(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.[6]
G10L 13/00
G10L 9/18

(45) 공고일자 2004년01월28일
(11) 등록번호 10-0406674
(24) 등록일자 2003년11월10일

(21) 출원번호 10-1996-0042737
(22) 출원일자 1996년09월25일

(65) 공개번호 10-1997-0017173
(43) 공개일자 1997년04월30일

(30) 우선권주장 95-250983 1995년09월28일 일본(JP)

(73) 특허권자 소니 가부시끼 가이샤
일본국 도쿄도 시나가와쿠 키타시나가와 6쵸메 7반 35고

(72) 발명자 니시구찌 마사유끼
일본국 도꾜도 시나가와구 기다시나가와 6쪼메 7방 35고 소니가부시끼가이샤내

마쯔모또 준
일본국 도꾜도 시나가와구 기다시나가와 6쪼메 7방 35고 소니가부시끼가이샤내

(74) 대리인 신관호

심사관 : 이승한

(54) 음성합성방법 및 장치

요약

정현파 합성기술을 이용하여 구성된 음성합성방법 및 장치는 정현파를 합성시 위상의 변환으로 야기되는 음질의 열화를 방지하도록 구성되어 있다. 복호화처리부는 부호화측으로부터의 데이터를 디코드한다. 복호화데이터는 배드프레임 마스크처리부를 통해서 유성음/무성음데이터로 변환된다. 그런다음, 무성음검출회로는 상기 데이터에서 무성음프레임을 검출한다. 두개이상의 연속하는 무성음프레임이 존재하는 경우, 유성음합성부는 기본파 및 그 고조파의 위상을 0 또는 π/2 등의 주어진 값으로 초기화시킨다. 이는 유성음프레임의 시점에서 무성음과 유성음사이에 변환된 위상을 초기화함으로써, 위상의 엇갈림으로 야기되는 합성음의 찌그러짐 등 음질의 열화를 방지할 수 있도록 해준다.

대표도

도 1

명세서

도면의 간단한 설명

도 1은 본 발명을 따르는 음성신호에 대한 분석/합성 부호화장치중 분석측(엔코드측)의 개략구성을 나타내는 기능블록도.
도 2a 및 2b는 윈도윙(windowing)처리를 설명하는 도면.
도 3은 윈도우처리와 윈도우함수간의 관계를 설명하는 도면.
도 4는 직교변환(FFT)될 시간축 데이터를 나타내는 도면.

도 5a, 5b 및 5c는 주파수축상의 스펙트럼데이터, 스펙트럼포락선 및 여기(勵起)신호의 파워스펙트럼을 나타내는 그래프.
도 6은 본 발명의 일실시예를 따르는 음성신호에 대한 분석/합성 부호화장치중 합성측(디코드측)의 개략구성을 나타내는 기능블록도.
*도면의 주요부분에 대한 부호의 설명
13. 피치추출부 14. 윈도우처리부
15. 직교변환(FFT)부 16. 고정밀피치탐색부
17. 유성음/무성음판별부 18V. 유성음진폭평가부
18U. 무성음진폭평가부 19. 데이터수변환부
20. 백터양자화부 21. 부호화처리부
22. 프레임 인터리브처리부 37. 유성음합성부
38. 무성음합성부

