□특 집□

대화음성 이해기술의 현황

한국전자통신연구소 자동통역연구실 이 용 주*

목 차

Ⅰ. 들어가는 말

인간끼리의 의사소통에 가장 편리한 수단인 음성을 인간과 기계간의 대화에도 이용하기 위한 음성인식의 연구가 컴퓨터의 발달과 더불어 그 가능성을 더해가고 있다. 음성인식의 형태는 구분 발성한 단어를 인식하는 고립단어인식(isolated word recognition)과 연속 발성한 음성을 인식하는 연속음성인식(Continuous Speech recognition)으로 나뉜다. 또한 이 연속음성인식은 비교적 소수의 어휘를 대상으로 하면서 단어레벨 이상의 지식을 이용하지 않는 연결단어 음성인식(Connected word recognition)과, 비교적 다수 어휘가 여러가지로 연속 발성된 음성을 대상으로 하여 단어레벨 이상의 구문, 의미 등 언어적 지식을 이용하는 회화음성인식(Conversational speech recognition)으로 나뉜다.

회화음성인식은 문장음성인식이라고도 부르며 특히 발성된 말의 모두를 정확하게 인식하지 않아도 발성자가 의도한 의미 내용만 정확히 인식하면 된다는 입장을 강조할 경우에는 음성이해(Speech Understanding)라고 부르기도 한다[1].

우리들이 일상회화에서 상대방의 말을 이해하는 경우에 단순히 음성의 물리적인 성질, 즉 어떤 음소가 발성되는가 하는 정보만을 이용하는 것이 아니고, 오히려 현재 말하고 있는 화제나 문법으로부터 다음에 할 말을 예측하면서 듣기 때문에 애매하게 발성되거나 잡음이 많을 경우에도 상대방의 말을 이해할 수 있다. 다수 어휘를 대상으로 하는 회화 음성 인식에 있어서는 단어가 연속해서 발성되므로 각 음소의 물리적 성질이 불명확해지고 단어 사이에도 조음결합이 일어나서 그 경계가 쉽게 구별되지 않으므로 단어음성에 비해 인식이 매우 어렵다. 또, 어떤 사람의 음성이라도 인식할 수 있는 불특정화자(speaker independent) 음성인식과, 그 시스템에 학습시킨 특정한 화자의 음성만 인식하는 특정 화자(speaker dependent) 음성인식으로도 나눌 수 있다. 낭독조의 음성(read speech)을 한자 한자 받아 적는 형태의 Dictation 시스템도 연구의 한 방향으로써 음성타이프라이터와 같은 응용을 목표

* 정회원

로 추진되어 현재는 단어사이에 약간의 포즈를 두고 발성한 음성을 인식하는 시스템이 상품화 되기도 하였다[2]. 최근에는 낭독조의 음성보다 대화 형태의 자연스런 음성을 이해한다는 관점에서 음성언어시스템(Spoken Language System) 또는 음성대화시스템(Speech Dialogue System)의 연구가 활발하다[3]. 본고에서는 음성대화 이해기술에 관하여 미국의 DARPA의 시스템 개발예를 중심으로 소개하고자 한다.

II. 대화음성 이해시스템

2.1 음성이해

기계에 의해 음성을 자동인식함에 있어서 고립된 단어의 인식에 비해 연속 음성의 인식이 어려운 이유는

- 단어 경계가 불명확
- 단어 경계 부근의 음이 선행 또는 후속단어의 영향으로 변형하여 조음결합을 일으킴
- 단어를 구성하는 각 음들의 지속시간이 짧고 발음도 애매
- 단어끝의 음운의 길이가 늘어나는 경우가 있음
- 간투사적인 단어가 문장중에 삽입
- 의미적으로는 옳으나 구문적으로는 틀린 경우가 많음
- 어순이 자유로운 점
- 어떤 단어를 빠뜨리거나 묵시적인 이해에 의한 불완전한 문장이 많음등을 들 수 있다[4].

회화음성을 연속단어 음성과 같은 발상으로 다루려면 여러가지 난점이 있으므로 고차 언어정보인 구문, 의미, Pragmatics, 운율정보 등 redundunt한 정보를 이용하면 다룰 수 있는 한정된 세계가 된다.

발성된 회화 문장의 정확한 인식을 꼭 요구하지는 않고 내용의 이해만으로 충분하다는 관점에서 종래의 음성인식과는 달리 음성이해라고 구별하여 불러왔다.

음성이해 시스템 개발을 위한 구체적인 계획으로 1971년부터 5년간 수행한 미국의 ARPA (Advanced Research Projects Ageney)의 SUS (Speech Understauding System) Project를 들 수 있다. CMU, SRI, BBN 등이 참가한 이 연구에서 CMU가 개발한 'Harpy'가 1000단어로 된 연속음성인식의 목표를 일단 달성하였고, 또 같은 CMU의 'Hearsay II'는 black board model이라는 새로운 시스템 구축개념을 제안하기도 하였으나 전체적으로는 컴퓨팅파워의 보완, 음성지식의 보완이 절실하다는 것을 확인하였다[5].

