

음성 합성 냉장고

方 鍾 建

三星電子株式會社 綜合研究所 研究員

I. 서 론

최근 컴퓨터 발전보급에 힘입어 컴퓨터 대중화 시대와 더불어 고도의 반도체 집적기술의 급속한 발전으로 인간과 기계사이의 인터페이스 수단으로써 음성을 이용한 음성 합성, 인식 기술의 연구가 기초, 응용 두 분야로 나누어져 활발히 추진되어져 왔다. 이러한 음성 합성, 인식 기술은 인간과 기계 사이의 직접적인 대화를 할 수 있는 계기를 만들어 주었다는 점에서 큰 의미를 가질 수 있다.

음성에 대한 연구는 고대 르네상스 시대부터 행하여져 기나긴 세월동안 연구가 추진되어져 왔지만 공학적인 면에서는 실용화된 음성 응답 system의 개발이 이루어 지지 않아 실제 기기의 응용에는 많은 문제점이 있었다.

그러나 1970년대에 들어서서 음성 연구가들의 끈질긴 노력과 더불어 반도체 기술의 급속한 발전으로 electronics의 소형화 및 음성 합성 chip 개발이 이루어 졌으며 1980년대부터는 정보 처리기술의 고도화, 음성분석 수법의 확립, 반도체 기술의 진보에 의한 주변 기술의 발전에 힘입어 음성 합성, 분석, 인식

시스템 개발이 이루어져 가전제품에 음성합성을 중심으로 응용 되기 시작하여 최근에는 각종 여러가지 기기에 활발히 응용 개발되는 방향으로 나아가고 있는 것이다.

본 고에서는 음성 합성 및 분석 기술의 기초원리 및 분류, 음성분석 처리기술의 간단한 설명을 II절에서 살펴보고 III절에서는 음성 합성 냉장고의 필요성 및 구조에 대해서 기술한다.

IV절에서는 결론을 곁하여 음성 합성 기술의 실용화를 위한 향후 전망에 대해 기술하겠다.

II. 본 론

1. 음성 합성의 기초

최근 컴퓨터의 급속한 발전으로 버턴 전화기를 사용하여, 각종 정보자료에 관한 문의를 computer에 입력시키며 전화기를 통하여 음성으로써 정보자료 문의에 대한 결과를 알려주는 음성 응답, 출력 장치는 computer 출력을 음성으로 변환시켜 인간에게 전달해 주는 편리한 장치이지만 대형 computer와 주변 기기로 이루는 대형 system이었다.

이러한 대형 system은 전자통신망에 응용되어져

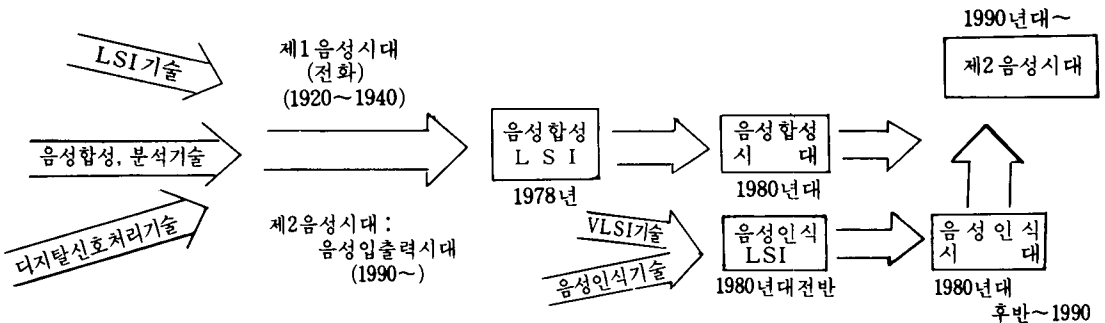


그림1.

오다가 반도체 집적 기술의 급속한 진전과 음성파형 분석의 통계학적인 기술 확립으로 음성합성 LSI (음성 합성 CHIP)의 탄생을 가져 오게 되었다.

음성 합성 및 인식 기술의 발전사를 간략하게 도시하면 아래의 그림 1 과 같다.

음성 합성기술은 음성파형의 정보를 세밀히 분석하여 음성파형을 digital (디지털) 신호로 변환하여 약 1/30~1/60정도로 정보 압축을 한후 LSI memory 에 저장해 음성 합성 LSI를 사용 원래의 음성 파형으로 재생하는 것이다.

음성 data (정보)의 저장 방법으로는 Analog tape 녹음, magnetic drum & disk 저장, 반도체 memory 에 저장하는 3 가지의 대표적 방법을 들 수 있다.

Analog tape 녹음이나, magnetic drum & disk 저장을 이용한 음성 합성 방식에 대해서는 본 고에서 서술하지 않고 반도체 memory 를 이용한 음성 합성 방식에 대해서만 서술 하겠다.

2. 음성 합성의 기초

현재 물리적 음향 이론에 의한 음성의 생성 과정은 아래 그림2와 같다.

음원의 생성, 조음, 방사의 3 작용을 거치므로써 음성파형이 발생하게 되는 것이다. 이러한 인간의 음성 생성과 음성합성 장치의 대응 관계를 아래 그림2를 통해 설명하겠다.

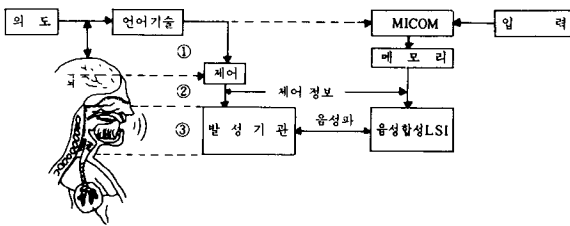


그림2.

① 부분은 대뇌에서 어떤 언어를 표현하려고 하는 생각을 하게되면 ② 부분의 제어 명령에 따라 ③ 부분의 성대의 진동과 공기 흐름의 복합 작용으로 음원이 발생된 후 혀와 입의 조음 작용으로 원하는 음성을 방사시키는 것이다.