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 소위 MBE(Muliband Excitation; 멀티비트 여기(勵起))부호화방식 및 고조파부호화방식 등의 정현파합성(sinusoidal synthesis)을 이용하여 음성을 합성하는 방법 및 장치에 관한 것이다.
오디오신호의 시간영역 및 주파수영역에서 오디오신호(음성신호와 음향신호를 포함)에 대한 통계적인 특성과 인간의 청각특성을 이용하여 신호를 압축하는 각종의 부호화방법이 제안되어 있다. 이러한 종류의 부호화방법은 대체로 시간영역에서의 부호화방법, 주파수영역에 대한 부호화방법, 오디오신호를 분석 및 합성처리를 통해 행해지는 부호화방법 등으로 분류된다.
음성신호에 대한 고효율의 부호화방법은 MBE(Multiband Excitation;멀티밴드 여기)부호화방법, SBE(Sigleband Excitation;싱글밴드 여기)부호화방법, 고조파부호화방법, SBC(Sub-band Coding;대역분할 부호화)방법, LPC(Linear Predictive Coding;선형예측부호화)방법, DCT(Discrete Cosine Transform;이산코사인변환)방법, MDCT(modified DCT;변형이산코사인변환)방법, FFT(Fast Fourier Transform;고속푸리에변환)방법 등을 포함한다.
이러한 음성부호화방법중에서, MBE부호화방법 및 고조파부호화방법 등과 같이 음성을 합성하는 데 정현파합성을 이용하는 방법에 있어서는 고조파진폭 및 위 상데이터 등 엔코더로부터 부호화되어 보내진 데이터를 기초로 하여 진폭과 위상에 대해 보간을 행한다. 그러한 보간된 파라미터에 따라서, 이러한 방법들은 시간에 따라서 주파수와 진폭이 변하는 1개의 고조파 시간파형을 산출하여서, 그러한 시간파형과 동일한 시간파형에 고조파의 수를 합하여 파형을 합성한다.
그렇지만, 종종, 위상데이터의 전송이 제한을 받게 되어 전송비트레이트를 감소시키게 된다. 이러한 경우, 정현파를 합성하기 위한 그러한 위상데이터는 프레임경계에서 연속성을 유지하기 위해서 예측된 값이 될 수도 있다. 이러한 예측은 각 프레임에서 실행된다. 특히, 그러한 예측은 유성음프레임에서 무성음프레임으로의 천이 및 무성음프레임에서 유성음프레임으로 천이에 있어 연속하여 행해지도록 되어 있다.
무성음프레임에는 어떠한 피치도 존재하지 않는다. 따라서, 어떠한 피치데이터도 전송되지 않는다. 이는 위상이 예측될때, 그 예측적인 위상값이 정확한 값에서 차이가 발생함을 의미한다. 따라서, 예측적인 위상값이 각각 본래 기대되고 있던 0위상가산 또는 π/2위상가산에서 점차 엇갈리게 되는 결과로 귀결된다. 이러한 위상차의 발생은 합성음의 음질을 열화시킬 수도 있다.

발명이 이루고자 하는 기술적 과제

본 발명의 목적은 정현파합성을 행하여서 음성을 합성하는 처리를 수행하는 경우, 이러한 위상차에 의해 야기되는 역효과를 미연에 방지하는 음성합성방법 및 장치를 제공하기 위한 것이다.
상기 목적을 수행하는데 있어서, 본 발명의 일태양에 따르면, 음성합성방법 은 음성신호에서 얻어진 입력신호를 프레임단위로 구분하는 단계와, 각 프레임의 피치를 구하는 단계와, 상기 프레임이 유성음이나 무성음중 어느 것을 포함하는 지를 판별하는 단계와, 전단계에서 얻어진 데이터로부터 음성을 합성하는 단계를 포함하며, 여기서, 상기 프레임이 유성음을 함유하는 것으로 판별되는 경우에는, 상기 유성음을 피치의 기본파 및 그 고조파상에 합성하고, 상기 프레임이 무성음을 함유하는 것으로 판별되는 경우에는, 기본파 및 그 고조파의 위상을 주어진 값으로 초기화하는 것을 특징으로 하고 있다.
본 발명의 또다른 태양에 따르면, 음성합성장치는 음성신호에서 얻어진 입력신호를 프레임단위로 구분하는 수단과, 각 프레임의 피치를 구하는 수단과, 상기 프레임이 유성음이나 무성음중 어느 것을 포함하는 지를 판별하는 수단과, 이전의 수단에 의해 얻어진 데이터로부터 음성을 합성하는 수단과, 상기 프레임이 유성음을 함유하는 경우에는 피치

의 기본파 및 그 고조파에 유성음을 합성하는 수단과, 상기 프레임이 무성음을 함유하는 경우에는 기본파 및 그 고조파의 위상을 주어진 값으로 초기화하는 수단을 포함한다.
두개이상의 연속하는 프레임이 무성음으로 판별되는 경우에는, 기본파 및 그 고조파의 위상을 주어진 값으로 초기화할 수도 있다. 또한, 상기 입력신호는 음성신호로부터 디지털변환된 디지털음성신호와 그 음성신호를 필터링하여 얻어진 음성신호뿐만 아니라 음성신호에 대해 선형예측부호화동작을 수행하여 얻어진 LPC잔차(殘差)신호이어도 좋다.
상술한 바와같이, 무성음으로 판별되는 프레임의 경우에 있어서는, 정현파합 성용의 기본파 및 그 고조파의 위상은 주어진 값으로 초기화된다. 그 결과 이러한 초기화를 통해, 무성음프레임에 있어서 위상의 엇갈림으로 야기되는 음성의 열화를 방지할 수 있다.
또한, 두개이상의 연속하는 무성음프레임의 경우에 있어서도, 기본파 및 그 고조파의 위상은 주어진 값으로 초기화된다. 이는 피치에 대한 오검출로 인해, 유성음프레임이 무성음프레임으로 잘못 판별되는 것을 사전에 방지할 수 있다.
계속해서, 본 발명의 목적 및 이점은 첨부한 도면을 이용해 도시설명한 본 발명의 적절한 실시예에 대한 다음의 설명으로부터 명백해질 것이다.

