

1988년도 춘계 일본음향학회 연구발표회 참관기

김 경 태

(한국전자통신연구소 음향연구실)

1988년 3월 23일부터 25일까지 3일간에 걸쳐 일본 도쿄의 마찌다시에 자리잡고 있는 타마가와 대학 공학부에서 일본음향학회 춘계 연구발표회가 개최되었다. 저자는 이 학회의 발표회에 계속 참석하여 발표해온 바 있으나 최근에는 참가하지 않고 문헌 및 다른 참가자를 통해 모니터만 했는데 이번에는 3년만에 다시 참석케 되어 그동안의 연구방향의 파악에 대한 많은 기대와 또 자신이 확실히 공부해보겠다는 마음으로 참석하였다.

일본음향학회는 올해(1988년도)로써 52주년이 되었으며 매년 춘계와 추계에 각각 전국대회를 개최한다. 이때 춘계 전국대회는 토쿄를 중심으로, 추계 전국대회는 지방에서 번갈아 열리고 있다. 또한 각 연구회별로 연구발표회가 매월 개최되고 있다.

이번 대회의 참가자수는 약 1000명 정도이며, 3일간의 발표논문수는 405편으로 8개의 회의장과 Poster Session 1회의장으로 진행되었다. 이러한 편수는 작년과 거의 비슷한 수준이라 할 수 있다.

부문별 발표편수로서는 음성A(인식) : 82, 음성B(분석, 합성, 부호화) : 55, 청각음성 : 17, 청각 : 10, 수중음향 : 12, 초음파 : 95, 건축음향 : 41, 소음진동 : 46, 전기음향 : 39, 음악음향 : 8 이었다. 따라서 음성청각 관계 논문이 164편으로 가장 많았다. 이들 중 음성관계, 소음진동관계, 초음파 관계 논문의 일부는 Poster Session으로 진행되었으며, 진행방법은 논문전체의 개황을 1건당 2분정도의 시간으로 설명을 듣고 Poster Session으로 들어 갔기 때문에 전체적으로 파악하는데 좋은 방법이라고 생각되었다.

발표기간 중 음향관계의 여러 회사들이 출품한 전시회도 병행하여 진행되었다. 주로 출품된 제품들은 주파수분석기, 음성분석기, 소음진동관계측정기 일반신호처리H/W와 S/W 그리고 일반 측정기들이 전시되었다. 주로 관심이 끌린 분야는 음성신호처리를 위한 H/W, S/W를 Package화 하여 IBM PC/AT 혹은 NEC PC-9801 Series 전용으로 상봉

하된 것들이었다.

2 일째는 오후 1:00부터 2:30까지 매년 일리는 행사로서 SATO 논문상과 구리아 학술 상려상 수상식이 있었다. SATO 논문상은 Dr. SATO의 기부금 이자로서 유지되고 있으며 음향학에 관한 학술 및 기술면에서 공헌한바가 큰 연구업적을 낸 자에게 수여하는 상이며 보통 2편 이하이다. 그리고 구리아 학술상려상은 전년도에 춘계 및 추계 전국대회의 발표를 통하여 내용이 뛰어난 사람에게 주어지는 상이다. 이어서 주취대학의 교수들 중 한분의 특별강연회가 개최되었는데 이번에는 [알 수 없는 생물시계의 매력]이라는 제목으로 SASAKI 조교수의 강연이 있었다.

오후 7시부터 구내 식당에서 간담회가 개최되었다. 칵테일 파티의 형식으로 진행되며 음향관련 모든 분야의 원로급들이 거의 모두 출석하여 서로 인사를 나누며 옛날 이야기와 최근의 연구방향, 활동근황 등을 서로 얘기하면서 다방면의 사람들과 교류가 가능한 기회였다.

다음은 본인이 참석한 음성 부문에 관련하여 좀더 자세한 내용을 소개하기로 한다. 앞에서도 언급했듯이 음성관계 논문편수는 164건으로 전체의 약 40%에 달하고 있다. 이중 음성관계 논문 중 인식이 82건으로 전체의 약 50% (82/164), 분석합성평가가 33% (55/104), 청각음성부문이 17% (27/164)를 차지하고 있다. 논문 발표한 기관별로 살펴보면 학교와 기타 기관(NTT연구소, ATR연구소, 전자기술종합연구소 및 각 기업체 포함)의 비율을 보면 음성인식의 경우 37건:45건(학교:기업), 음성분석합성 및 부호화가 35건:20건(학교:기업)음성청각의 경우 17건:10건(학교:기업)으로 나타났다. 이는 인식의 경우와 비교해볼때 음성청각, 음성분석합성 등 기초적인 분야에서는 학교가 기업보다 더욱 많이 연구하고 있다고 생각할 수 있다.

