《特 輯》

음성 처리 기술의 현황과 전망

김형순 · 김희동 · 임병근 · 은종관
(디지콤 정보통신연구소)

■ 차 례 ■

1 서 론

음성은 인간의 가장 기본적이며, 친밀할 뿐 아니라 가장 오래된 정보 전달의 수단으로서, 음성의 공간적 제약을 극복하기 위한 통신 기술의 발전은 1962년 미국 AT&T사가 PCM(Pulse Code Modulation)에 의한 디지탈 방식의 음성 통신 서비스를 시작한 이래 지난 20여년 동안 괄목할 만한 발전을 하여 왔다. 이러한 발전은 반도체 및 디지탈 신호처리기술 등 여러 분야에서 이루어친 급속한 발전에 힘입어, 인간과 기계 사이의 통신도 가능하게 되는 단계로 이어지고 있다.

현대는 정보화 시대로서 구미 선진 각국은 음성, 영상, 데이터 등 신호의 종류에 관계없이 이들을 디지탈화하여 같은 전송망을 통해 송수신할 수 있는 종합정보 통신망(ISDN)의 구축에 박차를 가하고 있다. 이러한 추세에 따라 국내에서도 디지탈 통신망의 증설이 빠른 속도로 이루어지고 있고 머지않아 우리도 종합정보 통신망을 갖게 될 전망이다.

향후 2000년대까지 통신 및 컴퓨터 기술이 합쳐져 이룩되는 정보통신 사업은 전체 산업중 가장 큰 비중을 갖는 유망한 산업이 될 것인바, 그 파급효과는 개인 생활은 물론 사업 전반의 발전에 지대한 영향을 줄 것이며 나아가 국가 산업 발전의 척도가 될 것이다.

현재 국내에도 음성처리 기술 및 관련 기술의

기술 축적이 진행되어, 이를 토대로 국내 기술에 의한 음성처리 시스템이 속속 개발되고 널리 활용되고 있으므로, 이 분야의 기술이 더욱 발전할 것으로 예견된다. 이러한 시기에, 이 분야의 현상 및 전망에 대한 개괄적인 정리를 해 두는 것도 의미있는 일로 여겨진다.

본고에서는 각종 통신 서비스중 가장 중요한 부분을 차지하는 디지탈 음성통신 기술에 관하여 살펴 보고자 한다. 본 논문에서는 음성 처리를 음성부호화, 음성인식, 음성합성의 3가지로 나누어서 그 기술의 현황 및 응용분야를 중심으로 살표본 후 향후 전망에 대하여 기술하고자 한다. 각 부분마다 매우 다양한 방법들이 연구되고 있기 때문에 상세한 것을 다룰 수는 없으나, 큰 연구의 방향 및 그 개론들에 대하여 전문성이 있도록 설명하고자 노력하였다.

표 1. 음성부호화 방식의 비교

방 식	알고리즘	전송 속도	음성 품질	복잡성
파 형 부호화	PCM	64kbps	대단히 좋다	소
	ADPCM	32kbps	좋 다	중
		16kbps	비교적 좋다	
	ADM	32kbps	좋 다	소
		16kbps	비교적 좋다	
혼 합 부호화	APC	16kbps	좋 다	대
		9.6~8kbps	비교적 좋다	
	SBC	16kbps	좋 다	중
	ATC	16kbps	좋 다	대
	RELP	16kbps	비교적 좋다	중
	CELP	9.6~7.2kbps	좋 다	매우복잡
		7.2~4.8kbps	비교적 좋다	
	MPLPC	9.6~7.2kbps	좋 다	대
보코더	LPC	8.0~2.4kbps	조금 나쁘다	대
	Formant	12.kbps	조금 나쁘다	매우복잡

[2] 디지탈 음성부호화 기술

2.1 개요

음성의 디지탈 부호화 기술은 크게 나누어 세가지로 분류할 수 있다. 이중 첫째는 음성 파형을 샘플링하여 양자화하는 파형부호화(waveform coding) 방식이고, 둘째는 음성의 주기와 성도의 계수 등 음성의 특징만 추출하여 전송해서 수신측에서 음성을 재생하는 분석 합성에 의한 보코딩(vocoding) 방식이며, 세째는 파형부호화 방식과 분석, 합성방식의 이점만 사용하는 혼합부호화(hybrid coding) 방식이다. 이들 부호화 방식들을 전송 속도에 따라 구분하면 표 1과 같다.

표 1에서 보는 바와 같이 파형 부호화 방식은 전송 속도가 16내지 64kbit/s(bps)로서 비교적 높지만 음질이 우수하여 일반 음성 통신에 많이 사용되고 있다. 혼합부호화 방식은 전송속도가 4.8 내지 16kbps로 비교적 낮기 때문에 모뎀을 사용해서 기존의 아날로그 회선으로 음성을 전송할 수 있는 이점이 있지만 음질은 일반적으로 파형부호화 방식보다 떨어진다.

한편 보코딩 방식은 전송 속도가 50bps에서 4.8kbps로서 매우 낮지만 부호기가 복잡하고 음질에 아직도 문제점이 있는 것이 단점이다. 혼합부호화나 보코딩방식은 현재 일반 상용 음성통신보다는 군통신 등 특수통신에 사용되고 있고, 많은 연구결과 음질이 점차 좋아짐에 따라 앞으로 그 사용 범위가 확대될 것으로 기대된다. 최근 디지탈 셀룰과 통신 방식이 차세대 음성통신의 총아로서 대두됨에 따라 혼합부호화 방식에의 관심이 한층 고조되고 있다.

계속해서 위에서 기술한 음성부호화의 세가지 방식들을 보다 자세히 기술한다.

2.2 파형 부호화 방식

파형부호화(waveform coding) 방식중 현재 가장 많이 사용되는 것은 PCM 방식이다. PCM은 음성신호를 부호화하는데 있어 개념적으로 가장 간단한 방식으로 제한된 대역폭(300~3400Hz)의 음성을 8kHz로 샘플링해서 2^8 레벨로 양자화한 뒤 코딩하여 64kbps로 송신한다. 양자기(quantizer)로는 음성의 진폭 분포에 의해 설계하여 대수(logarithm) 특성을 갖는 비선형

양자기가 주로 사용된다.

이러한 비선형 양자기가 선형 양자기에 비해서는 성능이 우수하지만 이들 모두 그 레벨이 고정되어 있기 때문에 입력 신호의 진폭이 클 경우 잘릴 가능성이 있다. 이러한 문제점은 적응양자기(adaptive quantizer)를 사용함으로써 해결할 수 있다. 한 예로 입력신호의 진폭에 따라 양자기의 최소 및 최고 레벨을 조절해 줌으로써 PCM 의 성능을 향상시킬 수 있는데 이를 APCM(Adaptive PCM)이라 한다.

PCM의 전송 속도는 64kbps로서 이는 대역폭의 사용면에서 볼 때 아날로그 통신 방법보다 훨씬 비경제적이다. 따라서 음성의 대역폭 축소에 관한 연구의 한 결과로 음성신호의 redundancy 를 이용한 예측부호화(predictive coding) 방식들이 제안되었는데, 대표적인 예로는 ADPCM(Adaptive Differential PCM)과 ADM (Adaptive Delt Modulation)을 들 수 있다.

예측 부호화 방식의 기본 원리는 과거에 들어온 음성 신호의 샘플들로 부터 다음에 들어올 신호의 크기를 예측하여 실제 입력 신호로부터 빼 줌으로써 오차 신호를 발생시켜 이 신호를 양자화하여 전송한다. 이 오차 신호의 진폭은 입력 음성 신호의 진폭보다 훨씬 작기 때문에 그만큼 양자화 레벨수도 줄어들게 된다.

대표적인 예측부호화기인 ADPCM의 블록도가 그림 1에 그려져 있다. ADPCM은 기본적으로 적응예측기와 적응양자기의 두개의 서브 시스템으로 나눌 수 있다. 예측기는 예측 필터의 형성을 위한 계수를 고정시키는냐 아니면 입력 신호에 따라 변화시키느냐에 따라 고정예측기(fixed predictor)와 적응 예측기로 구분된다. 또한 적응예측기의 경우, 필터의 계수를 매 샘플마다 변화시키느냐 아니면 일정 시간마다 변화시키느냐에 따라 순차 적응(sequential adaptation) 과 블록적응(block adaptation)으로 나누어진다.

예측 오차를 최소화하는 데에는 블록 적응 방식이 보다 효과적이나, 이 경우 필터계수를 수신측의 송신하기 위해 양자화된 오차 신호화의 다중화를 해야 하므로 시스템이 복잡해지는 단점이 있다. 반면 순차적응 방식은 계수를 매 샘플마다 update 시킴으로써 예측 오차를 최소화시키는데에는 문제점이 있지만, 필터 계수를 따로 송신할 필요가 없고 수신단에서 수신된 오차 신호로 송신부에서와 같이 계수를 update 시킴으로써 시스템의 구성면에서 훨씬 간단하다.

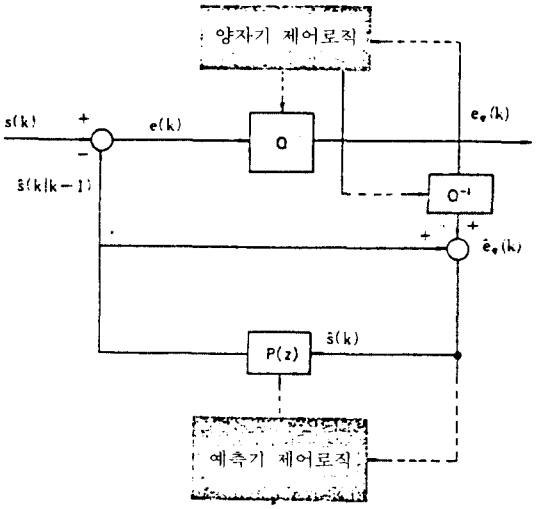


그림 1. ADPCM 송신기의 블록도

한편 ADPCM의 양자기는 적응 방식을 주로 사용한다. 양자기의 레벨크기를 적응시키는 방법은 매 샘플마다 레벨을 변화시키는 순간압신(instantaneous companding) 방식과 매 약 5 msec마다 변화시키는 syllabic companding 방식이 있는데, 전자가 후자보다 신호 대 잡음비 면에서 1.2dB 높으나 채널에 오차가 있을 때 성능이 급격히 떨어지는 단점이 있다.