1985년부터 재개된 DARPA 프로젝트에서는 불특정화자의 10000단어로 된 연속음성인식을 목표로 CMU, TI, MIT, BBN, NBS 등이 참여한 가운데 수행되었다[6]. 1989년부터 DARPA프로젝트의 주된 연구목표는 대어휘의 불특정화자 연속음성인식연구로부터 "음성언어시스템(Spoken language system)"이라고 부르는 음성언어 인터페이스연구로 옮겨갔다[7]. 지금까지의 연구결과인 정확도 높은 음성인식모델을 기초로, 음성이 가지고 있는 언어적 측면을 더욱 깊숙히 강조한 것이다. 음성언어시스템이라는 명칭은 여러가지 음성언어인터페이스의 핵이 되는 범용성이 있는 구조라는 의미와 함께 낭독조의 written language가 아닌 자유발화음성(spontaneous speech)를 의식하여 자연어처리와의 통합을 더욱 깊게 하고자하는 의미를 내포하고 있다. 이러한 시스템은 음성대화에 의한 데이타베이스의 검색, 인터렉티브한 문제해결 시스템, 자동통역시스템에 응용될 수 있다.

2.2 대화음성 이해시스템(음성대화시스템)

음성에 의한 인간과 기계의 대화를 목표로 한 새로운 관점의 음성이해시스템인 음성대화시스템은 종래의 연속음성인식 시스템과는 달리 자연언어처리와 멀티미디어기술도 포함된 복합시스템이다. 각 요소기술의 고도화와 함께 음성대화현상의 분석, 인간과 기계간의 대화모델, 이를

통합하는 방법, 통합시스템의 평가방법 등 여러 가지 연구과제가 복잡하게 얽혀있다. 특히 자연언어처리와 연속음성인식의 통합은 음성대화 데이타베이스와 함께 중요한 과제이다.

이 연구에는 다음과 같은 항목을 주 연구대상으로 하고 있다[8].

○ **자유발화(Spontaneous Speech)**

자연스러운 발성을 대상으로 해서 불필요한 말, 포우즈, 주변잡음도 포함한 음성인식에 도전한다. 또, 문법도 명확하지 않으므로 새로운 언어기술, 언어해석 기술이 필요하다.

○ **대화모델**

사람과 기계와의 대화로써, 기계로부터의 자연스런 응답(CRT 표시/음성응답)을 실현하기 위한 대화모델의 연구이다. 특히 이 대화모델의 연구에는 대화모델의 평가문제, 인식에러의 대처방법, 멀티미디어로서의 인터페이스 등 여러 가지가 포함되어 있다. 현재는 대상을 목적지향의 테스크로 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[8] 유럽에서도 ESPRIT의 일환인 SUNDIAL(Speech Understanding and Dialogue)프로젝트를 중심으로 연구되고 있으며[9], 일본에서도 이분야 연구가 각 대학, 연구소 및 기업을 중심으로 활성화 되고있다.

III. DARPA의 연구상황

3.1 시스템 개발현황

3.1.1 CMU(Carnegie Mellon)의 ATIS 시스템 (Phoenix)[10]

Phoenix는 DARPA SLS (Spoken Language System)프로젝트의 일환으로 CMU에서 연구하고 있는 자유발화 (Spontaneous Speech)의 인식을 목표로 한 음성대화시스템이다. 테스크로는 ATIS를 대상으로 한다. 시스템은 종래와 같이 명확하게 기술된 문법을 갖고 있지 않고, 입력음성도 낭독문장과 같이 유창하지도 않다. 이 Phoenix시스템의 음성인식부는 CMU가 개발한 이산 HMM을 기반으로 한 Sphinx시스템이 이용되었다. 또 인식결과는 의미프레임의 슬롯마다 인식되어 정보검색언어 SQL로 변환된다. 다른 ATIS시스템과 똑같이 OAG (Oficial Airline Guide) DB를 SQL로 검색해서 응답결과를 CRT로 표시한다.

Phoenix시스템에서는 대화이해 시스템으로서 다음과 같은 3가지 면에 촛점을 두어 연구되고 있다.

1) 자유발화에 대한 음향레벨에서의 대처

유저에 의한 잡음(포우즈, 숨쉬기, 불필요한 말, 바꿔말하기 등) 및 주변잡음(문여닫는 소리, 전화벨소리 등) 등에 대처하는 garbegie HMM 모델을 작성한다. garbegie 모델과 함께 단어 Bigram 모델(Word Pair Grammar)도 이용하여 Sphinx 시스템으로 입력음성을 인식한다.