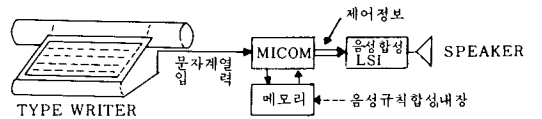
마찬가지로 음성 합성 장치에서도 원하는 선택 key 의 입력을 시키면 micro-computer 의 제어 명령에

따라 메모리에 들어 있는 음성 data를 꺼내어 음성 합성 LSI를 통해 음성파형을 재생 음성을 출력시키는 것이다.

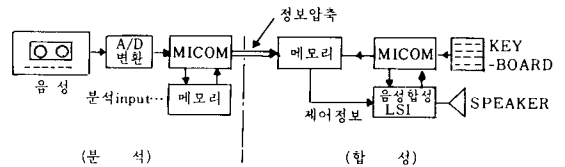
3. 음성 합성 방법 및 순서

먼저 음성 합성법에는 크게 2가지로 분류 할 수 있다. 규칙 합성법과 분석 합성법으로 나눌 수 있다. 여기서는 규칙 합성법에 관한 내용은 간단히 설명하고 분석 합성법에 대해서만 설명 하겠다.

규칙 합성법과 분석 합성법에 관한 내용을 간단히 도시한 그림이 그림 3 에 있다.



(a) 규칙합성



(b) 분석합성

그림3.

분석 합성법에 대한 방법 및 순서에 대해서 설명을 하겠다.

먼저 음성 합성을 하기 위해서는 맨 먼저 필요한 음성의 녹음을 해야 된다. 이 이 녹음을 할때는 잡음을 배제하기 위하여 studio에서 하는 것이 제일 바람직 하다. 녹음된 음성 파형은 A/D converter를 통해 digital 신호로 변환 분해되어 음성 파형의 분석을 하게 되는 것이다. 분석된 정보는 압축되어 지정된 memory내에 저장되어 음성 합성 LSI를 통해 음성파형으로 재생하게 되는 것이다. 음성 파형이 분석되어 data로 작성되는 과정과 음성 합성 LSI를 사용하여 음성을 합성하는 과정을 간략하게 FLOW-CHART로 표시 한것이 아래 그림 4, 5 에 나타나 있다.

1) 음성 합성 방식의 분류

음성 합성 기술은 아래 그림5에서 보는 바와 같이 크게 한정어 음성 합성과 임의어 음성 합성의 크게 2가지로 분류된다. 한정어 음성 합성은 수백어 정도 이하 어휘의 합성을 하는 방식으로써 분석 합성법 이라고도

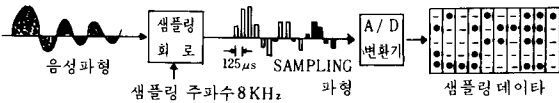


그림4.

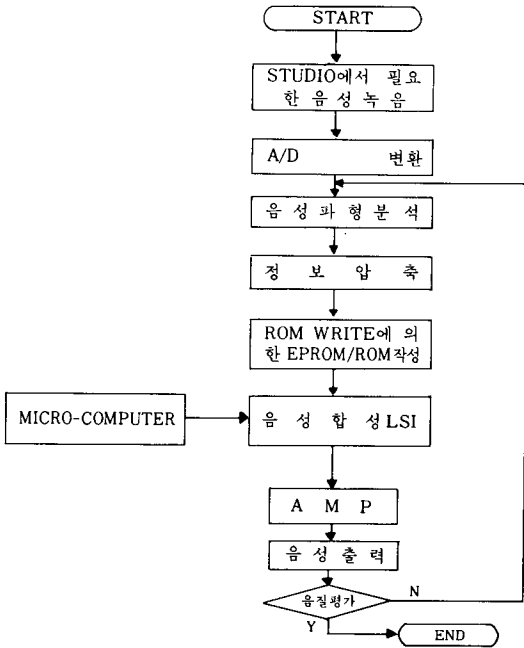


그림5.

한다. 임의어 음성 합성은 수천어 이상의 어휘를 합성하는 방식이며 규칙 합성법이라고도 한다.

음성 합성 방식에 대한 분류표는 아래 그림6 과 같다.

분석 합성 방식과 규칙 합성 방식의 비교를 위한 그림은 그림2 에 표시되어져 있다.

이러한 음성 합성 방식에 대한 구체적인 분류와 정보량, 합성장치, 제어정보, 음질, 등의 관점을 비교분석한 결과가 다음의 표1 에 나타나 있다. 상기의 여러가지의 음성 합성 방식중에서도 현재 파형 부호화 방식과 PARAMETER 방식의 전용 LSI화가 이루어져 음성합성 응용 제품에 널리 이용되고 있다.

(1) 파형 부호화 합성 방식

이 합성 방식은 자연 음성 파형을 디지털화(ENCODING : 부호화) 하여 메모리에 저장 하였다가 제어 정보에 따라 다시 복호화(DECODING) 되어 자연 음성 파

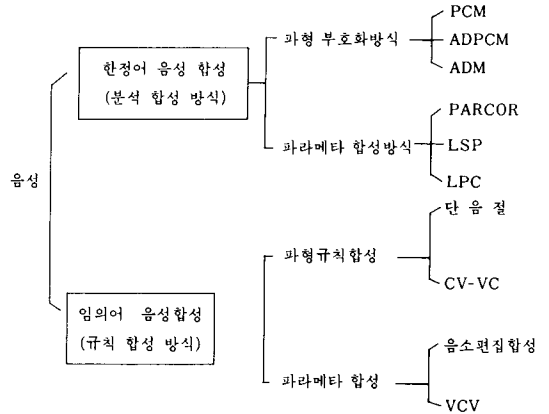


그림6.

형으로 재생되는 방식이다.