지난 20년동안 ADPCM에 관해 많은 연구가 이루어져 그 결과 여러가지 ADPCM 부호화 방식들이 제안되었다. 그중에서도 CCITT의 음성통신 연구 그룹에 의해 제안된 ADPCM은 앞으로 디지탈 음성통신에 큰 영향을 줄 것으로 기대된다. CCITT 표준 G.721 ADPCM의

성능을 몇가지 열거하면 다음과 같다.

- Codec이 전송속도 32kbps에서 운용되도록 설계되고,
- 음성신호 뿐 아니라 전화선을 사용하는 데이터 신호 및 톤(tone) 신호까지 부호화할 수 있도록 설계되며,
- 현재의 64kbps PCM과 직접 연결이 가능하고,
- 예측기와 양자기가 입력 신호에 적응하도록 설계되며,
- 채널 오차에 비교적 강한 특성을 갖고 있다.

이 ADPCM 시스템은 지금까지 연구 개발된 ADPCM중에서 가장 복잡한 시스템의 하나이지만 하드웨어를 구현하는데 있어서는 배율기(multiplier) 를 사용하지 않고도 구현할 수 있도록 설계되어 있다. 시스템의 성능면에서도 음성신호를 부호화할 경우의 신호 대 잡음비가 28 dB로서 다른 ADPCM보다 약 4dB 정도 우수하다.

예측 부호화기의 다른 한 부류로는 ADM을 들 수 있다. ADM도 입력 음성신호를 직접 양자화하지 않고, 오차신호를 양자화하는 점에서는 ADPCM과 같으나, 입력 음성을 PCM이나 ADPCM 에서 사용하는 나이키스트 속도(Nyquist rate)보다 훨씬 높은(보통 2~4배) 속도로 샘플링하고, 그대신 양자기의 레벨수는 단 2개(즉 1비트)를 갖는다는 면에서 다르다. 이렇게 ADM은 PCM이나 ADPCM과는 달리 각 샘플이 워드 단위가 아닌 비트 단위로 부호화되기 때문에 채널 에러에 강하다는 장점이 있다. 또한 ADM의 하드웨어 구현은 ADPCM보다 훨씬 간단하며, 입력 및 출력단에서 사용하는 필터도 간단하다.

ADM 방법의 대표적인 예로는 현재 상용의 LSI가 개발되어 가장 많이 사용하고 있는 CVSD (Continuously Variable Slope Delta Modulation)을 들 수 있는데 이는 양자기의 스텝크기를 입력 신호의 진폭변화에 따라 서서히 변화시키는 syllabic companding 방식을 채택하고 있다.

2.3 보코딩 방식

보코딩(vocoding) 방식은 음성파형을 직접 양자화하지 않고 음성파형을 분석하여 유/무성음 구별, 기본주기, 성도의 계수 등 음성의 특징만을 추출해서 전송하기 때문에, 전형적인 전송속도가 2.4~4.8kbps로 아주 낮은 반면 시스템의 구조는 상대적으로 복잡하다. 일반적인 보코더의 구조가 그림 2에 그려져 있다.

보코딩 방식의 대표적인 예로는 선형 예측 부호화(LPC : Linear Predictive Coding) 원리를 이용한 LPC 보코더(vocoder)을 들 수 있다. LPC 보코더의 원리를 간략하게 설명하면 다음과 같다.

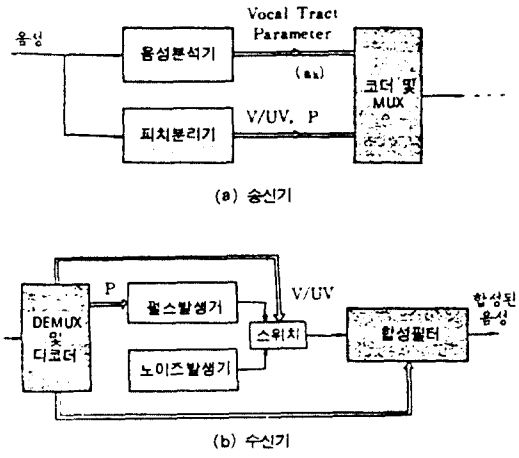


그림 2. 보코더의 일반적인 블록도

일반적으로 음성 신호와 같이 상호관계(correlation) 가 강한 신호는 일정한 수의 이전 샘플로부터 다음 샘플의 값을 예측할 수 있다. 이 예측되는 샘플의 값은 이전 샘플 값들의 선형 결합으로부터 얻어진다. 선형 예측에서의 문제의 초점은 이러한 선형 결합시의 계수(예측계수라고 부르며 보통 10~14개)들을 구한 일로, 이들은 음성샘플의 예측된 값과 실제 값의 오차를 최소

화시킴으로써 얻어진다.

예측 계수는 음성 신호의 non-stationary한 특성 때문에 20-30msec마다 얻어지는데, 이 계수들이 바로 음성 발생 기관인 성도를 특징지워 준다. 성도를 수학적으로 나타내는 데에는 폴(pole)의 계수만을 이용하는 all-pole 모델과 폴 및 제로(zero)의 계수를 모두 이용하는 폴-제로 모델이 있는데, 후자가 전자보다 비음 등의 자음을 묘사하는데 효과적인 반면, 하드웨어를 구현하는데는 계산량이 큰 제약 조건이 따르게 된다.

All-pole 모델은 예측 계수를 구한 과정에서 윈도우를 사용하느냐의 여부, 그리고 예측 오차를 어떻게 최소화하느냐에 따라 자기상관(autocorrelation) 방법과 covariance 방법으로 나눌 수 있다. 이중 후자는 윈도우를 사용하지 않고, 예측오차는 일정 구간에서만 최소화시킨다. 반면에 전자는 윈도우를 사용하되 예측 오차를 시간에 관계없이 최소화 시키며, 이 방법에 의하면 성도필터의 안정도를 항상 보장할 수 있기 때문에 실제로 후자보다 많이 사용된다.

보코딩에서 성도의 계수를 구하는 작업외에 또 하나의 중요한 일은 매 20msec 정도마다 음성이 유성음인지 무성음인지 판별하여 유성음일 경우 그 주기를 찾아내는 일이다. 유성음과 무성음의 구별은 유성음이 무성음보다 에너지는 크되 zero crossing rate는 작다는 특성 등을 이용하여 수행할 수 있다. 한편 유성음의 주기를 찾는 방법으로는 자기상관, AMDF(Averate Magnitude Difference Function) 등의 방법들이 많이 사용된다.

위에서 구한 여러가지 파라미터들, 즉 예측계수, 유/무성음 판별을 포함 한 주기에 관한 정보들은 부호화되어 수신기에 전달된다. 수신기에는 이들 정보를 받아 예측계수들로 인간의 성도에 해당하는 합성필터를 형성하고 주어진 시간 구간이 유성음일 경우에는 여기(excitation) 신호로 피치의 주기에 따른 펄스를, 무성음인 경우에는 랜덤(random) 잡음을 내보내어 합성필터를 여기시킴으로써 합성을 만들어낸다.

LPC 보코더는 전송 속도가 2.4kbp일 경우 다른 보코딩 방식보다 비교적 음질이 좋으나 아직도 음질 및 음색에 문제점이 있으며, 주변환경에 잡음이 심하거나 음파왜곡(acoustic distortion) 이 있을 경우 음질이 급격히 저하되는 단점이 있으며, 이러한 문제들을 최소화 하기 위한 연구들이 진행되고 있다. 또 LPC 계수들은 기본적으로 시간 영역에서 정의된 파라메타들로서 시간축상에서의 선형 보간의 성질이 나쁘기 때문에 주파수 영역의 파라메타를 시간 영역으로 변환한 LSP(Line Spectrum Pair) 계수를 활용하는 연구도 활발히 진행되고 있다.

한편 LPC 보코더 보다도 낮은 전송속도인 500~1,200bps에서의 음성 통신을 위해 formant 보코더가 개발되어 있다. 이 방식은 유/무성음 구별 및 피치의 주기를 찾는 면에서는 LPC 보코더와 동일하나, 선형 예측계수 대신 주파수 영역에서의 공진점(resonant point : 이를 formant 라고 함)의 주파수 및 그 진폭을 추출하여 전송함으로써 전송속도를 500bps까지 낮출 수 있다. 그 밖에도 channel 보코더, phase 보코더, homonorphic 보코더 등이 개발되었으나, LPC 보코더만큼 많이 사용되지는 않고 있다.

2.4 혼합 부호화 방식

혼합 부호화(hybrid coding) 방식은 크게 시간 영역의 혼합 부호화 방식과 주파수 영역의 혼합 부호화 방식으로 나눌 수 있다. 전자의 대표적인 예로는 RELP(Residual Excited Linear Prediction), APC(Adaptive Predictive Coder) 및 최근 각광을 받고 있는 CELP(Code-excited Linear Prediction), MPLPC(Multi-Pulse LPC) 와 RPE-LPC(Regular Pulse Excited LPC) 등을 들 수 있다. 그리고 후자의 대표적인 예로는 SBC(Sub-Band Coding)와 ATC(Adaptive Transform Coding)가 있다.

이들 혼합 부호화 방식들은 파형 부호화에 비해 전송 속도가 훨씬 낮지만 시스템이 복잡하다는 점이 큰 단점이다. 그러나 최근 VLSI 기술이 많이 발전하였고, 또한 특수한 디지탈 신호처

리 소자들이 개발되고 있기 때문에 하드웨어를 구성하기에 간단하고 가격도 점차 저렴화 되고 있다.

먼저 RELP 보코더를 소개하면, 그 블록도가 그림 3에 나타나 있다. 이 시스템은 원래 제안된 RELP 시스템을 약간 변형한 것으로 9.6kbps에서 작동되도록 설계된 것이다. RELP 보코더의 원리는 앞장에서 기술한 LPC 보코더의 원리에 기초를 두고 있으며, RELP 보코더의 핵심 부분의 하나는 수신단에서 여기(excitation)로 사용하기 위한 신호를 만들어 전송하는 부분이다.

여기서 여기신호는 LPC 역(inverse) 필터에 의해서 생긴 오차신호 또는 잔류(residual) 신호를 두개의 밴드로 bandpass filtering하여 sampling rate를 줄인 다음 이를 APCM으로 부호화하여 LPC 예측 계수와 함께 전송함으로써 전송속도를 파형 부호화에 비해 낮출 수 있다.

수신단에서는 역과정으로 bandpass filtering된 신호로부터 비선형 처리를 통해 본래의 full-band residual을 만들 수 있으며, 이 신호를 진폭조정하여 LPC 합성 필터의 입력 신호로 사용한다. RELP 보코더는 9.6kbps에서 다른 시스템보다 음질이 우수하며, 주위환경이나 잡음에 강한 장점이 있다.

