

# 눈으로만 보는 시대에서 귀로도 듣는 시대로!



양 재 수  
(한국통신 정보통신 사업본부)

DB Trend

—올해부터 하이텔의 학습정보에 음성지원서비스 개시,  
\* 이제 온라인 DB도 멀티미디어시대로 진입

**컴**퓨터 통신망의 초기에는 단순히 데이터를 전송하거나 편지 또는 화일을 송수신하는데 초점이 맞추어져 있다가, 근간에는 TEXT 형태의 문자서비스에 그래픽이 첨가된 그림정보가 부가되어 정보통신에 활기를 띠고 있다.

최근 미국에서도 교육용 컴퓨터 통신망 설치가 늘어나면서 이를 교육 현장에서 활용하는 예가 크게 늘고 있다. 이제까지 컴퓨터를 활용한 교육은 디스켓 보급에 근거를 두고 대부분 교육용 S/W를 이용해 교과과정을 컴퓨터로 반복, 학습하는 정도에 그쳤으나 컴퓨터 통신망을 활용한 통신교육은 학생들의 자율적인 학습과 탐구 자세를 길러주고, 다른학교 또는 폭넓은 세계로 확대되어 지능적인 교육과 정보교환 및 자원을 공유할 수 있게 되어가고 있다.

한때 각광을 받던 컴퓨터-Based-학습교육이 사양길에 접어 들었다가, 이제 다시 고속정보통신망이 교육용컴퓨터 통신망으로 이용되면서 고개를 들고 있다. 미국내 교육용 통신서비스를 살펴 보더라도 내셔널 지오그래픽 키즈네트워크, 프레

드메일, FIDO넷, AT&T 러닝네트워크 및 글로벌랩 등 붐이 일어나고 있다. 뿐만 아니라 미국과 유럽간의 K12 NET은 미국내는 물론 유럽 대륙간의 교육 프로그램을 위한 정보 교환망으로 유치원에서 고등학교 3학년까지의 모든 관련 과정을 네트워킹화하여 정보 교환을 시도하고 있다.

국내에서도 이제 한국통신에서 저명하고 인기 있는 학원이나 교사 또는 교육 관련 전문기관의 협조를 얻어 교육용 데이터베이스(DB)를 개발 서비스를 제공하고, 점차 이를 확대 보급할 것이다. 아직까지는 교육 DB개발이 수익성 없고 일반 VAN 사업자가 참여하기에는 상당히 모험과 무리가 따르지만 상업적인 여건이 조성될 때까지 한국통신이 교육 DB개발 등에 당분간 투자를 하여 정보통신망을 이용하여 수요를 창출시키고 역할을 맡을 것이다.

따라서, 고객 또는 이용자가 문자, 비데오텍스(VTX) 형태의 단순한 2차원적인 정보에서 음성 지원까지 확대되는 멀티미디어 서비스를 요구하고 있고 세계적인 추세를 보더라도 멀티미디어

형태로 발전함에 따라, 한국통신에서 기존의 하이텔서비스에 음성 기능을 추가 개발할 계획이다. 이를위해 음성압축 기술과 개발된 음성압축 방식, 알고리즘간 압축율 및 기술적인 특징을 비교 분석하고, 그중에서 통신에 적합한 방식으로 평가된 IMBE(Improved Multiband Excitation)음성 압축 방식에 대해 먼저 알아보려고 한다.

향후 HiNET-P를 이용한 하이텔망은 명실공히 멀티미디어 서비스를 제공함으로써 실질적인 협대역 종합정보통신망(ISDN)으로 계속 발전하리라 기대되어진다.

끝으로 정보통신을 이용한 학습 음성지원 서비스와 이의 통신망을 소개하고, 향후 DB구축 및 음성지원 서비스 개발 방향을 제시하고자 한다.

