

OA 기기의 음성 입출력 기술

이 병 수*

I. OA 기기

1.1 OA 기기의 분류와 기능

일반적으로 사무기기는 사무업무 처리의 신속, 정확을 기하고 생산성을 향상하기 위하여 사무처리 시스템에 사용할 수 있는 모든 기기나 장치를 총칭한다. 특히 사무자동화(OA : Office Automation) 업무에 적용되어 기업의 생산성 향상에 사용되는 각종 사무기기를 OA 기기라 부른다.

최근에는 전자기술의 발전에 따라 대부분의 OA 기기는 단기능 기기에서 기능을 복합화하여 하나의 시스템으로 여러 기능을 갖고 다목적으로 사용할 수 있는 복합적인 다기능 기기로 발전되고 있는 추세이다. 그 예는 워드프로세서와 개인용 컴퓨터가 복합된 장치, 광학문자 읽기장치 기능과 팩시밀리 기능을 복합한 장치 등이다.

이와 같이 다기능화 되고 있는 OA 기기를 업무의 유형에 따라 정확히 구분하는 것은 곤란하지만 정보처리 유형에 따라 분류해 보면 그림 1과 같이 정보의 수집 및 생성을 할 수 있는 데이터 준비기기, 데이터 계산업무 처리

를 위한 데이터 처리기기, 데이터의 전송과 교환을 위한 데이터 전송기기, 필요한 데이터를 획득하여 이를 보관하고 필요한 시기에 검색하는 데이터 축적·검색기기 등으로 분류할 수 있다.

1.2 OA화 기술 체계

OA란 컴퓨터기술, 통신기술, 시스템과학 및 행동과학 등에 의해 이루어지는 복합기술로서 사무실 조직이나 인적 요소는 물론이고 활용하고 있는 장비나 기기의 기능을 모두 흡수해야 하므로 그 범위가 넓고 종류가 매우 다양하다. 따라서 OA 시스템이 갖추어야 할 기본기능은 사무실에서 취급되는 모든 사무정보의 형태를 취급해야 하므로 OA 기술의 체계는 그림 2와 같이 사무실 업무에 따라 문서작성 및 맨/머신 인터페이스에 관련된 정보의 입출력 기술, 데이터 처리에 관련된 정보처리기술, 정보전송에 관련된 통신기술, 보존·검색에 관련된 축적검색기술 등으로 나눌 수 있다.

① 입출력기술

- 입력기술 : 키보드, 문자 및 문체인식, 음성인식
- 출력기술 : 소프트카피, 하드카피, 입체표시, 음성합성

* 순천향대학교 정보통신공학과

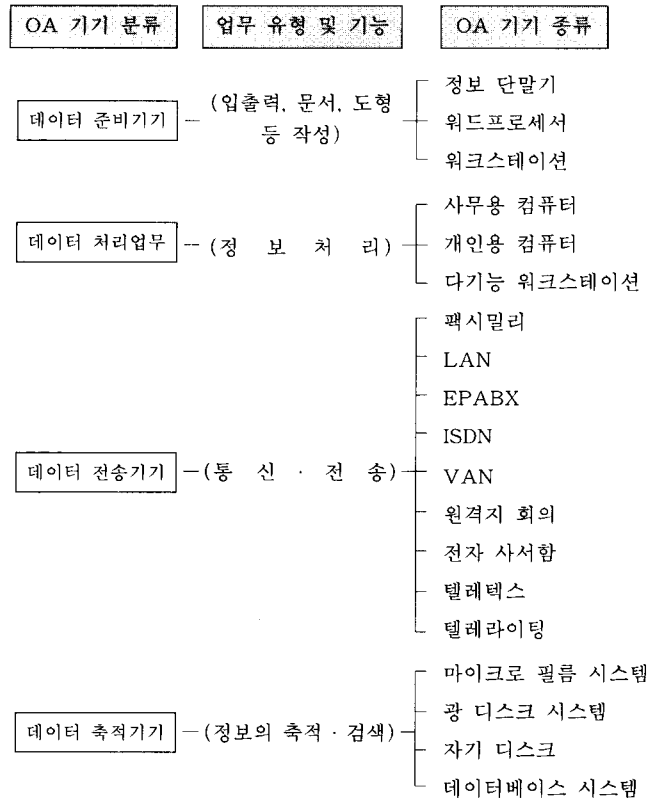


그림 1. OA 기기의 분류 및 기능

② 정보처리기술

- 본체기술 : 다중/병렬처리, 인공지능, 소프트웨어 기술
- 처리기술 : 문자정보처리, 음성정보처리, 영상정보처리

④ 축적·검색기술

- 파일링 시스템 : 광디스크, 대용량 자기디스크, 마이크로그래픽
- 데이터 베이스 : 정보검색과 정보안내

③ 통신기술

- 단말기술 : 음성, 비음성, 영상, 복합 단말기
- 전송기술 : 대용량화, 다기능화, 광대역화, 지능화
- 교환기술 : 디지털 교환방식, 회선/패킷 교환방식, 비동기식 교환
- 무선통신기술 : 이동통신, 위성통신, 휴대 통신

1.3 음성입력기술

음성 입출력 기술은 인간의 음성을 인식하여 컴퓨터 시스템에 입력시키고 그 결과를 사람의 음성으로 출력시켜서 인간에게 전달하는 기술로 전자를 음성인식, 후자를 음성합성 이라고 부른다.