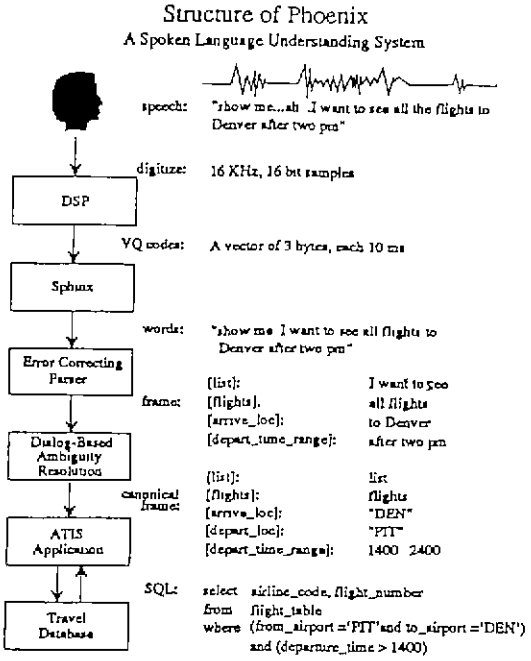
2) 미 등록어에 대한 대처

음운의 확률적 Bigram 모델을 이용

3) 유연한 파싱

프래임 기반의 slot filling이 원칙이며, phrase마다 문법을 가지고 있고 상태 Network로 표현되어 있다. 이들 상태 Network는 garbage 상태에 의해 결합되고, 이를 이용해서 인식결과를 해석한다. 해석에서는 키워드의 속성을 중시한 부분적인 의미해석(Partial Parsing)을 한다. 이 Phoenix 시스템의 구성을 (그림 3)에 보인다. 1991년에는 700단어의 사전과 Perplexity 55의 Bigram 단어 모델로 음성인식하고 제 1위의 결과 만으로 의미를 해석 하였다.

현재, 음성인식에 이용하고 있는 단어 Bigran



(그림 1) CMU의 Phoenix시스템의 구성[10].

모델을 단어천이 Network에 조합시켜서 제 1위의 인식결과와 파서에서의 성공률의 향상, 잡음모델의 개선, 미등록어 모델의 도입을 고려하고 있다. 또 반연속 HMM에 의한 음운모델의 정밀화, 화자적응의 검토, 대화상태의 Pragmatics 등 더 세밀한 의미정보를 이용하는 연구도 수행되고 있다.

3.1.2 SRI의 ATIS 시스템[11]

SRI도 DARPA의 Spoken Language System (SLS) 프로젝트의 일환으로 ATIS 태스크 대상으로 연구하고 있다. ATIS 텍스트는 종래의 DARPA의 RM(Resource Management)테스크 보다도 음향처리에 있어서 다음과 같은 점을 더 고려하였다.

1) 실제 사용 환경에서의 음성(주변잡음)

2) 미등록어에 대한 대처 (Out of Vocabulary)

3) 자유발화의 취급 (Spontaneous Speech)

특히 자유발화는 RM 테스크에서 다루어져온 낭독문(Read Speech)과 크게 다르다. 예를 들어, 자유발화 중에는 바꾸어 말하기(False Start) 불필요한 말(Non-Word)도 포함되어 있고 또, 조음도 불명확하다. 그리고 자연언어처리연구에서도 잘 정리된 문장이 아닌 대화언어를 해석해야 하므로 매우 도전적인 과제이다.

SRI의 ATIS 시스템은 단어열을 인식하는 음향처리부인 DECIPHER와, 에러를 포함한 단어열의 해석 및 응답부의 언어처리부인 Template Matcher로 구성된다.

(1) 음향처리(DECIPHER) DECIPHER에서는 FFT에 의한 멜 켑스트럼과 델타켑스트럼을 인식의 특징파라메터로 사용한다. 또, 단어의 발성변동을 고려한 음운네트워크를 작성하기 위해 조음규칙(Phonological rule)을 이용하고 있다. ATIS의 테스크에서는 평균적으로 한 단어당 75종의 발음기호로부터 triphone을 작성하여 이용하고, 음운 네트워크에서의 빈도에 따라 음운 HMM을 선택해서 작성한다. 현재는 이산 HMM (Tied-Mixture Continuous Density HMM)으로 바꾸어 음운모델을 개량 하고 있다. 또, 남녀별로 음운모델의 작성, 화자적응 기술의 이용, 식별학습 등에 의해 단어인식 에러율을 2/3 정도로 줄이고 있다.

음운처리에서의 언어 모델로써, Back-off Smoothing한 단어 Bigram모델을 이용하고 있다.

(2) 언어처리(Template Matcher)

자연언어처리 연구자에 있어서 ATIS테스크와 같이 간단하다고 생각되는 테스크라도 회화언어 입력에서는 시스템이 예기하지 못한 구문이 나타나는 경우가 있다. 이 언어처리(Template Matcher)에서는 단어열 전체를 해석할 수 없더라도 의미내용을 어느정도 파악할 수 있고(Partial Parsing), 해석이 애매성에 대해서도 스코어를 주는 메카니즘도 고려하고 있다. 기본적으로는 CMU의 Phoenix 시스템과 유사하다. 향후연구로는 에러를 더 많이 허용할 수 있는 음향처리, 언어처리의 개량 및 양자의 인터페이스 연구와 함께 실용적인 음성대화시스템의 구축을 목표로 하고 있다.

3.1.3 BBN의 ATIS 시스템 (HARC)[12]

BBN에서도 DARPA의 SLS 프로젝트의 일환으로 HARC(Hear and Respond to Continuous Speech) 음성대화시스템 연구를 추진하고 있다.

HARC는 음향처리 시스템으로서 연속음성인식 시스템인 BYBLOS를 이용하고, 언어처리시스템으로서는 DELPHI 시스템을 이용하고 있다. HARC의 블록도를 (그림 2)에 보였다. BYBLOS는 입력음성을 단어 Bigram 모델에 기반한 Forward-Backward Search로 N-best 후보를 출력한다.

DELPHI는 N개의 문장후보를 재평가해서 Mapping Unit라고 부르는 정식화에 의해 해석해서 OAG DB의 검색을 위한 검색언어 SQL로 변환한다.