발명의 구성 및 작용

본 발명을 따르는 음성합성방법은 MBE(Multiband Excitation;멀티밴드 여기)부호화방법, STC(Sinusoidal Transform Coding;정현파변환 부호화)방법 또는 고조파 부호화방법 등의 정현파합성 부호화방법일 수도, 그 정현파합성 부호화방법을 LPC(Linear Predictive Coding;선형예측부호화)잔차신호에 대하여 적용하는 방법일 수도 있으며, 여기서 부호화단위로써 기능하는 각각의 프레임이 유성음(V)인지 무성음(UV)인지가 판별되고, 무성음프레임을 유성음프레임으로 천이시, 정현파합성위상은 0이나 π/2등의 주어진 값으로 초기화된다. MBE부호화에 있어서는, 프레임을 복수의 대역으로 분할하며, 각각의 대역을 유성음나 무성음으로 판별한다. 모든 대역이 무성음으로 판별되는 프레임을 적어도 하나의 대역이 유성음으로 판별되는 프레임으로 천이시에, 정현파를 합성하기 위한 위상은 주어진 값으로 초기화된다.
이러한 방법은 무성음프레임으로부터 유성음프레임으로의 천이를 검출하지 않고, 단지 유성음프레임의 위상을 일정하게 초기화할 필요가 있다. 그렇지만, 피치에 대한 오검출은 유성음을 무성음으로 잘못 판별하는 것을 야기시킬 수도 있다. 이러한 것을 고려하여, 두개의 연속하는 프레임이 무성음으로 검출되는 경우나 3개의 연속하는 프레임 또는 3개의 프레임이상의 소정의 연속하는 수만큼의 프레임이 무성음으로 판별되는 경우에는 위상을 초기화하는 것이 바람직하다.
무성음프레임에 있어서, 피치정보를 보내는 대신에 다른 데이터를 보내는 시스템에 있어서는, 연속적인 위상예측이 곤란하다. 따라서, 그러한 시스템에 있어서는, 상술한 바와같이 무성음프레임에서 위상을 초기화하는 것이 훨씬 효과적이다. 이는 위상의 엇갈림으로 인해 음질이 열화되는 것을 방지한다.
이하, 본 발명을 따르는 음성합성방법의 구체적인 구성을 설명하기에 앞서 통상의 정현파합성을 행함으로써 실행되는 음성합성의 일예에 대한 설명을 기술한다.
부호화장치나 엔코더에서 음성합성을 위해 복호화장치나 디코더에 보내지는 데이터는 고조파와 스펙트럼포락선에 상응하는 진폭간의 간격을 나타내는 적어도 하나의 피치를 포함한다.
복호화측에서 정현파를 합성하기 위한 음성부호화방법으로는, MBE(멀티밴드 여기)부호화방법 및 고조파부호화방법이 알려져 있다. 여기서, MBE부호화방법을 아래에 간략히 기술한다.
MBE부호화방법은 음성신호를 각각의 주어진 샘플수(예를들어, 256샘플)마다 블록으로 분할하고, 그러한 블록을 FFT등의 직교변환처리를 통해 주파수축상에 스펙트럼데이터로 변환한 다음, 그 블록내의 음성피치를 추출하고, 그러한 피치에 의거한 간격으로 주파수축상의 스펙트럼데이터를 다수의 대역으로 분할하여, 각 분할된 대역이 유성음인지 무성음인지를 판별한다. 그 판별결과, 피치데이터 및 스펙트럼의 진폭데이터는 모두 부호화되어서 전송된다.
MBE부호화방법을 이용하는 음성신호용 합성 및 분석 부호화장치(소위 보코더(Vocoder))는 D.W. Griffin and J.S. Lim, 'Multiband Excitation Vocoder', IEEE Tran, Acoustic, Speech, and Signal Processing, vol.36, No.8, pp.1223 - 1235, Aug. 1988.에 개제되어 있다. 종래의 PARCOR(Partial Auto-Correlation;편자기상관)보코더는 음성을 모델화할 때, 각 블록이나 프레임마다에 유성음구간을 무성음구간으로 혹은 무성음구간을 유성음구간으로 절환하는 동작을 행한다. 반면에, MBE보코더는 음성을 모델화할 때, 유성음과 무성음을 (하나의 블록 또는 프레임내의) 주어진 시간의 주파수축영역에 확실하게 유지시킨다.
도 1은 MBE보코더의 개략적인 구성을 나타내는 블록도이다.
도 1에서, 음성신호는 입력단자(11)를 통해서 고역통과필터등의 필터(12)로 공급된다. 그러한 필터(12)를 통해서 DC오프셋성분과 최소한의 저역통과성분(200Hz이하)이 (예를들어, 200 3400Hz의 범위내의) 대역을 제한하기 위하여, 음성신호에서 제거된다. 필터(12)에서 출력된 신호는 피치추출부(13) 및 윈도우처리부(14)로 보내진다.
입력신호로는, 음성신호에 대해 LPC처리를 수행하여 얻어진 LPC잔차신호를 사용할 수도 있다. 이러한 처리에 있어서, 필터(12)의 출력은 LPC분석처리를 통해 구해진 α파라미터로 역필터링처리된다. 이러한 역필터링처리된 출력은 LPC잔차신호와 상응한다. 그런다음, 그러한 LPC잔차신호를 피치추출부(13)와 윈도우처리부(14)로 보낸다.
피치추출부(13)에서, 신호데이터는 각각의 블록이 소정수의 샘플(N)(예를들어 N=256)로 이루어진 다수의 블록으로 분할된다(혹은 신호데이터는 사각형의 윈도우로 절단된다). 그런다음, 각각의 블록내의 음성신호에 대한 피치가 추출된다. 도 2A에 나타낸 바와같이, 예를들어, 절단된 블록(256샘플)은 프레임사이에 L샘플(예를들어 L=160)로 이루어진 간격으로 시간축방향으로 이동된다. 인접하는 블록들간에 겹쳐진 부분은 (N-L)샘플(예를들어 96샘플)로 이루어진다. 또한, 윈도우처리부(14)는 1개의 블록(N샘플)에 대하여 허밍윈도우(humming window)등과 같은 소정의 윈도