이 중에서 음성인식(speech recognition)이란 아날로그 신호인 음성을 디지털 2진코드로

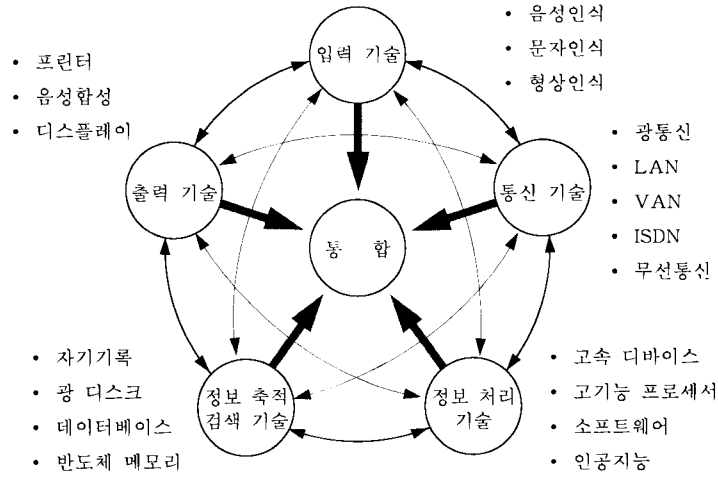


그림 2. OA 기술의 체계

변환하고, 이 디지털 값과 기억장소에 미리 저장된 단어 패턴(pattern)의 값들을 비교하여 같은 것을 찾아내는 과정을 말한다. 따라서 음성인식을 좋게 하기 위해서는 컴퓨터가 디지털로 표현된 어휘를 많이 저장하고 있어야 한다. 그러나 음성을 디지털 형태로 변환하여 저장하는데는 많은 기억용량이 소요되므로 컴퓨터에 저장된 어휘량은 일반적으로 매우 제한적이고 특정 화자에 국한될 수 밖에 없다.

음성인식 연구에서 가장 어려운 점은 사람마다 말을 할 때 억양이나 음정, 멈춤, 서로 다른 발음 등의 차이를 인식시키는 것이다. 음성인식 연구자들의 목표 중 하나는 음성신호를 보다 적은 디지털 2진코드로 변환하여, 같은 기억용량에 많은 어휘들을 기억시키고자 하는데 있다.

정보의 입력수단으로 음성을 직접 이용할 때 생각할 수 있는 유리한 점은 일반적으로 다음과 같다.

- ① 음성은 인간에 의한 가장 자연스러운 정보발생수단이며, 가장 편리한 입력수단이다. 즉,
 - 정보입력을 위해 특별한 훈련이 필요없다.
 - 예기치 않은 상황하에서의 정보입력도

용이하게 가능하다.

- ② 정보의 발생을 고속으로 행할 수 있다. 연속적인 자연회화언어의 입력이 가능하다면 키보드에 의한 정보입력 속도보다 4배 정도 빠르다.
- ③ 손이 부자유스러울 때도 자유롭게 입력할 수 있다.
- ④ 와이어리스 마이크를 사용하면 상당한 공간적 범위내에서 움직이면서 입력시킬 수 있다.
- ⑤ 입력의 결과를 바로 응답(표시 혹은 음성 응답)에 의해 재빨리 에러를 발견하여 바로 정정하므로써 입력의 신속성을 향상시킬 수 있다.
- ⑥ 원격단말(정보의 발생점)로부터 입력이 가능하여 전표를 중앙으로 모아 집중적으로 입력하는 것에 비해 시스템으로서의 정보입력 속도를 고속화할 수 있다.
- ⑦ 단순히 음성의 내용을 입력할 수 있을 뿐 아니라 말하는 사람의 확인도 가능하여 정보의 비밀유지를 동시에 실현할 수 있다.
- ⑧ 사용하는 쪽에서는 간략화에 의한 효과가

크다.

- ⑨ 장시간 주의깊게 긴장하여 음성을 들어야 할 때 피로하기 쉬운 사람보다는 기계에 의하는 편이 신뢰성이 높을 경우도 있다.