APC는 LPC 원리를 근거한 일종의 ADPCM이라고 생각할 수 있다. 음성 신호에는 두가지의 여분(redundancy) 요인이 있는데, 하나는 주기적인 피치에 의한 것이고, 다른 하나는 성도의 계수에 의한 것이다. 따라서 음성 신호의 기본 주기를 구하고 성도 계수를 매 일정한 시간마다 선형예측 방법으로 구하면, 이들 정보로써 ADPCM 과 같이 예측기를 형성할 수 있다. 이 예측기의 예측 신호를 입력 음성신호와 비교하여 만들어진 예측오차를 양자화하여 예측 계수 및 피치 정보와 함께 전송한다.

수신기의 구조는 송신기의 피드 백부분과 같고, 이들 전송된 정보들로부터 음성을 재생시킨다. APC에서는 LPC나 RELP와는 달리 단지 4개 정도의 예측 계수를 사용하며, 특히 16kbps

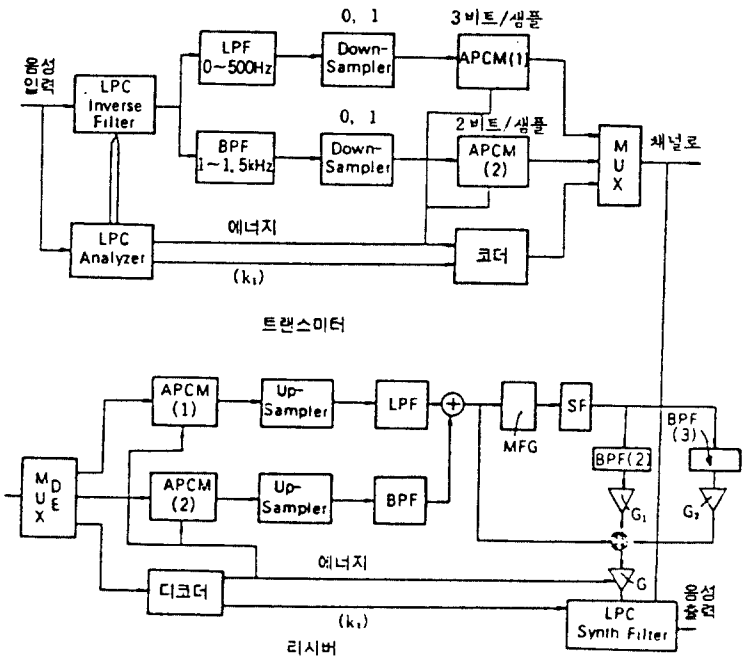


그림 3. RELP 보코더의 블록도

에서 우수한 음질을 갖는다.

지난 10년간에 걸쳐 파형 부호화기와 보코더 사이의 간격을 메꾸기 위한 많은 노력의 결과로 일련의 새로운 부호화기들이 개발되기에 이르렀는데, 이들이 CELP, MPLPC, 그리고 RPE-LPC 이다. 이들 부호화기는 그 구조상 공통점을 가지는데 이는 여기신호에 의한 출력 파형과 원래의 음성파형과의 차이에 해당하는 신호를 청각 기관의 특성에 의거한 perceptual weighting filter로 통과시킨 다음 이 결과를 최소화시키는 여기 신호를 선정하는 것이다. 그림 4에 LPC 보코더와 MPLPC, 그리고 CELP 부호화기의 구조 및 여기 신호에서의 차이점이 나타나 있다.

CELP 부호화기에서는 가능한 여기 신호 벡터들의 코드북(codebook)을 미리 구성한 뒤 부호화기 과정에서 최적의 여기신호 벡터를 선정하여 그 코드와 gain 값을 전송한다. MPLPC의 경우에는 여기 신호를 제한된 수효의 펄스로 묘사하며, 부호화 과정에서는 이를 펄스의 위치 및 크기 정보를 추출하여 이를 부호화하여 전송한다. RPE-LPC는 MPLPC와 유사하며 다만 펄스의 위치를 미리 고정시켜 놓고 크기 정보만을 부호화하여 전송하는 방법을 택한 것으로, 유럽의 디지탈 셀롤라 표준인 GSM(Group Special Mobile)의 음성부호화 방식으로 채택되었다.

CELP의 경우에도 디지탈 셀롤라에 사용되기 위한 최대 조건이 전송 오차에 인내성(robustness) 를 갖고 processing delay가 짧아야 한다는 2개의 조건을 만족하면서 또한 계산량이 적어야 구현 가능하다는 현실적인 문제를 해결하기 위한 연구가 활발히 진행되고 있다. 그중 미국 Motorola 연구진에 의해 제안된 VSELP(Vector Sum Excited Linear Prediction)라는 방식이 최근 미국 전기통신협회와 일본에서 디지탈 셀로라 통신에서의 음성 부호화 방식의 표준안으로 채택되었다.

다음으로 주파수 영역에서 음성을 부호화는 대표적 예인 서브밴드 코더의 구조가 그림 5에 그려져 있다. 이 방식은 음성 신호를 4~8개의 밴드패드 필터로 통과시킨 후, 이들을 APCM 이나 ADPCM 등 파형 부호화기를 사용해서 부호화·다중화하여 전송한다. 수신측에서는 이의 역과정을 거쳐 디코드된 각 밴드의 신호를 합하게 되면 원하는 음성 신호를 재생할 수 있다.

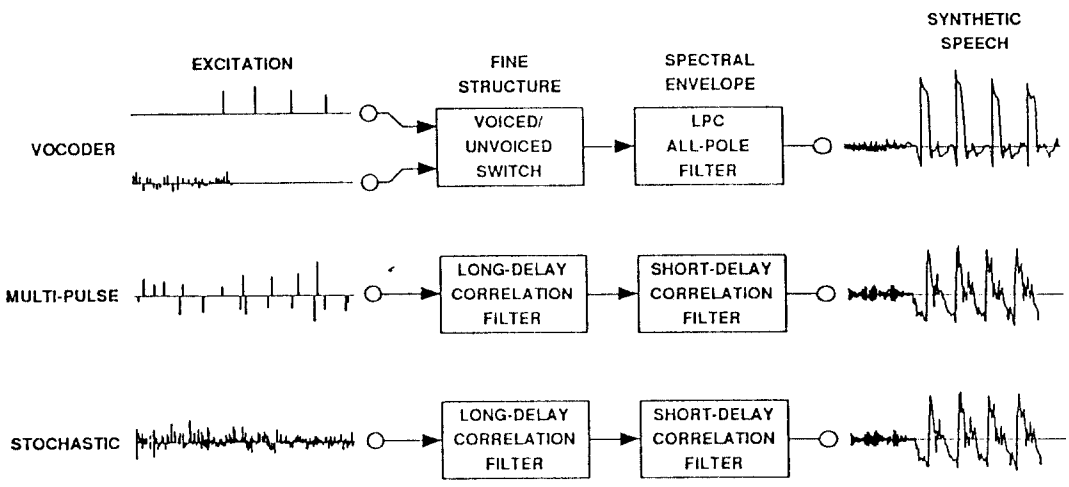


그림 4. LPC, MPLPC 및 CELP 부호화기의 차이점

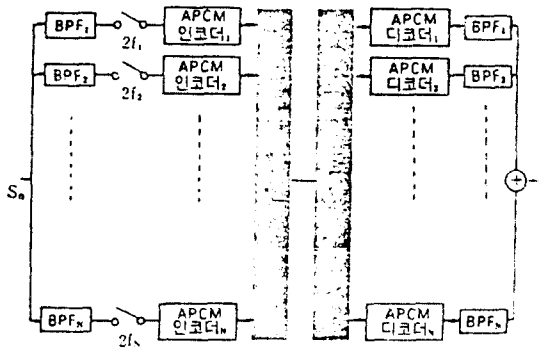


그림 5. 서브밴드 코더의 블럭도

[3] 음성인식 기술

음성인식 기술의 궁극적인 목표는 임의의 화자가 문법에 제한없이 일상적으로 자연스럽게 발음한 문장을 인식하여 그 의미를 파악하는 시스템의 개발이다. 그러나, 현재의 기술 수준으로 볼 때 이러한 목표 달성은 앞으로도 상당 기간동안 실현되기 어려울 것으로 보이며, 실제로 음성인식 기술에 대한 연구도 특정한 제약 조건하에서 제한된 목표를 가지고 이루어 지고 있는 것이 현실이다.

음성 인식은 인식하고자 하는 음성의 형태에 따라 단어의 시작과 끝을 명료하게 구분하여 발음한 음성을 인식하는 격리단어인식(isolated word recognition), 적은 수의 어휘를 대상으로 단어들을 연결지어 발음한 음성을 인식하는 연결 단어인식(connected word recognition), 그리고 비교적 많은 어휘를 대상으로 하여 일상 회화체로 자연스럽게 발음한 문장을 인식하는 연속음성 인식(continuous speech recognition)의 세 부류로 크게 나눌 수 있다. 이들 중 단어와 단어 사이의 경계가 불명료한 연결단어 인식이 격리단어 인식보다 어렵고, 문장 내용을 이해하기 위한 구문론, 의미론, 실용론 등 언어학적 지식들이 동원되어야 하는 연속음성 인식이 보다 더 어렵다. 또한 인식 대상 화자의 구분에 따라 훈련과정을 거친 특정화자의 음성만을 인식하는 화자종속(speaker-dependent) 음성인식과 임의의 다수 화자의 음성을 인식할 수 있는 화자독립(speaker-independent) 음성인식으로 나누어 진다. 동일한 화자의 발음이라 할지라도 발할 때의 감정 상태, 발음 속도 및 배경 잡음의 영향으로 차이가 발생하기 때문에 화자종속 음성 인식도 간단한 문제가 아니지만, 화자독립 음성인식의 경우 화자들 사이의 개인차가 크기 때문에 이를 효과적으로 흡수하는데 큰 어려움이 따르게 된다.

음성 인식 기술에 대한 많은 연구가 이루어진 구미 각국에서도, 상용화단계에 이른 대부분의 음성인식 시스템이 화자종속으로 격리 단어를 인식하는 기능을 가지며, 극히 일부의 제품만이 화자독립 기능 또는 연결단어 인식기능을 가지고 있는 실정이다. 따라서 연속음성 인식 기능을 갖는 제품이 나오기에는 앞으로도 상당 기간이 소요될 것으로 보인다.