이 방식은 명료한 양질의 음성을 얻을 수는 있으나 음성 파형 자체를 디지털 신호로 바꾸어 메모리에 저장해야 하므로 엄청난 메모리 용량이 요구되는 단점을 갖고 있다. 이 방식은 1960년대 최초 실용화 되었으며 다중응답능력이 뛰어나 대규모의 전화 통신망 system 에 많이 이용된 PCM(Pulse Code Modulation) 방식 이 주류를 이루어 발전되어 왔으나 음성파형을 디지털화 하여 기억시키는데 최저 8 KHz의 sampling 주파가 필요하다. 즉 다시 말하면 1 Sampling 당 8~7 bit가 필요한 것이다.

표 1.

구분방식	파형부호화 방식	PARAMETER 부호화 방식	순서합성 방식 (규칙합성)
합성방식	PCM	PAR COR	단음절 CV-VC
	ADPCM	LSP, LPC	음 소, VCV
정보량	PCM: 64K BIT/S ADPCM: 24K BIT/S	2.4, 4.8, 9.6K BIT/S	75 BIT/S
합성장치	PCM: 간단 ADPCM: 비교적 간단 (전용 LSI 필요)	복 잡 LSI 필요	전자계산기 SIMULATION 필요
제어정보	PCM: 간단 ADPCM: 간단 (MICOM 제어)	복 잡 4 BIT MICOM 제어가능	복 잡 대규모 PROGRAM 필요
음 질	매우양호	양 호	비교적 음절저하
음성사용유무	사람의 음성 이용	사람의 음성 이용	사람의 음성 사용 무 기계가 적절 음성제작
음성DATA 메모리 용량	메모리 용량: 대	메모리 용량: 대	없 음
COST	높 음	낮 음	대용량의 설비 필요

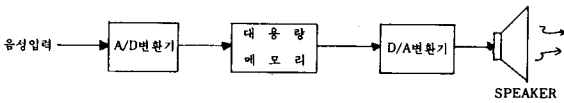
1초간의 파형을 기억시킬려면 약 60Kbit가 필요하여 100초간의 음성 파형을 기억시키는데는 6 Mbit라는 엄청난 메모리 용량이 요구되어 자기 (magnetic) disk나 자기(magnetic) drum의 메모리가 필요하여 대형 system 설비가 불가피 하였다.

이런 PCM 방식의 문제점을 보완하기 위해 음성파형의 분포 상관관계에 통계학적 이론을 도입하여 효율적인 부호화를 행해 음성 data의 정보량을 압축시킨 ADPCM(Adaptive Differential Pulse Code Modulation) 방식, ADM(Adaptive Delta Modulation) 방식, DPCM(Differential Pulse Code Modulation) 방식의 탄생을 가져오게 되었다. 이 중에서도 현재 가장 널리 응용되고 있는 방식이 ADPCM 방식이다.

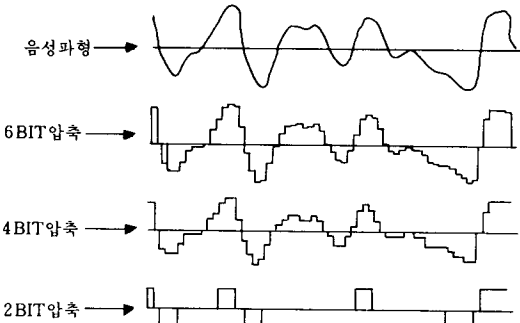
이 ADPCM 방식을 사용함으로써 음성 data 정보량을 4 bit의 ADPCM 방식으로 8 bit의 PCM방식과 동

질의 음성 부호화가 가능해 지는 것이다. 이렇게 함으로써 약 1/2 정도의 메모리 용량(bit)의 축소를 이룰 수 있는 것이다.

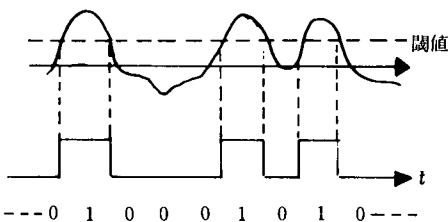
그림 7 은 파형 부호화 방식의 파형 처리 방법에 대해 각 방식별로 나타낸 것이다.



(a) PCM 합성방식

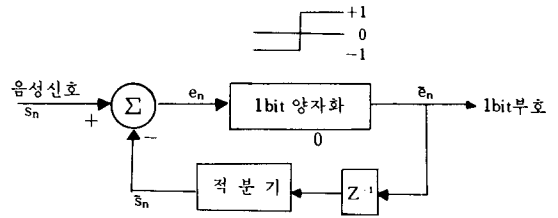


(b) 음성신호의 BIT 수에 따른 압축파형

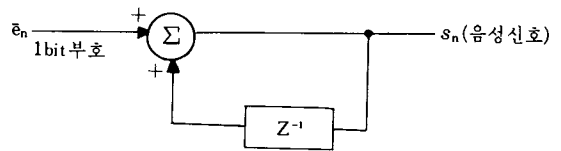


(c) PCM 녹음 재생파형

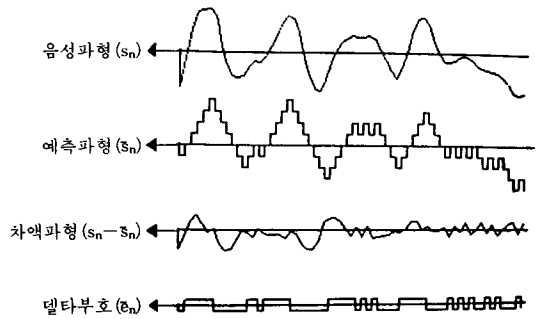
그림 7. PCM 방식



(a) 델타방식 부호화 원리도



(b) 델타방식 복호화 원리도

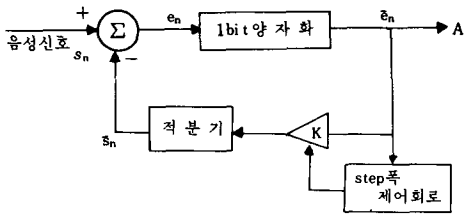


(c) 델타방식의 신호 파형

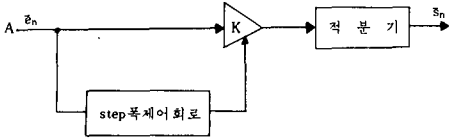
그림 8. DM(Delta Modulation) 방식

① ADPCM(Adaptive Differential Pulse Code Modulation)