- ① 하드카피를 남길 수 없으므로 불편하다.
② 인간에 밀착된 매체이기 때문에 사회제도 및 관습, 가치관 등에 의해 이용형태가 좌우된다.

1.4 음성출력기술

음성인식은 컴퓨터에게 사람이 하는 말을 이해하도록 하는 반면, 음성출력은 컴퓨터가 사람에게 말하도록 하는 기술분야이다. 즉, 인간이 들어서 이해할 수 있는 음성의 음향적인 파형을 인간의 직접적인 발성에 의하지 않고 인공적으로 만들어 내는 것을 음성의 합성 또는 출력이라 한다.

음성응답의 예로서 번호가 변경된 전화번호를 돌리면 전화국에 있는 컴퓨터로부터 “지금 거신 번호는 9-2-6-2-5-6-3으로 바뀌었습니다.” 라는 음성이 들려올 것이다. 이 응답의 일부는 이미 녹음 저장된 내용이고 새로운 전화번호는 컴퓨터 음성응답장치에서 발생한 응답이다. 이 음성은 다소 부자연스럽기는 하지만 이해하는 데는 별로 어려움이 없다.

정보의 출력수단으로 음성을 직접 이용할 때 생각할 수 있는 유리한 점은 일반적으로 다음과 같다.

- ① 이해가 쉽고 전달속도가 빠르다.
- ② 시선과 손발의 동작이 구속받지 않는다. 다수의 측정기 및 복잡한 일부를 상세하게 읽어 내용을 이해하지 않아도 되기 때문에, 동시에 다른 정보를 얻거나 동작의 확인을 할 수 있다.
- ③ 인간의 가장 강력한 정보교환 수단이다.
- ④ 복잡한 정보를 다루는 일에는 부적당하나, 화상 등과 같이 사용하면 효과가 크다.

이와같은 잇점이 있는 반면 다음과 같은 불리한 점도 있다.

II. 음성인식

2.1 음성인식 시스템

음성인식에서 실제로 시스템이 직면하는 문제의 본질을 정리해 보면 다음 세 가지로 요약된다. 첫째로 최첨단부(front end)에서 당면하는 음향학적 정보의 불완전성 문제가 있다. 음운학적 기호, 즉 음소등은 실제 발성에서는 여러가지로 변형되어, 길이가 변화하고 경계가 불확실하며 심지어 탈락되는 등 극히 왜곡된 음향정보의 형태로 제공된다. 이 신호로부터 어떻게 필요한 정보를 끌어내느냐가 문제인 것이다. 둘째로 반대쪽 측면에서 음성인식을 위한 사람의 지적과정에 대한 이해가 부족할 뿐 아니라 이를 기계화할 기술도 없다는 점이다. 즉 인간의 음성인식 과정에서는 자신의 지적능력이 총체적으로 작용하여 빠진 소리나 틀린 말도 채워 넣을 수 있는 일종의 생산적 과정이 형성되므로 이의 일부라도 기계화하는 것은 두뇌의 기능을 모형화하는 수준의 작업이 되는 것이다. 셋째로는 기계의 처리속도의 문제이며 이 세가지 문제의 해결이 음성인식을 개선하는 요체라 보고 그 역사적 접근방법에 대하여 살펴보기로 한다.

1959년 J.W. Forgie와 C.D. Forgie가 인식실험에 디지털 컴퓨터를 사용한 이후 디지털 컴퓨터는 아날로그 하드웨어 인식기의 많은 기능을 대신하게 되었는데 초기에는 프로그램 처리속도가 너무 늦어 컴퓨터를 인식실험에 사용하는데 큰 문제점으로 등장하였다. 이러한 문제점은 Cooley와 Tukey가 만든 'Fast Fourier Transform' 알고리즘의 탄생 이후 음

성의 주파수 분석에서 현저한 시간단축을 보임으로써 많이 보완되었으며 1970년대에 들어와 개발된 선형예측법(Linear Predictive Coding)은 FFT보다 속도가 빠르고 더욱 유용한 파라미터를 추출할 수 있으므로 계속 발전되어 지금까지 사용되고 있다. 또한 70년대 이래 동적계획기법(DP)과 HMM(Hidden Marker Model)기법이 도입되어 음성인식에 크게 공헌하여 왔다. 특히 HMM은 논리적 배경이 정연하고, 여러 계층으로의 확대 및 변형의 가능성이 다양하여 현재 가장 효과적인 방법으로 되어 있다.

80년대 이래 재조명을 받기 시작한 인공신경망(또는 연결주의망)은 새로운 가능성으로 받아들여지고 있다. 인공신경망의 동작은 본질적으로 분류기적 기능에 기초하고 있으며 여러 계층에서 분류기능을 필요로 하는 음성인식 분야에서 관심의 대상이 될 수 밖에 없었다. 인공신경망의 매력중 하나는 '학습'의 과정을 통하여 인식 대상의 내재적 구조를 스스로 파악하게 한다는 것이다.

