

國際會議 參席報告

ICASSP '86 및 自動通訊電話의 展望과 問題點에 관한 심포지엄 參席報告

金敬泰 · 李宣宅 / 音響研究室

I. 개요

IEEE의 Acoustics, Speech and Signal Processing Society가 매년 주최하는 음향, 신호처리 국제회의(약칭 ICASSP)는 금년으로 제11회째를 맞이하여 지난 4월 7일부터 4월 11일까지 동경의 케이오프라자 호텔에서 일본전자통신학회와 일본음향학회의 공동 주최로 개최되었다. 회의의 조직운영은 IEEE ASSP 일본지부장인 동경대학의 Hir-oya Fujiaki씨를 위원장으로 하는 위원회에서 운영되었다. 회기는 5일간이었고 첫날의 pre-conference로서 강연회가 있었으며 4일간의 회의중 발표논문수가 781건, 등록된 참가자의 수만 미국을 비롯한 33개국으로 1,386명 이었다(〈表 1〉 참조)

발표는 매일 오전과 오후에 3시간 30분씩의 session으로 일곱개의 회장에서 개최되었으며 이중 두개의 회장은 포스터 발표회장이었다. 포스터 발표는 1 session당 평균 23건으로서 자유롭게 토론할 수 있게 되어있어서

상당히 좋았으며 전 4권(총 3,171페이지)으로 된 논문집은 IEEE Service Center(445 Hoes Lane, Piscataway, NJ08854, U. S. A.)를 통해 입수할 수 있다.

〈表 1〉 ICASSP '86참가국 및 분야별 논문수

참 가 국	인 원	분야별 논문	건 수
Japan	661	음성 처리	313건(40%)
U. S. A.	367	신호 처리	117건(15%)
France	76	영상신호해석	93건(12%)
U. K.	47	VLSI	78건(10%)
Canada	30	Filter	74건(9.5%)
China	22	추정과 모델링	67건(8.6%)
Sweden	21	수중음향, 오디오, 기타	39건(5%)
Italy	19		
W. German	15		
Filand	12		
Korea	11		
Spain	10		
Australia	10		
etc	85		
합 계	1,386명		781건(100%)

〈表1〉에서 나타난 것과 같이 전체 발표수는 781건으로써 그중 음성신호처리 관련 40%, 신호처리 관련 15%, 화상처리 관련 12%, VLSI 관련 10%, filter 관련 9.5%, spectral estimation 및 system modeling 관련 8.6%, 수중음향 및 오디오 기타 관련 5% 정도로 구성되어 있으며 본고에서는 주로 음성처리 분야중 음성인식 분야에 대하여 기술하고자 한다.

그밖에 회기중에 32개사에 의한 전시회와 공식 리셉션파티, award luncheon이 있었으며 이 자리에서 음성처리를 위한 선형예측 부호화에 대한 공적으로 1986 Morris N. Liebmman Memorial Award가 B. S. Atal씨와 F. Itakura씨에게 주어졌다. 또한 회기중 3명의 강사에 의한 plenary lecture가 있었으며 일본 NTT, ETL, Hitachi, Fujitsu, Oki에 대한 technical visit가 회망자에 한해 견학할 수 있었다.

II. ICASSP '86

1. 음성처리 전반

음성처리에 관한 논문의 수는 전체 781건 중 313건으로서 40%를 차지했다. 이 숫자는 작년의 481건중 139건, 즉 전체의 29%였던 것보다 11%나 상회한 것으로서 건수로는 174건이 증가했다. 이는 상당히 중요한 의미를 갖고 있는 것으로서 두가지 측면에서 고려해 볼 수 있다. 즉, 음성신호처리 분야는 신호처리 분야의 다른 부분에 비해 아직까지 미개척 분야이기 때문에 개척자로서 도전하여 연구해 볼만한 가치가 있다는 인식이 점차 많아지고 있다는 사실과 현재의 산업화 동향이 이분야 기술의 응용을 크게 요구하고 있으며, 나아가서는 그 실제응용이 임박했다는 느낌이 팽배하고 있으므로 이에 도전하고자 하는 연구자가 증가된 점으로 미루어 볼 수 있다. 또한 개최지가 일본이었기 때문에

일본인들의 많은 참석이 가능했으므로 전체 발표 논문수가 많아졌고 따라서 음성처리 분야의 논문수가 자연스럽게 증가되었다고도 생각할 수 있다. 어떤 이유에서든지 이 분야의 연구의 중요성이 점차로 증대되고 있는 사실은 우리가 유의해야 할 점으로 생각된다.

음성처리 분야의 논문을 내용별로 보면 전체 313건중 제일 많은 분야가 음성인식 분야로서 114건(30%), 부호화 분야가 74건(24%), 분석합성 분야가 66건(21%), D/B 및 H/W Expert system이 27건(9%), 장애자 분야가 21건(7%), 화자인식이 11건(4%)를 차지하고 있다. 통신의 수단이라는 견지에서 볼 때 인식, 부호화, 합성이 81%를 차지한다 할 수 있겠다.