(1) 음향처리(BYBLOS)

RM 테스크로 개발되어 있던 BYBLOS는 ATIS 테스크를 위해 다음과 같이 개량했다.

a) Forward-Backwark Search의 효율 및 정확성을 높이기 위해 입력음성이 시작점과 끝점의 검출알고리즘을 개량

b) 단순한 단어 Bigram (word pair) 문법과 단어내의 HMM 음운 모델을 이용해서 N-best 후보를 forward-backward search로 구한 후, 단어 사이도 고려한 HMM음운모델과 통계적 단어 Bigram 모델을 이용해서 N-best의 결과를 재평가한다.

c) 단어삽입 Penalty 음향처리 스코어와 언어처리 스코어의 평가의 최적화 수행.

(2) 언어처리(DELPHI)

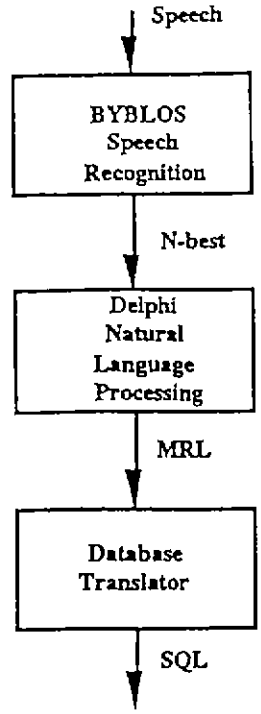
동사의 카테고리 분류를 주체로 한 Mapping unit라 부르는 일종의 unification 문법으로부터의 approach를 취하고 있다. 이 언어처리에서는 다음 3가지 특징을 가지고 있다.

a) 4가지의 기본적인 정보요소로써 문법의 격관계(주어, 직접목적어 등), 구문의 형태, 의미정보의 일치(unification)을 가지고 동사의 카테고리분류로부터 구조 mapping unit를 만든다.

b) mapping unit 와 phrase grammar의 관계를 Non-Constituent Predicate로 규정한다.

c) 언어의 바꿔 말하기에 어느 정도 대처할 수 있도록 Predicative Metonymy를 준비한다. 그 밖에 파싱(Partial Parsing)의 어프로치도 검토하고 있다.

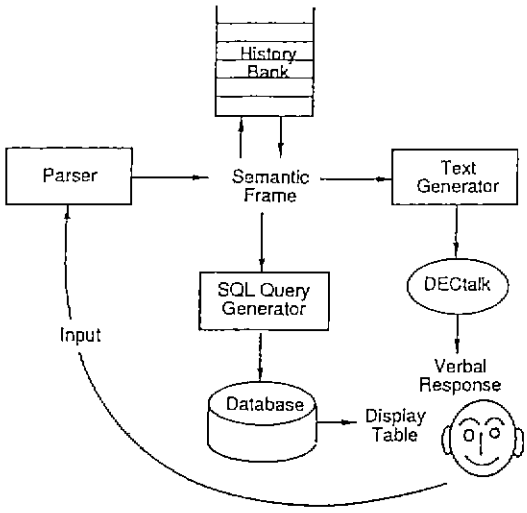
현재 DAPPA의 SLS 프로젝트에서 음향처리 기능은 가장 높은 성능을 나타내고 있으나 언어처리를 포함한 전체성능을 CMU, SRI 보다 떨어진다. 또 언어처리에 이용되고 있는 언어처리의 Unification (단일화) 알고리즘을 Robust하게 할 필요가 있다. 또, 음향처리에서는 화자적응화의 도입이 시도되고 있다.



(그림 2) BBN의 HARC 시스템[12].

3.1.4 MIT의 ATIS 시스템[13]

MIT는 이전부터 보스톤의 MIT 근처의 지리안내를 테스크로 한 Voyager라는 음성대화시스템을 개발해 왔다. 이 시스템의 특징은 음향처리에 HMM을 이용하고 있지 않다는 것이다. 음운의 Segment 후보검출 및 식별 구성된 소위 특징 Base의 시스템인 SUMMIT로 수행하고 있다. 이 Voyager를 ATIS 테스크로 변경하였다. SUMMIT는 Context 독립의 76개의 음운모델과 Perplexity 92의 단어 Pair문법을 이용하고 있다. 출력은 A* Search의 N-best 메카니즘으로, 복수개의 후보를 인식하여 언어처리로 보낸다. 언어처리는 Voyager에서 개발된 시스템인 TINA를 Modularity와 Portability의 관점에서 다시 구성하고 있다. 의미프레임 (Semamtic Frame)을 중



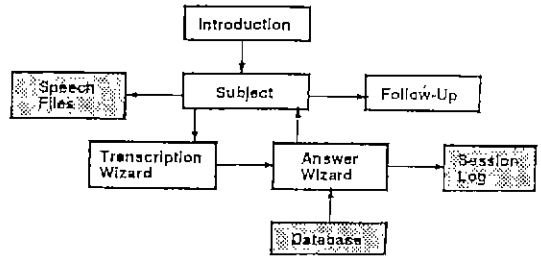
(그림 3) MIT의 음성대화시스템[13].

심으로 한 (그림 3)과 같은 구성으로 되어 있다. 우선 Parser에서 해석트리를 작성하고 그로부터 의미정보를 탐색해서 의미프레임에 써 넣는다. 또, 의미프레임 정보에 기반하여 검색언어인 SQL로 변환한다. SQL은 OAG DB를 검색하여 CRT에 출력한다. 응답문도 의미프레임으로부터, 대화이력을 참조해서 작성하고 Dectalk에 의해 음성으로도 출력한다.