음성인식 시스템은 그림 3과 같이 기본적으로 음성으로 부터 음성패턴(단어, 음소)의 특징을 추출하여 기준 음성패턴을 만드는 훈련과정과 1가지의 음성이 입력되면 저장된 기준 음성패턴의 특징과 비교하여 가장 유사한 기준 음성패턴을 찾아내는 인식과정으로 나눌 수 있다.

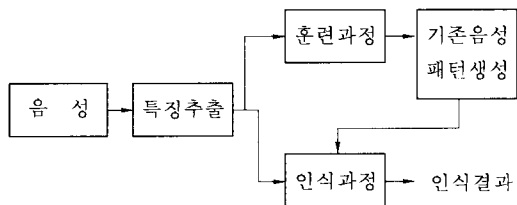


그림 3. 음성 인식 시스템

2.2 음성특징의 추출 또는 분석

음성신호로부터 적절한 음성특징을 추출해 내기 위한 연구는 크게 다음과 같은 세가지 분야로 나누어진다.

첫째는 음성의 발생과정을 모델링하는 방법이다. 발생과정을 모델링한다는 것은 사람이 음성을 발생시킬 때 각각의 발음이 성도로 부터 입술에까지 이르는 과정을 모델링하고 이것으로 부터 파라미터를 구하는 것이다. 이러한 방법의 대표적으로 잘 알려진 것이 선형예측부호화(LPC)기법이다.

둘째는 음성의 인지과정을 모델링하는 방법이다. 인지과정을 모델링한다는 것은 발음된 음성에 대하여 귀가 어떻게 그 특징을 구별해 내어 이를 두뇌로 전달 시키는지를 연구하고 이를 토대로 특징을 추출하는 방법이다. 가장 간단한 방법이 필터 뱅크(filter bank)를 이용하는 것이다.

세번째는 음성신호 자체를 주파수 영역에서 해석하는 방법이다. 음성신호 자체를 주파수 영역으로 변환하여 이것의 시간적인 변화를 관찰하여 이로부터 유용한 음성특징을 추출해 내는 방법이다. 이러한 음성특징으로는 에너지 정보, 포르만트(formant)정보, 피치(pitch)정보 등 여러가지가 있다.

① 선형 예측 부호화(LPC)

어떤 시각 $t=n$ 에 있어서의 음성값은 과거 몇개의 음성값에 계수를 곱하여 가산함에 따라 예측할 수 있다는 원리에 따라 이 예측계수를 산출하여 특징 파라미터로 사용하는 방법

② 필터 뱅크(filter bank)

선형 특성을 가지고 대역 필터군(bandpass filter group)의 음성 출력을 그대로 특징 파라미터로 사용하는 방법

③ 에너지(energy)정보

특정 구간내의 신호의 크기를 제공한 뒤 이를 더함으로써 계산, 이것을 유성음/무성음을 구분하는 자료로 삼음

④ 포르만트(formant)정보

음성(특히 모음)의 주파수 대역에서 에너지가 특히 큰 부분을 추출하여 공진 주파수, 이득, 대역폭 등으로 합성 또는 인식의 자료로 삼음

⑤ 피치(pitch)정보

음성 중 유성음은 성대를 이완진동적으로 개폐시켜 거의 주기적인 공기 펄스열로 생성되어 성도를 통해 외부로 방출된다. 이때 성대 진동의 기본주파수인 피치를 추출하여 특징파라미터로 사용한다.

2.3 음소의 분류 또는 인식

음성의 특징 파라미터를 이용하여 음소를 분류 또는 인식(classification)하는 과정은 크게 나누어 기존의 알고리즘의 접근방법과 신경회로망을 응용한 방법으로 나눌 수 있는데 전자의 경우 포먼트를 이용한 분류방법, DTW(Dynamic Time Warping)방법, HMM(Hidden Markov Model)을 이용한 방법등이 있으며, 후자는 MLP(Multilayer Perception), TDNN(Time Delay Neural Network), INN(Integrated Neural Network), TRACE model, SOFM 등이 있다.