2. 음성인식 Session

본 session의 논문수는 114건으로서 음성처리 분야중 36%를 차지하고 있으며 ICASSP 85의 34건(24%)에 비교하면 논문 편수로 3배가 넘는 수이다.

내용별로 보면 연속 음성인식이 53건으로 음성인식 분야중에서 46%를 차지하며 단어 음성 인식이 53건으로 음성인식 분야중에서 46%를, 단어음성 인식이 50건으로 44%, 나머지가 음소단위 인식으로 11건으로써 10% 정도 차지하고 있다. ICASSP '85와 비교해 보면 단어단위 인식에서 연속음성 인식으로 많이 옮겨졌다고 할 수 있다. 세계적으로 보면 영어, 프랑스어, 이탈리아어, 아랍어, 중국어, 독일어, 스웨덴어, 일본어 등 각국의 모국어에 대한 음성인식 연구가 진행되고 있으며 제일 활발하게 진행되고 있는 곳은 미국의 Bell Lab. 과 IBM 등이며 일본의 NTT연구소와 학교 및 민간연구소에서도 많은 연구가 행해지고 있다.

인식방법 또한 크게 대별해 보면 음성자체를 분석하여 음운의 성질을 규명하고 그들을 파라미터로 사용하는 음성학적 방법에 deci-

sion을 위하여 적절한 통계적 방법을 가미하는 혼합적 방법과 최적으로 카테고리를 분류하여 인식하기 위해 통계적 신호처리에 의해 decision하는 방법이 있다. 전자의 방법은 주로 일본에서 행하고 있는 방법이며 후자의 방법은 주로 미국에서 행해진다고 할 수 있다. 최근 몇년간 많이 행해지고 있는 수법인 Hidden Markov Model (HMM)이 후자의 그것으로서 물론 이 방법도 음성학적인 측면에서 인식대상 어휘수를 늘렸을 때 통상 패턴 매칭 방법으로는 음성의 세부 구조를 보기에는 한계가 있다.

여기에 ICASSP '86에서 인상적이었던 몇 가지 인식수행 방법 및 시스템에 대하여 소개하고자 한다.

- IBM

IBM에서 발표한 5000/2000 구분 발성 단어 실시간 인식장치의 VTR에 의한 데몬 스트레이션이 있었다. 특정화자 인식장치로서 새로운 사람이 등록하기 위해 20분 정도의 시간이 걸린다. IBM PC/AT와 IBM PC signal processor subsystem으로 구성되었다.

- 프랑스

J. L. Gauvain씨가 구분 발성단어 10,400 단어에 대하여 구분 발성한 2,498개의 음절을 기준패턴으로 한 음절단위 인식과 단어단위 인식을 비교 검토하여 각각 12%, 6%의 오 인식을 얻어졌다.

- 일본 마쯔시타

S. Hiraoka씨가 발표한 인식 시스템은 남녀 각 20인이 발성한 212단어에 대해 발성이 끝나고 0.8초 이내에 인식응답을 하며 95.5%의 인식을 보였다. (시스템을 3장의 PCB로 구성했음)

- 미국 ITT

ITT사의 B. Patrick Landell씨의 발표내용은 조용한 곳에서 등록된 기준패턴을 이용하여 비행기 소음이 있을 경우에 인식하는 방법에 대하여 발표하였다. 즉, 이미 등

록된 기준 패턴을 입력될 때의 소음에 적용시켜 가면서 인식을 수행하였다. 소음레벨이 112dBA의 경우 약 14% 정도의 인식을 개선을 발표하였다.

- 일본 NTT

NTT사의 Y. Tohkura씨는 보통의 LPC cepstral 거리보다 LPC cepstral계수에 적당한 가중치(계수분산의 역수)를 가함으로써 인식률을 3%정도 개선시키는 연구 결과를 발표하였다. 또한 불특정 다수 발성자에 강함을 보였다.

- 일본 NEC

NEC사의 H. Fujii씨는 연속음성 인식을 위하여 두 단어 사이에 나타나는 조음결합이라던가 변화성을 해결하기 위하여 Demi-word pair 기준패턴 방법을 제안하였다.

- 일본 SMARP

SMARP사의 F. Togawa씨는 자동학습기능을 가진 음성 입력형 워드프로세서를 개발 발표하였다.

- 일본 KDD

KDD사의 F. Yato씨는 지식공학적 기술을 이용하여 음성이해 시스템을 개발하였다.

- 미국 Carnegie-Mellon Univ.