3.1 격리단어 인식 기술

상용화된 음성인식 시스템을 비롯한 대부분의 격리단어 인식 기술은 패턴 매칭에 근거를 둔 인식 알고리즘을 사용하고 있으며 그 일반적인 구성도가 그림 6에 나타나 있다. 우선 디지탈

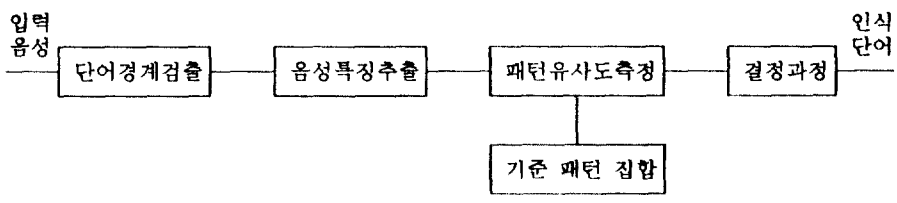


그림 6. 격리단어 인식 시스템의 구성도

신호로 변환된 입력음성이 들어오면 우선 단어경계 검출 과정을 통해 인식하고자 하는 단어의 시작과 끝을 판별한다. 이 과정의 정확도가 음성 인식 시스템의 성능에 미치는 영향이 매우 크기 때문에 높은 신뢰도가 요구된다. 이 과정은 잡음이 없는 입력 음성의 경우 단순히 구간별 에너지 및 영교차율 등의 특징을 이용하여도 비교적 정확한 결과를 얻을 수 있으나, 잡음이 섞어 있는 음성에 대해서는 보다 복잡한 방법을 사용하더라도 신뢰도 높은 음성 구간 검출에 많은 어려움이 따른다.

그 다음의 음성특징 추출과정은 일종의 데이타 감축 과정으로서 음성 신호에 포함된 음소정보를 잘 묘사해 주는 특징 계수들을 추출한다. 음성인식을 위한 특징계수로는 대역 필터군 에너지와 선형예측계수(LPC) 등이 주로 사용된다. 대역 필터군 에너지 추출방법에서는 각 대역필터의 중심 주파수 및 대역폭은 인간의 청각기관 모델에 근거하여 결정한 10~30개의 대역 필터군에 통과시킨 후 그 에너지들을 양자화하여 특징계수로 삼는 이 방법은 병렬 구현이 용이하며 잡음 환경에 대한 대처가 가능하여 대부분의 상용화 시스템에 사용되고 있다. 또한 선형 예측계수는 원래 저전송 속도의 음성부호와 방식에 사용되던 것인데 음성 발생기관중 성도의 특징을 잘 나타내 주기 때문에 음성 인식에도 널리 사용되고 있다.

인식하고자 하는 입력 단어의 특징계수들이 추출되면 그 다음 과정에서 어휘별로 이미 저장되어 있는 기준 패턴과의 유사도를 측정한다. 이 과정에서의 중요한 문제는 사람이 단어를 발음하는 속도에 따른 변화요인을 어떻게 보상하는가 하는 점이다. 발음속도 차이를 보상하기 위한 효과적인 방법으로 동적 프로그래밍 기법의 일종인 dynamic time warping(DTW) 알고리즘이 널리 사용되고 있다. 이 방법에 의하면 시간축상에서 입력 패턴과 기준 패턴의 비선형 신축을 허용하면서 최적 유사도를 계산할 수 있다. DTW 알고리즘은 일부 변형에 의해 연결 단어 인식에도 사용할 수 있으며, 인식 대상 어휘들이 음성학적으로 매우 유사하지만 않다면 높은 인식율을 얻을 수 있다. 최근에 이 방법에서 계산량이 많다는 문제를 하드웨어적으로 해결하기 위해 전용 VLSI를 개발하려는 시도들도 많이 이루어졌다.

이와 별도로 최근 각 단어의 발음상에 존재하는 각종 변화 요인들을 Markov 모델에 근거를 둔 통계적 기법으로 묘사하여 패턴 유사도를 측정하는 방법이 도입되어 각광을 받기 시작하고 있다. 이 방법은 Hidden Markov Modeling(HMM) 방법이라 불리우며, training 과정에서 Markov process에서의 상태 천이확률 및 출력 Symbol 관찰 확률을 추정한 다음, 인식 과정에서는 이들 확률로부터 Viterbi decoding에 의해 인식 단어를 결정하게 된다. HMM 방법은 DTW 방법에 비해 인식 소요 시간이 짧고, 음소 등 적은 수효의 음성 인식 단어들로부터 어휘들을 쉽게 모델링할 수 있다는 장점이 있어서 앞으로 널리 사용될 전망이다.

인식 과정의 마지막 단계는 패턴 유사도 측정의 결과를 이용하여 입력 음성이 어휘내의 어떤 단어인지를 결정하는 일이다. 어휘에 속한 각 단어마다 하나씩의 기준 패턴을 둘 경우에는 입력 패턴과 가장 유사한 기준 패턴에 해당하는 단어가 인식된 단어로 결정된다. 화자독립 음성인식의 경우에는 각 단어마다 복수의 기준 패턴을 두는 것이 상례이며, 이때는 좀더 복잡한 결정 방법이 사용된다. 또한 인식어휘 결정이 모호한 경우나 발음한 단어가 미리 정해진 어휘 집합에 포함되어 있지 않은 경우에 대한 대처도 이 과정에서 이루어 진다.

이상에서 언급한 격리단어 인식 방법은 패턴 매칭 기법에 근간을 두고 있다. 실제로 이 방법은 비교적 적은 수효의 어휘(수백 단어이내)를 대상으로 한 화자종속 음성인식의 경우 98-99 % 이상의 매우 높은 인식률을 나타낸다. 또한 이 방법은 알고리즘으로 기술하기가 용이하여 시스템 구현에도 큰 어려움이 없는 장점이 있다. 그러나 화자독립 및 연속음성 인식, 그리고 수천-수만 단어의 어휘를 대상으로 하는 대용량

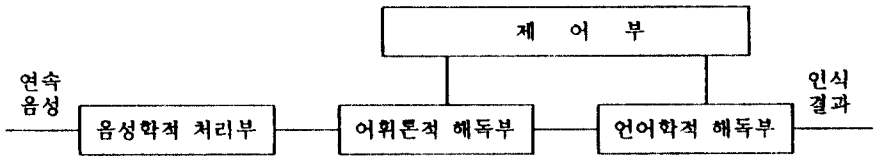


그림 7. 연속음성 인식 시스템의 기본적인 구성도

단어 인식을 목표로 할 때, 이상의 방법을 그대로 사용할 수 없음은 자명한 사실이다. 실제로 수십단어 이하의 단어 인식이라 할 지라도 이들 단어들이 음성학적으로 비슷하여 혼동되기 쉬울 경우에는 인식률이 급격하게 떨어지는 경우가 종종 있다. 이는 상기한 격리단어 인식 방법이 음성학 및 언어학적 지식을 거의 사용하지 못하고 있는데 기인한다.

3.2 연속음성 인식 기술

앞에서도 언급한 바와 같이 연속음성 인식이란 자연스럽게 발음한 회화체의 문장을 인식하는 것으로서, 그 문장에 속한 각 단어의 인식보다 발화하고자 하는 의미 파악에 촛점이 맞추어지기 때문에 음성이해(speech understanding)라고도 불리운다. 연속음성 인식을 위해서는 음성학 및 언어학적 지식이 총동원되어야 하며 그 기본적인 구성도가 그림 7에 나타나 있다.

연속음성 인식에서 하위계층에 해당하는 음성학적 처리부에서는 입력된 음성으로부터 추출된 특성들을 이용하여 미리 정해진 인식 단위들의 sequence로 바꾸는 과정을 수행한다. 여기서 인식 단위로는 음소, 이중음(diphone), 음절 등 단어보다 작은 단위들이 사용되는데, 이는 단어를 기본 단위로 할 경우 단어의 경계를 찾기가 어려울 뿐만아니라 단어와 단어 사이의 음운현상을 고려하여 방대한 단어 사전을 구성한다는 것이 현실적으로 불가능하기 때문이다. 현재의 음성학적 지식으로는 음성학적 처리부에서 오류가 발생하지 않기를 기대하기는 곤란한 형편이며 또한 이 경우 상위 계층에서 오류를 복구하는데에 어려움이 뒤따르기 때문에 일반적으로 음성학적 처리부는 하나의 해답보다는 복수의 후보 해답들을 제공하여 상위 계층에서 선택할 수 있도록 한다.

상위 계층은 어휘론적 해독부(lexical decoder)와 언어학적 해독부(linguistic decoder)로 나누어 진다. 어휘론적 해독부는 음성학적 분석기의 출력인 인식단위들의 sequence를 미리 저장된 인식 단어 사전(lexicon)과 비교하여 가장 합당한 단어를 추정한 다음 음운론적 규칙을 이용하여 이를 보완하여 문장 형태로 구성한다. 이 결과는 언어학적 해독부에 입력되어 인식하고자 하는 특정 언어의 문법에 맞는 지를 확인하는 구문론적 분석과 의미가 통하는 문장인지 검토하는 의미론적 분석, 그리고 그 의미가 문맥상 어울리는지를 검증하는 실용론적 분석 등을 거치게 된다. 또한 문장에서의 운율 및 강세에 대한 정보도 이 과정에서 활용될 수 있다.

연속음성 인식의 각 부분들은 또한 여러 개의 module들로 구성되며 각 module은 각자의 지식을 가지고 주어진 입력에서 필요한 정보들을 추출해 낸다. 이 때,(특히 상위계층에서의) 각 module사이의 상호 정보교환 작업이 중요한 역할을 담당하며 이를 제어하는 제어부가 필요하게 된다. 미국 Carnegie-Mellon 대학에서 개발한 Hearsay-II 시스템에서 사용된 blackboard 모델이 이러한 제어방식의 대표적인 예이며, 이 방식은 인공지능의 다른 응용 분야에서도 널리 사용되고 있다.

3.3 음성인식 기술 개발 동향

미국, 일본 및 유럽 등 선진 각국에서는 1970년대 이전부터 음성인식에 대한 연구를 계속

진행시켜 왔으며, 1970년대 이후 컴퓨터, 반도체 및 디지탈 신호처리 기술의 급속한 발전으로 인하여 이 분야에 대한 연구가 한층 가속화 되어 상당 수의 상용화 제품이 나오기에 이르렀다. 여기에는 이 분야의 중용성을 인식한데 따른 대규모의 국가 주도 프로젝트의 역할도 크게 작용했다.