3.2 대화 데이타베이스의 수집 및 시스템의 평가

음성대화시스템의 음성/텍스트 데이타베이스의 수집은 단순히 대화시스템의 평가만이 아니라 대화시스템연구의 장래를 위해서도 매우 중요하다. 또 다량의 대화 텍스트데이타베이스를 수집하는 것은 연속음성인식에서 좋은 성능을 보이고 있는 통계적 언어모델 수법을 대화시스템의 언어처리에도 도입하는데 불가결하다[14]. 여기서는 NIST(National Institute of Standard and Technology)를 중심으로 DARPA프로젝트에서 수행되고 있는 음성대화 데이타베이스의 수집에 대해 TI(Texas Instrument)에서의 상황과 DARPA의 최근동향을 중심으로 논한다[3].

TI의 대화데이타 수집의 경우, 테스크로는 ATIS(Air Travel Information System)을 대상



(그림 4) Wizard시스템에 의한 ATIS대화데이타의 수집 절차[3].

으로하며 기본적으로는 인간과 기계간의 대화의 텍스트데이타베이스 수집이 목표이다. (그림 4)에 데이타수집시스템의 블록도를 보였다. 수집시스템에는 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. 검색결과는 응답문과 함께 유저의 CRT상에 표시한다. 이 사이의 응답시간은 평균 20수초이다. 대부분의 유저는 기계와 대화를 하고 있다고 믿고 대화를 계속한다. 그래서 이와 같은 시스템을 Wizard 시스템이라고 부르며 사람과 기계간의 대화데이타 수집에 유력한 방법이다. Wizard시스템으로부터 음성 및 CRT출력으로 ATIS시스템이 소개된다. 거기에는 ATIS가 음성에 의한 정보검색실험시스템이라는 것과 음성입력의 조작방법, OAG의 개요가 설명된다. 그리고 OAG를 이용하여 주어진 여행시나리오에 따라 여행계획을 세우도록 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 데이타의 수집도 계속된다. 나중에 음성 및 단어열의 Transcription(NL_input)을 참고하여 숫자 등의 발성도 단어로 한 prompting-text와 불필요어도 써 넣은 SR_output을 작성한다. 이렇게 해서 수집된 데이타는 NIST에서 일괄 관리하고 ATIS의 연구기관들에 배포된다. 1991년 5월에는 이러한 ATIS테스크의 대화데이타베이스의 수집을 가속화 하기 위해 MADCOW

(Multi-site ATIS Data COllection Working group)이라는 워킹그룹이 구성되었다. 이 워킹그룹에는 AT&T, BBN, CMU, MIT, SRI, Paramax(옛 Unisys)가 참가하고 있다. 이 6개 기관은 ATIS데이타 수집과 이를 이용한 시스템의 작성 및 평가를 주요 임무로 하고 있다. NIST는 모아진 데이타의 체크, 배포, 평가를 위한 테스트 데이타의 선정 및 평가를 임무로 하고 있다. 또 NIST는 SRI의 레이블링(annotation)그룹과 연락을 취하면서 발화문의 응답에 따른 분류를 하고 있다. 1992년 2월 현재까지 5개 연구기관(Paramax제외)에서 총 280명의 화자로부터 10400 문장이 수집되었다. 그중에서 5600문장에는 바르다고 생각되는 응답문(database reference answer)과 그 타입이 붙여져 있다. 각각의 발화문은 그에 대한 응답문에 따라 다음 3가지 종류로 대분류된다.

1) 대화상황에 관계없는 응답(Context-independent, 타입 A)

2) 대화의 상황에 의존하는 응답(context-dependent, 타입 D)

3) 명확한 응답을 기대할 수 없는 것(unevaluable, 타입 X)

약 44%가 타입 A로, 32%가 타입 D로, 나머지 24%가 타입 X로 분류되었다.

각 연구기관에서의 대화데이타 수집방법은 다소 다르다. TI와 같이 Wizard시스템이 원칙이나, BBN, SRI와 같이 일부 음성인식시스템을 이용하는 기관도 있고 AT&T와 같이 음성응답만으로 대화를 하는 기관도 있다. 다섯군데에서 모아진 대화데이타의 일부를 이용하여 각 연구기관의 인식성능의 평가가 1992년초에 이루어졌다. 음향레벨처리의 평가는 각 연구기관에서, 클래스 A, D, X의 데이타(총 971문장)을 이용해서 하고 있다. 각 연구기관의 어휘수는 841단어(MIT)로부터 1881단어(BBN)까지이다. 언어모델로는 통계적인 단어Bigram모델이 이용되고 있다.

각 연구기관의 단어인식율을 〈표 1〉에 보였다. 또 클래스 A와 D의 데이타(총 687문장)을 이용하여 문장입력과 음성입력에 대한 언어처리도 포함한 시스템전체의 평가가 응답문의 적절성에 의해 수행되었다[8]. 그 결과를 〈표 2〉에 보였다.

〈표 1〉 ATIS시스템의 음성인식 성능[8].

System	att 3	bbn 3	cmu 4	mit 4	sri 3
Word recognition rate	86.2%	93.8%	88.2%	86.4%	91.6%

〈표 2〉 ATIS시스템의 언어처리(NL) 및 통합시스템(SLS)의 성능[8].

