

(19) 대한민국특허청(KR)
(12) 특허공보(B1)

(51) Int. Cl.⁵
G10L 7/02

(45) 공고일자 1994년04월04일
(11) 공고번호 특1994-0002854

(21) 출원번호	특1991-0019617	(65) 공개번호	특1993-0010850
(22) 출원일자	1991년11월06일	(43) 공개일자	1993년06월23일
(71) 출원인	한국전기통신공사 이해욱 서울특별시 종로구 세종로 100번지		
(72) 발명자	이종락 서울특별시 노원구 월계동 13번지 박용규 경기도 부천시 춘의동 142-14		
(74) 대리인	이병호, 정상구		

심사관 : 강해성 (책자공보 제3590호)

(54) 음성 합성시스템의 음성단편 코딩 및 그의 피치조절 방법과 그의 유성을 합성장치

요약

내용 없음.

대표도

도1

명세서

[발명의 명칭]

음성 합성시스템의 음성단편 코딩 및 그의 피치조절 방법과 그의 유성을 합성장치

[도면의 간단한 설명]

제1도는 음성단편조합방식 무제한 음성합성시스템의 구성도.

제2도는 음성단편조합 서브시스템의 구성도.

제3a 내지 t도는 본발명에 의한 주기파형분해 및 단위파형재배치 방식의 설명도.

제4도는 본발명에 의한 주기파형분해 방법의 블록선도.

제5a 내지 e도는 블라인드 디콘벌루션 과정의 설명도.

제6a 및 b도는 음성단편저장부에 저장되는 음성단편정보의 코드포맷.

제7도는 본발명에 의한 유성음 합성부의 구성도.

제8a 및 b도는 본발명에 의한 지속시간 및 피치조절 방법의 설명도.

* 도면의 주요부분에 대한 부호의 설명

- | | |
|--------------------|----------------|
| 1 : 음성학적 전처리 서브시스템 | 2 : 운율발생 서브시스템 |
| 3 : 음성단편조합 서브시스템 | 4 : 음성단편 선정부 |
| 5 : 음성단편 저장부 | 6 : 유성음 합성부 |
| 7 : 무성음 합성부 | 8 : D-A 변환부 |
| 9 : 디코딩부 | 10 : 지속시간 조절부 |
| 11 : 피치조절부 | 12 : 에너지 조절부 |
| 13 : 파형조립부 | |

[발명의 상세한 설명]

본발명은 음성합성시스템(speech synthesis system)과 그의 합성방법에 관한 것으로서, 특히 합성음(synthesized speech)의 음질을 현저히 개선할 수 있게하는 음성단편코딩 및 피치조절방법에 관한 것이다.

문자열(text)을 음성으로 변환함으로써 무제한의 어휘를 합성할 수 있게 하는 무제한 음성합성시스템(text-to-speech synthesis system)을 실현시키기 위한 음성합성방식에는 몇 가지가 있으나, 그중 실현이 용이하여 가장 보편적으로 쓰이고 있는 방식은 사람의 음성을 채취(sampling)하여 반음절(demisyllable)이나 다이폰(diphone)과 같은 음성단위(phonetic unit)로 분할하여 얻은 짧은 음성단편(speech segment)들을 코딩(coding)하여 메모리에 저장해 두었다가, 문자열이 입력되면 그것을 발음기호(phonetic transcription)로 바꾼 후 그에 맞는 음성단편들을 메모리에서의 차례대로 꺼내어 재생(decoding)시킴으로써 입력문자열(input text)에 해당하는 음성을 합성해내는 음성단편합성(segmental synthesis) 방식인데, 일명 연쇄합성(synthesis-by-concatenation) 방식이라고도 한다.

이와같은 음성단편합성방식에 있어서 합성음의 음질을 좌우하는 가장 중요한 요소중하나인 음성단편의 코딩방법이다. 기존의 음성단편합성방식을 위한 음성합성시스템에 있어서 음성단편을 저장하기 위한 음성코딩(speech coding)방식으로서 음질이 낮은 보코딩(vocoding) 방식을 주로 사용하고 있는데, 이것이 합성음의 음질을 나쁘게하는 가장 중요한 원인중의 하나가 되고 있다. 이하 기존의 음성단편 코딩방법에 대해 간략히 살펴보기로 한다.

음성코딩방식은 음질이 좋은 파형코딩(waveform coding) 방식과 음질이 낮은 보코딩 방식으로 크게 나누어 볼 수 있다. 파형코딩방식은 음성파형(speech waveform)을 충실히 그대로 전달하려는 방법이기에 때문에 피치주파수(pitch frequency)와 지속시간(duration)을 변화시키는 것이 극히 어려워 음성합성시 억양과 발성속도등을 조절할 수 없을 뿐아니라, 음성단편간을 원활하게 접속(concatenation)시킬 수 없어 근본적으로 음성단편의 코딩용으로 적합하지 않다.

이에 비해, 분석-합성방식(analysis-synthesis method)이라고도 불리는 보코딩 방식을 쓸 경우에는 피치패턴(pitch pattern)과 음성단편의 지속시간(duration)을 임의로 변화시킬 수 있고, 스펙트럼포락(spectral envelope)추정 파라미터(estimation parameter)의 내삽(interpolation)에 의해 음성단편간을 매끄럽게 접속시킬 수 있어 무제한 음성합성용 코딩 수단으로서 적합하기 때문에 현재 거의 대부분의 음성합성 시스템에 선형예측코딩(Linear Predictive Coding : LPC) 또는 포먼트 보코딩(formant vocoding)등의 보코딩 방식이 채택되고 있다. 그러나, 보토딩 방식으로는 음성을 코딩할 경우에는 재생되는 음성의 음질이 낮기 때문에, 저장된 음성단편을 디코딩(decoding)하여 연쇄(concatenation)시킴으로써 만들어지는 합성음도 이들 보토딩 방식이 가지는 음지르이 한계 이상 좋은 음질을 가질수 없다.

보코딩 방식이 가지는 음질의 한계를 극복하기 위한 기존의 시도는 여기신호로 사용되는 임펄스열(impulse train)을 보다 덜 인위적인 파형으로 대체하는 것이었다. 그 한가지는 임펄스 대신 그보다 첨예도(peakiness)가 낮은 파형, 예컨대 삼각파나 반원파 또는 성문펄스(glottal pulse)와 유사한 모양의 파형을 사용하는 것이었다. 다른 한가지는 역필터링(inverse filtering)에 의해 얻어진 잔여신호(residual signal)의 피치펄스들 중의 어느하나 또는 몇개를 표준피치펄스(sample pitch pulse)로 선정하여 한 표준펄스를 전 시간 또는 상당히 긴 시간구간에 걸쳐 임펄스 대신 사용하는 것이었다. 그러나 임펄스를 다른 파형의 여기펄스로 대체하려는 기존의 이와같은 시도는 음질을 전혀 개선하지 못했거나 단지 약간(slightly) 개선시킬 수 있었을 뿐이며 자연음에 가까운 음질을 가진 합성음을 얻을 수는 없었다.

본발명의 음질이 좋으나 피치조절이 불가능한 파형코딩방식과 피치조절이 가능하나 음질이 낮은 보코딩방식의 장점을 결합시켜 얻어진, 음질이 좋으면서도 피치조절을 가능케하는 새로운 음성단편코딩 방법을 제공함으로써 인간의 발성과 같은 정도의 자연성(naturalness)과 명료성(intelligibility)을 가진 고품질의 음성을 합성하는데 목적이 있다.

본발명의 특징은 원음성(original speech) 중의 유성음 구간의 신호를 각 성문펄스(glottal pulse)에 의해 만들어지는 한 주기분 음성파형에 해당하는 단위파형(unit waveform or wavelet)들로 분해하고 그것들을 각각 코딩하여 저장하는 코딩방식인 '주기파형분해방식'(periodic waveform decomposition method)과, 저장된 단위파형들중 배치시키고자 하는 위치에 가장 가까운 단위파형을 선택하여 디코딩(decoding)하고 그것들을 서로 중첩(superposition)시킴으로써 원음성의 음질을 그대로 가지면서도 음성단편의 지속시간(duration)과 피치주파수(pitch frequency)를 임의대로 조절할 수 있게 하는 파형합성방식인 '시간왜곡식 단위파형 재배치방식'(time warping-based wavelet relocation method)을 사용하는데 있다.

이하, 첨부된 도면을 참조하여 본발명을 상세히 설명하기로 한다.

음성단편합성방식 무제한 음성합성 시스템의 일반적인 구조는 제1도에 도시된 바와같이 음성학적전처리 서브시스템(phonetic preprocessing subsystem)(1), 운율발생 서브시스템(prosodics generation subsystem)(2), 그리고 음성단편조합 서브시스템(speech segment concatenation subsystem)(3)이라는 3개의 서브시스템으로 구성된다. 키보드, 컴퓨터 또는 다른 어떤 시스템으로부터 무제한 음성합성 시스템으로 문자열(text)이 입력되면 음성학적전처리 서브시스템(4)은 그것의 구문(syntax)을 분석한 후 음운변동규칙(phonetic recoding rule)을 적용하여 문자열을 발음기호(phonetic transcriptive symbol)의 열(string)로 바꾸는데, 운율발생 서브시스템(2)은 그것에 적당한 억양(intonation)과 강세(stress)를 붙일 수 있도록 구문분석정보(syntactic analysis data)를 발생시켜 음성단편조합 서브시스템(3)에 제공한다. 운율발생 서브시스템(2)은 또한 각 음소의 지속시간(duration)에 관한 정보를 음성단편조합 서브시스템(3)에 제공한다. 이들 3가지의 운율정보(prosodic data)는 발음기호열과 별도로 음성단편조합 서브시스템(3)으로 전해질 수도 있으나, 대개 발음기호열내에 삽입되어서 음성단편조합 서브시스템(3)으로 보내어진다. 음성단편 조합 서브시스템(3)에서는 발음기호열에 의거하여 적합한 음성단편을 메모리(도시안됨)에서 차례로 꺼내어 재생시킴으로써 연속된 음성을 만들어 낸다. 이때 음성단편조합 서브시스템은 운율정보에 의하여 각 음성단

편의 에너지(세기)와 지속시간(duration) 그리고 피치주기(pitch period)를 조절(control)함으로써 운율발생단계가 의도한 억양, 강약 및 발생속도를 가진 합성음성을 만들어 낼수 있게 된다.

본발명은 이와같은 음성단편조합 서브시스템에 있어서 음성단편을 저장하기 위한 코딩방법을 개선함으로써 기존의 코딩방법에 의한 합성음성에 비해 음질이 현저히 개선되도록 하는 것이 목적이다. 이 하 음성단편조합 서브시스템의 동작을 제2도에 도시된 내부 구성도에 의하여 설명하기로 한다.

음성단편조합 서브시스템(3)으로 발음기호열과 운율정보가 입력되면 음성단편선정부(speech segment selection block)(4)는 그 발음기호열을 음성으로 합성해 내는 데 필요한 음성단편들을 차례로 선정 하게 되는데, 그결과 선정된 음성단편들의 기호(index)가 출력되어 음성단편저장부(speech segments storage block)(5)로 입력된다.

상기 음성단편 저장부(5)에는 음성단편들이 후술할 본 발명에 의한 방법으로 코딩되어 메모리에 저장되어 있는데, 상기 음성단편선정부(4)로부터 선정된 음성단편의 기호가 입력되면 음성단편 저장부 (5)는 해당 음성단편정보를 꺼내어(fetch) 그것이 유성음 또는 유성마찰음(voiced fricative)이면 유성음합성부(voiced sound synthesis block)(7)로, 무성음이면 무성음합성부(unvoiced sound synthesis block)(6)로 각각 보내어져 합성되게 한다. 그 결과 만들어진 디지털 합성음성신호 (digital synthesized speech signal)는 D-A변환기(digital-to-analog converter)(도시안됨)와 애널 로그 저역통과필터(analog low-pass filter)(도시안됨) 그리고 애널로그 증폭기(analog amplifier) (도시안됨)로 구성된 D-A 변환부(D-A conversion block)(8)로 보내어져서 애널로그 신호로 변환됨으 로써 원하는 합성음성(synthesized speech sound)이 얻어지도록 되어 있다. 상기 유성음합성부(6)와 무성음합성부(7)는 음성단편들을 연쇄시킬때 운율정보를 이용하여 음성단편의 지속시간과 세기 그리 고 피치주파수를 적절히 조절함으로써 합성음성이 제1도의 운율발생 서브시스템(2)에 의해 의도된 대로의 운율을 가지게 한다.