Carnegie대학의 F. Alleva씨는 불특정 화자 연속음성 인식 시스템을 개발하기 위한 소프트웨어 environment를 보고하였다. (A-GORA)

- 이태리 CSFLT

M. Cravero씨는 HMM을 이용하여 음소단위의 음성인식 평가를 하였다. 즉, 22개의 정상상태 음소와 101개의 과도상태 음소를 나누어서 실험하였다.

- 프랑스 IBM

B. Meraldo씨는 한 사람의 speaker에 대해 대어휘의 자동 dictation시스템에 대하여 보고하였다.

3. 기 타

그밖에 미국의 BBN, Bell Lab. 등에서 화자 인식에 대한 보고가 있었으며 Bell Lab.의 F.K. Song씨가 보고한 내용으로서 화자 인식을 위해 순시스펙트럴과 과도 스펙트럴을 사용하였다. 특히 재미있는 사실은 개인성을 추출하여 화자인식을 가지는 목적의 경우에도 cepstral 계수에 가중치를 두고 개인성을 제거함으로써 불특정 화자용 단어음성 시스템에서도 cepstral 계수에 가중치를 두어 좋은 경과를 얻은 일본 NTT의 Y. Tohkura와 서로 상반되는 결과라 하겠다.

장애자들의 보조기기를 위한 신호 해석에 관한 보고사항들로서 청각 장애자들을 위한 음의 처리문제 (J. Genin, 프랑스), 청각기관중의 달팽이관의 제작 (T. Ifukube, 일본), 언어장애자들을 언어 훈련기 (L. E. Bernstein, 미국), 성도의 형태를 이용한 언어 훈련기 (M. Shigenaga, 일본), 소경을 위한 음성 독서기 (Y. Mitome, 일본) 등 약 20편의 논문이 발표되었다.

음성인식 시스템이 보급됨에 따라 시스템의 성능평가를 위한 공통 데이터베이스와 음성학적 연구를 위한 연구용 데이터베이스의 요구가 급격히 증가됨에 따라 데이터베이스 통일화에 관한 논문이 많았다. 고립단어 데이터베이스에 대해서는 일본 쓰쿠바 대학의 S. Itahashi씨가 발표한 내용으로 323 단어를 남녀 각 75명이 4 번씩 발성한 공통음성 데이터베이스가 있다. 이 작업은 4 년간에 걸쳐 15개의 연구기관과 기업연구소가 참여하였다.

이외에 미국 Air Force Army의 데이터베이스 시스템 (4,000명의 남자와 400명의 여성 데이터)과 미국 National Bureau of Standard의 미국 TI음성 데이터베이스의 표준화와 저장방식에 관한 발표가 있었다. 특기할 것은 이들 모두가 음성을 PCM녹음을 통하여 VCR에 수록한다는 것이다.

III. 통역전화의 전망과 문제점에 관한 국제 심포지엄

일본전기통신발전재단의 후원아래, 일본음향학회, 전자통신학회, 정보처리학회, 일본언어학회 및 IEEE ASSP Society가 주최하여 열린 심포지엄으로서, 언어권이 다른 지역의 사람들 간에 원활한 의사소통을 하기 위한 통역전화의 실현은 음향공학, 전자통신공학, 정보처리공학이 종합적으로 밀접한 상호정보교환 아래 이루어져야 하는 궁극의 과제라는 관점에서 해결해야 할 문제제들을 확실히 밝히고, 각 관련 학문분야로부터 의견을 수렴하여 이의 가능성을 높이는 동시에 국제간의 협력을 도모하기 위해 개최된 것이다.

강연자 및 강연내용은 다음과 같다.

- Fujisaki (도쿄대학) : 통역전화-언어장벽의 극복을 위하여
- L. Rabiner (ATT Bell Lab.) : 음성처리기술의 입장에서 본 통역전화
- Nagao Makoto (교토대학) : 자연언어 처리의 입장에서 본 통역전화
- P. Cohen (SRI International) : 대화모델을 통해 본 통역전화의 전망과 문제점
- Iwao (게이오대학) : 사회학적 입장에서 본 국제통역 전화

강연후에는 관련분야 인사를 초청한 panel discussion이 있었으며 panelist로서는 강연자 외에 5 명이 더 참석하여 여러가지 문제점을 토론하였다.

NEC의 Dr. Kato가 H/W 구현 및 프로토타입 구성의 가능성에 관하여 의견을 피력하였고, KDD의 Dr. Kurematsu씨가 국가적 차원 연구기관의 입장에서 통역전화를 구현해야 할 필요성 및 장래의 계획 그리고 시스템 spec. 을 제시하였다.

그밖에 와세다 대학의 Dr. Shirai는 technical problem에 관한 의견을 내놓았으며, 영국 P. A. Technology의 Dr. J. Holmes는 음성이 해에 앞선 음성합성의 중요성에 관한 견해를 밝혔고 동경대학의 Dr. Kunihiro가 언어학자로서 여러가지 대화 상황의 categorization의 어려움을 피력하였으며 그의 다수의 상이한 견해 및 입장을 밝혔다.