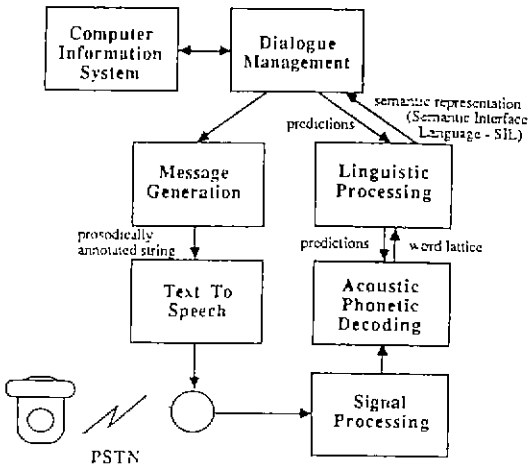
	System	att 1	bbn 1	cmu 1	mit 2	sri 1
NL	True	55%	77%	85%	80%	78%
	False	30%	11%	15%	13%	9%
	No answer	15%	13%	0%	7%	14%
	System	att 2	bbn 2	cmu 2	mit 1	sri 2
SLS	True	44%	72%	67%	69%	65%
	False	34%	15%	33%	19%	10%
	No answer	22%	13%	0%	12%	25%

음향처리는 AT&T에서의 16혼합 가우스분포 연속HMM(16 Gaussian Mixture Continuous Density HMM), CMU와 SRI의 반연속 HMM(Tied Mixture Continuous Density HMM)을 이용해서 HMM을 개량하고 있지만 그 유효성은 명확하지 않다. 오히려 불필요어의 모델, 잡음모델 등의 설정이 중요해지고 있다. 언어처리는 ATIS와 같이 단순테스크에서도 문장입력일지라도 정답율이 80% 정도밖에 얻고 있지 못하므로 더 robust한 대화모델, 회화언어의 해석수법에 대한 연구의 강화가 필요하다고 보고 있다.

IV. 그밖의 나라들의 연구동향

4.1 유럽

유럽에서도 ESPRIT프로젝트의 일환으로 음성대화시스템 SUNDIAL(Speech Understanding and Dialogue)연구가 Logica, CNET, Siemens, CSELT 등을 중심으로 수행되고 있다[9]. 이 프로젝트에서는 영어, 불어, 독어, 이태리어를 대상으로 하여 전화에 의한 음성대화시스템 구축을 목표로 하고 있다. 인식대상으로는 열차의



(그림 5) SUNDIAL 시스템의 구조[9].

안내 및 예약, 비행기정보 안내 및 예약, 그리고 호텔예약을 다루고 있다.

4.2 일본

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

4.2.1 ATR 자동번역전화연구소

화자적응 기능을 가진 문절음성인식시스템과 일본어 대화음성의 번역을 결합한 일영 음성언어번역 실험시스템 SL-TRANS2와 이를 확장한 ASURA를 개발하였고[15], 후속의 음성번역통신연구소에서는 멀티미디어 음성번역을 계획하고 있다[16][17].

4.2.2 전자기술총합연구소의 음성대화시스템

동경공업대학과 공동으로 불특정화자 연속음성인식을 전용 하드웨어 없이 거의 실시간으로 수행하며, 테스크는 쓰쿠바로부터 토쿄로 가는 교통안내를 간단한 문의에 의해 응답해주는 음성대화시스템을 개발하였다[18].

4.2.3 NTT 기초연구소

여러가지 음성인터페이스의 핵이 되는 범용적인 구조를 목표로 하며 지식베이스(언어모델, 테스크월드)의 내용을 변경하므로써 특정영역으로 포팅할 수 있는 음성언어연구의 테스트베드로써 '음성언어시스템'이라고 부르는 연속음성인식 및 대화음성의 이해를 위한 기본적인 구조가 연구되고 있다[19].

4.2.4 오사까대학의 음성대화시스템

음성이해, 음성합성, 대화관리의 3개 모듈로 구성된 음성인터페이스가 구축되고 있다[20].

4.2.5 토요하시 기술과학대학의 음성이해시스템

UNIX에 대한 질의응답을 테스크로 한 대화음성이해시스템으로 SPOJUS-SYNO (SPOken Japanese Understanding System-SYNtax Oriented)이 개발되어 이를 계속 보완 중에 있다[21].

4.2.6 쿄토 공예섬유대학

유저가 시스템과 대화해서 도시의 관광플랜을 작성하는 것을 테스크로 한 음성대화시스템이 연구되고 있다[22].

4.2.7 NEC

반음절을 인식단위로 하고, 구문네트워크제어에 의한 음성인식을 통해, 입력음성의 의미를 나타내는 개념표현을 출력하는 불특정화자 연속음성인식 이해시스템을 개발했다. 또, 이 시스템에 다언어 기계번역, 음성합성을 통합하여 일영쌍방향 자동통역실험시스템 INTERTALKER을 개발했다. 이것은 음성입출력의 일본어와 영어간 쌍방향 자동통역을 하며 동시에 불어, 스페인어에 대해서도 통역결과를 음성출력할 수 있다[23].

4.2.8 HITACHI

건축을 계획하고 있는 사람이 자신이 희망하는 건물, 인테리어 등의 조건을 음성으로 입력하면 검색조건에 맞는 샘플의 컬러사진을 검색하여 화면에 출력해 주는 음성대화시스템을 개발하고 있다[24].