상기 음성단편저장부(5)에 저장된 음성단편을 준비하는 과정은 다음과 같다. 먼저 사용할 합성단위 (synthesis unit)를 결정한다. 합성단위로는 음소(phoneme), 변이음(allophone), 다이폰(diphone), 음절(syllable), 반절음(demisyllable), CVC, VCV, CV, VC단위(여기서 C는 자음, V는 모음음소를 표 시) 또는 이들의 혼합사용 등이 있는데, 이 중에서 음성단편합성방식의 음성합성시스템에서 현재 가 장 많이 채택되고 있는 합성단위는 다이폰과 반음절이다. 그 합성단위 집합을 이루는 원소(element)의 각각에 해당하는 음성단편을 사람이 실제로 발음한 음성샘플로부터 분할(segmentation)한다. 따라서 합성단위집합의 원소의 수와 음성단편의 갯수는 같게된다. 예를들면, 영어에 있어서 반음절 을 합성단위로 택할 경우 반음절의 종류가 약 1000가지이므로 음성단편의 수도 약 1000개가 된다. 일반적으로 이와같은 음성단편은 무성음 및 유성음 구간별로 다시 분할하여 얻은 무성(unvoiced) 및 유성 음성단편(voiced speech segment)을 기본적인 합성 단위로 사용한다.

무성 음성단편은 제2도의 무성음 합성부(7)에서 재생(decoing)되는데, 무성음의 재생시에는 인위적 인 백색 랜덤잡음신호(white random noise signal)를 합성필터의 여기신호(excitation signal)로 사 용하여도 재생된 음성의 음질이 나빠지지 않는 것으로 알려져 있다. 그러므로 무성음단편의 코딩과 재생에 있어서는 백색잡음을 여기신호로 쓰는 기존의 보코딩 방식을 그대로 사용하면 된다. 예컨대 무성음 합성시에는 난수(random number)발생 알고리즘에 의해 백색잡음신호를 발생시켜 여기신호로 쓰든가, 미리 발생시켜 얻은 백색잡음신호를 메모리에 저장시켰다가 합성시 꺼내서(fetch)쓰든가, 또는 실제음성의 무성음 구간을 역스펙트럼포락필터(inverse spectral envelope filter)(도시안됨)로 필터링(filtering)해서 얻은 잔여신호(residual signal)를 메모리에 저장시켜 두었다가 합성시 꺼내 서 쓰는 방법중 어느 방법이나 쓸수 있다. 만일 무성음성단편의 지속시간을 변화시킬 필요가 없다고 하면 무성음 부분을 PCM(Pulse Code Modulation)이나 ADPCM(adaptive Differential PCM)과 같은 파 형코딩방법에 의해 코딩하여 저장해 두었다가 합성시 디코딩하여 그대로 쓰는 극히 간단한 코딩방법 도 쓰일수 있다.

본발명은 합성음의 음질을 좌우하는 유성 음성단편의 코딩과 합성방법에 관한 것으로 그 기능을 수 행하는 제2도중의 음성단편 저장부(5)와 유성음합성부(6)에 대해 중심으로 설명하기로 한다.

음성단편저장부(5)의 메모리(도시안됨)내에 저장되어 있는 음성단편들 중에서 유성 음성단편들은 본 발명에 의한 주기파형 분해방식(periodic waveform decomposition method)에 의해 미리 피치주기별 성분인 단위파형들로 분해되어 저장되어 있다. 유성음 합성부(6)는 시간왜곡식 단위파형 재배치 방 식(time warping-based wavelet relocation method)에 의해 그 단위파형들을 적절히 선정 및 배치함 으로서 원하는 피치와 지속시간을 가진 음성을 합성한다. 이들 방식의 원리를 첨부된 도면에 의해 설명하기로 한다.

유성음성 $s(n)$ 은 성대에서 발생하는 주기적인 성문파(glottal wave)가 구강(oral cavity)과 인두강(pharyngeal cavity) 그리고 비강(nasal cavity)으로 이루어진 음향학적 성도필터(vocal tract filter) $V(f)$ 를 거치면서 필터링됨으로써 만들어진 주기적 신호이다. (여기서 성도필터 $V(f)$ 에는 입 술복사현상에 의한 주파수 특성이 포함된 것으로 생각한다.) 그것의 스펙트럼 $S(f)$ 는 주파수 f 에 대 해 빨리 변하는 미세구조(fine structure)와 천천히 변하는 스펙트럼포락(spectral envelope)을 특 징으로 하는데, 전자는 유성음성신호의 주기성(periodicity)에 기인하는 것이며 후자는 성문펄스(glottal pulse)의 스펙트럼과 성도필터의 주파수 특성을 반영한다. 유성음성의 스펙트럼 $S(f)$ 는 피 치주파수 F_0 의 정수배마다 존재하는 고조파성분(harmonic component)으로 인한 임펄스열 형태의 미세 구조가 스펙트럼포락함수 $H(f)$ 에 의해 변조되어 곁해진 것과 같은 형태를 취하고 있다. 따라서 유성 음 $s(n)$ 은 그것의 스펙트럼포락 함수 $H(f)$ 와 동일한 주파수응답 특성을 가진 시변필터에 그것과 동 일한 주기와 평탄한 스펙트럼포락을 가진 "주기적 피치펄스열신호"(periodic pitch pulse train signal) $e(n)$ 이 입력되었을 때의 출력신호로 볼수 있다. 이것을 시간 영역(time domain)에서 본다면 유성음 $s(n)$ 은 필터 $H(f)$ 의 임펄스 응답 $h(n)$ 과 주기적 피치펄스 열신호 $e(n)$ 과 콘볼루션이다.

이와같은 시변필터의 주파수응답 특성인 $H(f)$ 는 유성음 $s(n)$ 의 스펙트럼포락함수에 해당하기 때문에 이 시변 필터를 스펙트럼포락 필터(또는 합성필터(synthesis filter)라고도 함)라고 한다. 제3a도는

성문파형(glottal waveform)의 3주기분 신호를 도시한다. 성문파형을 구성하는 성문펄스(glottal pulse)들의 파형은 서로 유사하나 완전히 동일하지는 않으며, 인접한 성문펄스 간의 시간간격도 대략 같으나 완전히 같지는 않은 것이 보통이다. 전술한 바와같이, 제3c도의 유성음성파형 $s(n)$ 은 제3a도에 도시한 성문파형 $g(n)$ 이 성도필터 $v(f)$ 에 의해 필터링됨으로써 생성된다. 성문파형 $g(n)$ 은 시간적으로 서로 구분되는 성문펄스 $g_1(n)$, $g_2(n)$, $g_3(n)$ 등으로 이루어져 있으며 그것들이 성도 필터 $V(f)$ 에 의해 각각 필터링되면 제3b도에 도시된 바와같은 단위파형(wavelet) $s_1(n)$, $s_2(n)$, $s_3(n)$ 등이 만들어진다. 제3c도의 음성파형 $s(n)$ 은 이들 단위파형들이 중첩됨으로써 형성된 것이다. 만일 유성 음성파형 $s(n)$ 을 분해하여 그것을 구성하는 단위파형들을 찾아낼 수 있다고 하면 그것들을 저장하였다가 합성시 그것들의 세기와 그것들 간의 시간간격을 바꿈으로써 원하는 임의의 액센트와 억양을 가진 음성을 합성할 수 있다는 것이 본 발명의 기본개념이다.

제3b도에 도시된 바와같이 유성음성파형 $s(n)$ 은 시간간격에서 서로 중폭(overlap)되는 단위파형들이 중첩되어 만들어진 것이기 때문에 음성파형 $s(n)$ 으로부터 그것을 구성하는 단위파형들을 쉽사리 다시 분리해낼 수는 없다. 각 주기분 파형들이 시간영역에서 서로 겹치지 않도록 하기 위해서는 그것들이 에너지가 한 시점에서 집중된 첨예한(Spikey) 파형으로 되어야 한다. 이와같은 파형의 첨예함은 주파수 영역(frequency domain)에서 평탄한(flat)스펙트럼포락을 가지는 것을 의미한다. 유성 음성파형 $s(n)$ 이 주어졌을 때 그것의 스펙트럼 $S(f)$ 의 포락(envelope)을 추정하여 그 포락함수 $H(f)$ 의 역수를 주파수 특성으로 가지는 역스펙트럼포락 필터 $1/H(f)$ 에 입력시키면 제3f도에 도시된 바와같은 스펙트럼 포락이 평탄한 주기적 피치펄스 열신호 $e(n)$ 을 출력으로 얻을 수 있다. 제3f도에 도시된 바와같이 주기적 피치펄스열신호 $e(n)$ 을 구성하는 각 주기분 피치펄스파형들은 시간영역에서 서로 겹치지 않으므로 분리해낼 수 있다. 분리된 "한주기분 피치펄스신호" $e_1(n)$, $e_2(n)$ 등을 스펙트럼포락 필터 $H(f)$ 에 다시 입력시키면 제3b도에 도시된 단위파형 $s_1(n)$, $s_2(n)$ 등을 각각 구할 수 있다는 것이 주기파형분해방식의 원리이다.

제4도는 유성음 음성단편을 단위파형들로 분해하는 본발명에 의한 주기파형분해 방법의 블록선도로서, 애널로그 형태의 유성음성 신호 또는 악기음 신호를 저역통과 필터(low-pass filter)로써 대역제한(band-limit)하고 A-D 변환(analog-to-digital conversion)시킴으로써 얻어진 디지털신호(digital signal) $s(n)$ 을 여러 비트(bit)씩 묶어서 PCM(Pulse Code Modulation) 코드(code) 포맷으로 디스크 등에 저장시켰다가 꺼내어 처리한다. 주기파형 분해방법에 의한 단위파형준비 과정의 첫 단계는 주기적 신호 $s(n)$ 을 그것의 스펙트럼포락함수 $H(f)$ 의 시간영역 함수인 임펄스응답 $h(n)$ 과, $s(n)$ 과 주기가 같고 평탄한 스펙트럼포락을 가진 주기적 피치펄스열신호 $e(n)$ 으로 디콘볼루션시키는 "블라인드 디콘볼루션(blind deconvolution)"이다.

전술한 바와같이, 블라인드 디콘볼루션을 수행하기 위해서는 $s(n)$ 으로부터 그것의 스펙트럼포락함수 $H(f)$ 를 추정하는 스펙트럼 추정기법이 필수적이다.

기존의 스펙트럼 추정기법은 분석구간(analysis interval)의 길이에 따라 블록별 분석법, 피치동기식 분석법 그리고 순차적 분석법의 3가지로 크게 분류될 수 있다. 블록별분석법(block analysis method)은 음성신호를 10-20ms 정도의 일정지속시간의 블록들로 분할하여 각 블록내에 존재하는 일정갯수의 음성샘플에 대하여 분석을 행하는 방법으로서, 블록당 한 세트(10-16개 정도)씩의 스펙트럼포락 파라미터를 구하는 방법인데, 준동형 분석법과 블록단위의 선형예측 분석법이 대표적이다. 피치동기식 분석법(pitch-synchronous analysis method)은 제3c도와 같이 피치주기를 단위로 분할된 각 주기분 음성신호에 대해 분석을 행하여 주기당 한 세트(set)씩의 스펙트럼포락 파라미터를 얻는 방법인데, '합성에 의한 분석법'(analysis-by-synthesis method)이나 피치동기식선형 예측분석법이 대표적이다. 순차적분석법(sequential analysis method)은 매 음성 샘플마다 스펙트럼포락 파라미터를 얻는 방법인데, 적응여파법(adaptive filtering)의 일종인 회귀최소자승법(Recursive Least Squares)등이 대표적이다.

제3d도에 대표적으로 순차적 분석법에 의해 구해진 스펙트럼포락 파라미터 세트인 14개의 반사계수 k_1, k_2, \dots, k_{14} 중 처음 4개의 시간적 변화를 예시하였다. 이 그림에서 알 수 있는 것과 같이 스펙트럼포락 파라미터들의 값은 조음기관의 연속적인 움직임으로 인해 계속 변화한다. 이것은 스펙트럼포락 필터의 임펄스응답 $h(n)$ 이 계속적으로 변함을 의미한다. 여기서는 설명의 편의상 한주기 구간 내에서는 $h(n)$ 이 변하지 않는다고 가정하고 제3e도에 나타난 바와같이 첫째, 둘째, 셋째주기 동안의 $h(n)$ 을 차례로 $h(n)_1, h(n)_2, h(n)_3$ 으로 각각 표시하기로 한다.

