조시작표시기호>	::= '↑'

<성조끝 표시기호>	::= '↓'
<성조>	::= \ /
↑	::= 끝이 올라가는 형태의 성조
↓	::= 끝이 내려가는 형태의 성조

4.3.2 transcription에 의

갑:안녕하세요(Q)↑가 행렬씨(P)
 을:예(P) 안녕하세요(Q) 도현씨- 잘 지내고 계셨어요(Q)↑#C#
 갑:예(P) 잘 지냈습니다(P) (그) 삼*3*월 둘*2*째주나
 세*3*째주 쯤 어떻게 <<만날려>> 만나고 싶는데 시간
 좀 있으십니까(Q)↑가
 을:예(P) 바쁜 일정이네요(P) 그 중간중간에 쯤 쉬는 시간이
 없네요(가)(P) (그) 어떠십니까(가)↑(Q) 팔*8*월
 둘*2*째주하고 세*3*째주 <<자>> 시간이 쯤 있으십니까
 (Q)
 갑:을- 저도 둘*2*째주와 세*3*째주에 바쁜 일정이
 있는데(가) 서로- 시간을 미쳐와야겠조(가)↑(P)
 을:제가 삼*3*월 팔*8*일예요(가) (어) 회의를 있는데요(P)
 회의가 오전 열*10*시에서 열한*11*시까지 이루어요(가)↑(P)
 (어) 오후에도 두*2*시에서 네*4*시까지 있는데 그쪽은
 어떠세요(가)↑(Q)
 갑:삼*3*월 팔*8*일에는- (을) 제가- 회의가- (어) 아홉*9*
 시에서 열한*11*시(C) 열두*12*시에서 두*2*시(C)
 세*3*시에서 네*4*시 계속있으니까- 그때는 좀 어렵겠구
 요(가)↑(P) 제가 삼*3*월 7*9*일부터 십일*11*일까지
 출장이거든요(가)↑(P) (을) 그래서- (을) 그날쯤 시간
 있으세요(가)↑(Q)

5. 결론

지금까지 음성변역, 음성대화에 관한 연구동향과 이러한 연구
 에서 사용되는 대화데이터의 확보를 위한 동향, 스케들링 메스
 크에서의 대화음성 방법 및 절차의 검토, 수집시스템의 구성,
 데이터의 수집 및 수집된 데이터의 transcription에 관한 중
 간 결과 등을 기술하였다.
 자유발화음성에 대한 코퍼스 수집은 처음이므로 트랜스크립션
 기준 등을 비롯하여 여러가지 시행착오가 있을 것으로 예상되
 나 계속 진행하면서 보완할 예정이다. 이번의 시도는 그 양으
 로 보아 본격적인 코퍼스 구축을 위한 초기단계에 지나지 않는
 다. 관련 연구자들의 공동노력이 필요한 시점이다.

参 考 文 献

[1] 이용주, "자동통역전화의 기술현황 및 과제" 전자공학회지 제20권 제5호, 1993
 [2] Stentiford, Steer, "Machine translation of speech", British Telecom, Technology Journal Vol. 6 No. 2 (April 1988)
 [3] Preprints of International Symposium on Prospects and Problems of Interpreting Telephony, The Telecommunications Advancement Foundation of Japan (April 12, 1986)
 [4] 이용주, "일본 ATR연구소의 첨단연구 소개", 한국전자통신 연구소 전자통신동향분석 (93.1.)
 [5] 우정성 자동번역전화시스템 개발추진협의회 보고서, 일본 우정성 자동번역전화시스템 개발추진협의회 (86.7.)
 [6] 김경호, "ATR의 자동통역전화 연구현황", 한국정보과학회 1993년도 춘계학술발표회 특강 요약집 (93.4.)
 [7] Waibel, Jain, et al. "JAXUS: A speech to speech translation using connectionist and symbolic