우함수를 수행하는 동작을 행하여, 윈도우처리된 블록을 각각 1샘플(L샘플)로 이루어진 간격으로 시간축방향으로 연속해서 이동시킨다.
이러한 윈도우처리는 다음식으로 표현할 수 있다.

수학식 1

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>xw(k,q) = x(q) w(kL-q)</FONTID>

여기서 k는 블록수를 나타내고, q는 데이터의 타임인덱스를 나타낸다. 이 수학식 1은 k번째 블록의 윈도우처리함수(w(kL - q))를 처리전의 입력신호의 q번째 데이터(x(q))에 대해서 처리하여서 데이터(xw(k, q))가 얻어지는 것을 나타내고 있다. 피치추출부(13)에 있어서, 도 2a에 나타낸 바와같은 사각형 윈도우는 다음 의 윈도우처리함수(wr(r))에 의해 실현된다.

수학식 2

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>wr(r) = 1 0≤ r < N

= 0 r < 0, n≤ r</FONTID>

윈도우처리부(14)에서, 도 2b에 나타낸 바와같은 허밍윈도우용의 윈도우처리함수(wh(r))는 다음식으로 나타낼 수도 있다.

수학식 3

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>wh(r) = 0.54 - 0.46 cos (2π r/(N-1))

0≤ r< N</FONTID>

=0 r < 0, N≤ r

윈도우처리함수(wr(r) 및 wh(r))를 사용하는 경우에는, 수학식 1에 나타낸 윈도우처리함수(w(r) (w=(kL-g))의 0이 아닌 구간은 다음과 같다.
0 ≤ kL - q < N
이식을 변형하면, 다음과 같은 식이 유추된다.
kL - N < q ≤ kL
따라서, 사각형 윈도우에 있어서, 도 3에 나타낸 바와같이 kL - N < q ≤ kL일 때, 윈도우처리함수(wr)(kL - q) = 1이 된다. 또한, 상기 수학식 1 3은 N(=256)샘플의 길이를 갖는 윈도우가 L샘플씩 L(=160)샘플쪽으로 이동하는 것을 나타내고 있다. 수학식 2 또는 3에 의해 표시된 윈도우처리함수로 절단된 각 N점(0 ≤ r < N)에서의 0이 아닌 샘플열은 xwr(k, r), xwr(k, r)로 표시되어 있다.
윈도우처리부(14)에 있어서는, 도 4에 나타낸 바와같이, 수학식 3에 표시된 허밍윈도우가 적용된 1블록에 대한 256샘플의 샘플열(xwh(k, r))내로는 1792샘플의 0데이터가 삽입된다. 그 결과 시간축상의 데이터열은 2048샘플을 포함한다. 그런다음, 직교변환부(15)는 이러한 시간축상의 데이터열에 대하여 FFT(고속푸리에변환)등의 직교변환을 수행하는 동작을 행한다. 어떠한 0데이터도 삽입되지 않은 256샘플의 본래의 샘플열에 대한 FFT를 수행하기 위해서는 또다른 방법이 제공될 수 있다. 이 방법은 처리량을 줄이는 데 효과적이다.