## ■ 음성압축방식의 현황

영상압축 기술은 MPEG, JPEG등, 그리고 이미지 라이브러리 및 그래프 라이브러리 등이 개발되어 어떤 형식으로 화상을 압축하여 화일 또는 비트 스트림으로 변환시키는 기술 등 연구가 활발히 수행되어 왔고, 여러 형태로 상업용으로 이용되고 있어 멀티미디어에 활기를 더해 주고 있다. 그러나 데이터 통신을 이용한 멀티미디어 서비스는 아직도 제공되고 있지 않다. 그리고 비록 협대역(Narrowband)-ISDN이 있기는 하나 현재의 음성서비스는 음성DB를 구축한 서비스가 아니라 전화급 음성을 의미하고 있어, 팩킷 교환 회선등의 서비스에는 실현되고 있지 않다.

국내의 음성 합성에 대한 연구는 그 역사가 10년 이내 정도로 매우 짧으며, 일부의 언어 처리와 일부의 시스템 측면에서 구축되어 있으나 특히 언어학적인 기반이 약하며 데이터베이스 구축에도 어려움이 많다.

현재 음성지원이 가능한 음성압축 방식의 카드는 PCM, ADPCM, LPC, 그리고 IMBE 방식들을 들 수 있다. 음악지원이 가능한 카드는 옥소리, 사운드마스터, 사운드 블래스터 등의 예가 있으며, 이들의 녹음 방식은 PCM과 ADPCM을 쓰고 있다. 현재 PCM이나 ADPCM은 압축을 조정

불가능하고 동일 내용에 대한 단위 생성 데이터 량 즉, 기록용량(byte/s)은 PCM이 8,000 byte/s, ADPCM이 4,000 byte/s로써, IMBE방식의 300 byte/s에 비해 부가적으로 생성되는 데이터량이 너무 많아 통신으로 데이터를 보내는데 많은 시간이 요하게 되어 부적합하다고 볼 수 있다.

다음으로, LPC(Linear Prediction Code) 방식은 선형 예측 코딩방식으로 1967년 처음으로 발표되어 그동안 많은 발전을 해왔다. 음성 압축율은 26:1정도로 상당히 높으나 적절한 데이터율(Data rate)은 2400 bps에서 적합하고, 그 비율이 고정되어 있다. 이 방식의 알고리즘 처리 영역은 시간영역(Time Domain)이고, 화자 인식 능력(화자 음색 유지도)이 대체로 낮다.

IMBE 음성 압축 시스템(Speech Compression System)은 낮은 데이터율(Data Rate)에서 우수한 음질을 나타낸다. Digital Voice Systems, INC(DVSI)에 의하면, 이 시스템은 Background 잡음과 채널 에러에 매우 강하다고 보고되고 있다. 이 보고서에 의하면 채널 에러면에서는 DVSI의 IMBE가 가장 우수하고, 다음으로 모토롤라의 VSELP, Aware의 STC, 그리고 DOD의 CELP 순으로 채널 에러를 보여주고 있다.

알고리즘 또는 처리의 복잡성은 MIPS에서는 역시 IMBE가 제일 낮은 값으로 우수하였고, STC, VSELP와 DOD의 세가지 방식은 상호 근소한 차이를 보인 반면, 가장 복잡한 것은 IMBE의 거의 2배에 가까운 15MIPS를 나타냈다.

## ■ 음성지원 방식에 따른 특징

### 방식별 음성지원 카드 비교

국내에서 사운드 카드로 활용되거나 개발중인 사운드 지원 카드는 크게 옥소리, 사운드마스터, 사운드블래스터, 그리고 YES로 크게 분류할 수 있다. 압축 방식에 따른 주요 업체명은 PCM방식이 삼호전자, ADPCM이 한국로보트, LPC가 중앙전자 그리고 IMBE가 NAT회사가 주축을 이루고