준동형 분석법에 의해 구해지는 파라미터인 렉스트럼(cepstrum) $c_l(i)$, 회귀최소자승법 또는 선형예측분석법에 의해 구해지는 파라미터인 예측계수(prediction coefficient) 집합 $\{a_i\}$ 또는 반사계수(reflection coefficient) 집합 $\{k_i\}$ 또는 그것의 변형(transform)인 대수 면적비(Log Area Ratio), 선스펙트럼쌍(Line Spectrum Pair)등 각종 스펙트럼 추정기법에 의해 얻어지는 스펙트럼포락 파라미터들은 그것들으로써 스펙트럼포락 필터의 주파수 특성 $H(f)$ 나 임펄스응답 $h(n)$ 을 만들어 낼 수 있기 때문에 $H(f)$ 나 $h(n)$ 과 동일하게 취급될 수 있다. 따라서 이후부터는 임펄스응답까지도 포함하여 스펙트럼포락 파라미터(또는 성도파라미터라고도 함)으로 칭하기로 한다.

제5a 내지 e도는 블라인드 디콘볼루션의 구체적인 방법들을 예시하고 있다.

제5a도는 선형예측 분석법 또는 회귀최소자승법에 의한 블라인드 디콘볼루션(blind Deconvolution) 방법을 도시한다. 제3c도와 같은 유성 음성파형 $s(n)$ 이 주어지면 전술한 선형예측 또는 회귀최소자승법을 이용하여 스펙트럼포락 필터의 주파수특성 $H(f)$ 또는 임펄스응답 $h(n)$ 을 나타내는 스펙트럼포락 파라미터들인 예측계수들 $\{a_1, a_2, \dots, a_N\}$ 또는 반사계수들 $\{k_1, k_2, \dots, k_N\}$ 을 구한다. 예측의 차수(order) N 은 보통 10-16정도로 충분하다. 이들 스펙트럼포락 파라미터들은 이용하면 스펙트럼포락 필터의 주파수특성 $H(f)$ 의 역수인 $1/H(f)$ 를 주파수 특성으로 가지는 역스펙트럼포락 필터(inverse spectral envelope filter)(간단히 역필터(inverse filter)라고도 함)를 쉽게 구성할 수 있다. 선형예측분석법 또는 RLS 기법에서 선형예측 오차필터(Linear Prediction error filter)라고도 불리는 이 역 스펙트럼포락 필터에 유성 음성파형 $s(n)$ 을 입력시키면 예측오차신호(prediction error signal) 또는 잔여신호(residual signal)로 불리는 평탄한 스펙트럼포락을 가진 제3f도와 같은 형태

의 주기적 피치펄스열신호를 출력으로 얻을 수 있다.

제5b 및 c도에 도시된 방법은 준동형분석법(homomorphic analysis method)을 이용하여 블라인드 디콘벌루션 방법인데, 제5b도에 쿼퍼런시 분할(quefrequency division)에 의한 방법을, 제5c도에 역필터링(inverse filtering)에 의한 방법을 각각 보였다.

먼저 제5b에 대해 설명하면 다음과 같다. 유성음성신호 $s(n)$ 에 10-20ms 정도의 지속시간을 가진 해밍창함수(Hamming window)와 같은 양단감쇠형 창함수(tapered window function)를 곱하여 한 블록(block)의 분석용 음성샘플을 얻고 그것을 제5d도와 같이 DFT(Discrete Fourier Transform), 복소대수함수(complex logarithm), 역(inverse) DFT로 이어지는 일련의 준동형처리(homomorphic processing) 과정을 통하여 켈스트럼 열(cepstral sequence) $c(i)$ 를 얻게 된다. 켈스트럼(cepstrum) $c(i)$ 는 시간을 유사한 단위인 쿼퍼런시(quefrequency)에 대한 함수인데, 원점을 중심으로 위치하는 낮은 쿼퍼런시 켈스트럼(low-quefrequency cepstrum) $c_L(i)$ 는 유성음성 $s(n)$ 의 스펙트럼포락을 나타내며, 높은 쿼퍼런시 켈스트럼(high-quefrequency cepstrum) $c_H(i)$ 는 주기적 피치펄스열신호 $e(n)$ 을 나타내는데, 이들은 쿼퍼런시 영역(quefrequencydomain)에서 서로 분리될 수 있다. 즉 켈스트럼 $c(i)$ 에 낮은 쿼퍼런시 창함수(low-quefrequency window)와 높은 쿼퍼런시 창함수(high-quefrequency window)를 곱함으로써 $c_L(i)$ 과 $c_H(i)$ 를 각각 구할 수 있다. 이들을 제5e도와 같은 과정으로 각각 역준동형처리(inverse homomorphic processing)함으로서 임펄스응답 $h(n)$ 과 피치펄스열신호 $e(n)$ 을 얻는다. 이 경우에 $c(i)$ 를 역준동형처리하면 피치펄스열신호 $e(n)$ 이 바로 구해지지 않고 시간창함수 $w(n)$ 이 곱해진 한 블록의 피치펄스열신호가 구해지기 때문에 $w(n)$ 의 역수에 해당하는 역시간창함수 $1/w(n)$ 을 다시 곱하여 $e(n)$ 을 얻는다.

제5c도의 방법은 제5b도와 같으나, 단지 주기적 피치펄스열신호 $e(n)$ 을 구하는데 있어서 $c_H(i)$ 대신 $c_L(i)$ 를 이용하는 점이 다르다. 즉 $c_L(i)$ 의 음수를 취하여 만들어진 $-c_L(i)$ 를 역준동형 처리하면 $h(n)$ 의 주파수특성 $H(f)$ 의 역수인 $1/H(f)$ 에 해당하는 임펄스응답 $h^{-1}(n)$ 이 얻어진다는 성질을 이용하여, 역스펙트럼포락 필터, 즉 $h^{-1}(n)$ 을 임펄스응답으로 하는 FIR(finite-duration impulse response) 필터(filter)를 구성하여 그것에 창함수가 곱해지지 않은 원음성신호 $s(n)$ 을 입력시킴으로써 출력으로 주기적 피치펄스열신호 $e(n)$ 을 얻는 방법이다.

이 방법은 제5a도와 근본적으로 같은 역 필터링 방법으로서, 차이점은 제5c도의 준동형 분석에서는 역스펙트럼포락 필터의 임펄스응답 $h^{-1}(n)$ 를 구하여 역스펙트럼포락 필터 $1/H(f)$ 를 구성하는데 비해 제5a도에서는 선형예측 분석법에 의해 구한 예측계수 $\{a_i\}$ 또는 반사계수 $\{k_i\}$ 로써 직접 역스펙트럼포락 필터 $1/H(f)$ 를 구성할 수 있다는 점이다.

준동형 분석법에 의한 블라인드 디콘벌루션에 있어서는 전술한 바와같이 스펙트럼포락 파라미터로서 임펄스응답 $h(n)$ 을 사용할 수도 있고, 제5b도 및 제5c도에 점선을 나타낸 바와 같이 낮은 쿼퍼런시의 켈스트럼 $c_L(i)$ 를 사용할 수도 있다. 임펄스응답 $\{h(0), h(1), \dots, h(N-1)\}$ 을 사용할 때는 N이 90-120정도로서 파라미터의 수가 매우 많은데 비해 켈스트럼 $\{c_L(-N), c_L(-N+1), \dots, 0, \dots, c_L(N)\}$ 을 사용할 때는 N이 25-30정도로서 파라미터의 갯수는 50-60개로 줄어든다.

이상과 같이 제5a 내지 5e도의 과정에 의해 유성음성파형 $s(n)$ 은 스펙트럼포락 필터의 임펄스응답 $h(n)$ 과 주기적 피치펄스열신호 $e(n)$ 으로 디콘벌루션된다.

제4도의 블라인드 디콘벌루션 과정에 의해 피치펄스열신호와 스펙트럼포락 파라미터들을 얻었으면 다음으로 이포크 검출 알고리즘(epoch detection algorithm)과 같은 시간영역(time-domain)에서의 피치펄스 위치검출 알고리즘을 이용하여 주기적 피치펄스열신호 $e(n)$ 이나 음성파형 $s(n)$ 으로부터 피치펄스들의 위치 P_1, P_2 등을 구한다. 그다음에 제3f도에 예시한 것과 같이 피치펄스가 한 주기구간당 하나씩 포함되도록 피치펄스열신호 $e(n)$ 을 주기적 분할(periodic segmentation)함으로써 제3h도, 제3k도 및 제3n도에 보인 $e_1(n), e_2(n), e_3(n)$ 과 같은 피치펄스신호들을 얻는다. 절단위치는 피치펄스들 간의 중첩 또는 각 피치펄스들간의 중첩 또는 각 피치펄스의 앞쪽으로 일정시간되는 시점으로 정해도 되나, 제3a도 및 제3f도를 비교하여 알 수 있는 바와같이 피치펄스들의 시간적 위치는 성문펄스들의 끝부분과 일치하므로 제3f도의 점선으로 나타낸 것과 같이 각 피치펄스의 뒷쪽으로 일정시간 되는 시점으로 택하는 편이 좋다. 그러나 청각상 가장 큰 효과를 주는 것은 피치펄스이므로 어느 경우나 실제로 합성음의 음질에 있어서는 별로 차이가 없다.

이와같은 방법으로 얻어진 피치펄스신호들 $e_1(n), e_2(n), e_3(n)$ 등을 그 주기구간 동안의 임펄스응답인 제3e도의 $h_1(n), h_2(n), h_3(n)$ 과 각각 다시 콘벌루션시키면 제3i, l 및 o도 등과 같은 목적하는 단위파형들이 얻어진다. 이와같은 콘벌루션은 실제로는 제4도에서와 같이 스펙트럼포락 파라미터를 필터계수로 사용하는 스펙트럼포락 필터 $H(f)$ 에 각 피치펄스신호를 입력시켜 행하는 것이 편리하다. 예컨대, 선형예측분석법에서와 같이 스펙트럼포락 파라미터로서 선형예측 계수나 반사계수 또는 선스펙트럼쌍을 사용하는 경우에는 그것들을 직접 필터 계수로 하는 IIR(infinite-duration impulse response) 필터를 구성한다. 준동형분석법과 같이 임펄스응답을 스펙트럼포락 파라미터로서 사용할 경우에는 임펄스응답을 탭 계수(tap coefficient)로 하는 FIR 필터(filter)를 구성한다. 스펙트럼포락 파라미터가 대수면적비나 켈스트럼이면 그들을 직접 필터계수로 하는 합성필터를 구성할 수 없으므로, 반사계수 및 임펄스응답으로 각각 다시 변형시킨다음 IIR 및 FIR 필터의 계수로 사용하면 된다. 이와같이 구성된 스펙트럼포락 필터에 한주기분 피치펄스신호를 입력시키면서 필터계수를 피치펄스신호의 각 샘플과 같은 순간에 해당하는 스펙트럼포락 파라미터와 같이 변화시키면 그 주기분의 단위파형이 출력된다.

이와같은 이유 때문에, 각 주기분 피치펄스신호를 얻기 위해 $e(n)$ 을 절단했던 것과 동일한 시점에서 "스펙트럼포락 파라미터들의 시간함수파형(time function waveform)들"을 잘라낸다. 예컨대 순차적 분석법의 경우 제3d도에 예시한 것과 같은 스펙트럼포락 파라미터들의 시간함수들로부터 제3h도에 보인 첫주기분 피치펄스신호 $e_1(n)$ 과 같은 시간구간에 해당하는 스펙트럼포락 파라미터들을 잘라내어 제3g도와 같은 첫주기분 스펙트럼포락 파라미터들을 얻을 수 있다. 제4도에 스펙트럼포락 파라미

터로서 대표적으로 반사계수를 k_1, k_2, \dots, k_N 과 임펄스응답 $h(0), h(1), \dots, h(N-1)$ 을 표시하였는데, 이들이 시간에 대한 함수라는 점을 강조하기 위해 각각 $k_1(n), k_2(n), \dots, k_N(n)$ 및 $h(0, n), h(1, n), \dots, h(N-1, n)$ 으로 표기하였다. 캡스트럼 $c_L(i)$ 가 스펙트럼포락 파라미터로 쓰이는 경우에는 마찬가지로 $c_L(i, n)$ 으로 표기될 것이다.