피치추출부(피치검출부)(13)는 xwr(k, r)로 나타낸 샘플열(1블록의 N샘플)을 기초로 하여 피치를 추출하는 동작을 행한다. 피치를 추출하는 여러종류의 방법이 알려져 있는 바, 각각의 방법은 예를들어 각각 시간파형의 주기성, 스펙트럼의 주기적 주파수구조 또는 자기상관함수를 사용한다. 본 실시예에 있어서, 피치추출방법은 센터클립파형의 자기상관방법을 사용한다. 1블록내의 센터클립레벨은 1블록에 대해서 1클립레벨로 설정될 수 있다. 실제로, 클립레벨은 1블록을 서브블록으로 분할하고, 각각의 서브블록의 신호에 대한 피크레벨을 검출하고, 인접하는 서브블록들간의 피크레벨차가 커지게 되면 1블록내의 클립레벨을 점차적으로 혹은 연속적으로 변화하는 방법에 의해 설정된다. 피치주기성은 센터클립파형에 대해서 자기상관데이터의 피크위치상에서 결정된다. 구체적으로, 복수의 피크는 현재프레임에 대해서 (데이터(1블록의 N샘플)로부터 얻어진) 자기상관데이터로부터 구하여진다. 이러한 피크중에서 최대의 피크는 소정의 역치와 같거나 보다 클때, 최대의 피크위치가 피치주기성으로 설정된다. 그 외일때는, 현재프레임 이외의 다른 프레임 예를들면, 이전 또는 이후의 프레임에서 구해진 피치에 대하여 소정관계를 충족하는 피치범위내의, 예를들면 이전 프레임의 피치를 중심으로 ±20%의 범위내에 있는 또다른 피크가 구해진다. 그렇게 구해진 피크에 따라서, 현재프레임의 피치가 결정된다. 피치추출부(13)에서, 피치는 개루프에서는 대강 탐색된다. 그렇게 추출된 피치데이터는 고정밀피치탐색부(16)로 보내어지고, 여기서 피치에 대한 정밀한 탐색이 폐루프에서 실행된다. 또한, 센터클립파형대신에, 입력파형에 대하여 LPC분석을 행하여 얻어진 잔차파형의 자기상관데이터를 사용하여 피치를 구할 수도 있다.
고정밀피치탐색부(16)는 피치추출부(13)에 의해 추출된 정수치의 조악한 피치데이터와 직교변환부(15)에 의해 예를들어 고속푸리에변환된 주파수축상의 데이터를 수신한다. 고정밀피치탐색부(16)에 있어서는, 몇몇의 최적의 플로팅 정밀데이터가 조악한 피치데이터값을 중심으로 +측과 -측으로 준비된다. 이러한 데이터는 0.2 0.5의 단계로 구성된다. 이러한 고정밀탐색방법은 소위 합성방법에 의한 분석방법을 이용하며, 여기서 피치는 원음의 파워스펙트럼에 가장 근접한 스폿에 합성된 파워스펙트럼을 위치시키도록 선택된다.

이제, 피치를 정밀하게 탐색하는 고정밀탐색에 대한 설명을 기술한다. MBE보코더에 있어서는, 다음과 같이 직교변환(예를들어, 고속푸리에변환)된 스펙트럼데이터(S(j))를 주파수축상에 나타내고 있는 모델을 가정한다.