기존의 알고리즘적 접근방법은 음성인식 시스템에서 많이 사용되는 기법이였다. 그러나, 이러한 방법들의 가장 큰 취약점은 우리가 음성에 대하여 하나의 모델 혹은 가정을 세우고, 이를 기반으로 분류방법을 도출하였다는 점이다. 만약 이러한 가정이 잘 맞는 경우에는 그 인식률이 높지만, 실제 음성들이 이러한 가정을 잘 만족하지 않는 경우에는 그 인식률이 많이 떨어지게 된다. 이러한 문제점을 해결할

수 있는 하나의 방법이 신경회로망을 이용한 방법이다. 여기에서는 많은 음성샘플로 부터 인식하는 방법을 학습을 통하여 스스로 배우도록 한 뒤, 그 학습결과를 바탕으로 시험 음성샘플을 분류하게 한다. 즉 음성에 대한 특별한 모델을 설정하지 않고 그 안에 숨은 규칙을 신경회로망으로 하여금 찾아내게 하는 방법이므로, 앞에서 설명한 기존의 방법이 갖는 취약점을 보완할 수 있다. 그러나 이 방법은 아직 원숙하지 않은 방법으로서 H/W 구현시의 어려움이 가장 큰 문제점으로 남아 있다.

① DTW(Dynamic Time Warping)

미지의 입력 특징 패턴의 시간열과 저장된 음성단위들의 대표패턴을 직접 비교하는 경우 발음속도의 차이에 따른 영향을 보상해 주기 위한 시간축 정규화 기술

② HMM(Hidden Markov Model)

통계적인 정보를 확률모델 형태로 저장하고 미지의 입력패턴이 들어오면 각각의 모델에서 이 패턴이 나올수 있는 확률을 계산함으로써 이 패턴에 가장 적합한 음성단위를 찾아내는 방법

③ MLP(Multi Layer Perception)

전처리 과정의 결과를 다층 퍼셉트론의 입력에 인가하고 학습을 시킴으로써 음성인식 능력을 가지도록 학습시키는 방법

④ TDNM(Time Delay Neural Model)

시간축 상에 퍼져있는 음소의 정보를 병렬 회로망을 이용하여 모으는데 착안한 계층적 신경회로망

⑤ INN(Integrated Neural Network)

구분하고자 하는 여러 음소를 유사한 정도에 따라 그룹화한뒤 각 프레임이 어느 그룹에 속하는지, 또 그 그룹중 어떤 음소인지를 따로 판단하여 종합하는 음성인식용

신경회로망

⑥ TRACE model

사람의 음성인식 과정을 여러개의 계층으로 나누고, 각 계층에서의 인식과정이 아래 계층으로 부터 전해지는 데이터와 위 계층으로 부터 보내지는 피드백 데이터를 이용하여 수행된다고 보는 신경회로망 모델

⑦ 코호넨 네트워크(Kohonen Network)

단지 두개의 층으로 이루어져 있으며, 자기 조직화를 통하여 비교적 단순하면서도 매력적인 결과를 도출해내는 모델로 백 프로 퍼게이션 모델과 같은 지도학습모델 보다 는 인간의 두뇌에 가깝다고 볼 수 있다.

피하고, 그러한 정보를 조직화하여 가능한 한 법칙을 추출하여 그것을 이용하는 데에 중점을 두고 있는 방식을 법칙합성방식이라 말하고 정보의 변환, 압축을 행하고 있지만, 인간의 음성의 정보를 직접 이용하고 있는 방식을 분석합성방식이라 부르고 있다.

현재 시장에 나와있는 음성합성용 LSI의 음성합성 알고리즘은 대부분 분석합성방식이다. 이 방식은 음성신호의 특징을 이용한 것으로 음성정보를 처리하여 정보량을 압축시키고 이 압축된 데이터로부터 음성신호를 재생하는 것이다. 이 때 음성신호를 시간영역에서 처리를 행하는 방법과 주파수영역으로 변환시켜 처리하는 방법이 있는데 전자를 파형부호화방식, 후자를 파라미터 합성방식이라 부른다. 두 방식 모두 정보량의 압축을 행하므로 분석처리 공정을 필요로 하여 압축 데이터를 만드는 시간이 필요하다.

III. 음성합성

3.1 음성합성 시스템

음성합성의 연구가 본격적으로 시작된 것은 1950년대에 이르러서이며 최초에는 실용적인 목적보다 오히려 음성의 정보요소에는 무엇이 있는가를 규명하기 위한 학구적인 연구가 주목적이었다. 따라서 음성합성 방법들은 실제적으로 다양하여 이를 계통적으로 정리하고 상호관계의 적절한 비교에 의해 장단점을 판정하는 것이 대단히 어렵다.

먼저 하드웨어적으로 분류하면 인간의 음성을 소재로 하지 않은 방식을 순수합성방식, 사전에 녹음, 기록되어 있는 인간의 음성을 직접 재생하는 방식을 녹음편집방식, 인간의 음성을 간접적인 소재로서 사용하고 있는 방식을 편집합성방식이라 부르고 있다. 이 관계는 직접 음성의 합성에 관계하는 하드웨어 모두가 처리장치인가, 처리와 기억의 병용인가, 모두 기억장치인가 하는 관점에서 볼 수도 있다.