4.2.9 TOSHIBA의 멀티모달 음성대화시스템 (TOSBURG)

햄버거집에서의 주문시스템을 상정하여 불특

정의 사용자가 일상의 언어로 계산기와 자연스럽게 대화하는 것을 목표로 한 실시간 음성대화 시스템인 TOSBURG(Task-Oriented dialogue system Based onspeech Undestanding and Response Generation)를 개발하였다[25].

4.3 우리나라

국내에서의 대화음성이해를 목표로한 시스템의 구체적인 개발예는 아직 없다. 그러나 이러한 기술들을 포함한 자동통역전화연구가 한국통신과 ETRI를 중심으로 수행되고 있고[26, 27, 28], 그 일환으로 학계와 함께 호텔예약을 위한 텍스트베이스의 질의응답시스템 및 이의 고도화 연구[29], 텍스트DB의 연구[30], 대화체기계번역연구[31] 등 기초적인 연구가 수행되고 있다.

V. 전망 및 과제

대부분의 음성대화 시스템은 음성인식부, 언어처리부, 대화관리부로 되어 있다. 대화처리부 등에서는 복잡한 언어처리가 필요하므로, 지금까지 연속음성인식의 언어처리에 이용되어온 통계적인 방법만으로는 부족하다.

지금까지의 연속음성인식에서는 다음에 찾아낼 단어를 통계적으로 예측하기 위해 단어big-ram/trigram모델, 상태천이네트워크, 문맥자유문법 등이 이용되어왔다. 단일화(unification)도 부분적으로 이용되고 있다. 이들 수법은 단어퍼플렉시티로 평가할 수 있어서 평가방법이 확립되어 있다. 이러한 통계적 언어처리는 수년동안 텍스트데이타베이스의 증대와 함께 큰 성공을 거두어오고 있어 음성대화시스템의 음성인식을 위한 언어처리로써 통계적 언어처리 모델연구는 계속될 것으로 보인다. 아울러 대화텍스트데이타베이스를 이용한 통계적 어프로치에 의한 개념 및 대화상태 추이의 연구, 운율정보의 이용에 의해 음성인식에서의 탐색을 고속화 하거나, 운율정보에 포함되어 있는 정보를 적극적으로 이용해서 언어처리의 애매성을 줄이는 방법의 연구도 활발해질 전망이다. 또, 불필요단어와 운율, 초점(focus) 및 화제(topics)와 운율과의 관계도 자유발화데이타베이스를 이용해서 정량적으로 검토할 필요가 있다.

전체적으로 볼 때 대화음성의 이해연구를 고도화하기 위해서는

- ○대규모 대화텍스트데이타베이스 작성.
  적절한 정보검색 등 비교적 단순한 테스크 설정. 사람과 기계와의 대화를 고려해서 대화데이타베이스를 수집. 대화 텍스트데이타베이스의 레이블링(품사, 생략, 의미개념, 대화상태).
- ○음성대화시스템의 평가방법 확립
  음운레벨, 의미파악레벨, 응답문레벨, 대화의 자연성.
- ○유연하고 robust한 파서
  유연한 정식화, robustness, 효율, 부분적 의미해석, 생략의 적절한보간
- ○대화모델
  적절한 테스크의 설정, 적절한 응답문, 멀티미디어에 의한 대화.

등에 대한 심도깊은 연구가 필요한 것으로 지적되고 있다.

또한, 대화음성이해의 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[32]. MIT가 음성과 제스츄어를 이용하여 휴먼인터페이스를 대폭적으로 향상시켜 개발한 "Put that There"[33]나, 다양한 미디어정보의 상호이용에 의해 언어해석의 애매성을 보완한 연구들도 그 좋은 예가 될 것이다. 이밖에도 음성미디어와 시각미디어에 의한 독순(lip reading)의 통합연구가 있다. 이러한 연구는 입모양에 의한 자모음인식, 세그먼테이션 성능의 향상, 잡음환경에서의 음성인식에 응용이 기대되고 있다. 또한 눈의 움직임(시선)과 음성인식의 통합연구도 시도되고 있으며, 대화음성이해의 좋은 응용예인 자동통역 시스템으로로 멀티모달음성번역통신이 국내외에서 제안되고 있다[34].

VI. 맺는말

지금까지 음성대화의 이해기술에 관하여 각국의 시스템의 개발예를 중심으로 소개하였다. 음성이란 음향신호를 통해 고도의 언어정보를 전달하는 수단이므로 인간의 지식처리를 기반으로 음성처리와 언어처리가 통합된 연구가 절실히 요청되고 있다. 각국이 대부분 그러하듯이 우리나라도 음성처리연구자와 언어처리연구자가 많은 교류없이 개별적으로 연구해온 것이 현실이다. 음성신호처리와 언어처리가 통합된 대화음성이해와 같은 음성언어(spoken language)연구를 위해서 두 분야 연구자의 공동작업이 절실하며, 이를 인터페이스하기 위한 음성언어통합처리에도 관심이 모아져야 할 것이다. 또한 최근의 음성및 언어처리기술은 각종 통계적수법의 도입으로 대량의 데이타에 의존하는 바가 크다. 이를 국가적인 차원에서 체계적으로 확보하기 위하여 연구소, 학계, 산업계를 연계한 "음성 및 텍스트(사전포함)데이타베이스를 위한 컨소시움[가칭]"의 구성이 시급하다고 생각된다.