미국의 경우 1971년부터 국방성의 주도 아래 음성이해 시스템의 개발을 위한 5개년 프로젝트가 당시 1500만불의 예산을 투입하여 수행되었으며 Carnegie-Mellon 대학의 Hearsay-II 및 HARPY 시스템, 그리고 BBN사의 WHIM 시스템들이 이때 개발되었다. 또한 1985년부터 국방성의 Strategic Computing 프로젝트의 일환으로 제 2차 음성 이해 시스템 개발 연구가 진행되고 있다. 이 프로젝트의 장기적인 목표는 10,000 단어의 어휘를 대상으로 한 연속음성 인식이며, 단기적으로는 1000단어의 어휘를 갖는 연속음성을 인식하되 95% 이상의 단어 인식률을 얻는 것을 목표로 하고 있다. 여기에는 Carnegie-Mellon 대학, MIT, BBN, SRI 및 TI 등 대학, 연구소, 그리고 기업체가 참여하여 분야별로 팀을 구성하여 연구를 수행하고 있다. 최근에 carnegie-Mellon 대학에서 개발된 SPHINX 시스템은 연구소 개발 수준이기는 하나 1000 단어 어휘의 화자독립 연속음성 자식율을 95.8%까지 높이는데 성공하여 지금까지 개발된 음성인식 시스템중 가장 우수한 성능을 기록하였다. 이 시스템은 조음현상(coarticulation effect)를 고려한 1000개의 유사 음소에 의한 HMM 알고리즘을 사용하였는데, 문법의 경우에는 아직 제한적 문법을 사용하고 있어서 음성 인식을 위한 task domain을 미리 설정할 필요가 있다.

격리단어 인식 분야에서는 AT&T Bell 연구소 및 IBM 연구소 등을 중심으로 활발한 연구가 이루어져 왔으며, IBM의 경우 2만 단어의 어휘를 인식하는 시스템의 개발에 성공하였다. 또한 Dragon System, Kurzweil AI 등 여러 회사에서 상용화된 음성인식 시스템들을 내놓고 있다.

일본에서도 일찍부터 음성인식에 대한 연구가 진행되어 왔으며 상당수의 음성인식 시스템 및 음성인식용 IC들을 개발하였다. 1982년부터 시작된 제 5세대 컴퓨터 프로젝트에서도 지능을 가진 interface module 개발의 일환으로 음성인식 연구가 검토되었으며, 1987년부터 우정성이 주관하는 자동통역 전화시스템 개발 프로젝트에서 음성인식 기술이 핵심 과제로 연고되고 있으며 이 프로젝트를 위해 관민합동으로 ATR(자동통역전화) 연구소가 설립되었다. 이 프로젝트의 최종 목표는 간단한 일상회화나 국제무역을 위한 교섭에 이용 가능한 자동 번역 전화 시스템의 구현으로서 개발 기간 15년에 900억원 규모의 예산이 투입되고 있다.

유럽에서는 유럽 공동체내의 정보통신 분야 연구 개발의 공동추진을 목표로 하여 1984년에 시작된 ESPRIT 프로젝트의 일환으로 음성인식에 관한 공동 연구가 추진되고 있으며, 영국의 Alvey 프로젝트 및 프랑스의 GRECO 프로젝트 등 각 나라별로도 국가 주도 과제로 음성인식 연구가 진행되어 왔다.

음성인식 분야에 대한 국내의 연구는 1980년대에 이르러서야 비로서 기지개를 켜기 시작했다. 현재 한국과학기술원과 한국전자통신연구소, 그리고 여러 대학의 연구소 및 일부 기업부설 연구소에서 이 분야에 대한 연구가 진행되고 있다. 아직까지는 대부분의 연구가 화자종속 격리단어 인식 단계에 머무르고 있는 형편이나, 한국과학기술원과 한국전자통신연구소 등에는 대용량 어휘 인식을 위한 기초 연구들이 수행되고 있다. 기업부설 연구소인 디지콤 정보통신연구소에서는 복수화자 인식 기능을 갖는 200단어 용량의 실시간 음성인식 시스템을 개발하였는데, 이 시스템은 host 컴퓨터와 일반 단말기 사이에 설치되어 keyboard 대신 음성으로 문자를 입력할 수 있는 음성 단말기(voice terminal)의 기능을 가는 제품이다.

3.4 앞으로의 전망

선진각국에서 음성인식 분야에 관한 연구가 수십년간에 걸쳐 진행되어 왔음에도 불구하고

임의의 화자가 발음한 연속음성의 인식이라는 최종목표의 달성은 아직 요원한 실정이다. 물론 패턴 매칭 방법에 기반을 둔 격리단어 인식 시스템의 경우 상당수의 상용화 제품이 나오기에 이르렀지만, 이 방법은 음성 및 언어에 관한 지식들을 적용하기 곤란하기 때문에 보다 넓은 응용 분야에의 확장이 곤란하다. 따라서, 인공지능 기법을 이용하여 음성학 및 언어학적 지식에 기반을 둔 음성인식 방법의 연구가 불가피하다. 그러나 불행하게도 현재로서는 이들 지식의 부족과 효과적인 지식 표현 방법의 부재로 인하여 지식에 기반을 둔 인식 방법이 제한된 응용분야에서 조차 패턴 매칭 방법보다 성능면에서 뒤떨어진다. 앞으로의 연구는 이들 지식의 보완 및 효과적인 표현 방법에 초점이 맞추어질 것이다.

음성학적 지식연구에서는 적절한 음성 기본단위 선정과 이의 신뢰도 높은 추출 방법, 화자의 개인차이나 감정상태 변화 등을 정규화 시키는 방법, 그리고 잡음 환경에 대처하는 방안 등이 연구되어야 한다. 또한 언어학적 지식 연구에서는 음운 변동 규칙, 구문론 및 의미론 등의 활용 방안, 그리고 운율 정보의 이용 방법 등이 심도깊게 연구될 필요가 있다. 특히 한국어의 경우 이들 분야에 대한 연구가 상당히 낙후되어 있으며, 이 분야에 관한 한 외국의 연구 결과를 도입하여 사용할 수 없기 때문에 국내에서의 독자적인 연구가 시급한 실정이다.

앞으로 일반 회화체의 음성을 인식하기 위해서는 자연언어처리(natural language processing) 분야와의 병행 연구가 요청된다. 음성인식 및 합성연구가 자연언어처리 연구와 결합될 때, 자동통역 시스템과 같은 제품도 탄생할 수 있다. 기호에 의한 표현과 구술적인 지식에 기반을 둔 현재의 인공지능 기법의 한계를 극복하기 위한 한 방법으로서, 인간의 뇌에서 수행되는 정보처리 과정을 공학적으로 모델링한 신경망(neural network) 기법에 대한 연구도 기대를 모으고 있다.

음성인식 시스템의 개발은 방대한 과제로서 단 기간에 한 두 그룹의 연구팀에 의해 이루어질 수 있는 성질의 것이 아니다. 음성에 의한 manmachine interface의 시대를 열기 위해서는 이 분야 연구 인력의 노력과 더불어 보다 과감한 정책 투자와 연구 자원의 효율적인 뒷받침 되어야 할 것이다.

4. 음성 합성 기술

4.1 개요

음성합성이란 입력 문장을 사람이 청취할 수 있는 음성 신호로 변환시키는 과정을 말한다. 이 경우 입력 문장은 keyboard 입력, 광학문자인식(optical character recognition) 결과, 또는 컴퓨터에 수록된 데이타베이스 등 여러 가지 형태가 있을 수 있다. 이렇게 입력된 문장을 음성으로 변환시키는 가장 손쉬운 방법은 합성하고자 하는 문장에 음성파형을 컴퓨터에 디지탈화하여 미리 저장시켜 놓았다가 이들 저장된 데이타를 토대로 간단한 단어 조합 등을 이용하여 음성 파형을 재생시키는 방법이다. 이 방식은 조합할 음성 기본단위(단어, 구 또는 문장) 및 디지탈화 시킨 음성의 억양 등에 의해 제약을 받을 뿐만 아니라, 사용 어휘수 및 문장 형태에 제한이 가해지게 되므로 제한적 음성합성 방식이라 불리운다. 이러한 제한적 음성합성 방식은 말하는 장난감이나 기계장치 등의 경고(warning) 시스템, 그리고 은행잔고나 증권시세 등을 숫자 조합으로 제공해 주는 음성 응답 시스템 등 단순한 음성합성 기능이 요구되는 응용분야에 널리 사용되고 있다.

이에 반하여 어휘나 문자 형태에 아무런 제약없이 임의의 문장을 음성으로 합성해 내는 방식을 무제한 음성합성 또는 text-to-speech 합성방식이라 불리운다. 넓은 의미에서의 음성합성은 제한적 음성합성과 무제한 음성합성의 두가지를 다 포함하지만 실제로 제한적 음성 합성은 단순히 음성 부호화 기술의 연장이라고 볼 수 있으며, 엄밀히 말해 무제한 음성합성 만이 진정한 의미에서의 음성 합성이므로 본 고에서는 무제한

표 2. 제한적 음성합성 방식과 무제한 음성합성 방식의 비교

제한적 음성합성 방식(Digitized Speech)	무제한 음성 합성 방식(Text-to-speech)
1. 문장전체, 또는 단어 및 구절 등을 디지탈 방식으로 녹음하여 컴퓨터 기억장치에 저장시킨 뒤 필요에 따라 이들을 조합하여 재생시키는 방식.	1. 음소 등 말소리의 기본단위에 해당하는 정보들을 저장시킨 다음 합성하고자 하는 문장을 분석하여 말소리의 기본 단위로부터 음성을 합성해 내는 text-to-speech 합성방식.
2. 사람이 발음한 문장을 토대로 하므로 녹음기에 문장을 녹음시킨 것과 같이 자연스러운 음질을 가짐.	2. 합성음이므로 음성의 자연스러운 면에 있어서 사람이 발음한 문장에 비해 뒤떨어짐.
3. 미리 정해진 형태의 문장만을 출력시킬 수 있음.	3. 어휘와 문장에 제한없이 어떠한 형태의 문장도 출력시킬 수 있음.
4. 출력시키고자 하는 문장은 모두 미리 컴퓨터 기억장치에 음성으로 저장시켜야 하므로 기억용량의 제한에 따른 한계점을 지님.	4. 음성의 기본단위만 컴퓨터 기억장치에 저장시키면 되므로 기억 용량에 제한없이 무제한 음성합성이 가능함.
결 론:	
1. 단어나 문장형태가 매우 제한적인 응용분야에서는 digitized speech에 의한 기존의 합성방법으로 사용 가능함.	
2. 시시각각으로 변하는 정보를 합성하거나 합성해야 할 문장용량이 클 경우 text-to-speech가 필수적으로 요구됨.	
3. Text-to-speech의 음질이 향상됨에 따라 모든 음성합성 응용분야에 digitized speech를 대체하여 사용될 것임.	

음성합성 기술만을 다루고자한다. 참고적으로 표 2에 제한적 음성합성 방식과 무제한 음성합성 방식의 비교가 나타나 있다.