소프트웨어적으로는 인간의 음성을 분석하여 얻어진 정보를 직접적으로 이용하는 것을

3.2 파형부호화방식

파형부호화방식은 사논의 표본화정리(sampling theorem)에 근거를 두고 신호파형을 2치 또는 다치 펄스열로 변환하는 방식이다. 역사적으로는 이 같은 파형부호화방식을 총칭해서 펄스부호변조방식이라 불렀으나 오늘날에는 사용하는 양자화방식에 따라 펄스부호변조(PCM)방식, 차동펄스부호변조(DPCM)방식, 델타변조(DM) 및 예측부호화(predictive coding)방식인 ADPCM방식, ADM방식 등으로 구분되고 있다. 파형부호화방식에는 위에서 열거한 시간영역부호기(time-domain coder) 이외에 sub-band coding 방식, ATC 방식 등의 주파수영역부호기(frequency-domain coder)가 있다. 그림 4는 파형부호화방식의 블록도로 신호파형에 표본화 및 양자화라고 하는 2가지 조작을 가하므로써 신호파형의 디지털 표현이 얻어지지만, 디지털 량으로 저장 또

는 전송하기 위해서는 부호화 조작에 의하여 2진수열로 변환시켜야 한다. 이와같이 압축 저장된 디지털 신호를 D/A변환기를 통하여 원음을 재생해내는 방식을 파형부호화에 의한 음성합성방식이라 할 수 있다.

⑦ BBC(Base Band Coding)

상기의 APC 방식에서 잔차신호에 대해서는 기저대역 성분의 내용을 기억하여 합성시에 그 고주파 성분을 재생함. (9.6kb/sec)

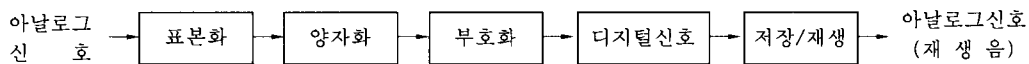


그림 4. 파형부호화 방식

① ADM(Adaptive Delta Modulation)

델타변조에 양자화 잡음을 감소시키기 위해 양자화 폭을 가변시킴(bit rate 24 ~ 32kb/sec)

② APCM(Adaptive PCM)

신호레벨에 따라 PCM 방식의 양자화 특성을 변화시킴(32kb/sec 정도)

③ ADPCM(Adaptive Differential PCM)

DPCM에 양자화 잡음을 줄이기 위해 양자폭을 미분치의 크기에 따라 가변(16 ~ 32kb/sec)

④ APC(Adaptive Predictive Coding)

신호의 선형예측계수와 잔차신호에 의해 신호파형을 재생(9.6kb/sec 정도)

⑤ SBCL(Sub Band Coding)

신호를 대역분할 한 후 각 채널당의 예측계수 및 잔차신호를 기억(9.6~16kb/sec)

⑥ ATC(Adaptive Transform Coding)

신호를 주파수 영역의 데이터로 변환하여 주파수 영역에서 양자화를 행하는 방식의 일종으로 스펙트럼의 형태에 각 계수의 양자화 비트의 할당, 양자화 폭은 변화시킴(9.6kb/sec)

⑧ CSM(Composite Sinusoidal Modeling)

음성과형을 10~30m/sec 정도의 단구간마다 수개의 정현파의 합으로서 표현하여 합성하는 것. 분석을 위한 처리량은 LPC 방법 또는 LSP방법과 같은 정도이지만 합성절차가 간단하고 저 비트 영역에 적합하다.(1.2~2.4kb/sec 정도)

3.3 파라미터 합성방식

아날로그 신호의 부호화에는 파형부호화 외에 신호가 그 발생원(source)에서 생성될 때의 메카니즘에 착안하여 부호화를 행하는 방법이 고안되어 있는데 이것을 원부호화(source coding)방식이라 한다. 이 원부호화방식의 구체적인 예가 음성의 생성 메카니즘에 관한 연구에서 얻어진 음성부호기(vocoder)이다. 음성정보의 전송에는 반드시 음성파형 그 자체를 전송할 필요는 없고 음성을 적당한 구간으로 분할해서 각 구간마다 유성음, 무성음의 구별, 만일 유성음이라면 그 피치, 성도 필터의 파라미터 여진 파라미터등의 소위 여진 및 성도정보를 수신측에 보내주면 수신측에서는 원음을 합성할 수가 있는데 이와같은 음성합성방식을 파라미터 합성방식이라 하며 그림 5는

이것을 블록도로 나타낸 것이다. 이때 음성의 특징 파라미터를 추출하는 방법은 음성인식에서 설명한 바와 같다.

60%의 정보량에서 같은 품질을 확보할 수 있다(2.4kb/sec).