구 분	PCM	ADPCM	LPC	MBE	비 고
Bit rate	64 Kbps	32 Kbps	24 Kbps	24~9.6Kbps	생성데이터량
압축을 조정	불가능	불가능	불가능	가능 (100bps단위)	
음 명료도	상	중	하	상	
소요 기록량	8,000	4,000	386	300	byte/s
재생 활용시간	150초	300초	3,109초	4,000초	12 MB
하드웨어 구성	음악 + 음성재 생/입력+MDI	음 성 재 생 ADPCM 칩+ 스피커	음성재생 DSP (NEC)칩 + 알 고리즘 S/W	음성재생 DSP칩+ 알고리즘 S/W	

[표 1] 음성지원카드 비교표

있다. 표[1]에 나타난 바와 같이 압축을 조정은 IMBE만이 가능하며, 단위 시간당 생성되는 데이터량은 PCM이 8,000byte나 생성되나, LPC나 IMBE는 각 386byte와 300byte로 Redundant한 데이터 생성량이 훨씬 작아 데이터를 전송하는데 시간이 적게 걸린다. 1.2Mbyte이 메모리 용량을 갖는다고 볼 때, 이용 또는 디스플레이하여 음성을 들을 수 있는 양은 PCM이 150초밖에 안되는 것에 비해 IMBE 방식을 적용했을 때는 4,000초 동안 활용이 가능하다.

참고로 옥소리나 사운드마스터, 사운드블래스터카드는 녹음방식에서 PCM이나 ADPCM방식을 쓰고 있으나 이들 대부분은 디스켓 보급형 Stand alone으로 음악을 대부분 지원하고 있다. PC통신에 올려 활용되는 사운드 카드의 이용 및 보급현황은 다음절에서 설명하고자 한다.

## 사운드 카드 이용 현황

[표2] 사운드카드 활용 하이텔 동호회 현황

구 분	옥 소 리	사운드블래스터	사운드 마스터	PC 라운드	계
일평균이용시간	450 시간	50 시간	25 시간	85 시간	610
일평균이용자수	2,000 명	600 명	230 명	600 명	3,430 명
일평균이용횟수	3,500 회	800 회	300 회	800 회	5,400 회
회 원 수	14,730 명	5,120 명	1,530 명	1,770 명	

\* 음악(노래방)서비스만 제공

하이텔에서는 통신을 이용한 노래방을 호스트에 올려 사운드카드를 보유한 회원을 대상으로 음악지원 서비스를 실시하여 하이텔내 사운드카드 활용 동호회가 구성되는등 이용자 수가 날로 증가하여 PC통신의 위력을 과시하고 있다. 표[2]는 '93년 11월 현재 사운드카드별 이용현황을 일평균 이용시간, 일평균 이용자수, 일평균 이용횟수 및 회원수별로 집계한 것이다. 이표에 의하면 옥소리카드의 경우 회원수가 약 15,000명으로 가장 많고 최근에 개발된 사운드카드인 PC라운드가 약 2,000명선에 이르고 있다.

한국통신에서는 1차적으로 공익적인 차원에서 학습정보를 개발, 호스트에 올려 학습용 음성지원 서비스를 '94년 상반기부터 실시하게 된다. 이는 기존의 노래방 목적의 사운드 카드가 압축없이 전송하는 것에 비해 한국통신이 제공하는 것은 고압축 알고리즘 방식에 의해 제공함에 따라 전송에 따른 많은 시간을 절약할 수 있어, 이용

자가 많은 비용을 절감할 수 있고, 신상품인 Voice 서비스를 지원받게 되는 것이다.

뿐만 아니라 한국통신에서는 금년 1월 음성서비스 개시와 함께 상반기 중에는 음성/음악 통합 서비스를 (음성:IMBE방식, 음악:사운드블래스터 호환)를 실시할 예정인데 이용자의 입장에서는 별도의 카드 구입없이 한장의 카드만으로 전송형(Down Loading) 및 온라인 학습서비스, 노래방 서비스 등을 받을 수 있게 되어 이용자 수는 지원되는 DB의 확충과 함께 상당히 늘어나리라 예상된다. 이는 기존의 노래방서비스 한가지만으로도 급격히 확대되고 있는 실정을 감안한다면, 두가지 기능이 한꺼번에 제공될 경우에는 상당히 전망이 밝은 서비스로 기대되어 진다.