음성인식의 연구방향을 보면 특정화자 인식은 거

의 없고 대부분 불특정화자 총용량(수백 단어) 혹은 대용량의 음성자료를 대상으로 하고 있다. 인식 방법 또한 음소단위 혹은 음절로 이루어지며 기본적인 Approach 방향은 음성이 가지고 있는 본래의 성질을 이용하는 Bottom-up 방식과 Top-down의 방식을 결합하여 인식의 기본 Flow로 하고 있는게 많았다. (미국의 경우는 주로 통계학적 수법에 의존하는 경향이 짙음). 특히 지금까지와 다른 경향을 보이고 있는 것은 신경회로망을 이용한 음소 Segmentation, 음소분류, 음소인식이 많이 늘고 있다. 원인은 Computer 계산 능력이 향상되어 계산량 때문에 종래에는 생각지 못했던 학습이 이제는 약 2~3 개월의 실험을 거쳐서 가능케 되었기 때문이라 생각된다. (종래는 한번 학습하는데 1년 정도 걸리는게 보통임). 이러한 Approach를 하는 기관은 주로 ATR 자동 번역선화연구소와 NTT연구소이며, 학교에서는 아직도 생각치 못하고 있는 실정인 것 같다. 이렇게 신경회로망을 이용하게 되면 인식을 등이 훨씬 개선될 희망과 확실한 근거가 있어서라기 보다는 아직 미개척 분야, 즉 해보지 않았기 때문에 한나라는 어쩌면 미지에 대한 호기심으로 approach 하고 있는게 아닌가 하는 느낌을 받았다.

여기에 학회에서 발표된 논문 중에서 인상 깊었던 몇가지 연구방법 및 시스템에 대하여 소개하고자 한다.

음소인식과 음소의 Auto-labeling을 위하여 Spectral Vector Field Pattern을 이용하는 법(ETL), 동기분석 혹은 사후 확율을 이용하는 방법(후지쯔) Spectrogram 지식을 이용하는 방법(ATR), Time Spectral Pattern 이용(동북대) 등이 발표되었다. 이는 전혀 새로운 아이디어라기 보다는 종래의 방식에 대한 보완의 의미가 크다고 생각된다.

단어음성의 경우는 Stochastic DP법을 이용(보요하시 기술대), 단어철이 Network를 이용한 전화 음성인식을 16단어에 대하여 남성 75명의 음성자료

로서 평가하였다. 이는 현재 전화 서비스용 음성인식 응답 시스템에 사용되고 있는 수법이다.(TOSHIBA). 사후화블에 근거를 둔 Top-down과 Bottom-up을 최적으로 결합시켜 인식하는 방법으로 212단어를 남겨 각 10명이 발성한 음성자료로 평가하였으며(동복대), 음성지각 과정에 착안한 인식수법(동경대) 등이 발표되었다.

불특정화자 인식을 위하여는 많은 음성자료를 사용하여 분산이 큰 표준패턴을 작성하는 방법을 탈피하여 적극적으로 화자 적응하는 방법을 제안한 논문들도 여러 편 있었다. 그중에서도, 특히 Fuzzy V. Q에 의한 화자간 Spectrogram의 정규화를 위하여 2가지의 방법을 제안했는데 첫째는 변환 Code Book을 사용하여 입력 Vector를 변형하는 방법, 둘째는 화자 적응의 변환 Code Book 생성에 Fuzzy V. Q 이용하는 방법을 제안하여 실제로 그 유효성을 증명했다(ATR).

화자 인식에 관해서는 Text 독립형(NTT), 양자화 분포도 이용, V. Q 오차이용법 등이 소개되었다.

인식 시스템에 관한 논문으로써 Taxi 배차 예약 접수를 하는 자동전화접수장치를 구현하였다. 이는 공중전화 회선을 이용하여 12단어에 대하여 평가했다(마쓰시타), 내선번호 안내를 대상으로 한 음성 Q. A 시스템 구축(KDD), 그리고 오오사카 대학에서는 문절 발성을 전제조건 음성인해 시스템 STU-

RT 1을 구현하고 1000단어 문절에 대한 성능평가를 행한 결과를 제시하였다. 이 시스템은 3부분으로 나누어져 있으며 음향 처리부에서 음소군별로 인식하며, 문절후보생성부, 언어처리부로 이어져서 인식결과가 나오도록 되어 있다. Symbolics사의 Lisp 머친상에서 구현되었으며, 특히 음향처리부에서는 그래픽으로 표시된 음성의 특징 파라메타의 시간변화를 관측함으로써 음운의 인식이 가능한 음성 전문가의 지식을 이용한 지식 베이스 시스템이다. 즉 연속음성인식을 위한 지식 공학적 Approach로써 적응 Rule Induction기구를 실현했다.

한편 토요바시 기술대학에서는 초극저 Bit(100bit/s) 부호화 합성 방식을 제안했으며 발성 혹은 청각 장애자들을 위한 연구 및 보조훈련장치의 개발도 동경대를 비롯해서 여러 곳에서 제안되었다.

마지막으로 학회의 분위기는 발표시간 도중에는 어느 누구와 만나 인사하고 얘기를 나눌 수 없었다 왜냐하면 원로교수 등 소위 각 분야의 원로급들이 발표회장 내에서 제일 앞 자리에 진을 치고 앉아서(?) 열심히 경청하고 질문하고 발표자에게 유익한 충고와 Comment를 하며, 모두가 배운다는 입장과 논문 뒤에 숨겨진 정보를 얻기 위해 혈안이 되어 있기 때문이다. 이러한 모습이야말로 참다운 학회의 발표 분위기 혹은 풍경이라 생각된다.