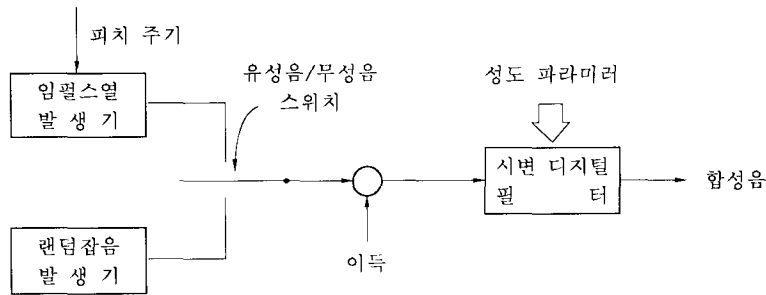


그림 5. 파라미터 합성 방식

① 채널 보코더(vocoder) 방식

신호를 15~20채널의 대역으로 분할하여 스펙트럼 포락 정도를 추출하고 음원의 유·무성의 판정 및 피치 추출을 하여 그 정보를 기억(2.4kb/sec)

② VEV(Voice Vocoder) 방식

고역의 스펙트럼 정보는 채널 보코더의 수법에 의해서 압축하지만 음원정보의 추출은 행하지 않고 기저대역성분을 그대로 부호화하여 기억한 후 합성시 고역부분의 음원을 재생(9.6kb/sec)

③ LPC(Linear Predictive Coding) 방식, PARCOR(Parital Auto Correlation) 방식

신호 LPC 계수(혹은 PARCOR 계수)와 음원 파라미터를 추출하여 기억함. 현재의 분석합성방법의 주류(4.8kb/sec)

④ LSP(Line Spectrum Pair) 방식

음성의 스펙트럼 포락을 비교적 소수의 선 스펙트럼 성분의 배치에 조밀에 따라 근사시킴. PARCOR 방식보다도 파라미터의 양자화 왜곡 혹은 보간왜곡에 의한 품질의 열화가 적고, 따라서 PARCOR 방식의 약

⑤ 소편편집방식

유성음의 1주기 정도의 파형(음성소편)을 잘라내어 기억하고 있다가 필요에 따라 접속하여 이용하므로써 분석합성의 절차를 줄인다. 음성소편으로는 자연음성에서 추출한 임펄스응답파형, 자연음성을 LPC 분석합성하여 얻어진 임펄스응답파형, 성도 모형에서 구한 임펄스응답파형, 포르만트에 대응한 감쇠정현파 등이 있다.

3.4 법칙합성(synthesis by rule) 방식

음성신호는 음운(phoneme) 등과 같은 이산적인 언어기호가 인간의 음성발생과정에 의해서 연속적인 파형으로 변환된 것이라고 볼 수 있다. 법칙합성방식은 이러한 변환을 기계적으로 실현하는 것이다. 여기서 입력은 음운 등의 기호계열로 주어지는데, 이는 기호계열로 표현되는 어떤 단어나 문장도 음성으로 변환이 가능하여 합성가능한 어휘는 무한하다는 것을 의미한다. 그림 6은 이것을 블록도로 나타낸 것이다.

실제적으로는 자연스럽고 육성에 가까운 합

성음을 얻는 데에는 곤란한 문제가 많다. 가장 문제가 되는 것은 변화법칙이다. 또한 자연음성에서는 음운간의 상호작용에 의한 음운특징 파라미터의 변형이 오는데 법칙합성에서는 이러한 조음결합의 처리를 어떻게 하는가가 법칙합성음의 품질을 좌우한다. 이러한 여러가지 어려운 점이 있음에도 불구하고 법칙합성 방식이 궁극적인 합성방식으로 판단되어 활발한 연구가 진행되고 있다.

라서 구성된 구조물이므로 그 음향적 성질을 한개의 기호로 표현할 경우 동일한 음이라도 단독으로 발성된 때와 단어나 문장중에서 발성되었을 때가 다르며 의미를 지닌 내용이나 발성자의 성별, 연령 및 발성시의 감정, 기분 등에 의해서도 복잡하게 변화한다. 이와같은 양태변동의 원인이 음성합성, 인식장치의 실용화를 어렵게 하고 있다.