## IMBE 알고리즘

IMBE는 Frequency Domain에서 Waveform Coding 방식으로 압축한다. 음성을 유성음/무성음, 즉 Voiced/Unvoiced Sound Decision Base으로, 한 프레임 내에서도 여러 Band에 대해 구분하여 음색을 유지한다. 또한 Waveform Coding 방식을 쓰므로 계산량이 LPC에 비해 적다.

## 압축절차

압축, 즉 Encoring하는 절차를 간략히 설명하자면 아래와 같이 8단계로 나눌 수 있다.

1. 아날로그 신호를 디지털 신호로 변환(A/D 변환).
2. 일정시간에 대한 Speech Segment로 분할.
3. Segment의 Original 파형을 FFT(Fast Fourier Transform)로 변환처리.
4. Dynamic 프로그램을 이용, Pitch 측정을 하여 기초 주파수를 구함.
5. Spectral Envelop를 산출.
6. 각 주파수 대역별로 Voice/Unvoice 판정.
7. 구해진 Spectral Envelop와 기초 주파수, Voice/Unvoice 판정 내용 기록.
8. 다음 Speech Segment를 구하고 (3)으로 가서 반

복 수행함.

## 복원절차

복원, 즉 Decoding하는 절차를 아래와 같이 4단계로 크게 나눌 수 있다.

1. Voice/Unvoice 판정을 Band별로 적용하여 유성음이면 기초 주파수의 스펙트럼을, 무성음이면 White Noise를 적용하여 주어진 프레임에 대한 Excitation 스펙트럼을 만듦.
2. Excitation 스펙트럼에다 Spectral Envelope를 곱하여 Original 스펙트럼을 합성함.
3. Inverse FFT 변환하여 합성음 데이터를 생성.
4. 마지막으로 D/A 변환하여 음성으로 출력함.

## 음성지원 서비스와 통신망

한국통신에서는 음성지원 서비스를 위해 자체 연구소인 소프트웨어 연구소에 저작도구용 음성처리 장치 시험 평가를 의뢰한 결과 IMBE 방식이 우수한 방식으로 평가되어, 이를 사업화하기로 하고 데이터통신에 음성지원 서비스를 추진하고 있다.

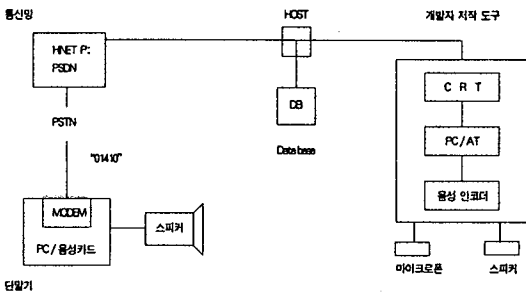
데이터통신용 음성지원 서비스를 소프트웨어적인 측면으로 고려한다면 음성(Voice)입력 및 코스웨어화하기 위한 Authoring Tool(EDSCOP), 음성 및 문자, 이미지 등의 정보를 갖고 있는 데이터베이스(DB), 그리고 이용자가 음성카드를 통하여 스피커로 들을 수 있는 카드구동용(Driver) S/W의 세가지를 들 수 있다.

이용하기 위한 필요장비나 이용방법은 다음과 같다. 음성카드가 내장된 PC 본체와 모니터, 소리를 들을 수 있는 헤드폰, 또는 오디오 앰프나 일반 소형 스피커중 어느 것이든 사용이 가능하며, 2,400bps 통신용 모뎀이 있으면 된다. 이용방법은 하이텔 이용 접속번호 01410을 돌려 하이텔 메뉴에서 음성 교육 정보를 선택하여 학습하고자 하는 화일을 본인이 갖고 있는 PC에 Down load 받아 학습도구 실행 화일을 동작시키면 문자 및

이미지와 함께 음성으로 최고의 학습효과를 올릴 수 있다.