무제한 음성합성 방식과 제한적 음성합성 방식의 구성상의 가장 현저한 차이는 언어학적 처리부의 유무에 있다. 무제한 음성합성 방식의 언어학적 처리부는 입력 문장을 분석하여 일련의 세부적인 음성 기본 단위(음소, diphone, 음절 등)들로 바꾸어 주며, 음성의 고저나 장단 등 운율에 관한 정보들을 결정하는 기능을 가진다.

4.2 무제한 음성합성의 원리

Text-to-speech 합성 시스템의 기본 구성도는 그림 8과 같다. 그림에서 보는 바와 같이 합성 시스템은 크게 언어학적 처리부와 음성신호처리부의 두 부분으로 나눌 수 있다. 이들 중 언어학적 처리부는 다시 text 전처리, 문장분석 및 발음표기 변환의 세 과정으로 나누어진다. 입력문장이 언어학적 처리부로 들어오면, 먼저 전처리 과정을 통해 약어나 특수기호 등의 표기를 구술적인 표현으로 대체시킨 다음, 문장분석 과정에서 문장의 구조를 분석하게 된다. 이 과정에서 얻어지는 정보는 음성신호 처리부로 넘어가고,

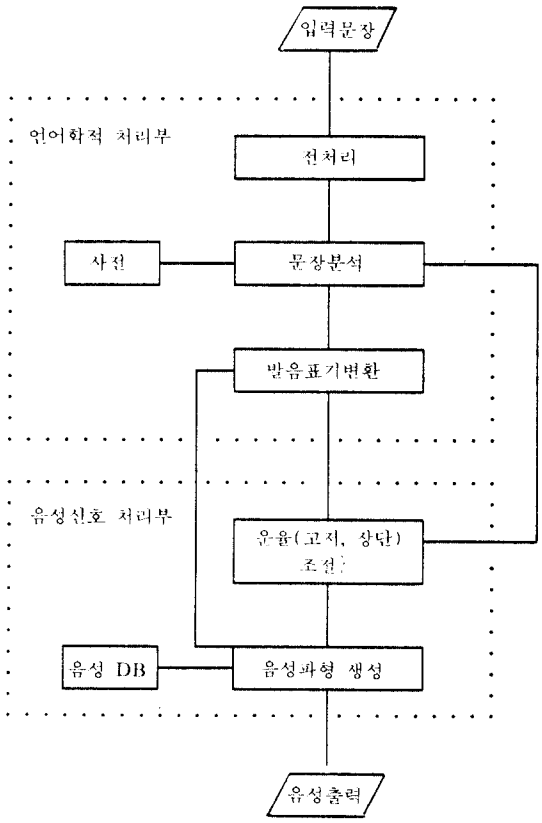


그림 8. Text-to-speech 합성 시스템의 구성도

고저, 장단 및 휴지기간 등 운율 조절에 사용된다.

언어학적 처리부의 마지막 단계인 발음표기 변환 과정에서는 발음법칙에 의해 입력 문장을 사람이 발음하는 형태의 표기로 변환시키게 된다. 이 언어학적 처리부의 구현에는 규칙으로 표현하기 어려운 많은 예외 조항이 따르므로 사전을 이용하여 이러한 문제들을 처리하도록 한다.

음성 신호 처리부는 운율조절 과정과 음성 파형 생성과정으로 나누어진다. 운율 조절 과정에서는 언어학적 처리부의 결과를 토대로 하여 음성의 고저, 장단, 강세 등 운율을 사람이 발음하는 것과 같은 자연스러움에 접근하도록 조절한다. 끝으로 음성파형 생성 과정에서는 미리 구축된 음성 데이타베이스로부터 음성 기본 단위들에 대한 정보를 제공받아 운율 정보를 고려한 합성 파형을 생성시킨다.

4.3 음성합성 단위

무제한 음성합성을 위한 가장 보편적인 접근 방법은 먼저 음절, 반음절, diphone, 음소 등 음성의 기본 단위를 컴퓨터에 데이타베이스로 저장시킨 후, 문장을 합성할 때 필요한 데이타베이스를 꺼내어 이들을 알맞은 규칙에 따라 연결시킴으로써 합성음을 만들어 내는 것으로 이러한 방식은 concatenative synthesis라 부른다. 이때 데이타베이스를 어떤 음성의 기본 단위로 구축하는가에 따라서 합성 알고리즘의 복잡성, 데이타베이스의 갯수 및 데이타베이스가 차지하는 기억 용량 사이에 trade-off가 있을 뿐만아니라 합성음의 음질에도 큰 영향을 준다. 특히 이러한 음성의 기본 단위 선정은 음성합성을 하고자 하는 언어의 음운체계 등과 밀접한 연관을 맺고 있으므로 언어에 따라 적절한 음성 기본 단위의 선정이 이루어질 필요가 있다.

물론 음성의 기본 단위로 문장, 구, 단어 등을 이용하는 것도 한 방법이며 이들은 그 내부에 음성이 갖는 거의 모든 정보를 포함하고 있으므로 좋은 음질의 합성음을 기대할 수 있다. 그러나 이들을 이용하여 임의의 문장을 합성하기 위해서는, 데이타베이스를 구축하는데 기억 용량이 엄청나게 많이 필요하게 된다. 따라서 문장, 구, 단어를 데이타베이스의 기본 단위로 삼아 음성을 합성하는 방법은 일기예보, 증권안내 등 특성 용도에 국한된 제한된 어휘의 음성합성 시스템에 적용할 수 있을 뿐이며 임의의 문장을 합성하는 무제한 음성합성 시스템에 이들을 이용하기는 적절하지 못하다.

따라서 본 절에서는 음절, 반음절, diphone, 음소 등의 음성 기본 단위들이 한국어 음성합성 시스템에 사용되었을 때의 장·단점을 간략히 검토해 본다.

4.3.1 음소

음소의 정의에는 여러가지 학설이 있으나 "둘 이상의 음성이 말의 뜻을 표현하는 과정에서 하나의 소리로 뭉쳐질 때, 그 뭉쳐진 하나의 소리"로 정의될 수 있다. 이에 근거하여 한국어의 음소를 특성에 따라 분류한 것을 표 3과 표 4에 도시했는데, 한국어의 경우 19개의 자음과 21개의 모음을 합하여 40개의 데이타베이스만 있으면 무제한 음성합성을 할 수 있다. 또한 변이음(allophone)을 고려해도 데이타베이스의 수가 많지 않으므로, 기억 용량을 적게 필요로 한다는 장점이 있다.

그러나 데이타베이스의 기본 단위로 음소를 이용하여 문장을 합성하는 경우에는 음소와 음소를 연결시켜야 하는데 이 과정에서 연결 부분에 불연속이 생겨 합성음의 명화도가 떨어진다. 또한 음성이 갖고 있는 정보중에서 understanding 에 관련된 많은 정보가 음소와 음소 사이의

표 3. 한국어에서 사용되는 모음(vowel) 분류표

모음의 구성		모음의 종류
단모음	단자 단모음	ㅏ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ
	복자 단모음	ㅐ, ㅔ, ㅚ, ㅟ
복모음	단자 복모음	ㅑ, ㅕ, ㅛ, ㅠ
	이중자복모음	ㅒ, ㅖ, ㅘ, ㅝ, ㅢ
	삼중자복모음	ㅙ, ㅞ

표 4. 한국어에서 사용되는 자음(consonant) 분류표

소리내는 방법	소리내는 위치	입술소리	윗잇몸소리	센입천장 소리	여린입천장 소리	목청소리
터짐소리	예사소리	ㅂ	ㄷ		ㄱ	
	된 소 리	ㅃ	ㄸ		ㄲ	
	거센소리	ㅍ	ㅌ		ㅋ	
터짐같이 소리	예사소리			ㅈ		
	된 소 리			ㅉ		
	거센소리			ㅊ		
갈이소리	예사소리		ㅅ			ㅎ
	된 소 리		ㅆ			
콧 소 리		ㅁ	ㄴ		ㅇ	
흐름 소리			ㄹ			

천이 부분에 있으며, 한 음소가 인접한 음소에 영향을 미쳐 음소의 음향학적 특성을 변형시키는 상호 조음 현상도 고려하여 음소를 연결해야 하므로, 음소와 음소를 연결시키는데 필요한 규칙을 찾기 위해서는 한국어에 대한 광범위한 지식과 과학적 연구가 선행되어야 한다.

4.3.2 Diphone

Diphone의 정의는 "한 단음(phone)의 정적인 중심부분에서 다음 단음의 정적인 중심 부분까지의 음성 단편"이라고 표현할 수 있다. 예를 들어 '가나'란 단어에서 하나의 diphone은 'ㅏ'의 정적인 중심 부분에서 'ㄴ'의 정적인 중심 부분까지라고 말할 수 있다. 이와 같은 diphone의 정의에 따르면 하나의 diphone은 diphone을 구성하고 있는 단음과 단음 차이의 천이 부분을 포함하게 되고, 음소인 경우에는 그들을 연결할 때 천이부분을 규칙에 의해 연결해야 하는 것과는 달리, diphone을 연결하여 한 문장을 구성할 때는 diphone의 경계가 정적인 중심 부분이므로 diphone 끼리의 연결이 간단하다는 장점을 갖게 된다. 따라서 diphone을 연결하여 단어나 문장을 구성할 때 연결 부분에서의 불연속성이 적어 음질이 손상되지 않는다.

그러나 diphone으로 한국어 데이타베이스를 구축할 경우에는 음운학적으로 사용되지 않는 것을 제외하더라도 diphone의 수가 약 1,400개로 음소의 수에 비해 상당히 많아지고, 따라서 기억 용량이 많이 필요하며 데이타베이스를 구축하는 데에도 많은 시간이 소요된다.

4.3.3 음절

음절은 음소의 결합으로 구성되어 있는데, 한국어의 경우 초성, 중성, 종성이 어울려 한 음절을 이루는 독특한 특징이 있다. 음절의 구성은 C^iVC^f(C^i=초성자음, V=중성모음, C^f=종성자음)으로 나타낼 수 있다. 여기서 C^i의 위치에는 'ㅇ'을 제외한 18개의 자음과 C^i가 없는 경우를 포함하여 모두 19개의 자음이 올 수 있고, V의 위치에는 10개의 단모음과 11개의 복모음 등 모두 21개의 모음이 오며, C^f의 위치에는 7개의 대표 자음과 종성이 없는 경우 등 모두 8개의 자음이 올 수 있다. 따라서 한 국어의 음절 수는 산술적으로 약 3,200개 정도가되나, 실제 사용되는 음절의 수는 음소와 음소가 연결될 때의 제약성과, 현대 국어에 사용되지 않는 음절을 고려하여 1,096개의 음절이 현재 사용된다고 알려져 있다. 한국어의 음절의 구성은 다음 4가지로 분류할 수 있다.