수학식 4

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>S(j) = H(j) | E(j) | 0 < j < J</FONTID>

여기서, J는 ωs / 4π = fs / 2에 대응하고, 샘플링주파수(fs = ωs / 2π) 가 8kHz일 경우에는, 예를들어, J는 4kHz에 대응한다. 수학식 4에서, 주파수축상의 스펙트럼데이터(S(j))가 도 5a에 나타낸 바와같은 파형을 갖는 경우, H(j)는 도 5b에 나타낸 바와같은 원래의 스펙트럼데이터(S(j)에 대한 스펙트럼포락선을 나타낸다. E(j)는 도 5c에 나타낸 바와 같이 동일한 레벨상의 주기적인 여기신호 즉, 소위 여기스펙트럼을 나타낸다. 즉, FFT스펙트럼(S(j))은 스펙트럼포락선(H(j))과 여기신호의 파워스펙트럼(｜E(j)｜)과의 곱으로써 모델화된다.
피치상에 결정된 주파수축상의 파형의 주기성을 고려하여, 여기신호의 파워스펙트럼(｜E(j)｜)은 주파수축상의 대역에서 1대역의 파형에 상응하는 스펙트럼파형을 반복적으로 배열함으로써 형성된다. 1대역의 파형은 1792샘플의 0데이터가 가산된, 즉, 1792샘플의 0데이터가 삽입된, 256샘플의 허밍윈도우함수로 이루어진 파형, 다시말해서, 시간축상의 신호로써 가정되는 파형에 대해 FFT를 수행하고, 결과의 주파수축상에 주어진 대역폭을 갖는 임펄스파형을 각각의 피치에 따라 절단함으로써 형성할 수 있다.
각각의 분할된 대역에 대해서는, H(j)의 대표값 즉, 각각의 분할된 대역에 대한 오차를 최소화하는 일종의 진폭(｜Am｜)을 구하기 위한 동작이 행해진다. m번째 대역 즉, m번째 고조파의 대역의 하한점과 상한점이 각각 am 및 bm을 지칭하는 것이라 가정하면, m번째 대역의 오차( )는 다음과 같이 표시된다.

수학식 5

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>??m= $\sum_{j=am}^{bm}$ {| S(j) | - | Am | | E(j) |} 2</FONTID>

오차( )를 최소화하는 ｜Am｜의 진폭은 따라서, 다음과 같이 표시된다.

수학식 6

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2> $\frac{????m}{??|Am|}$ = -2 $\sum_{j=am}^{bm}$ {| S(j) | - | Am | | E(j) |} | E(j) |

</FONTID><FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2≥ 0

</FONTID><FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>?? | Am | = $\sum_{j=am}^{bm}$ | S(j) | | E(j) | / $\sum_{j=am}^{bm}$ | E(J) | 2</FONTID>

이 수학식 6의 진폭(｜Am｜)은 오차( )를 최소화한다.
이 진폭(｜Am｜)을 각 대역에 대해서 구한다. 그런다음, 상기 수학식 5에 규정된 각 대역의 오차( )는 그러한 진폭(｜Am｜)에 의해서 구해진다. 이후, 모든 대역의 오차( )의 합( )을 구하기 위한 동작이 행해진다. 모든 대역의 오차합( )은 서로다른 비트인 몇몇 피치에 대하여 구해진다. 그런다음 그러한 피치들의 합( )을 최소화하는 피치를 구하는 동작이 행해진다.
구체적으로, 피치추출부(13)에 의해 구해진 거친 피치를 중심으로 하여, 상측과 하측의 여러종류의 피치가 0.25의 간격으로 준비된다. 서로다른 비트인 각각의 피치에 대해서는, 오차합( )이 구해진다. 수학식 6에 따르면, 수학식 5의 오차( )는 주파수축상의 데이터의 전력스펙트럼(｜S(j)｜)과 여기신호 스펙트럼(｜E(j)｜)을 사용하여 구해진다. 그런다음, 오차( )로부터 모든 대역의 오차합( )이 얻어진다. 이러한 오차합( )은 각각의 피트에 대해서 구 해진다. 최소의 오차합을 위해 피치는 최적의 피치로 결정된다. 상술한 바와같이, 고정밀피치탐색부는 예를들어, 0.25의 간격에서 최적의 고정밀한 피치를 구한다. 그런다음, 최적의 피치에 대한 진폭(｜Am｜)이 결정된다. 그러한 진폭치의 계산은 유성음의 진폭평가부(18V)에서 행해진다.
설명의 간략화를 위해, 상기 피치에 대한 고정밀탐색에 대한 설명에 있어서, 모든 대역은 유성음인 것으로 가정하고 있다. 그렇지만, 상술한 바와같이, MBE보코더는 무성음영역이 동일시간대의 주파수축에 존재하고 있는 모델을 채용하고 있다. 따라서, 각 대역에 대해서, 그 대역이 유성음인지 무성음인지를 판별해야 할 필요가 있다.
고정밀피치탐색부(16)로부터의 최적의 피치와 진폭평가부(유성음)(18V)로부터의 진폭(｜Am｜)은 유성음/무성음판별부(17)로 보내지며, 여기서 각 대역은 유성음인지 무성음인지가 판별된다. 이러한 판별은 NSR(잡음 대 신호비)를 사용한다. 즉, m번째의 NSR 즉, NSRm은 다음과 같이 표시된다.