피치동기식 분석법이나 블록별분석법의 경우에는 순차적 분석법의 경우와 달리 스펙트럼포락 파라미터들임 시간함수가 구해지지 않으며 분석구간에 대해 일정한 스펙트럼포락 파라미터값들이 구해지므로 그것들을 가지고 먼저 스펙트럼포락 파라미터의 시간함수를 만든다음 그 시간함수를 주기적으로 분할하여 한주기분 스펙트럼포락 파라미터들을 만들어야 하나 실제로는 시간함수를 구성하는 대신 다음과 같이 처리하는 것이 편리하다. 즉 피치동기식 분석법의 경우에는 제8b도의 선택과 같이 각 피치주기구간에 대해 일정한 값을 가진 한 세트씩의 스펙트럼포락 파라미터들이 대응되기 때문에 이 경우에는 주기적 분할을 해도 변화가 없으며 따라서 버퍼에 저장될 한주기분의 스펙트럼포락 파라미터는 시간함수가 아니고 시간에 무관한 상수(constant)가 된다. 또 블록별분석 방법의 경우에는 블록당 한 세트의 일정한 스펙트럼포락 파라미터값들이 구해지므로 한 블록에 속하는 모든 한주기분 스펙트럼포락 파라미터들, 예컨대 $k_1(n)_1, k_1(n)_2, \dots, k_1(n)_N$ 의 값은 시간에 무관한 상수일 뿐 아니라 모두 동일한 값이 된다. (여기서 $k_1(n)_j$ 는 j번째 주기구간에 k_1 의 시간함수를 의미하며 M은 블록에 속하는 피치주기구간의 수를 표시함) 단지 블록 경계점을 기준으로하여 그 전 및 후 신호분에 대해 각각 전 및 후 블록의 스펙트럼포락 파라미터 값을 사용해야 함을 유의해야 한다.

제3i도에서 볼수 있듯이 단위파형의 지속시간은 반드시 한주기와 같지는 않다. 따라서 주기적 분할로 얻어진 하주기길이의 피치펄스신호와 스펙트럼포락 파라미터들을 스펙트럼포락 필터에 인가하기 전에 그것들의 지속시간이 단위파형의 유효지속시간(effective duration) 이상이 되도록 하기 위해 제4도에 보인 영샘플 추가(zero appending)와 파라미터 연장(parameter trailing) 과정이 필요하다. 영샘플 추가 과정은 한주기분 피치펄스신호의 뒤에 영의 값을 가진 샘플들을 추가하여(append) 전체 지속시간을 필요한 길이가 되도록 하는 것이다. 파라미터 연장과정은 한주기분 스펙트럼포락 파라미터의 뒤에 그 다음 주기분 스펙트럼포락 파라미터의 앞부분을 추가하여 전체 지속시간을 필요한 길이가 되도록 하는 것인데, 끝값을 반복시키든가 그 다음 주기분 스펙트럼포락 파라미터의 첫값을 반복시켜 추가하는 간단한 방법을 써도 합성의 음질은 별로 나빠지지 않는다.

스펙트럼포락 필터에 의해 만들어질 단위파형의 유효지속시간은 스펙트럼포락 파라미터들의 값에 따라 좌우되며 미리 추정하기 어렵다. 그러나 대개의 경우 단위파형의 유효지속시간은 2주기 이내로 보아 실용상 지장이 없기 때문에 영샘플 추가에 의해 만들어진 '연장된 피치펄스신호'와 파라미터 연장에 의해 만들어질 '연장된 스펙트럼포락 파라미터'의 지속시간이 2주기 길이가 되도록 하는 것이 편리하다. 제3g도에 주기적 분할로 얻어진 첫주기구간(a-b)의 스펙트럼포락 파라미터 다음에 점선으로 표시된 그 다음 주기구간(b-c)의 스펙트럼포락 파라미터를 추가하여 만들어진 2주기구간(a-c)의 연장된 첫 주기분 스펙트럼포락 파라미터를 예시하였다. 또 제3h도에 주기적 분할로 얻어진 첫 주기구간(a-b)의 피치펄스 신호의 다음 주기구간(b-c)에 영샘플들을 추가하여 만들어진 2주기구간(a-c)의 연장된 첫 주기분 피치펄스신호를 예시하였다.

이상과 같은 경우에, 영샘플추가와 파라미터연장 이전의 피치펄스신호와 스펙트럼포락 파라미터의 지속시간은 1주기인데 비해 영샘플추가와 파라미터 연장 이후의 지속시간은 2주기로 늘어나게 되므로 제4도에서와 같이 그 중간에 버퍼를 두어 주기적 분할로 얻어진 피치펄스신호와 스펙트럼포락 파라미터를 일단 버퍼에 저장했다가 꺼내어 씌으로써 시간적 완충(temporal buffering)이 되도록 하였다.

제4도에서 영샘플추가와 파라미터연장에 의해 연장된 피치펄스신호와 연장된 스펙트럼포락 파라미터가 얻어졌으면 최종적으로 스펙트럼포락 필터 H(f)에 제3h도의 구간(a-c)과 같은 연장된 첫주기분 피치펄스신호를 입력시키면서 그에 동기적으로 계수들을 제3g도의 구간(a-c)과 같은 연장된 첫주기분 스펙트럼포락 파라미터와 같이 변화시키면 제3i도의 구간(a-c)과 같은 2주기구간 길이의 첫주기분 "단위파형신호" $s_1(n)$ 을 얻을 수 있다. 둘째 및 셋째 주기분 단위파형신호 $s_2(n), s_3(n)$ 도 마찬가지로 방법으로 얻을 수 있다.

이상 설명한 바와같이 제4도의 과정에 의해 결과적으로 유성음성파형 $s(n)$ 은 그것을 구성하는 단위파형들로 분해된다. 명백히, 분해되어 얻어진 제3i, 제3j 및 제3o도의 단위파형들을 원래의 시점에 다시 배치시키면 제3b도와 같이 되어, 그것들을 중첩시키면 제3c도와 같은 원음성파형 $s(n)$ 이 도로 만들어질 것이다. 만일 제3i, 제3j 및 제3o도의 각 단위파형을 제3p도에 보인것과 같이 상호간 간격을 변화시켜 재배치한 후 중첩시키면 제3q도와 같이 다른 피치패턴을 가진 음성파형이 얻어질 것이다. 이와같이, 분해되어 얻어진 단위파형들 간의 시간간격을 적절히 변화시키면 임의의 원하는 피치패턴 즉 억양을 가진 음성을 합성할 수 있게 된다. 또한 단위파형들의 에너지를 적절히 변화시키면 임의의 원하는 강세패턴을 가진 음성을 합성할 수 있다.

제2도의 음성단편저장부에는 제4도와 같은 방법에 의해 피치펄스 갯수만큼의 단위 파형들로 분해된 상태의 각 유성음성단편이 제6a도와 같이 포맷(format)으로 저장되어 있는데 이와같은 것을 "음성단편정보"(speech segment information)라 한다. 음성단편정보의 앞부분인 헤더필드에는 음성단편에 있어서의 중요시점(important time point)들인 경계시점(boundary time point)들 b_1, b_2, \dots, b_L 과 각 단위파형을 합성할때 사용되었던 각 피치펄스신호의 피치펄스 위치(pitch pulse position)들 P_1, P_2, \dots, P_M 이 저장되었는데, 첫피치펄스신호 $e_1(n)$ 의 첫 샘플 위치를 0으로 삼아 그 각 시점들에 해당하는 샘플수가 기록된다. 경계시점은 음성단편을 편의상 몇개의 소단편(subsegment)들로 분할해 볼때 생기는 그들간의 경계점의 시간적 위치이다. 예컨대 저후에 자음을 가진 모음은 느린 발성속도에서는 중간부분의 정상상태구간(steady-state interval)과 그 전후에 위치하는 두개의 천이구간(transitional interval)으로 구분될 수 있기 때문에 3개의 소단편으로 되어 있다고 볼수 있는데, 음성단편 헤더필드에는 각 소단편의 끝점 3개가 경계시점으로 저장된다. 그러나 빠른 발성속도에서 채취(sampling)한 경우에 천이구간이 하나의 시점으로 되어 모음의 음성단편은 2개의 소단편으로 되어 있다고 볼수 있으므로 2개의 경계시점이 헤더정보 내에 저장되게 된다.

음성단편정보의 뒷 부분인 단위파형 코드필드에는 각 주기에 해당하는 단위파형을 파형코딩하여 얻어지는 코드들인 "파형코드"(waveform code)들을 저장한다. 단위파형들은 PCM과 같은 간단한 파형 코딩방법으로 코딩해도 되지만 큰 단기적 및 장기적 상관성을 가지기 때문에 피치예측루프(pitch-predictive loop)를 가진 ADPCM이나 적응예측코딩(Adaptive Predictive Coding) 또는 디지털(digital) 방식의 적응델타코딩(Adaptive Delta Modulation)등을 이용하여 효율적으로 파형코딩하면 저장에 필요한 메모리 양을 상당히 줄일 수 있다. 분해되어 얻어진 단위파형들을 파형코딩하고 그 코드들을 저장했다가 합성시 디코딩하여 재배치 및 중첩시켜 합성음을 만드는 이와 같은 방법을 "파형코딩 저장방식(waveform code storage method)"이라고 한다.

피치펄스 신호와 그것에 대응하는 스펙트럼포락 파라미터들은 단위파형을 만들 수 있는 재료들이므로 단위파형과 동일한 것으로 취급될 수 있다. 따라서 그들을 각각 코딩하여 얻어진 "파원코드"들을 저장했다가, 합성시 디코딩하여 얻어진 피치펄스신호와 스펙트럼포락 파라미터들을 가지고 단위파형들을 만든다음 그것들을 재배치 및 중첩시켜 합성음을 만드는 방법도 가능하다. 이와 같은 방식을 "파원코드저장방식(source code storage method)"이라고 한다. 이 방식은 제4도에 있어서 출력으로 얻어지는 단위파형을 대신 버퍼(buffer)들에 저장된 피치펄스신호들과 스펙트럼포락 파라미터들을 각각 같은 주기 구간의 것끼리 짝지어 음성단편저장부에 저장하는 것과 같다. 따라서 파원코드저장 방식에서는 제4도의 버퍼다음의 처리과정들 즉, 파라미터연장(parameter trailing)과 영샘플 주기(zero appending)과정 그리고 합성필터 H(f)에 의한 필터링 과정을 제7도의 파형조립부내에서 처리하여야 한다.

파원코드 저장방식의 경우에 음성단편정보의 포맷은 제6b도와 같은데, 제6a도와 마찬가지로 단위 파형 코드필드의 내용만이 다르다. 즉, 제6a도에서의 각 주기분 단위파형이 저장될 위치에 단위파형 대신 그 단위파형을 합성하는데 필요한 피치펄스신호와 스펙트럼포락 파라미터들을 코딩하여 저장한다.

스펙트럼포락 파라미터들은 기존의 스펙트럼포락 파라미터의 양자화(quantization)방법에 의해 코딩되어 단위파형 코드필드에 저장된다. 이때 스펙트럼포락 파라미터를 적당한 방법으로 변형(transformation)시켜 양자화하면 코딩을 효율적으로 할 수 있다.

예컨대 예측계수는 선스펙트럼쌍 파라미터로 변형시키고 반사계수는 대수면적비로 변형시켜 양자화 하는 것이 좋다. 또 임펄스 응답은 인접 샘플상호간 및 인접 임펄스응답간에 상관성이 크므로 차분식(differential) 코딩 방법으로 파형코딩하면 저장에 필요한 데이터 양을 크게 줄일 수 있다. 캡스 트럼 파라미터의 경우에도 변형하여 데이터 양을 상당히 줄일 수 있는 코딩방법이 알려져 있다.

한편 피치펄스신호들은 적당한 파형코딩 방법에 의해 코딩되어 그 코드가 단위파형 코드필드에 저장된다. 피치펄스신호들은 단기적 상관성(short-term correlation)은 거의 가지지 않으나 상호가 큰 장기적 상관성(long-term correlation)을 가지므로 피치예측루프를 가진 피치예측시(pitch-predictive) 적응(adaptive) PCM 코딩과 같은 파형코딩 방법을 사용하면 저장에 필요한 메모리양을 샘플당 3비트(bit) 정도로 줄여도 고음질의 합성음을 얻을 수 있다. 피치예측기의 예측계수는 자기 상관법에 의해 피치주기마다 구한 값일 수도 있고, 일정한(constant) 값일 수도 있다. 코딩의 첫 단계에서 코딩할 피치펄스 신호를 샘플당 평균 에너지의 평균근 G로써 나누어 규준화(normalization)함으로써 피치예측 효과를 올릴 수 있다. 디코딩은 유성음합성부에서 행해지는데, 디코딩 끝단계에서 다시 G를 곱하여 피치펄스 신호를 원래의 크기로 복원시키면 된다.