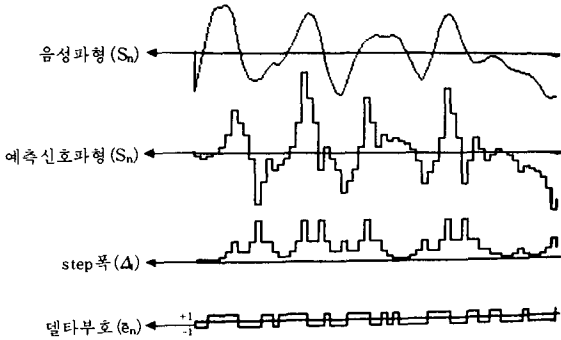
DPCM 방식에서 양자화 폭을 적당히 변화 시키는 방식으로써 양자화폭 Δ 의 최대치 Δ_{max} 와 최소치 Δ_{min} 을 설정하여 필요이상의 적은 신호나 필요이상의 큰 신호를 제거하는 방식이다. 이 ADPCM방식을 사용함으로써 4 bit의 ADPCM 방식으로써 8 bit PCM 방식과 같은 수준의 음질을 얻을 수 있다.



(a) ADM 방식의 부호화 원리도



(b) ADM 방식 복호화 원리도



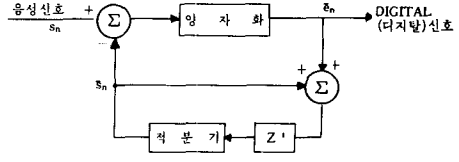
(c) ADM 방식의 신호파형

그림9. ADM(Adaptive Delta Modulation)

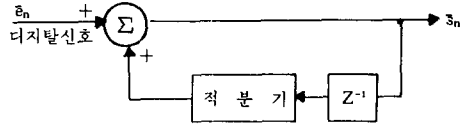
(2) 파라메타 (PARAMETER) 부호화 방식

앞에서 서술한 파형 부호화 방식은 정보(음성 data) 압축률에서 한계가 있어 대용량의 음성 출력을 사용하는데는 문제가 있다. 이런 문제를 해결하기 위해 음성 파형 그 자체의 음성 data(정보량)를 압축시키는 방식이 아니고 음성 파형이 가지고 있는 특성 즉 parameter 분석을 하여 정보를 압축시켜 합성시키는 방식으로써 파형 부호화 방식에 비해서는 메모리 용량을 줄일 수 있어 cost적인 측면에서는 유리하나 음질적인 면에서는 파형 부호화 방식에 비해서 저하되는 단점을 갖고 있다.

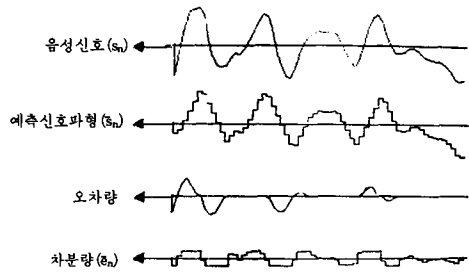
이 방식에는 LPC방식, LSP방식, PARCOR 방식이 있는데 이론적인 측면에서는 모두 선형 예측 분석 합성법을 채택하여 선형예측계수와 음원 정보와의 양자화를 실시하여 부호화(ENCODING)한후 메모리에 저



(a) DPCM 방식의 부호화 원리도



(b) DPCM 방식의 복호화 원리도



(c) DPCM 방식의 신호파형

그림10. DPCM(Differential Pulse Code Modulation) 방식

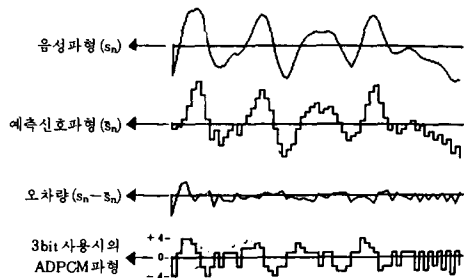
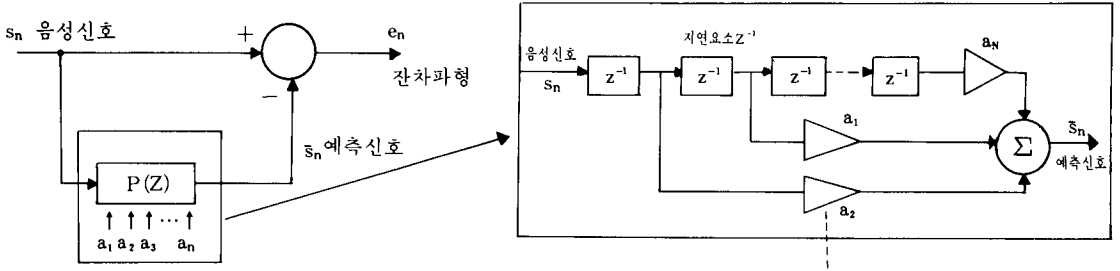
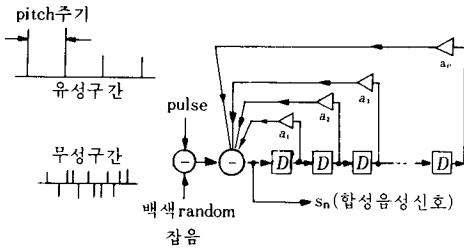


그림11.