수학식 7

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>NSR= $\frac{\sum_{j=am}^{bm} \{| S(j) | - | Am | | E(j) |\}^2}{\sum_{j=am}^{bm} | S(j) |^2}$ </FONTID>

만일 NSRm이 소정의 역치(Th $_1$)(예를들어, Th $_1$=0.2)보다 크다면, 즉 오차값이 주어진값보다 크다면, 그 대역에서의 |Am| |E(j)|의 |S(j)|에 대한 근사치가 적절하지 않은, 즉 다시말해 여기신호(|E(j)|)가 베이스로써 적당하지 않다는 것 이 판별된다. 이러한 대역은 무성음으로 판별된다.
입력음성신호가 8kHz의 샘플링주파수를 갖는 경우, 총대역폭은 3.4kHz(여기서 유효대역범위는 200 3400Hz)이다. 여성의 고음성에서 남성의 저음성까지의 피치래그(즉 피치주기성에 상응하는 샘플의 수)는 20 147의 범위에 있다. 그러한 피치주파수는 8000/147 54Hz에서 8000/20 = 400Hz로 변화한다. 이는 약 8 63피치의 펄스(고조파) 즉, 고조파수가 음성레벨(피치의 크기)에 따라서 8 63의 범위에서 변화한다.
본 실시예에 있어서, 각각의 고정된 주파수대역폭에서 분할된 각 주어진 수의 대역에 대해서는, 유성음/무성음판별의 결과가 수집(혹은 제거)되고 있다. 구체적으로, 유성음대역을 함유한 주어진 대역폭(예를들어, 0 4000Hz)을 N $_B$(예를들어 12)대역으로 분할하는 동작이 행해지며, 겹쳐지는 평균치를 소정의 역치(Th $_2$)(예로, Th $_2$=0.2)로 판별하여 그 대역이 유성음인지 무성음인지를 결정한다.
이어서, 무성음진폭평가부(18U)에 대한 설명이 기술될 것이다. 이 평가부(18U)는 직교변환부(15)로부터 주파수축상의 데이터, 피치탐색부(16)로부터의 정밀한 피치데이터, 유성음진폭평가부(18V)로부터의 진폭(|Am|)데이터 및 유성음/무성음판별부(17)로부터의 유성음/무성음판별에 대한 데이터를 수신한다. 진폭평가부(무성음)(18U)는 진폭에 대한 재평가를 행하는 동작을 행하며, 그 결과 진폭은 무성음으로 판별된 대역에 대해서 또다시 구해진다. 무성음대역에 대한 진폭(|Am|uv)은 다음으로부터 구해진다.

수학식 8

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2> $|Am|uv=\sqrt{\sum_{j=m}^{bm}|S(j)|^2/(bm-am+1)}$ </FONTID>