제6b도에서는 스펙트럼포락 파라미터로서 14개의 반사계수를 사용하는 선형예측 분석법을 채택한 경우의 음성단편정보를 예시하였다. 선형예측 분석의 분석구간이 피치주기이면 한 피치펄스신호에 대해 14개씩의 반사계수가 대응되어 저장된다. 분석구간이 일정길이의 블록이면 한 블록 내의 여러 피치펄스에 대해 반사계수는 모두 같은 값이 되므로 단위파형 코드의 저장에 필요한 메모리 양이 적어진다. 이 경우에는 전술한 바와 같이 두 블록의 경계에 걸쳐있는 피치펄스신호에 대해서는 그 신호의 샘플이 그 경계점의 앞 또는 뒤에 있는가에 따라 합성시 각기 앞 또는 뒷블록의 반사계수들을 적용하여야 하므로, 헤더필드(header field)에 블록간 경계점들의 위치를 추가로 저장하여야 한다. 만일 회귀최소자승법과 같은 순차적 분석법을 사용하였다면 제3d에 예시한 바와 같이 반사계수 k_1, k_2, \dots, k_{14} 는 시간변수(time index) n 의 연속함수가 되어 이들 시간함수 $k_1(n), k_2(n), \dots, k_{14}(n)$ 를 저장하기 위해서는 많은 메모리가 요구된다. 제3a 내지 t도의 경우를 예로들면 단위파형 코드필드의 첫주기분으로 제3g 및 h도의 구간(a-b), 둘째 주기분으로 제3j 및 k도의 구간(b-c), 셋째 주기분으로 제3m 및 n도의 구간(c-d)의 파형들이 단위파형 코드필드에 저장된다.

파형코드 저장방식과 파원코드 저장방식은 근본적으로 같은 방식으로서, 실제로 파형코드 저장방식에서 단위파형들을 적응예측코딩(APC)과 같은 효율적인 파형코딩방법에 의해 코딩하였을때 얻어지는 파형코드는 파원코드 저장방식에서 얻어지는 파원코드와 내용상 거의 같아진다. 파형코드 저장방식에서의 파형코드와 파원코드 저장방식에서의 파원코드를 통틀어 단위파형코드(wavelet code)라 칭한다.

제7도는 본 발명에 의한 유성음합성부의 내주 구성도이다. 음성단편저장부(제2도의 5)로부터 받은 음성단편정보중의 단위파형 코드필드에 저장된 단위파형 코드들은 디코딩부(decoding subblock)(9)에 의해 그것들을 코딩했을때의 역과정으로 디코딩된다. 파형코드저장방식에서의 파형코드가 디코딩되어 얻어진 단위파형 신호(wavelet signal)들, 또는 파원코드 저장방식에서의 파원코드가 디코딩되어 얻어진 피치펄스 신호들과 그것들과 짝지어진 스펙트럼포락 파라미터들을 단위파형 정보(wavelet information)라 칭하는데 파형조립부(13)에 제공된다. 한편 음성단편정보중의 헤더필드에 저장된 헤더정보(header information)는 지속시간조절부(10), 피치조절부(11) 그리고 에너지조절부(12)에 입력된다.

제7도의 지속시간조절부(duration control subblock)(10)는 운용정보중의 지속시간 정보와 음성단편 헤더정보에 포함된 경계시점들을 입력으로 받아 그것들을 이용하여 시간왜곡정보(time warping information)를 만들어 파형조립부에 제공한다. 음성단편의 전체 지속시간이 길어지거나 짧아지면

그 음성단편을 구성하는 소단편들의 지속시간도 그에 따라 길어지거나 짧아지는데, 그 신장 또는 수축의 비율은 각 소단편의 성격에 따라 다르다. 예컨대, 전후에 자음을 가진 모음의 경우에는 양쪽의 천이구간에 비해 가운데에 있는 정상상태(steady state) 구간의 지속시간의 변화율이 훨씬 크다. 지속시간조절부는 저장되었던 원래의 음성단편의 지속시간과 지속시간정보가 지시하는 합성될 음성단편의 지속시간을 비교하여 원래의 각 소단편의 지속시간에 규칙에 의해 구함으로써 "합성음의 경계시점"들을 구한다. 원래의 경계시점들과 그에 대응하여 짝지어진 합성의 경계시점들을 합친 것을 시간왜곡정보라 하는데, 파형조립부(13) 뿐 아니라 피치조절부와 에너지조절부에도 제공된다.

제7도의 피치조절부(pitch control subblock)(11)의 기능은 합성음이 억양패턴정보가 지시하는 대로 억양패턴을 가지도록 피치펄스 위치정보(pitch pulse position information)를 만들어 파형조립부(13)에 제공하는 것이다. 피치조절부(11)는 음소별 목표 피치주파수 값(target pitch frequency value)들인 억양패턴 정보를 입력으로 받아서 이들을 매끄럽게 연결함으로써 시간에 대한 피치주파수의 연속적 변화를 나타내는 피치 윤곽함수(pitch contour)를 발생시킨다. 피치조절부(11)는 저해음(obstruent)에 의한 미세억양(micro intonation)현상을 피치윤곽함수에 반영시킬 수도 있는데, 이 경우 피치윤곽함수는 저해음 음소와 인접한 다른 음소와의 경계점에서 피치주파수 값이 시간에 대해 급격히 변하는 불연속 함수가 된다. 이 피치윤곽함수로부터 임의의 한 시점에서의 피치주파수를 구하고 그것의 역수를 취하여 피치주기를 구하여 그 주기만큼 진행된 후의 시점을 피치펄스 위치로 정하고, 그 시점에서의 피치주파수로부터 다시 피치 주기를 구하여 다음 피치펄스 위치를 구하는 과정을 반복해나가면 합성음의 피치펄스들의 위치를 모두 구할 수 있다. 상기 피치조절부(11)는 이렇게 하여 구해진 합성음의 피치펄스 위치들과 음성단편 헤더정보속에 포함된 원래의 피치펄스 위치들을 함께 묶어서 파형조립부와 에너지조절부로 보내는데, 이것을 피치펄스 위치정보라 한다.

제7도의 에너지조절부(energy control subblock)(12)는 강세패턴정보가 지시하는 대로 가세 패턴을 합성음이 가지도록 이득정보(gain information)를 만들어 파형조립부(13)에 제공한다.

에너지조절부(12)는 음소별 목표 진폭값(target amplitude value)들인 강세패턴정보를 입력으로 받아서 이들을 매끄럽게 연결함으로써 시간에 대한 진폭의 연속적 변화를 나타내는 에너지 윤곽함수(energy contour)를 발생시킨다. 각 음소별 에너지의 상대적 차이를 반영하기 위해 음성단편들은 저장시 음성단편의 종류에 따른 상대적 에너지를 가지도록 미리 규준화(normalization)되어 있다고 가정한다. 예컨대 모음에 있어서는 개모음(low vowel)일 수록 단위시간당 에너지가 크며, 비음은 모음에 비해 단위 시간당 에너지가 절반 정도 밖에 되지 않는다. 또한 파열음(plosive)의 폐쇄구간(closure interval)의 에너지는 매우 약하다. 따라서 이들을 음성단편으로 저장할때 이와 같은 상대적 에너지를 가지도록 미리 에너지를 조정(adjustment)한 다음 코딩하여야 한다. 이 경우 에너지조절부에서 만들어진 에너지 윤곽함수는 합성될 파형에 곱할 이득이 된다. 에너지조절부(12)는 에너지 윤곽함수와 피치펄스 위치정보를 이용하여 각 합성음 피치펄스 위치에서의 이득값들을 구하여 파형조립부(13)에 제공하는데, 그것을 이득정보라 한다.

제7도의 파형조립부(waveform assembly subblock)(13)는 전술한 단위파형정보, 시간왜곡 정보, 피치펄스 위치정보 그리고 이득정보를 입력으로 받아서 최종적으로 유성 음성신호를 만들어 내는 역할을 한다. 파형조립부(13)는 디코딩부(9)로부터 받은 단위파형정보를 이용하여 운율정보가 지시한 대로 억양패턴, 강세패턴 그리고 지속시간을 가진 음성을 만들어내야 한다. 이때 어떤 단위파형은 반복되고, 어떤 단위파형은 생략되기도 한다.

운율정보 내의 포함된 지속시간정보, 억양패턴정보, 그리고 강세패턴정보는 서로 독립적인 지시정보(indicative information)들인데 비해 단위파형정보를 가지고 파형을 합성함에 있어서는 그 3가지 정보 상호간 연관성이 있기 때문에 그것들을 연계해서 취급하지 않으면 안된다. 파형조립에 있어 가장 중요한 문제중 하나는 합성음의 각 피치펄스 위치에 배치할 단위파형으로서 어느 단위파형을 선정하느냐 하는 것인데, 적절한 단위파형을 선정하여 배치하지 못하면 좋은 음질의 합성음을 얻을 수 없다.

다음에 음성단편 저장부(제2도의 5)로부터 받은 음성단편 정보를 이용하여 합성음을 합성함에 있어 높은 음질을 얻을 수 있는 단위파형 재배치방법인 본 발명에 의한 시간왜곡식 단위파형 재배치방법을 이용한 파형조립부의 동작을 설명하기로 한다.

상기 파형조립부(13)의 유성음파형 합성과정은 시간왜곡함수(time warping function)를 이용하는 단위파형 재배치단계와 배치된 단위파형들을 중첩하는 단계의 2단계로 이루어진다.

즉 파형코드저장방식의 경우에는 단위파형정보로서 입력받은 단위파형신호(wavelet signal)들 중에서 합성음의 피치펄스 위치에 가장 적합한 것을 선정하여 배치하고 이득조정후 그것들을 중첩시키므로써 합성음을 만든다.

파원코드저장방식에서는 단위파형정보로서 피치펄스신호들과 그에 대응되는 각 주기분 스펙트럼포락 파라미터들을 입력받는다.

이 경우에는 2가지의 합성음 조립방법이 가능하다. 첫번째 방법은 단위파형정보를 가지고 제4도의 버퍼의 우측부분에 해당하는 과정들, 즉 전술한 파라미터 연장(parameter trailing)과 영샘플추가(zero appending)를 각각 행하여 만들어진 2주기 구간 길이의 스펙트럼포락파라미터와 피치펄스신호를 합성필터에 인가하여 각 단위파형을 얻고, 그 단위파형들을 가지고 전술한 파형코드 저장방식에서와 동일한 과정으로 합성음을 조립하는 방법인데, 이 방법은 기본적으로 파형코드 저장방식에서의 합성음 조립과 동일하므로 별도의 설명을 생략하기로 한다. 두번째 방법은 피치펄스신호들 중에서 합성음의 피치펄스 위치에 가장 적합한 것을 선정하여 배치하고 이득을 조정후 그것들을 서로 중첩시킴으로써 평탄한 스펙트럼포락을 가지나 원래의 주기적 피치펄스열신호와 다른 피치패턴을 가진 "합성 피치펄스열신호"(synthetic pitch pulse train signal) 또는 "합성여기신호"(synthetic excitation signal)를 만들고 그것을 구성하는 각 피치펄스신호에 스펙트럼포락 파라미터들을 대응시켜 만든 "합성 스펙트럼포락 파라미터"(synthetic spectral envelope parameter)들을 만들어 그

합성여기신호와 합성스펙트럼포락 파라미터들을 합성필터에 인가하여 합성음을 만드는 방법이다.

이 두 방법은 합성음 조립에 있어 합성필터와 중첩과정간의 순서만 바뀌어 있을 뿐 근본적으로는 동일한 방법들이다.

이들 합성음 조립방법을 제8a 및 제8b도에 의하여 설명하기로 한다. 이 단위파형 재배치 방법은 기본적으로 파형코드 저장방식의 경우이나 파원코드저장방식의 경우이나 똑같이 적용될 수 있으므로 두 방식에서의 합성음파형 조립과정을 제8a 및 제8b도에 의해 동시에 설명하기로 한다.

제8a도는 원래의 음성단편과 합성하려고 하는 음성단편간의 대응관계를 예시하였다. 원래의 경계시점들 B1, B2 등과 합성음(synthesized sound)의 경계시점들 B1, B2등 그리고 점선으로 표시된 그들간의 대응관계는 지속시간조절로부터 받은 시간왜곡정보에 포함되어 있다. 또 원래의 피치펄스 위치들 P1, P2 등과 합성음의 피치펄스 위치들 P1, P2 등은 피치조절부(11)로부터 받은 피치펄스 위치정보 내에 포함되어 있다.