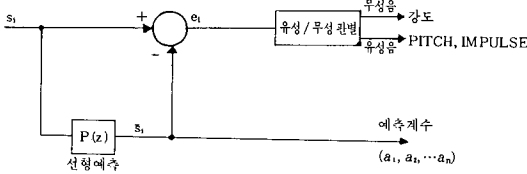
장하는 LPC방식이 저 bit(5bit 이하)에서 양자화를 실시하면 합성 음성의 spectrum이 원래 음성의 spectrum과의 차이가 많이 생겨 음질이 현저히 저하되며 또 합성회로가 불안정하면 발진을 일으키는 점등이 실용상의 문제에서 치명적인 단점으로 대두되는 문제점을 개량하여 보완한 방식이 PARCOR 방식이며 PARCOR 방식의 선형보간특성을 개량한 방식이 LSP방식 이므로 이 3 가지 방식의 기초는 모두 선형예측 분석



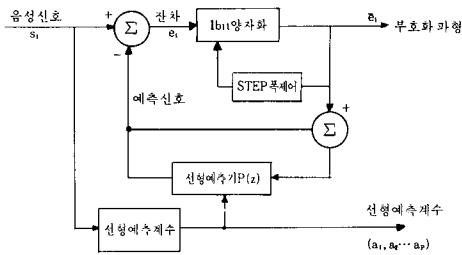
(a) 선형 예측법 구성



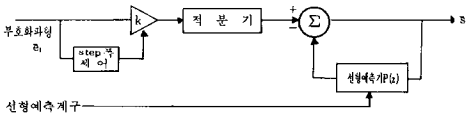
(b) LPC방식의 음성 합성 원리도



(c) LPC방식의 파라메타 추출



(d) LPC방식의 부호화 회로



(e) LPC방식의 복호화 회로

그림12. LPC(Linear Predictive Code) 방식

합성 이론에 기인한다고 볼 수 있는 것이다.

그림12는 선형 예측 방식의 구성 및 합성회로, 파형 처리 방법을 그림으로 표시하였다.

그림14의 LSP 방식은 PARCOR 방식의 선형 보간 특성을 개량한 방식으로써 같은 정보량으로써 음성합성을 행하면 PARCOR 방식 보다는 양질의 음성을 얻을수 있다. 또한 음질을 같은 수준으로 가면 이 LSP 방식은 PARCOR 방식의 비해 80~60%의 정보량으로 행할수 있어 메모리 용량면에서 유리한 점이 있다.

LSP 방식은 선형 예측 분석 합성계에서 최종 목표치라고 이야기 할 수 있다.

III. 음성합성 냉장고

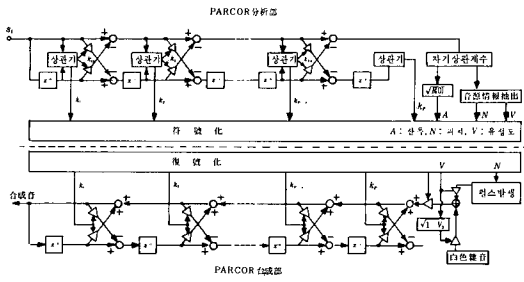
1. 음성합성 냉장고의 필요성

1) 도아 개폐시 에너지 손실

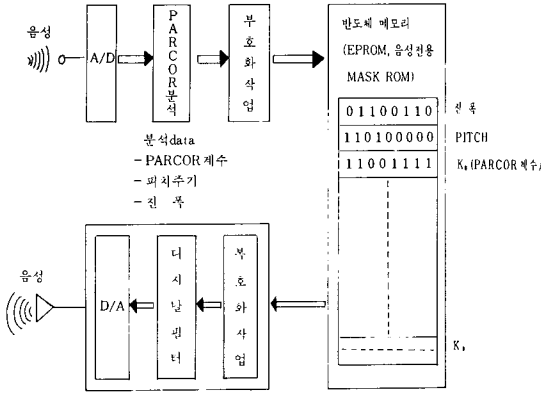
식품을 꺼낼때 냉기가 유출되는 것보다 반도아시(도아가 완전히 닫혀지지 않은 상태)에서 냉기 유출이 심해 소비 전력량의 증가가 이외로 큰편이다.

왜냐하면 식품을 꺼낼때는 냉장고 문이 완전히 열린 상태가 되어 팬(FAN)이 정지하여 있기 때문에 자연 순환되는 냉기만이 유출되나 반도아시에는 팬이 계속 돌아 냉기를 냉장고 문이 열린 틈새로 계속 뿜어내어 고내의 온도 상승을 초래하기 때문이다. 특히 야간에 냉장고 문이 완전히 닫혀지지 않은 상태가 되면 냉기 유출이 심해 냉장고내의 온도상승으로 인한 식품변질의 원인이 될뿐 아니라 계속 냉장고가 운전됨으로써 막대한 소비전력 상승을 가져오게 된다.

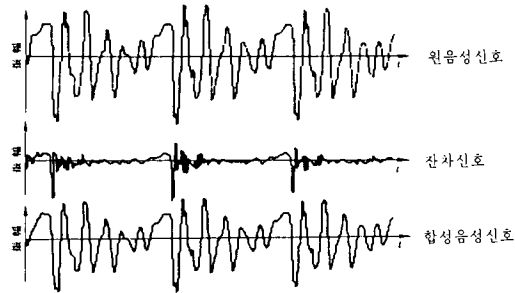
그러면 반도아시(냉장고 문이 완전히 닫혀지지 않은 상태)의 소비 전력량 증가 및 냉장고 문 개폐 횟수에 따른 소비 전력량 증가에 대해 실험치를 가지고 고찰해 보기로 한다. 먼저 냉장실 문이 10mm 정도 열려 있다면 1일의 소비 전력량을 100%로 본다면 이에 대



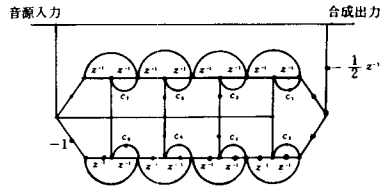
(a) PARCOR 방식의 분석 합성계



(b) PARCOR 방식의 음성분석 합성 개략도



(c) PARCOR 방식의 신호 파형



(LSP 방식 합성회로 구성)

그림14. LSP (Line Spectrum Pair) 방식

한 소비 전력량의 증가비율을 나타낸 표가 아래 표 2에 있다.