C^i, C^f가 모두 있는 경우 : CVC형
C^i는 있고 C^f는 없는 경우 : CV형
C^i는 없고 C^f는 있는 경우 : VC형
C^i, C^f가 모두 없는 경우 : V형

음절은 음절을 구성하고 있는 음소들 사이에

존재하는 천이 부분의 정보를 포함하고 있으므로 단음절의 경우에는 좋은 음질의 합성음을 기대할 수 있다. 그러나 음절이 이울려 단어를 구성할 때 음절과 음절사이에도 상호 조음현상이 존재하고, 이 천이 부분을 규칙에 의해 연결해야 하는데 이러한 규칙 개발에 많은 어려움이 따르게 된다.

4.3.4 반음절(semisyllable)

반음절은 음절의 steady한 중앙 부분을 중심으로 음절을 둘로 나눈 단위를 말하며, 음절중 CV형, V형, VC형의 데이타베이스만 가지는 셈이다. 따라서 CVC형의 음절은, CV형과 VC형의 음절을 모음 부분에서 연결하여 만든다. 반음절을 음성 기본 단위로 이용할 경우 데이타베이스의 수가 산술적으로 계산할 때 570개 정도로 음절 데이타베이스보다 적은 기억 용량이 필요하고, 음절에 비하여 데이타베이스를 제어하기 쉬운 장점이 있다. 다만, 음절 경계에서의 상호조음 현상의 처리는 음절의 경우와 마찬가지로 적절한 규칙에 의해 처리될 필요가 있다.

4.4 음성합성 방식

일반적으로 text-to-speech 합성 시스템에 사용되는 음성합성 방식으로는 formant 합성 방식과 LPC(Linear Predictive Coding : 선형예측 부호화) 합성 방식이 주로 사용된다.

4.4.1 Formant 합성방식

Formant란 사람의 음성 발생 기관중 음성 스펙트럼 형성에 가장 큰 역할을 담당하는 성도(vocal tract)의 공명 주파수를 의미하며, 음성인지(speech perception) 연구 결과에 의하면 말소리(특히 모음 등 유성음)의 변별에 있어서 가장 중요한 특징이라고 알려져 있다. Formant 합성 방식은 이러한 formant의 특징을 이용하여 음성 발생 기관을 formanf 주파수에 따른 일련의 공명기(resonator)들로 모델링한 것이다. 공명기들의 구성방법에 따라 여러가지 형태의 formant 합성기가 구현될 수 있으며, 그림 9는 이러한

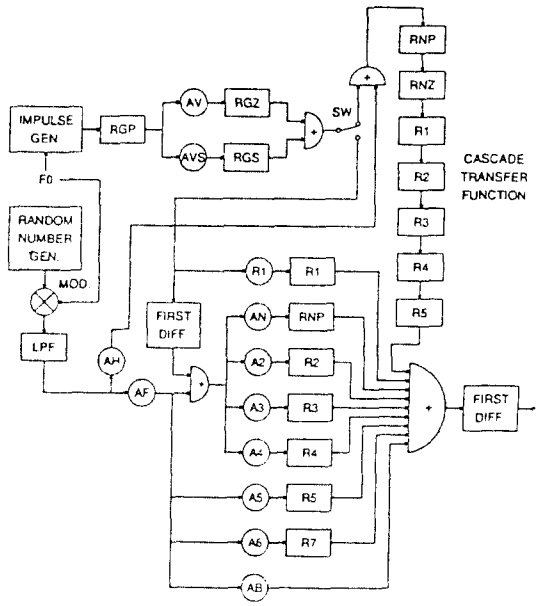


그림 9. Cascade / parallel formant 합성기의 구성도

formant 합성기의 대표적인 형태인 Klatt의 cascade-parallel formant 합성기의 구성도를 나타낸 것이다.

Formant 합성기의 여기 신호(excitation signal)로는, 유성음의 경우 주기적인 pulse들이, 무성음의 경우 pseudo-random noise가, 그리고 유성 마찰음의 경우에는 주기적인 형태를 띤 noise가 사용된다. 성도는 일반적으로 2차 디지탈 공명기의 cascade 또는 parallel 구조로 모델링하게 되며, 이들 각각의 formant 합성기는 모음의 음성 스펙트럼을 효과적으로 표현할 수 있으며, 각각의 formant amplitude를 별도로 조절 가능하다는 점이 장점이다. 우수한 음질을 얻기 위해서는 주파수가 낮은 쪽으로부터 처음 4개의 formant 주파수 및 처음 3개의 formant 대역폭을 시간에 따라 가변시켜 줄 필요가 있는데, 이는 고차 formant로 옮아갈수록 음성인지에 미치는 영향이 작아지기 때문이다. 보다 단순한 시스템의 경우에는 처음 3개의 formant 주파수만 가변시키고 대역폭은 일정한 값으로 유지시키기도 하는데, 대역폭을 일정하게 했을 때 가장

음질 저하가 현저한 비음의 경우에는 제 1 formant 의 대역폭을 증가시켜 줌으로서 그 영향을 최소화 시킬 수 있다.

Parallel 구조의 formant 합성기는 마찰음 등의 합성에 효과적인 것으로 알려져 있다. 유성 마찰음을 합성하기 위해서는 parallel 구조의 formant 합성기의 입력으로 주기적인 파형에 의해 변조된 noise를 사용한다.Parallel 구조를 이용하여 모음이나 유음 등을 생성해 낼수도 있지만, 이 경우 각 공명기들 사이의 multiplication 효과로 각각의 formant amplitude를 구하는데 많은 연산이 소요된다. parallel 및 cascade 구조를 함께 갖는 formant 합성기에서는 여기 신호가 voicing source인가 noise source인가에 따라 각각 cascade 구조와 parallel 구조로 입력시킴으로써 보다 자연스러운 음의 생성을 도모한다.

비음은 구강과 비강을 모두 사용하는 소리로서 비강을 거치는 경로가 일반적으로 모음을 발음할 때 성도를 거치는 경로에 비해 길다. 이 때문에 비음 생성에는 모음의 경우보다 공명기를 하나 더 사용하게 된다. 또한 비음의 스펙트럼에 존재하는 영점(zero)들을 모델링하기 위해, 가장 낮은 주파수에 존재하는 영점을 2차 anti-resonator 에 의해 구현한다. Cascade 구조의 formant 합성기에서 비음을 구현하는 방법으로는 한 쌍의 추가적인 공명기와 anti-resonator 를 두고, 비음이 아닐 경우 이들을 서로 상쇄시키는 방안이 주로 사용된다.

4.4.2 LPC 합성방식

LPC(Linear Predictive Codint)는 저전송속도의 음성부호화에 널리 사용되는 방식으로 음성 발생 기관을 all-pole 디지탈 필터로 모델링한 것이다. LPC 합성 방식은 formant 합성 방식에 비해 단순한 구조를 갖는 장점이 있으며, 이는 음성신호의 스펙트럼 정보가 단지 몇개의 LPC 계수안에 모두 포함될 뿐 아니라 이들 계수들이 입력 음성의 분석과정에서 자동적으로 계산될 수 있다는 데에 기인한다. LPC 합성기는 보통 lattice 필터의 구조로 구현되는데, lattice 필터는 필터의 안정성을 보장하면서 계수를 linear interpolation에 의해 조정할 수 있다는 장점을 갖는다.

LPC 합성 방식의 단순한 구조는 그 자체가 구현상의 장점으로 부각되면서도 이로 인한 음질 저하가 문제점으로 지적되어 왔으며, 이 문제점들을 해결하기 위한 여러가지 시도들이 이루어졌다. LPC 합성 방식에서 음질에 영향을 주는 문제점들로는 all-pole 모델 자체의 한계, LPC 분석과정에서의 부정확성, Pitch 결정 및 유성음/무성음 판별 과정에서의 오류, 여기신호의 선정, 그리고 계수들을 양자화하는데서 비롯되는 오차 등을 들 수 있다. 이중에서 Pitch 결정 및 유성음/무성음 판별 과정에서의 오류는 음성합성의 경우 규칙에 의해 제공되므로 음성부호화의 경우와는 다른 양상을 띈다.

LPC 합성 방식에서 유성음일 때의 주기적인 여기 신호 패턴의 선택은 합성음의 음질과 자연스러움을 결정하는 중요한 요소이다. 가장 간단한 형태는 단일 펄스에 의한 여기 신호 패턴으로서 이 패턴을 사용할 경우 합성음에 울리는(buzzy) 소리가 들리게 된다. 이는 여기 신호패턴이 존재하는 기간이 단지 한 sample로 매우 짧기 때문에 합성음의 amplitude의 감쇄가 매우 빨라지게 되며, 따라서 합성음의 파형 형태가 여기 신호 펄스의 위치에서만 날카로운 peak를 갖고 그 다음 여기신호 펄스가 나타나는 한 주기 동안은 거의 zero에 가까운 값을 갖는 형태를 가지는 데에 기인한다. 반면에 고주파 성분이 상대적으로 적은 smooth 한 파형을 여기 신호 패턴으로 사용하게 되면 고음역이 줄어들어 음성 스펙트럼에 왜곡을 주게 된다. 그러므로 바람직한 여기신호 패턴은 어느 정도 폭넓은 펄스 형태를 가지면서도 평탄한 주파수 특성을 지닐 필요가 있다. 음성 합성과 관련해서 여기 신호 패턴에 미치는 또 다른 제약 요건으로는 음성의 기본 주파수(Pitch) 변화에 따른 여기신호 패턴의 truncation 문제이다. Pitch 주기가 여기신호 패턴보다 짧아질 경우 truncation에 의해 스펙트

럼의 왜곡이 생길 수 있으므로, 여기 신호 패턴의 에너지가 대부분 처음 2-3ms내에 존재하도록 하든지 또는 여기신호 패턴을 all-pass 신호의 embedded 형태로 만들어 스펙트럼에 대한 truncation 의 영향을 최소화시켜야 한다. 그러나, 실제로 일정한 여기신호 패턴을 모든 음성신호에 적용할 경우 단일 펄스에 의한 여기 신호의 경우에 비해 음질개선의 폭이 현저하지는 못한 것이 현실이다.