파형조립부(13)는 먼저 원래의 경계시점들과 합성음의 경계시점들 그리고 그들간의 대응관계를 이용하여 제8b도에 예시한 것과 같은 시간왜곡함수를 구성한다. 시간왜곡함수의 횡축은 원래의 음성단편상의 시간 t 를, 종축은 합성될 음성단편상의 시간 t' 를 표시한다. 제8a도의 경우를 예로들면, 원래의 음성단편의 첫번째 및 마지막 소단편은 2/3배와 2배로 각각 축소 및 신장되어야 하므로 이들 대응관계는 제8b도의 시간 왜곡함수에서 기울기가 각각 2/3 및 2인 선분으로 나타난다. 또, 2번째 소단편은 지속시간의 변화가 없으므로 시간왜곡함수에서 기울기가 1인 선분으로 나타난다. 합성될 음성단편의 두번째 소단편은 원래의 음성단편의 경계시점 B1이 반복되어서 생기는 것이며, 반대로 원래의 음성단편의 세번째 소단편은 합성될 음성단편에서는 하나의 경계시점 B3으로 변했다.

이와 같은 경우 이들 대응관계는 시간왜곡함수에서는 각각 수직과 수평선분으로 나타난다. 시간왜곡함수는 이와 같이 원래의 음성단편의 경계시점과 그에 대응하는 합성될 음성단편의 경계시점과를 하나의 점으로 나타내고 그들간의 직선으로 연결함으로써 얻어진다. 경우에 따라서는 이 점들 사이를 매끄러운(smooth) 곡선으로 연결하여 소단편간의 대응관계를 실제와 더 가깝게 표현되게 할 수도 있을 것이다.

파형코드 저장방식의 경우, 파형조립부(13)는 합성음 피치펄스 위치에 해당하는 원래의 시점을 시간 왜곡 함수를 이용하여 찾고 그것에 가장 가까운 피치펄스 위치를 가지는 단위파형을 찾아 합성음 피치펄스 위치에 배치한다. 합성음 음성단편의 각 피치펄스 위치에 이와 같은 방법에 의해 배치된 단위파형의 번호를 제8b도에 예시하였다. 이 그림에서 볼 수 있듯이 원래의 음성단편을 구성하는 단위 파형중에서 어떤 것은 소단편의 수축으로 인해 탈락되며, 어떤 것은 소단편의 신장으로 인해 반복 사용되기도 한다.

그다음 단계로, 파형조립부(13)는 재배치된 각 단위파형신호에 그것의 피치펄스 위치에 해당하는 이득을 이득정보에서 찾아 곱한다음, 이득조절된 이들 단위파형신호들을 단순히 가산(addition)하여 중첩함으로써 최종적으로 원하는 합성음을 얻는다. 제3q도에 제3i, 3j 및 3o도의 단위파형들이 제3p도와 같이 재배치된 경우 이와 같은 과정에 의하여 만들어진 합성음성을 예시하였다.

파원코드 저장방식의 경우에도 마찬가지로, 파형조립부(13)는 합성음 피치펄스 위치에 해당하는 원래의 시점을 시간왜곡함수를 이용하여 찾고 그것에 가장 가까운 피치펄스 위치를 가지는 피치펄스 신호를 찾아 그 합성음 피치펄스 위치에 배치한다.

파형코드 저장방식의 경우에도 마찬가지로, 파형조립부(13)는 합성음 피치펄스 위치에 해당하는 원래의 시점을 시간왜곡함수를 이용하여 찾고 그것에 가장 가까운 피치펄스 위치를 가지는 피치펄스 신호를 찾아 그 합성음 피치펄스 위치에 배치한다.

파형코드 저장방식에서의 단위파형들의 중첩은 파원코드저장방식에서는 피치펄스신호의 중첩에 해당한다. 따라서 파원코드 저장방식의 경우에 파형조립부(13)는 재배치된 각 피치펄스 신호의 피치펄스 위치에 해당하는 이득을 이득정보에서 찾아 각 피치펄스 신호에 곱한다음 이득조절된 이들 피치펄스 신호들을 중첩함으로써 합성 여기신호를 만든다. 그러나 이 경우에는 피치펄스에 에너지가 집중되어 있으므로 일정한 이들을 피치펄스신호에 곱하여 중첩시키는 대신 먼저 피치펄스신호를 중첩시켜 이득조절이 되지 않은 합성 여기신호를 만들고 그것에 에너지조절부에서 발생된 에너지 윤곽함수를 곱하여 이득조절된 합성 여기신호를 얻을수도 있다. 제3r도에 이와 같은 과정에 의하여 제3h, 3k 및 3n도의 피치펄스신호들을 피치패턴이 제3p도의 경우와 같도록 재배치하였을 때 얻어진 합성여기신호를 보였다.

파원코드 저장방식의 경우에 파형조립부(13)는 또한 합성 스펙트럼포락 파라미터를 만들어야 한다. 만일 스펙트럼포락 파라미터들이 시간에 대한 연속함수이며 음성 스펙트럼의 포락을 충실히 표현한다면 제8a도에 예시한 바와 같이 원래의 스펙트럼포락 파라미터를 소단편별로 단순히 시간적으로 압축 또는 신장시킴으로써 합성 스펙트럼포락 파라미터를 얻을 수 있을 것이다. 그러나, 블록별분석법 또는 피치동기식 분석법을 쓸 경우에는 스펙트럼포락 파라미터를 얻을 수 있을 것이다. 그러나, 블록별분석법 또는 피치동기식 분석법을 쓸 경우에는 스펙트럼 합치성(spectral match)이 나쁘거나 스펙트럼포락 파라미터의 시간적 변화가 불연속적이기 때문에 그와 같은 "시간축 압축 및 신장(temporal compression and expansion)방법"을 써서는 좋은 합성음질을 얻을 수 없으며, 제8b도에 예시한 바와 같이 피치펄스 신호별로 스펙트럼포락 파라미터를 대응시켜 조립하는 방법을 쓰는 것이 가장 좋다. 즉, 파형코드 저장방식에서의 한 단위파형은 파원코드 저장방식에서는 같은 피치주기간의 피치펄스신호와 스펙트럼포락 파라미터를 그 피치펄스 신호에 동기적으로(synchronously) 배치함으로써 합성 스펙트럼포락 파라미터를 만든다. 제8b도에 블록별 분석법과 피치동기식 분석법으로 얻어진 스펙트럼포락 파라미터중의 하나인 k_1 에 대하여 대표적으로 이와같은 방법에 의해 조립된 합성 스펙트럼포락 파라미터 k_1 를 각각 실선과 쇄선으로 예시하였다. 제8a도에 보인 순차적 분석법에 의해 구해진 스펙트럼포락 파라미터에 대해서도 물론 제8b도와 같은 방법으로 합성 스펙트럼포락 파라미터를 조립할 수 있는데, 예를들어 각 주기분 피치펄스 신호들이 제3r도와 같이 재배치되었다

면 각 주기분 스펙트럼포락 파라미터들은 그 피치펄스 신호들에 대응되어 제3s도와 같이 배치되게 된다.

파원코드 저장방식에 있어서는 합성여기 신호와 합성 스펙트럼포락 파라미터의 조립시 합성음의 피치주기가 원래의 피치주기보다 긴 경우에는 8a 및 8b도에 빗금으로 나타낸 바와같이 인접한 두 피치주기구간 사이에는 공백구간(blank interval or gap)이 생기며, 합성음의 피치주기가 원래의 피치주기보다 짧은 경우에는 인접한 두 피치주기구간이 겹치는 중복구간(overlap interval)이 생기게 된다. 제3r 및 3s도에 중복구간(a-b)과 공백구간(c-d)를 예시하였다. 전술한 바와같이, 재배치된 피치펄스 신호들은 중복시 중첩되어야 한다. 그러나 그것들에 대응하여 재배치된 스펙트럼포락 파라미터들은 중복시 중첩되는 대신 두 값의 평균값을 취하는 것이 합리적이다. 따라서 공백구간과 중복구간을 고려한 합성여기 신호와 합성 스펙트럼포락 파라미터의 조립방법은 다음과 같다.

합성여기 신호의 조립시 공백구간에는 영샘플(zero sample)들을 삽입하면 된다. 유성마찰음의 경우에는 공백구간에 고역통과 필터링된 잡음신호를 삽입하면 좀더 자연스러운 음을 합성할 수 있을 것이다. 중복구간에는 중복된 피치펄스 신호들을 배치된 상태에서 그대로 더하면 된다. 이와같은 가산방식(addition method)은 번거롭기 때문에 실제로는 중복구간에서 중복된 두 피치펄스 신호중의 어느 한쪽 신호만을 선택하는 잘라버리기방식(truncation method)을 쓰는 것이 간편하며, 합성음의 음질도 별로 열화되지 않는다. 제3r도에서는 공백구간(c-d)에는 영샘플들로 채우고 중복구간(a-b)에서는 앞쪽 피치펄스 신호를 선택하였다. 즉 중복이 발생할 경우 각 피치펄스 신호의 겹치는 앞쪽 구간분을 잘라버렸는데, 전술한 바와 마찬가지로 이 방법은 피치펄스들의 직전에서 절단하여 피치펄스 신호들을 만들고 합성시 중복되면 피치펄스 신호의 뒷쪽 중복구간분을 잘라버리는 방법에 비해 물리적으로 더 의미가 있다. 그러나 실제로 합성음의 음질에 있어서 양자간 큰 차이는 없다.

합성 스펙트럼포락 파라미터의 조립시 공백구간에는 앞주기분 스펙트럼포락 파라미터의 끝 시점의 값에서 뒤 주기분 스펙트럼포락 파라미터의 첫 시점의 값으로 직선적으로 변하게 만든 값들로 채워 넣고, 중복구간에는 중복된 두 스펙트럼포락 파라미터의 평균을 구하는 내삽법을 이용하여 시간에 대해 직선적으로 변하는 가중치를 가지고 앞주기분 스펙트럼포락 파라미터 값으로부터 뒤주기분 스펙트럼포락 파라미터의 값으로 서서히 옮겨가게 하는 것이 이상적이다. 그러나 이와같은 방법들은 번거롭기 때문에 음질의 열화(degradation)가 그리크지 않으면서 더 간편한 다음의 방법들을 쓸수 있다. 즉 공백구간의 스펙트럼포락 파라미터로서는 제8b도의 예와같이 앞주기분 스펙트럼포락 파라미터의 끝 시점의 값을 반복시켜 쓰든가, 뒤 주기분 스펙트럼포락 파라미터의 첫시점의 값을 반복시켜 쓰든가, 공백구간 내의 적당한 점을 경계로 전후에 각각 앞 및 뒤 주기분 스펙트럼포락 파라미터의 끝 및 첫 시점의 값을 반복시켜 쓰든가, 아니면 그 두 스펙트럼포락 파라미터값의 산술평균 값을 쓰든가 하면 된다. 또 중복구간에서의 스펙트럼포락 파라미터로서는 단순히 선택된 어느 한쪽의 피치펄스에 대응되는 부분을 선택하면 된다. 제3c도에서는 중복구간(a-b)에서 합성여기신호로서 앞주기구간의 피치펄스 신호를 택했으므로 합성 스펙트럼포락 파라미터로서도 앞주기 구간의 값을 택하였다. 제8b도와 제3s도의 공백구간(c-d)에서는 앞주기 구간끝에서의 스펙트럼포락 파라미터 값을 반복하여 사용하였다. 물론 스펙트럼포락 파라미터가 시간에 대한 연속 함수인 제3s도의 경우에는 공백구간 동안 앞주기 구간 끝값 또는 뒤주기 구간 첫값을 반복 사용하는 방법이나 두 값을 직선적으로 변화시켜 사용하는 방법 모두가 같은 결과로 된다.

합성여기 신호와 합성 스펙트럼포락 파라미터들이 모두 조립되었으면, 파형조립부(13)는 통상 인접한 음성단편간에 스펙트럼포락 파라미터의 변화가 매끄럽도록, 조립된 음성단편의 합성 스펙트럼포락 파라미터의 양단부분을 내삽법을 이용하여 원활화(smoothing)시킨다. 이상과 같이 조립된 합성여기신호와 합성 스펙트럼포락 파라미터들이 파형조립부 내의 합성필터에 각각 여기신호와 필터계수로써 입력되면 원하는 합성음성이 최종적으로 합성필터로부터 출력된다. 제3r도에 제3h, 3k 및 3n도의 피치펄스 신호들을 제3p도와 피치패턴이 같도록 재배치하였을때 얻어진 합성여기신호를, 제3s도에 그 합성여기신호에 제3g, 3j 및 3m도의 한주기분 스펙트럼포락 파라미터들을 대응시켜 만들어진 합성 스펙트럼포락 파라미터들을 예시하였다. 제3s도와 같이 변화하는 반사계수들을 필터계수로 하는 시변 합성필터를 구성하고 그것에 제3r도와 같은 합성여기신호를 입력으로 가하면 제3p도의 합성음과 거의 같은 제3t도의 합성음이 만들어진다.