본 음성 합성 냉장고는 15초 이상(시간 조절 가능) 문이 열리면 "빨리 냉장고 문을 닫아 주세요"라고 경고 음성 메시지를 전달하여 냉장고 문을 닫아 주게 되는 것이다.

표 2.

냉장고문이 10mm 열린후 경과시간	소비 전력량의 증가율
30 초	0.3%
1 시간	16%
3 시간	23%

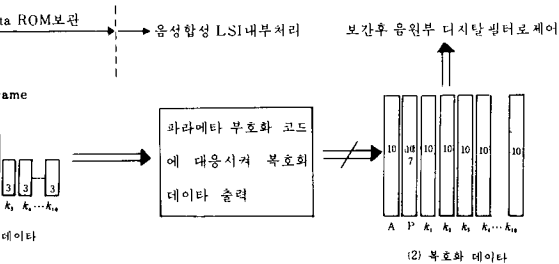
(조건: 주위온도 30℃, 전원: 100V/60Hz, 온도조절기: 냉동실, 냉장실 중-NOTCH) 그리고 식품을 꺼낼 때의 냉장고 문이 열린 시간과 소비 전력량과의 관계에 대해서도 알아 보면 실험을 통한 조사표에 따르면 냉장고 문을 열고 식품을 꺼낼때 냉기가 외부로 유출되는데 사용자의 입장에서는 단시간내에 식품을 꺼내고 문을 닫아 주는 것이 성능에서 측면에서 가장 양호한 것이다. 일반 가정에서의 냉장고내의 식품을 꺼낼 때 시간 분포에 대한 조사표가 아래 표 3에 나타나 있다.

표 3.

도아열린시간	5초 이내	5초-10초	10초-15초	15초이상
비율	14%	58%	21%	7%
누계 비율	14%	72%	93%	100%

상기표에서 보듯이 15초 이내에 사용하는시간이 93%를 차지하고 있다. 이렇게 냉장고 문을 열고 닫는 시간이 15초 이하이면 냉장고에서는 별 문제가 없는 것으로 나타나는 것이다.

그러나 냉장고 문이 15초 이상 열려 있으면 고내온도 및 소비 전력량에 영향을 미치게 된다. 냉장고 문이 15초에서 30초 사이의 시간이 경과 할때의 소비 전력량 분포에 대한 시험 결과표가 아래 표 4에 표시 되



(d) 음성합성용 파라메타 부호화 데이터와 부호화 데이터 관계

그림13. PARCOR(Partial Auto Correlation) 방식

어져 있다. 아래 도표는 냉장고 문이 15초 이내에 열린 경우 소비전력량 증가율은 무시한 것이다.

표 4.

냉장고문 개폐조건			1 일 소비전력 증 가 율
개 방 시 간	개 폐 회 수	개 방 도 아	
15초	15회/일	냉 동 실	0%(기준치)
	50회/일	냉 장 실	
30초	15회/일	냉 동 실	약 10%증가
	50회/일	냉 장 실	

이상으로써 냉장고 문(DOOR) 개폐와 에너지 손실과의 관계를 고찰한 결과 본 말하는 냉장고(음성 합성 냉장고)는 음성메세지를 사용자에게 전달하여 줌으로써 귀로 듣고 눈으로 확인할 수 있게 하여 절전효과에 크게 기여할 수 있다는 점에서 개발의 의미가 있는 것이다.

2) 제상기능시의 음성경보

냉장고가 일정한 주기가 되면 성에를 제거하는 작업에 들어 가게 되는데 이때는 Heater 발열로 성에를 제거 하게 되므로 사용자들이 성에 제거 기능에 들어간 것을 모르고 냉장고 문을 자주 열게 되면 냉장고 고내 온도 온도 상승으로 식품 변질의 원인이 되므로 사용자에게 성에 제거중인 것을 음성 메세지로 경보를 해 주어 냉장고의 올바른 사용을 유도하여 사용의 편리성 강화를 할 수 있는 것이다.

3) 냉동실, 냉장실 온도상태 감지

센서를 이용하여 냉장실, 냉동실 온도를 감지하여 냉장고내의 온도가 높은 경우(냉장실: +9℃ 이상, 냉동실: -10℃ 이상)이면 음성 메세지(음성경보)를 사용자에게 알려주어 냉장고의 이상상태 유무를 진단하거나 냉장고 내부의 정리정돈을 하여 냉기 순환을 원활히 하도록 함으로써 전력낭비의 감소 및 사용의 편리성을 추구 할 수 있도록 하였다.

냉장고내의 온도 분포상태는 냉장고내의 식품 보관량과도 관계가 있다. 냉장고내에 적량의 식품을 투입했을 때와 많은량의 식품을 넣을 때의 식품의 냉각속도 관계를 나타낸 그래프가 아래 그림15에 나타나 있다(여기서 식품의 적량이란 냉장실 내용적 1/3의 식품의 보관량을 말한다). 그림15에서 보는 바와 같이 식품의 온도가 안정되는 시간까지에는 2 개의 그래프가 상당히 차이가 있는것을 알 수 있다.