4.4.3 합성방식의 선정

음성합성 방식으로 formant 합성 방식을 사용할 것인가 LPC 합성방식을 사용할 것인가 하는 결정은 trade-off에 의해 좌우된다. LPC 합성방식은 formant 합성방식에 비해 단순한 구조를 가지며 분석 과정을 자동적으로 수용될 수 있는 장점을 가진다. 그러나 formant 합성 방식이 음성신호의 음향학적 경계 부분에서 음성 특징 계수들을 용이하게 조절할 수 있는데 반하여, LPC 계수들의 변화는 음성 스펙트럼 전반에 걸쳐 복잡한 영향을 미치기 때문에 LPC 계수들을 interpolation 하는데에는 한계가 있다.(LPC 계수의 변형 형태인 line spectral Pair(LSP) 계수의 경우에는 이들 계수가 formant 주파수와 밀접한 관계를 갖게 되므로 조절이 비교적 용이하다.) 더우기 formant 합성 방식은 여러 사람의 음성을 표현하기 위해 음색을 변경시킬 필요가 있을 때 이에 대처하기가 용이하다. 즉, formant 합성 방식에서 formant 주파수를 변경시키면 합성음의 음색이 쉽게 바뀌는데 반해서, LPC 합성 방식에서 이러한 작업을 하기 위해서는 LPC 계수들을 Pole 위치들로 변환시켜야 한다. 또한 LPC 합성 방식은 all-pole 모델이 갖는 특성상 음성 스펙트럼의 peak에는 민감하나 valley 묘사에는 적절하지 못하므로 formant 합성 방식에 비해 formant의 대역폭 이 부정확하게 결정되는 것이 보통이며, anti-formant를 갖는 비음의 합성에 불리하다. 결론적으로, LPC 합성 방식과 formant 합성방식 사이의 선택 기준은 간단한 구조에 계수분석이 자동적으로 이루어지는 방식을 택한 것인가, 아니면 복잡하고 특징 추출에 방대한 시간이 소요되나 보다 우수한 음질 구현의 가능성이 큰 방식을 택한 것인가에 달려있다고 할 수 있다.

4.5 음성합성 시스템의 개발동향

미국과 일본을 비롯한 선진 각국에서는 1970년대 이전부터 음성합성에 대한 연구를 진행시켜 왔으며 1980년대에 들어서 컴퓨터 및 반도체, 그리고 디지탈 신호처리 기술의 급속한 발전에 힘입어 이 분야에 대한 연구도 한층 가속화되었다. 미국의 경우 MIT 및 AT&T Bell 연구소 등을 중심으로 활발한 연구가 이루어져서 현재 상당수의 상용화 제품이 나오기에 이르렀으며, 음성합성 제품의 시장 규모가 매년 30% 이상 증가하고 있는 추세이다.

대표적인 음성합성 시스템으로는 Digital Equipment 사의 DECtalk 시스템, Speech Plus 사의 PROSE 4000 시스템, Street Electronics 사의 ECHOPC+ 등을 들 수 있다. 일본의 경우 NTT, NEC를 비롯한 여러 회사들이 음성합성 시스템을 개발 중에 있으며, 비교적 단순한 언어체계의 장점을 활용하여 모음-자음 및 자음-모음 형태의 음절을 기본 단위로 하는 합성방식이 주종을 이루고 있다. 이미 상용화된 대표적인 합성 시스템으로는 NTT사의 VCS-II 시리즈 등이 있다.

Digital Equipment사의 DECtalk 시스템은 MIT에서 연구용으로 개발한 Klattalk 시스템 및 MITalk 시스템을 상용화시킨 것이다. 이 시스템은 formant 합성 방법에 근간을 두고 있으며, 성인 남성, 성인 여성 및 어린이를 포함한 7명의 음색을 표현할 수 있으며 전화망 interface 가 내장되어 있어 음성응답 시스템에의 응용이 용이하다. 최초의 제품은 단일회선 규모의 독립형 시스템이었으나 그 이후 최대 8회선까지 확장 가능한 제품이 나왔으며, 최근에는 DECtalk 기능에 음성인식 기능까지 포함시킨 DECVoice response system을 내놓았다.

Speech Plus사의 PROSE 4000 시스템은

Telesensory Systems사의 PROSE 2000 시스템의 상위 기종으로서 역시 klattalk의 formant 합성기를 이용한 제품이다. 이 제품은 IBM PC의 plug-in card 형태로 개발된 제품중 가장 음질이 뛰어난 것으로 알려져 있으며, 3명의 남성 음성을 묘사할 수 있다.

Street Electronics사의 ECHO 시리즈는 저가형 음성합성 시스템으로서 가장 널리 보급된 제품이다. Apple과 IPM PC Plug-in card 형태의 제품들이 있으며, Naval Research Laboratory 에서 개발된 algorithm을 토대로 입력 문장을 allophone 부호로 바꾸어 준 다음 LPC 합성 방식으로 합성음을 생성시킨다. 최근에는 범용 DSP chip을 이용하여 음성부호화, 음성합성 및 음성인식을 모두 수행할 수 있는 제품을 내놓고 있다.

NTT의 VCS-II 시리즈는 NEC PC의 내장형 board로서 음성합성 방식으로는 LPC의 변형 형태인 LSP(Line Spectral Pair) 방식을 사용하고 있다. 이 제품은 남성 및 여성의 합성음을 생성시킬 수 있으며 한자/가나 변환 사전을 가지고 있다.

지금까지 언급한 제품들 이외에도 많은 음성합성 시스템들이 개발되고 있으며, 이들 제품들의 가격은 미화로 $180에서 $6,000 사이로 매우 큰 격차를 나타내고 있는데, 대체적으로 가격과 음질이 비례하는 양상을 보이고 있다.

국내에서는 1980년대 후반에 들어서서야 무제한 음성합성에 대한 연구가 진전되기 시작했으며, 서울대를 비롯한 여러 대학, 그리고 한국과학기술원 및 한국전자통신연구소에서 무제한 음성합성 기술을 연구하고 있다. 또한 삼성, 금성 그리고 디지콤 등의 기업부설 연구소에서도 이 분야의 연구가 진행되고 있어서 머지않아 상용제품이 나올 것으로 기대된다.

4.6 앞으로의 전망

음성 합성의 결과로 얻어지는 합성음의 음질을 평가하는 두가지 기준은 명료성(intelligibility)과 자연스러움(Naturalness)이다. 이미 많은 연구가 이루어진 선진 외국의 기술 수준도 합성음의 명료성 면에서는 어느 정도 기대 수준에 육박하고 있으나 자연스러움 면에서는 아직도 해결해야 할 많은 문제점을 안고 있는 실정이다. 합성음의 자연스러움을 개선시키기 위해서는 억양, 장단, 강세 등의 운율이 잘 구현되어야 하는데, 이를 위해서는 입력 문장을 분석하여 이해하는 언어학적 처리 능력의 개선이 절실히 요청된다. 최근 기계 번역 등을 위한 자연 언어 처리 연구가 많은 진전을 보이고 있으나, 섬세한 운율 묘사를 위한 정보를 제공하기 위해서는 더 많은 노력이 경주되어야 한다.

음성신호 처리 분야에서도 vocoder 수준이 아닌 보다 고음질을 가지면서도 다양한 운율 제어가 가능한 음성 합성 방식이 개발될 필요가 있다. 이와 병행하여 음색이나 감정 등을 변화시킬 수 있는 합성 방식의 개발이 요청되며, 현재 일본의 자동통역 전화 프로젝트 등에서 이들 분야의 연구가 심도있게 진행되고 있어 앞으로는 특정 화자의 목소리를 어느정도 입력시키면, 이 목소리에 의한 무제한 음성합성이 이루어지는 날도머지 않을 것으로 전망된다.

5 결 론

지금까지 디지탈 음성통신에서 사용되는 각종 부호화 기술, 음성인식 기술, 음성합성 기술에 관하여 그 기술의 주요 내용 및 현황, 장래 전망에 대하여 살펴보았다. 이러한 음성처리 기술들은 음성에 의한 입출력을 요구하는 분야가 계속 증가됨에 따라 컴퓨터 기술, 반도체 기술들과 결합되어 다양한 형태로 실용화되어 가고 있다.

다행히 음성처리 분야의 국내 기술개발도 상당히 진척되어 있어, 최근에 크게 활성화되고 있는 음성정보 서비스 시스템이 완전 국내 개발되었으며, 무제한 음성합성 기술도 곧 상용화 제품이 출현되는 단계에 있다. 한국어의 음성처리는 국내 기술진에 의해서만 해결될 수 밖에 없는 현실을 볼 때, 이러한 국내의 연구 노력은 고무

적인 것으로 평가할 수 있다. 향후 정보화 사회에서 음성 처리 기술이 주요한 역할을 담당할 것임은 재차 강조할 필요가 없는 바 이 분야에 대한 꾸준한 연구 노력을 기대해 본다.

김 형 순

저자약력

- 1960년 8월 21일생
- 1983년 2월 : 서울대학교 전자공학과 학사
- 1984년 2월 : 한국과학기술원 전기 및 전자공학과 석사과정 수료
- 1989년 2월 : 한국과학기술원 전기 및 전자공학과 박사
- 1987년 1월 ~ 현재 : 디지콤 정보통신연구소 선임연구원

임 병 근

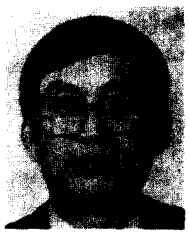
저자약력

- 1962년 10월 20일생
- 1984년 2월 : 한양대학교 전자공학과 학사
- 1986년 2월 : 한국과학기술원 전기 및 전자공학과 석사
- 1991년 8월 : 한국과학기술원 전기 및 전자공학과 박사
- 1987년 1월 ~ 현재 : 디지콤 정보통신연구소 선임연구원

김 희 동

저자약력

- 1957년 3월 생
- 1982년 2월 서울대학교 전자공학과 학사
- 1984년 2월 한국과학기술원 전기 및 전자공학과 석사
- 1987년 8월 : 한국과학기술원 전기 및 전자공학과 박사
- 1987년 1월 ~ 현재 : 디지콤 정보통신연구소 책임연구원



은 종 관

저자약력

- 1940년 8월 25일생
- 1964년 6월 : 미국 University of Delaware 전자공학과 졸업 전자공학학사 학위
- 1966년 6월 : 동대학원 졸업 전자공학석사 학위
- 1969년 6월 : 동대학원 졸업 전자공학박사 학위
- 1969년 9월 ~ 1973년 5월 : 미국 University of Maine 전자공학 조교수
- 1973년 5월 ~ 1977년 6월 : 미국 스탠포드연구소(SRI) 책임연구원
- 1977년 6월 ~ 현재 : 한국과학기술원 전기 및 전자공학과 교수. IEEE Fellow