이러한 어려움 속에서도 음성합성장치 및

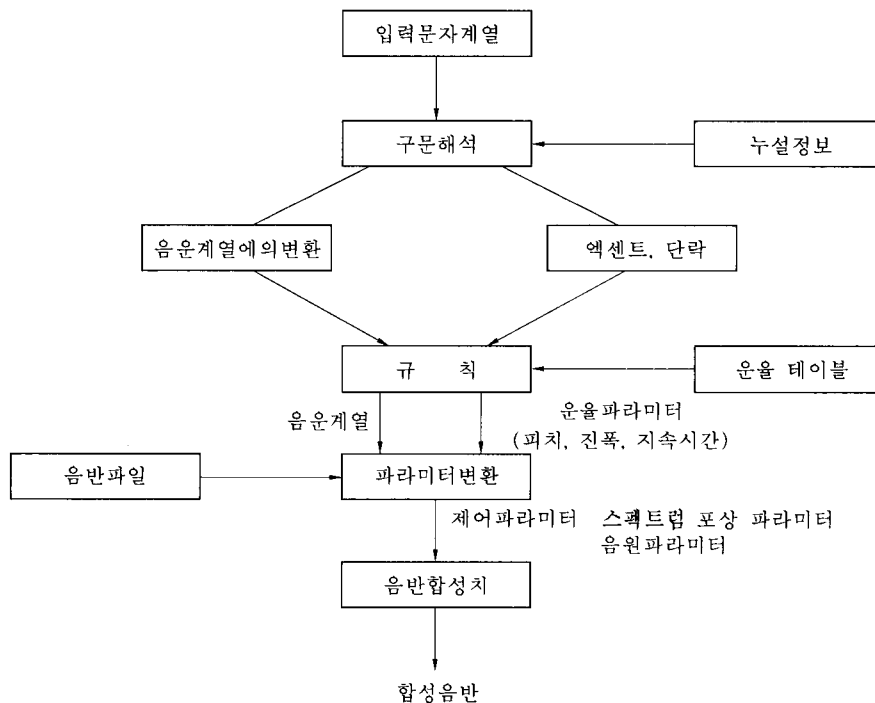


그림6. 법칙 합성 방식

IV. 결론

음성이 인간과 인간 사이의 통신매체로서 인간의 일상생활에 사용되고 있는 한 인간과 기계 또는 기계와 기계와의 대화의 방법을 생각하고 실현하려는 노력은 지극히 당연한 것이다.

그러나 음성은 단순한 음향적 부호가 아니라 다양한 구성요소가 언어로서의 규칙에 따

인식장치에 관한 연구가 구체화되기 시작하여 그 연구개발의 성과에 많은 기대가 모아지게 되었고, 음성정보처리분야가 크게 주목받고 있다. 그것은 다음과 같은 이유 때문이다.

- ① 음성정보처리, 특히 그 기초가 되는 음성 분석의 이론적 진보
- ② 이러한 이론적인 진보를 현실적으로 의미

가 있도록 하는 디지털 신호처리의 발달
 ③ LSI화에 의한 장치면에서의 비약적인 발전 등이다.

이 중에서 ①, ②는 음성정보처리 자체의 내부적인 진보, 발전에 의한 것이지만 ③은 현대 전자공학 기술전반의 진보, 발달에 의한 것으로 말하자면 외부적 요인이다. 이들 내·외의 요인이 조화를 이루어 머지않아 음성 타자기를 비롯한 모든 OA 기기에도 음성 입출력 기술이 도입될 것이다.

참 고 문 헌

- [1] J.W. Forgie and C.D. Forgie : Results obtained from a vowel Recognition Computer program, J.Acoust. Soc. Am., Vol. 31, pp. 1480-1489, 1959.
- [2] J.W. Cooley and J.W. Tukey : An Algorithm for the Machine Calculation of Complex Fourier Series, Math. of comp., Vol. 19, pp.297-301, 1965.
- [3] John Makhoul : Linear Prediction, A Tutorial Review Proc. IEEE, vol. 63, No.4, 1975.
- [4] H. Sakoe and S. Chiba : Dynamic Programing Algorithm Optimization for Spoken Word Recognition, IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-26, pp. 43-49, Feb. 1978.
- [5] L.R. Rabiner and B.H. Juang : an Introduction to Hidden markov Models, IEEE ASSP Magazine, pp. 4-16, Jan. 1986.
- [6] R.P. Lipmann : An Introduction to Computing with Neural Nets, IEEE ASSP Magazine, pp. 4-22, Apr. 1987.
- [7] D.E. Rumelhart and J.L. McClelland : Parallel Distributed Processing: Experiments in the Microstructure of cognition, vol. 1, pp.318-362, Cambridge, M.I.T. Press, MA 1986.
- [8] 이병수 : 신경회로망을 이용한 음소인식, 순천향대학 논문집, 제17권 제3호, 1994.
- [9] 강민호외 6인 : 전기통신기술개론, 청문각, 1990.
- [10] 박두석외 1인 : 사무정보기기, 도서출판 동광, 1996.

□ 著者紹介

이 병 수(李秉洙)

1975년 2월 한양대학교 전자공학과 졸업

1977년 2월 건국대학교 대학원(공학박사)

1982년 3월 ~ 1988년 2월 대우공업전문대학 전자통신과

1988년 3월 ~ 현재 순천향대학교 정보통신공학과 부교수

※ 관심 분야 :