적량의 식품을 냉장고에 보관할시에는 약8시간 정도

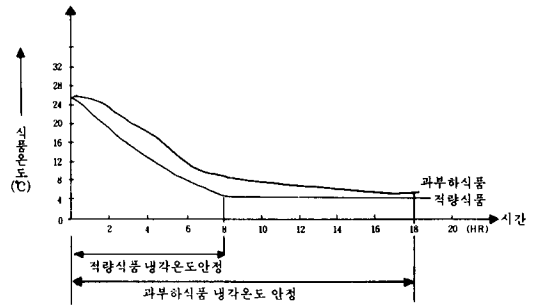


그림15.

되면 식품의 냉각이 안정되지만 적당량 이상의 식품을 냉장고에 보관할시에는 약18시간 정도가 소요되어 적당량 식품의 2 배 이상이 되어야 식품의 냉각이 안정되어 냉장고 냉각기의 구동으로 인한 소비전력의 상승이 많이 이루어 진다.

본 음성합성 냉장고는 상기와 같이 냉장고내의 과다한 식품이 보관되어 적정시간 동안에 설정한 냉각속도까지 온도가 내려가지 않으면 음성경보를 사용자에게 하여 주어 냉장고내의 식품의 정리정돈 및 적정량의 식품 보관이 이루어 지도록 하여 준다.

이외에도 본 음성 합성 냉장고는 사용자가 급속냉동 기능을 선택할시 음성 메세지를 전달해 주어 냉장고 냉각기의 연속구동 및 냉동실 문을 열지 않도록 주의를 환기시켜 주며 처음 냉장고 전원을 투입하거나 정전후 다시 전원이 공급될때도 음성 메세지를 전달하여 냉장고에 전원이 공급되고 있다는 것을 알려주어 냉장고 이상유무 상태도 판단할 수 있도록 하였다.

2. 음성 합성 냉장고 구조

본 냉장고의 외형 및 컨트롤부의 구조는 다음 그림 16과 같다.

그러면 본 냉장고의 음성 합성부에 대해서 서술하겠다.

1) 음성 합성부

본 냉장고에서 채용하고 있는 음성 합성 방식은 파라메타(PARAMETER) 합성방식중의 PARCOR 방식을 사용하고 있다. 음성합성 방식에 대한 분류는 앞의 II 절에서 서술하였다. 본 음성 합성 냉장고의 PARCOR 방식은 음성파형 data 압축률이 높아 메모리 용량이 적어지는 이유로 음성합성 방식 중에서 지금 현재 주류를 이루고 있는 편이다.

이 PARCOR 방식은 음성신호 파형 자체를 분석 합성하는 방식이 아니고 음성이 가지고 있는 특성 파라

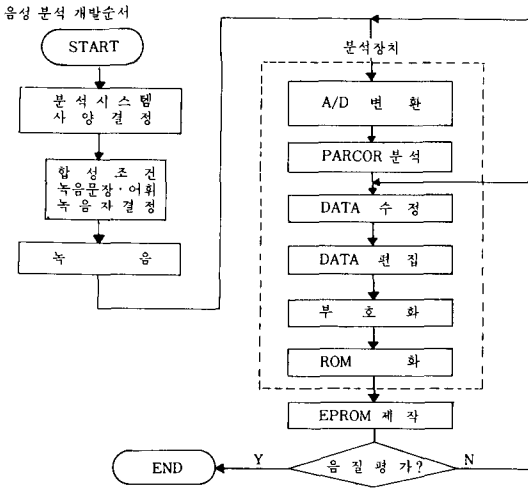


그림18.

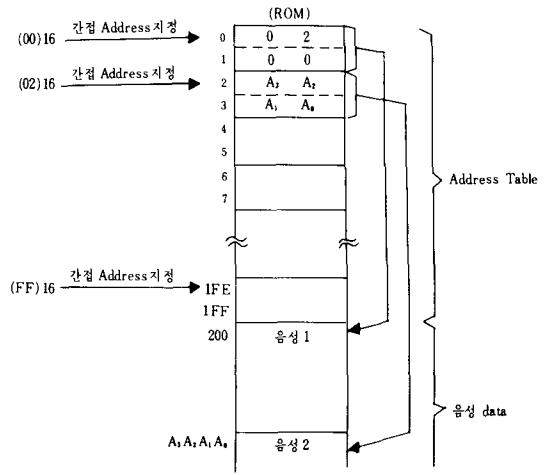


그림19.

음성 합성 LSI로 들어간 음성 data는 내장된D/A 변환기를 통해 아날로그 신호로 변환되어 $f_0 = 4 \text{ KHz}$ 의 저역필터를 통해 filtering된 후 증폭기를 통해 스피커를 통해 음성이 재생되는 것이다.

본 냉장고에서 사용한 음성합성의 어휘는 아래에 있는 표5와 같다.

표 5.

구 분	합 성 어 휘
1. 최초전원 공급시 또는 정전후 전원공급시	안녕하세요, 저는 삼성전자의 말하는 냉장고 입니다. 앞으로 많이 애용해주세요
2. 냉장고 문이 열린 경우 (냉동실, 냉장실문이 열리고 30초후)	빨리 냉장고 문을 닫아주세요.
3. 냉장실 온도가 높은 경우(+9℃ 이상)	냉장실 온도가 높습니다.
4. 냉동실 온도가 높은 경우(-10℃ 이상)	냉동실 온도가 높습니다.
5. 냉장실, 냉동실 온도가 모두 높은 경우	냉장실, 냉동실 온도가 높습니다.
6. 급속 냉동 기능시	지금은 급속 냉동중입니다.
7. 성에 제거시	지금은 제상 중입니다.

음성data 지정방식에 대한 그림은 그림19에 나타나 있다.

간접지정 방식

ROM ADDRESS	BIT8	7	6	5	4	3	2	1	0
DATA LINE	D 0	D7	D6	D5	D4	D3	D2	D1	

IV. 결 론

1. 음성 합성 냉장고의 실용화 및 향후 추진 방향
본 냉장고의 음성분석, 합성방법이 파형 그자체를 분석, 합성하는 것이 아니고 음성파형이 가지고 있는 파라미터를 추출 분석하여 합성을 하는 방식이므로 음질 저하는 불가피하나 코스트적인 측면에서는 유리 한 것이다. 이러한 코스트적인 측면도 메모리 대용량화에 따른 COST DOWN에 의해서 점차 저하되고 있다.