- processing strategies", Proc. ICASSP-91(April 1991)
- [8] Sagayama, et al., "An Experimental Interpreting Telephone System 'ASURA'" ASJ Spring Meeting 3-4-17, (1993-3)
- [9] W. Walster, "First results of Verbomobil: Translation Assistance for Spontaneous Dialogues, IWST'93 at ATR, Nov. 93.
- [10] 자동통역전화를 위한 요소기술 개발(I)(II)(III), 한국전 자동산연구소 연구보고서 (1991, 1992, 1993)
- [11] 이종락 "한국통신의 자동통역연구현황", 제9회 음성통신 및 신호처리 워크샵 논문집, 한국음향학회(92.8)
- [12] 김경태 외, 휴대형 통역기 개발에 관한 조사연구사업, 과학기술처 신경계 5개년 계획을 위한 미래 복합형 기술 개발 사업을 위한 사전조사 보고서 (1994.7.)
- [13] 이용주, "대화음성 이해기술의 현황", 한국정보과학회지 제11권 제5호(93.10.)
- [14] .., 음성인식 처리기술에 관한 조사연구, 일본정보처리개발협회 보고서, 1992.6.
- [15] Furui, "The present status of DARPA Spoken Language Processing projects" Technical report of IEICE, SP 92-35, 1992
- [16] Peckham, "Speech understanding and dialogue over the telephone: An overview of progress in the SUNDIAL project", Eurospeech '91, 1991
- [17] W. Ward, "Understanding Spontaneous Speech: The phoenix system", Proc. ICASSP91, SS.29, 1991-5
- [18] H. Murveit, et al., "Speech recognition in SRI's Resource management and ATIS systems" Proceedings of the DARPA speech and natural language workshop, 1991-02
- [19] M. Bates et al., "The BBN/HARC spoken language understanding system", ICASSP '93, 1993
- [20] Goodine, et al., "Full integration of speech and language understanding in the MIT spoken language system", EURO-SPEECH '91, 1991
- [21] Kurematsu, "Perspective view of multimedia cross language communication" Proc. International workshop on Advanced communications and applications for high speed networks, Germany, March 1992
- [22] Y. Yamazaki, "ATR research activities on Speech Translation" IWST'93 at ATR, Nov. 93.
- [23] Hayamizu et al., "A spoken language dialog system for spontaneous speech collection" Technical report of IEICE, SP91-101, Japan, 1991
- [24] S. Matsunaga, et al., "Task adaptation in stochastic language models for continuous speech recognition", Proc. ICASSP-92, S25.3, 1992
- [25] Yanamoto et al., "Dialog management system MASCOTS in speech understanding system", ICSLP'90, 1990
- [26] Nakagawa, et al., "Comparison of syntax-oriented spoken Japanese understanding system with semantic-oriented system" Tran. IEICE Vol. E74 No.7, 1991-07
- [27] Kobayashi, et al., "SUKIT-II --- a speech understanding system based on robust phone spotting" Tran. IEICE Vol. E74 No.7 1991-07
- [28] Watanabe et al., "자동통역을 위한 블록정화자 연속음 성인식시스템", Technical report of IEICE, SP91-115, Japan, 1991-12
- [29] Komatsu et al., "Conversational speech understanding based on cooperative problem solving", Proc. ICSLP, 1990-11
- [30] Takebayashi et al., "Noisy spontaneous speech understanding using noise immunity keyword spotting with adaptive speech response cancellation" Proc. ICASSP '93, 1993
- [31] T. Nitta, "Trends in spoken multimodal dialogue" 음성 언어처리와 대화이해에 관한 공회연구회 자료집, 일본, 1992-7
- [32] Takebayashi, "Human-computer dialogue using multimedia understanding and synthesis functions" Technical report of IEICE, SP 92-37, Japan, 1992
- [33] 이용주, "통신서비스의 고도화를 위한 음성언어처리기술", 대한전자공학회지 제30권 제8호, 1993
- [34] Ito, "자연스런 발화"에 대한 의논, 일본 전자정보통신학회 및 정보처리학회, 음성인식 이해의 금후의 과제에 관한 공회연구회 자료집, 1994년 2월
- [35] Kudo, Morimoto, "A Keyboard-Dialogue Collecting System for Building a Bilingual Dialogue Corpus", Tran. IEICE Vol. J75-D-11 No.4 (1992.4.)