참 고 문 헌

1. D. O'shoughnessy, Speech Communication, Addison-Wesley Pub. Co., 1987.
2. Engelien, MacBryde, Natural language markets, Ovum Ltd. 1991.
3. ——, 음성의 지적처리에 관한 조사연구, 일본정보처리개발협회 보고서, 1992. 6.
4. ——, Speech recognition and synthesis manual, Japan Industry engineering center, 1980.
5. D. H. Klatt, Review of the ARPA speech understanding project, JASA Vol. 62, No. 6, 1977.
6. 이용주, 김경태, "음성이해 연구의 동향" 전자통신 9권 1호, 1987년 3월.
7. M.Okada, "Recent trends in continuous speech recognition and spoken language system" J. of the Acoustical society of Japan, Vol. 48 No. 1, 1992.
8. Furui, "The present status of DARPA Spoken Language Processing projects" Technical report of IEICE , SP 92-35, 1992.
9. Peckham, "Speech understanding and dialogue over the telephone: An overview of progress in the SUNDIAL project", Eurospeech '91, 1991
10. W.Ward, "Understanding Spontaneous Speech: The phoenix system", Proc. ICASSP91. S5. 29, 1991-5.
11. 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.
12. M. Bates etal., "The BBN/HARC spoken language understanding system", ICASSP '93, 1993.
13. Goodine, et al., "Full integration of speech and language understanding in the MIT spoken language system", EUROSPEECH '91, 1991.
14. 최기선, "언어모델의 통계모델과 지식모델의 융합", 음성통신 및 신호처리 워크샵 논문집, 한국음향학회, 1993년 8월.
15. Sagayama, et al., "An Experimental Interpreting Telephone System 'ASURA'" ASJ Spring Meeting 3-4-17, Japan 1993-3.
16. Kurematsu, "Perspective view of multimedia cross language communication" Proc. International workshop on Advanced communications and applications for high speed networks, Germany, March 1992.
17. 김경호, "ATR의 자동통역전화 연구현황", 한국정보과학회 1993년도 춘계학술발표회 특강 요약집 1993-4.
18. Hayamizu et al., "A spoken language dialog system for spontateous speech collection" Technical report of IEICE, SP91-101, Japan, 1991.
19. S. Matsunaga, et al., "Task adaptation in stochastic language models for continuous speech recognition", Proc. ICASSP-92, S25.3, 1992.
20. Yamamoto et al., Dialog management system MASCOTS in speech understanding system, ICSLP '90, 1990.
21. Nakagawa, et al., "Comparison of syntax-oriented spoken Japanese understanding system with semantic-oriented system" Tran. IEICE Vol. E 74, No. 7, 1991-07.
22. Kobayashi, et al., "SUSKIT-II --- a speech understanding system based on robust phone spotting", Tran. IEICE Vol. e74, No.7, 1991-07.
23. Watanabe et al., "자동통역을 위한 불특정화자 연속음성인식시스템", Technical report of IEICE, SP91-115, Japan, 1991-12.

24. Komatsu, et al., "Conversational speech understanding based on cooperative problem solving", Proc. ICSLP, 1990-11.
25. Takebayashi et al., "Noisy spontaneous speech understanding using noise immunity keyword spotting with adaptive speech response cancellation" Proc. ICASSP '93, 1993.
26. 한국전자통신연구소, 자동통역전화를 위한 요소기술 개발(I)(II), 한국전자통신연구소 연구보고서, 1991, 1992.
27. 이용주, "자동통역전화의 기술현황 및 과제" 전자공학회지 제 20권 제 5호, 1993.
28. 이종락 "한국통신의 자동통역연구현황" 제 9회 음성통신 및 신호처리 워크샵 논문집, 한국음향학회 1992-8.
29. 김영길, 최병욱 외, "한국어질의응답시스템 설계와 관한 연구", 제 1회 ETRI 음성, 언어 및 음향정보처리워크샵 논문집, 1993-4.
30. 이용주, 임연자 외, "ETRI의 음성 및 텍스트 데이터 베이스의 구축 현황 제 1회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, 1993-4.
31. 최기선, 대화체 한영기계번역 연구, 위탁과제 중간보고서, 한국전자통신연구소, 1993-8.
32. T. Nitta, "Trends in spoken multimodal dailogue" 음성언어처리와 대화이해에 관한 공최연구회 자료집, 일본, 1992-7.
33. Takebayashi, "Human-computer dialogue using multimedia undestanding and synthesis functions" Technical report of IEICE, SP 92-37, Japan, 1992.
34. 이용주, "통신서비스의 고도화를 위한 음성언어처리기술", 대한전자공학회지 제 30권 제 8호, 1993.

이 용 주

1976 고려대학교 전자공학과 (학사)
1987 고려대학교 대학원 전자공학과 (석사)
1992 고려대학교 대학원 전자공학과 (박사)
1976 ~ 1980 공군통신장교 근무
1985 ~ 1986 일본 토호쿠대학 응용정보학연구센타 (연구생)
1980 ~ 현재 한국전자통신연구소 자동통역연구실 실장(책임연구원)
관심 분야 : 음성, 언어 및 음향정보처리, HCI 등