향후 본 냉장고의 실용화를 위해서는 음질을 거의 원래의 음성과 같은 수준으로 가면서 가격적인 면도 점차 저하하고 있는 ADPCM방식도 깊게 검토 해볼 필요가 있다.

2. 음성 합성 기술의 향후 전망

음성 합성의 연구는 20년 이상의 역사를 가지고 있어 본격적인 실용화는 비교적 최근에 와서 이루어진 사실이다. 음성 합성인 경우 최초로 미국의 TI(TEXAS INSTRUMENT)에서 LPC 파라미터 음성 합성 LSI가 개발에 성공하여 1978년도에 단어 학습기에 응용 판매되자 선진국 각사에서 음성 합성 LSI개발에 박차를 가해 급속적인 발전을 보게 되었다.

이러한 추세로 음성합성 장치에 대한 코스트가 대폭 저하됨과 동시에 소형화가 이루어져 각종 제품에 음성 출력 기능을 부가 하게 되었다.

그러나 실제 음성 합성 LSI 응용에는 코스트, 음성, 음성 출력시간, 사용어휘수, 사용어휘의 변경 유무 등이 필요한 조건으로 대두되어 위의 조건들을 완전히

만족하는 음성 합성 LSI가 요구 되겠지만 이 모든 조건의 만족은 사실상 불가능 하므로 사용조건에 따라 가장 적당한 LSI를 선택하는 것이 최선의 방법이다.

사용 메세지 수가 적고 고 품질의 음질을 원한다면 파형부호화 방식이 가장 좋으며 사용 메세지 수가 많고 음성의 질이 다소 떨어지는 사용조건에는 파라메타 방식의 음성 합성 LSI 선택이 적절한 방식인 것이다.

음성 합성의 응용에는 퍼스날 컴퓨터를 이용한 게임, 교육등의 프로그램 작성시 경보용으로 응용이 가능하며 인터폰에 연결하여 방문객과의 자동응답 기능에 응용 할 수 있다. 또한 음성 합성 및 음성 실시간 분석 기술의 조합으로 정보, 압축 디지털, 데이터 전화 개발이 가능 해 질수 있다. 음성 합성 및 음성 실시간 분석 기술의 조합에 의한 디지털, 데이터 전화 시스템 그림이 그림20에 표시되어 있다.

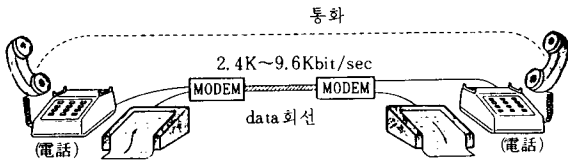


그림20.

그리고 음성분석, 합성, 인식기술의 3위 일체에 의한 전자 메모 기기(포켓용)의 탄생도 기대할 수 있다. 각종 회의시간, 약속시간, 전화연락 등의 메모내용을 음성으로 입력시키면 음성을 분석하여 메모리에 저장해 두었다가 디지털 시계가 지정된 시간이 되면 음성 출력을 내주게 되는 것이다. 이런 제품은 음성기술의 복합적인 발전과 더불어 음성합성 LSI의 개발도 활발히 이루어져야 될 것이다.

음성 합성 LSI에 대한 발전과정이 아래 표에 나타나져 있으며 음성 합성의 기술의 응용 분야에 대한 현재 및 미래의 동향을 그림으로 표시하면 그림21과 같다.

이상으로써 음성 합성 분석 기술의 개념 및 합성 LSI의 개요, 응용분야 등을 서술 하였다. 이 분야는 장기

1978년도	상용 MOS 기술 5~3 마이크로	(15,000Gate) (3~5ns/gate)	음성합성 1CHIP⇒음성출력
1980년대 전반	최신 MOS 기술 3~2마이크론	(7,000Gate) (2~3ns/gate)	음성분석 1~2CHIP⇒음성전달, 기억
1980년대 후반~	VLSI 기술 2마이크론 이하	(20K Gate) (1.5ns~1ns/gate)	음성인식 1CHIP⇒음성출력

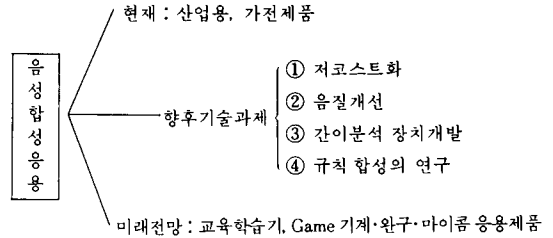


그림21.

간의 연구 개발의 역사를 이어 오면서 궁극적인 목표로는 인간과 동일한 음성출력을 요구하는 것이므로 앞으로 계속 음성합성에 대한 연구 개발이 활발히 이루어지고 있으며 또한 반도체 집적기술의 집적도 및 처리 속도의 향상으로 계속적인 진보 발전이 되어 응용 분야 또한 넓어 질것이라 믿어 의심치 않는다.

參 考 文 獻

- [1] 大泉(監修) : 음성과학, 東大出版會
- [2] 市川 : 음성의 합성출력, 정보처리 vol. 19, no. 7, pp. 657 (1978)
- [3] 음성デバイス 활용과 고찰 : 中田和男 (1982)
- [4] 컴퓨터 음성처리 : 전자과학 series (1980)
- [5] X. Rodet et al : Time domain speech synthesis-by -rules using a flexible and fast signal Management system
- [6] 음성 인식, 합성 : 일본공업기술 협회 (1980)
- [7] 板倉, 東倉 : 음성의 특징추출 및 정보처리, 정보압축 vol. 19, no. 7, pp. 652 (1978) *