여기서 파형코드저장방식과 파원코드저장방식을 비교해보면, 두 방식은 원리적으로 같다고 볼 수 있으나 단지 접속성이 나쁜 음성단편들을 서로 연쇄시킬 때 후자에 있어서는 내삽법에 의해 저장된 스펙트럼포락 파라미터를 원활화(smoothing) 시킴으로써 매끄럽게 연결된 음을 합성할 수 있으나 전자에서는 그것이 불가능한 점이 차이점이다. 또한 파원코드저장방식은 유성음 합성부의 기능과 전술한 무성음 합성부의 기능을 통합하기 쉽다는 장점을 가진다. 준동형분석법을 사용할 수 있으나 파원코드저장방식에서는 캡스터럼을 사용하는 것이 사실상 불가능하다. 본 발명에 의한 파원코드저장방식은 한주기분 피치펄스를 여기펄스로 사용하지만 그것과 그것에 대응하는 한주기분 피치펄스를 여기펄스로 사용하지만 그것과 그것에 대응하는 한주기분 스펙트럼포락 파라미터가 합쳐 각 주기분 단위 파형을 만든다는 점에서 단순히 임펄스를 일정한(regular) 피치펄스로 대체하려는 기존의 표준 피치펄스 여기방식과는 다르다.

지금까지 상술하여 알 수 있는 바와같이 본 발명은 음성절편 합성방식의 무제한 음성 합성시스템의 음성절편 코딩 및 재생용으로 적합하다. 또한 본발명은 음소, 반응절, 다이폰, 소단편등 음성을 구성하는 임의의 음성단위(phonetic unit)의 전체적 및 부분적 지속시간과 패치패턴을 독립적으로 자유롭게 변화시킬 수 있는 방법이기 때문에 성악음성 합성시스템이나 음성의 억양 패턴을 변화시키지 않고 발성속도를 일정한 비율로 원래의 속도보다 빠르게 또는 느리게 변화시키는 음성속도변환시스템에도 쓰일 수 있으며, 미리 저장된 표준 음성단편(template speech segment)들의 지속시간과 피치를 변화시켜 음성을 전송하는 포네틱 보코더(phonetic vocoder) 또는 세그먼트 보코더(segment vocoder)와 같은 저전송률 음성코딩 시스템(very low rate speech coding system)에도 쓰일 수 있다.

본발명의 또다른 용도는 샘플링 방식의 전자악기와 같은 악기음 합성시스템이다. 기존의 샘플링 방

식 전자악기에 있어서는 그 전자악의 음역이내의 거의 모든 음을 디지털 방식으로 파형코딩하여 저장하였다가 키보드 등으로부터 요구가 있을때 재생하기 때문에 악기음 저장에 필요한 메모리 양이 많은 단점이 있었다. 그러나 본 발명에 의한 주기파형분해 및 단위파형 재배치 방식을 이용하면 단 몇 종류 피치의 음(tone)만을 샘플링하여, 여러다른 피치의 음들을 합성해 낼수 있기 때문에 저장에 필요한 메모리 양을 크게 줄일 수 있다. 악기음은 전형적으로 개시부(attack)와 지속부(sustain) 그리고 감쇠부(decay)의 3부분으로 구성되어 있다. 이들 3부분 간에는 물론이고 지속부 내에서도 부분적으로 스펙트럼포락이 서서히 변하기 때문에 음색도 이에 따라 변한다. 따라서 스펙트럼 변화가 큰 이들내의 적당한 점들을 경계시점으로 삼고, 악기음 단편(musical sound segment)을 전술한 주기파형분해 방식에 의해 코딩하여 저장하였다가 키보드 등으로부터 요구가 발생할 때 전술한 시간왜곡시 재배치방식에 의해 음을 합성하면 원하는 임의의 피치를 가진 악기음을 합성할 수 있다. 다만, 악기음을 선형예측분석법으로 디콘벌루션하는 경우에는 스펙트럼 포락이 정확하게 구해지지 않고 피치펄스가 첨예하게 되지 않는 경향이 있으므로 샘플링 주파수를 올리든가 피치동기식 분석법을 사용하는 것이 바람직하다.

본발명의 원리는 음성합성 뿐만아니라 악기음(musical instrumental sound)이나 성악음성(singing vocie)등 음성과 유사한 성질을 가진 음의 합성이나 저전송률 음성코딩(very low rate speech coding) 또는 음성속도변환(speech rate conversion) 등에도 그대로 이용될 수 있다.

(57) 청구의 범위

청구항 1

음성합성 시스템의 음성단편 코딩방법에 있어서, 유성음성(voiced speech)이나 악기음(musical instrumental sound)과 같은 주기적(periodic) 혹은 준주기적인(quasi-periodic) 디지털 신호를 스펙트럼 추정기법(spectrum estimataion technique)을 써서 분석하여 각 분석 시간구간에 있어서의 스펙트럼포락을 나타내는 파라미터들을 구하고, 역필터링(inverse filtering) 방법을 이용하여 원래의 신호를 스펙트럼포락 파라미터가 나타내는 임펄스 응답과 평탄한 스펙트럼포락을 가지는 주기적 혹은 준주기적인 피치펄스열(pitch pulse train) 신호로 디콘벌루션(deconvolution)시킨 다음에, 각 주기마다 피치펄스가 하나씩 포함되도록 그 피치펄스열 신호를 주기별로 분할(segmentation)하여 얻어진 각 한주기분 피치펄스 신호의 뒤에 영샘플들을 추가(append)하여 만든 여기신호와, 그것과 동일한 시간구간(time interval)에서의 스펙트럼포락 파라미터에 해당하는 임펄스응답과를 시변 필터를 이용하여 다시 콘벌루션(convolution)시킴으로써 만들어지는 각 주기분 단위파형(wavelet)들을 각각 파형코딩(waveform coding)하여 메모리에 저장하였다가, 합성시 그 단위파형들을 디코딩하여 원하는(desired) 피치패턴을 가지도록 적절한 시점에 각각 재배치(relocation)시키고 중첩(superposition)시켜 음을 만듦으로써 지속시간(duration)과 피치주파수의 조절이 가능하면서 원음과 거의 같은 고음질의 음을 합성할 수 있게 하는 것을 특징으로 하는 음성합성 시스템의 음성단편 코딩방법.

청구항 2

제1항에 있어서, 단위파형을 코딩하여 메모리에 저장하는 대신, 분할하여 얻어진 각 한주기분 피치펄스 신호를 파형코딩한 정보와 그것과 동일한 시간구간 동안의 한 주기분 스펙트럼포락 추정 파라미터들 또는 그에 해당하는 임펄스 응답을 코딩한 정보를 짝지어 함께 메모리에 저장하였다가, 합성시 그것들을 디코딩하여 얻어진 각 한주기분 피치펄스 신호의 뒤에 영샘플들을 추가(append)하여 만든 여기신호와, 그것과 동일한 시간구간(time imerval)에서의 스펙트럼포락 파라미터에 해당하는 임펄스응답과를 시변 필터를 이용하여 다시 콘벌루션(convolution)시킴으로써 만들어지는 각 주기분 단위파형(wavelet)들을 원하는(desired) 피치패턴을 가지도록 적절한 시점에 각각 재배치(relocation)시키고 중첩(superposition)시켜 음을 만듦으로써 지속시간(duration)과 피치주파수의 조절이 가능하면서 원음과 거의 같은 고음질의 음을 합성할 수 있게 하는 것을 특징으로 하는 음성합성 시스템의 음성단편 코딩방법.

청구항 3

제1항에 있어서, 단위파형을 코딩하여 메모리에 저장하는 대신에, 분할하여 얻어진 각 한주기분 피치펄스 신호를 파형코딩한 정보와 그것과 동일한 시간구간 동안의 한주기분 스펙트럼포라 추정 파라미터들 또는 그에 해당하는 임펄스 응답을 코딩한 정보를 짝지어 함께 메모리에 저장하였다가, 합성시 그것들을 디코딩하여 원하는 피치패턴을 가지도록 적절한 시점들에 배치시키되 원하는 피치주기가 원래의 피치주기보다 길때 생기는 공백구간(blank interval)에는 영(zero)샘플들로 채우고 원하는 피치주기가 원래의 피치주기보다 짧을때 생기는 중복구간(overlap interval)에서는 상호간 단순히 가산(additon)시키거나 그중 어느 한쪽 신호만을 택함으로써 만들어진 합성여기신호(synthetic excitation signal)와, 공백구간에서는 구간 양끝에서는 스펙트럼포락 파라미터들 중의 어느 한쪽 값 또는 양쪽의 사이값 또는 양쪽을 매끄럽게 연결한 값으로 채우고 중복구간에서는 중복되는 스펙트럼포락 파라미터들 중의 어느 한쪽 값을 택하거나 상호간의 사이값을 사용함으로써 만들어진 합성 스펙트럼포락 파라미터(synthetic spectral envelope parameters)에 해당하는 임펄스 응답과를 시변 필터를 이용하여 콘벌루션시켜 음을 만듦으로써 지속시간(duration)과 피치주파수의 조절이 가능하면서 원음과 거의 같은 고음질의 음을 합성할 수 있게 하는 것을 특징으로 하는 음성합성 시스템의 음성단편 코딩방법.

청구항 4

음성합성 시스템의 피치조절 방법에 있어서, 시작과 끝, 정상상태위치(steady-state point)등 음성 또는 오디오파형 단편내 중요 경계시점(boundary time point)들과 각 단위파형 또는 한주기분 피치펄스 신호의 피치펄스 위치들을 코딩하여 각 음성 또는 오디오파형 단편 저장시 메모리에 함께 저장하였다가, 합성시 원하는 경계시점들과 그에 대응되는 저장된 원래의 경계시점들을 상호 비교하여

시간왜곡함수(time warping function)를 얻고, 그 시간왜곡함수를 이용하여 원하는 각 피치펄스 위치에 대응하는 원래의 시점을 찾아서, 그것에 가장 가까운 피치펄스 위치를 가진 단위파형 또는 한 주기분 피치펄스 신호 및 그에 해당하는 스펙트럼포락 파라미터들을 선택하여 원하는 그 피치펄스 위치에 배치하고, 여기신호와 동일한 시간구간에서의 스펙트럼포락 파라미터에 해당하는 임펄스 응답과는 시변 필터를 이용하여 콘벌루션 시킴으로써 얻어지는 단위파형을 중첩시켜 합성음을 만듦으로써 원음과 같은 고음질의 음을 합성할 수 있게 하는 단위파형 재배치 방식에 의해 음성 또는 오디오 파형단편이 지속시간 및 피치를 조절할 수 있는 것을 특징으로 하는 음성합성 시스템의 피치조절 방법.

청구항 5

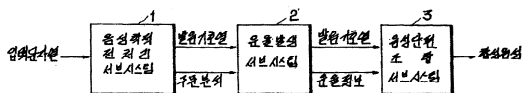
제4항에 있어서, 단위파형 대신 피치펄스 신호들을 중첩시켜 합성음을 만드는 것을 특징으로 하는 음성합성 시스템의 음성단편 코딩방법.

청구항 6

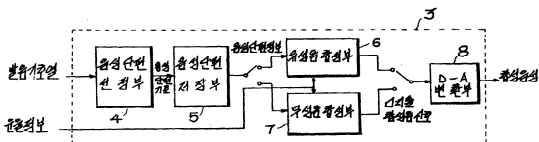
음성합성 시스템의 유성음 합성장치에 있어서, 음성단편 저장부(5)로부터의 단위파형 코드를 디코딩하여 단위파형 정보를 생성하는 디코딩부(9)와, 운율발생 서브시스템(2)으로부터의 지속시간정보 및 상기 음성단편 저장부(5)로부터의 헤더정보에 포함된 경계시점들을 입력으로하여 시간왜곡정보를 생성하는 지속시간 조절부(10)와, 상기 음성단편 저장부(5)로부터의 헤더정보, 상기 운율발생 서브시스템(2)의 억양패턴정보 및 상기 지속시간 조절부(10)의 시간왜곡정보를 입력으로하여 합성음의 억양패턴정보가 지시하는 대로의 억양패턴을 가지도록 피치펄스 위치 정보를 생성하는 피치조절부(11)와, 상기 음성단편 저장부(5)로부터의 헤더정보, 상기 운율발생 서브시스템(2)의 상세패턴정보, 상기 지속시간 조절부(10)의 시간왜곡정보 및 상기 피치조절부(11)의 피치펄스 위치정보를 입력으로하여 강세패턴 정보가 지시하는 대로의 강세패턴을 합성음이 가지도록 이득정보를 생성하는 에너지 조절부(12)와, 상기 디코딩부(9)의 단위파형정보, 상기 지속시간 조절부(10)의 시간왜곡정보, 상기 피치조절부(11)의 피치펄스 위치정보 및 상기 에너지 조절부(12)의 이득정보를 입력으로하여 유성음 성신호를 생성하는 파형조립부(13)로 구성되는 것을 특징으로 하는 음성합성시스템의 유성음 합성장치.