음성서비스를 위한 통신망 구성도는 [그림1]과 같으며, 크게 세분야로 나눌 수 있다. 음성카드를 내장한 PC단말기와 공중통신망을 구성하는 공중전화망 PSTN(Public Switching Telephone Network) 및 패킷교환망 PSDN(Packet Switching Data Network), 그리고 호스트 및 DB로 구성되어 있다.

멀티미디어 서비스로서의 역할을 제대로 다 하자면, 음성카드의 지속적인 기능 보강 및 알찬 DB와 전송로의 품질향상과 함께 모뎀의 속도향상(2,400bps → 4,800 또는 9,600bps)을 들 수 있다.



[그림1] 하이텔 음성지원 서비스 구성도

## ■ 멀티미디어서비스 확대될 것

정보통신의 고속화 추세와 발맞추어 서비스에 있어서도 단순한 문자 형태의 Text 서비스에서 이미지 형태로 변천해 왔으며, 근간에는 멀티미디어에 대한 고객 욕구가 상당히 대두되어 정보화사회에 있어 음성서비스가 지원되어야 하는 시대적 요청에 있다. 현재는 음성지원 서비스가 PC용 Down load 형태로 개발되고 있으나 '94년 말까지는 하이텔단말기 및 PC를 표적으로 한 On-Line용 음성 데이터베이스(DB)도 개발될 예정이다.

뿐만아니라 음성용 카드도 다양화시켜 음성에 음악 기능을 추가한 복합적인 음성지원 카드도 개발될 예정이며, 모뎀을 갖지 않은 PC 이용자를 위해 음성 + 음악 + 모뎀의 다기능을 갖는 카드로 시판되리라 전망된다.

앞으로 통신의 전송속도를 높일 수 있는 모뎀이 2,400bps에서 4,800bps 또는 9,600bps로 전환될 경우 On-Line용 음성지원 서비스도 PC 또는 하이텔 단말기상에 지연없이 실시간(Real-time)으로 서비스가 가능하고 모뎀의 속도 상향 조절에 대한 고객의 요구가 높아 조만간에 멀티미디어 서비스가 활성화 되리라 기대된다.

음성DB도 학습정보에만 국한되는 것이 아니라 일반생활정보, 공공정보 등에도 이용이 가능하여 정보 전달의 효과를 배가할 수 있다. 한국통신이 채택한 IMBE방식은 배경 잡음에 대한 영향이 없고 압축율이 PCM의 1/26 내지 1/8 정도로 뛰어나며, 음색 구분이 가능하고, 직접 음성입력으로 어느 국가 언어도 사용이 가능하여 앞으로 각 기관이나 여러 국가에서 표준으로 채택할 가능성이 크며, 이미 INMARSAT(국제해사위성통신기구) 등에서도 이를 표준으로 채택하고 있다. 응용분야는 이동무선(Mobile Radios), 셀룰라 폰(Cellular Phone), 안전통신(Secure Communications), 음성다중(Voice Multiplexing), 음성메일(Voice Mail), Dictation Equipment 및 멀티미디어 응용분야 등으로 이용할 분야가 많을 것으로 예측된다.

결과적으로 향후 DB서비스는 문자, 비디오, 음성정보를 동시에 제공하는 멀티미디어 서비스로 매체의 전달효과를 극대화시키며, 실질적으로 협대역 정보통신망(N-ISDN)으로 계속 발전하리라 전망된다.

\* 필자는 59년생으로 항공대 통신공학과와 건국대 전자공학과(석사)를 거쳐 미국 New Jersey Institute of Technology에서 박사학위를 취득한 후 현재는 한국통신 정보통신사업본부 하이텔 사업2부장을 맡고 있다.