도면

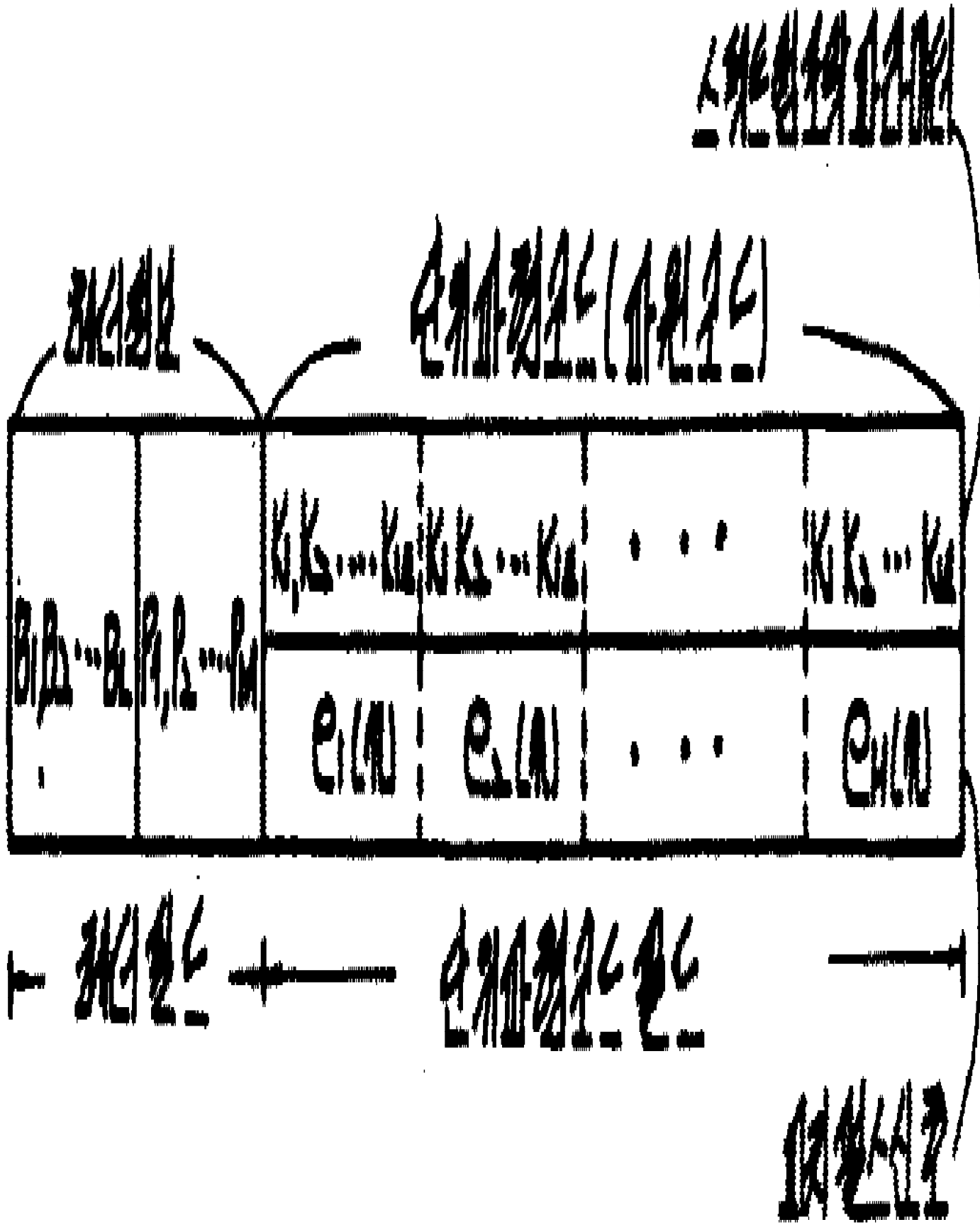
도면1



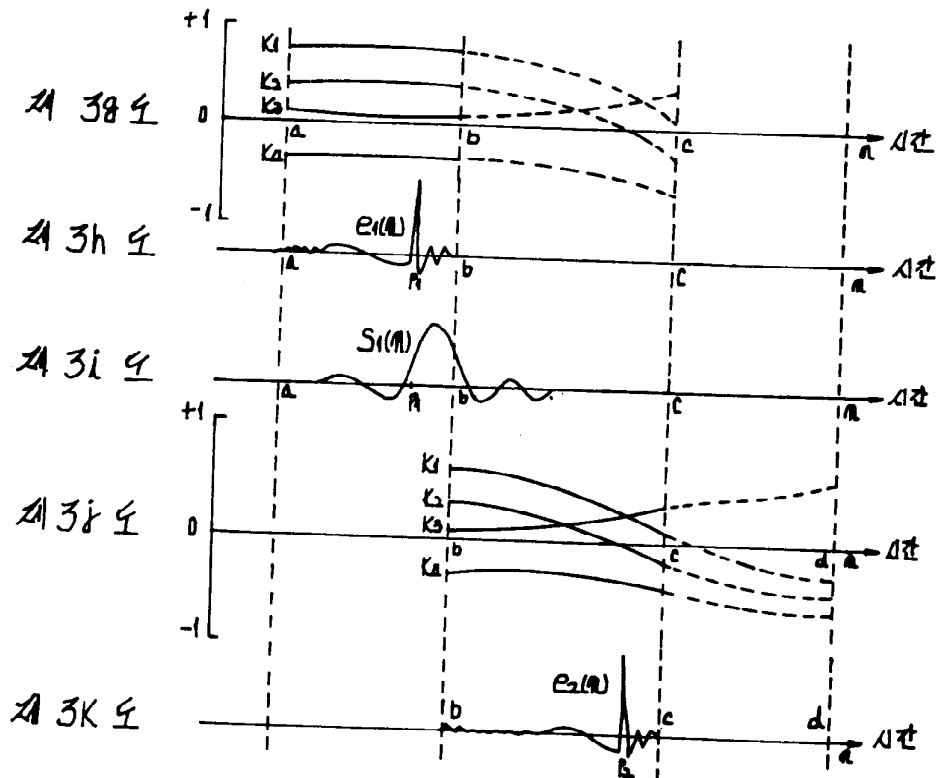
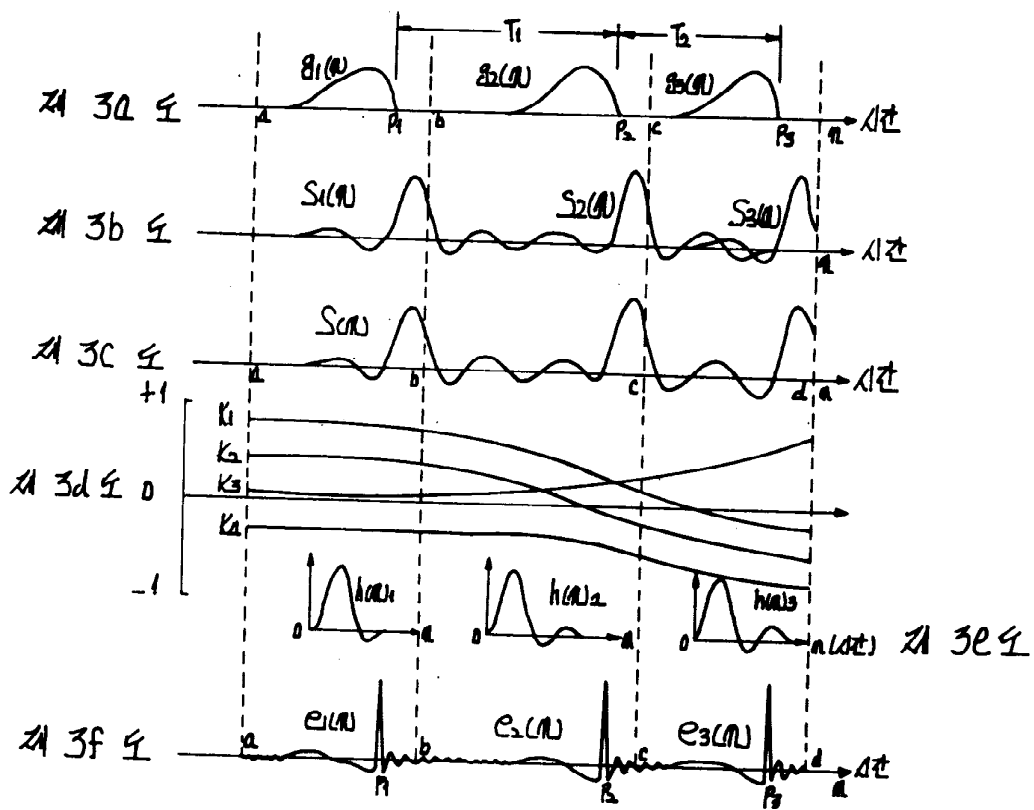
도면2



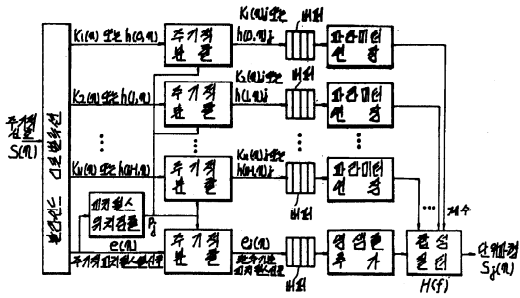
도면3-계속



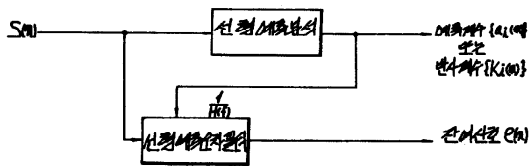
도면3



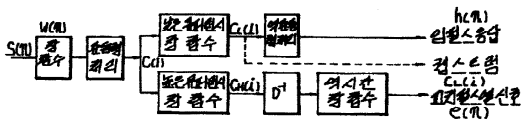
도면4



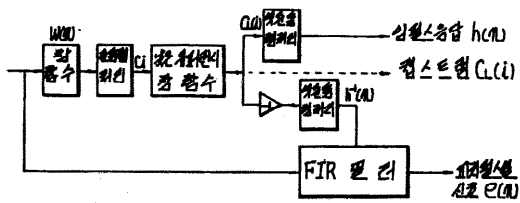
도면5a



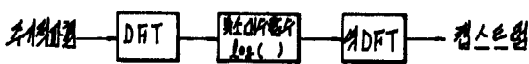
도면5b



도면5c



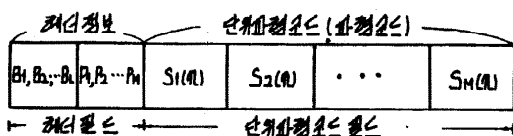
도면5d



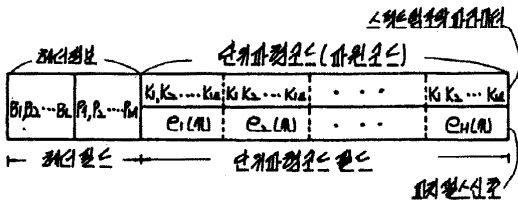
도면5e



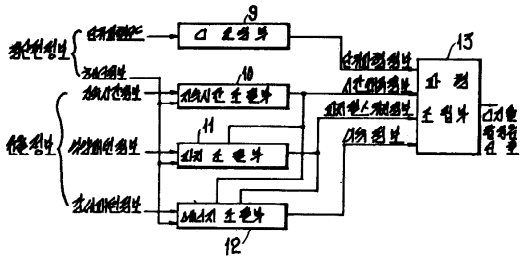
도면6a



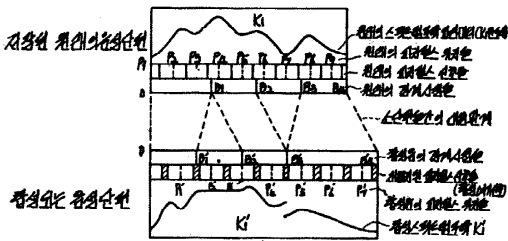
도면 6b



도면 7



도면 8a



도면 8b