진폭평가부(무성음)(18U)는 데이터를 데이터수변환(일종의 샘플링변환)부(19)에 보내는 동작을 행한다. 이 데이터수변환부(19)는 피치에 따라서 주파수축상에 다른 분할대역수를 갖는다. 데이터의 수 특히, 진폭데이터의 수가 서로 다르기 때문에, 변환부(19)는 그 수를 일정하게 유지하는 동작을 행한다. 즉, 상술한 바와같이, 그 유효대역이 3400kHz까지의 범위에 있으면, 그 유효대역은 피치에 따라서 8 63으로 분할된다. (무성음대역의 진폭(|Am|uv)을 함유하는) 진폭데이터의 개수(mMX+1)는 8 63의 범위로 가변된다. 데이터수변환부(19)는 가변개수(mMX+1)의 진폭데이터를 일정개수(M)(예를들어, M=44)의 데이터로 변환하는 동작을 행한다.
본 실시예에 있어서는, 더미데이터(dummy data)를 주파수축상의 유효대역내의 1블록의 진폭데이터에 가산하는 동작을 수행하여 그 블록의 내측에서 최종의 데이터로부터 제 1의 데이터까지의 값을 보간하여서, 데이터수를 N $_F$로 확대하고, 그 확대된 데이터에 대해서 대역제한형의 O $_S$배의 오버샘플링처리를 수행하여 O $_S$배의 개수의 진폭데이터를 얻는다. 예를들면, O $_S$=8이 제공된다. O $_S$배의 개수의 진폭데이터, 즉 (mMX+1) × O $_S$ 진폭데이터는 선형보간처리되어서 그 진폭데이터의 개수를 N $_M$으로 확대시킨다. 예를들어, N $_M$=2048이 제공된다. N $_M$ 데이터를 솎아냄으로 써, 데이터는 일정개수(M)의 데이터로 변환된다. 예를들면, M=44가 제공된다.
데이터수변환부(19)로부터의 데이터 즉, 일정개수(M)의 진폭데이터가 벡터양자화부(20)로 보내지며, 여기서 주어진 수의 데이터가 벡터로 모여진다. 벡터양자화부(20)로부터의 양자화출력(의 주요부)와, 고정밀피치탐색부(16)로부터의 P 또는 P/2선택부(26)를 통해서 구해진 정밀한 피치데이터 및 유성음/무성음판별부(17)로부터의 유성음/무성음판별에 대한 데이터는 모두 부호화처리부(21)로 보내져서 이러한 데이터를 부호화하게 된다.
각각의 이러한 데이터는 N샘플, 예를들어 블록내의 256샘플의 데이터를 처리함으로써 얻어질 수 있다. 블록은 시간축상을 L샘플의 프레임단위로 전진하게 된다. 따라서, 전송될 데이터가 프레임단위로 얻어진다. 즉, 피치데이터, 유성음/무성음판별에 대한 데이터 및 진폭데이터가 모두 프레임주기에서 갱신된다. 유성음/무성음판별부(17)로부터의 유성음/무성음판별에 대한 데이터는 필요에 따라 12대역으로 감소 또는 제거된다. 모든 대역내에는, 유성음영역과 무성음영역사이에 1개이상의 구분위치가 제공된다. 일정조건이 충족되는 경우, 유성음/무성음판별에 대한 데이터는 저역측의 유성음이 고역측으로 확대되는 유성음/무성음판별데이터패턴을 나타낸다.
그런다음, 부호화처리부(21)는 예를들어 CRC부가 및 레이트 1/2콘벌루션부호부가 처리동작을 행한다. 즉, 피치데이터의 주요부, 유성음/무성음판별에 대한 데이터 및 양자화데이터가 CRC부호화가 되어 콘벌루션부호화가 된다. 부호화처리부(21)로부터의 부호화데이터는 프레임 인터리브처리부(22)로 인가되며, 여기서 데이터는 벡터양자화부(20)로부터의 일부(덜 중요한 부분)데이터와 함께 인터리브처리된다. 이어서, 인터리브처리된 데이터는 출력단자(23)로 출력되어서 합성측(디코딩처리측)으로 전송된다. 이 경우, 전송은 통신매체를 거치는 송/수신 및 기록매체로부터의 기록/재생을 포괄한다.
계속해서, 부호화측으로부터 전송된 상기 데이터를 기초로 하여 음성신호를 합성하는 합성측(디코드측)의 개략구성을 도 6을 참고로 하여 설명한다.
도 6에 있어서, 전송에 의해 야기된 신호열화 즉, 송/수신 및 기록/재생으로 인해 야기된 신호열화를 무시하면, 입력단자(31)는 도 1에 나타낸 바와같은 엔코더의 출력단자(23)로부터의 데이터출력신호와 대체로 동일한 데이터신호를 수신한다. 입력단자(31)로 전송되는 데이터는 프레임 디인터리브처리부(31)로 보내진다. 프레임 디인터리브처리부(31)는 도 1에 나타낸 바와같은 인터리브처리에 반대되는 디인터리브처리를 수행한다. 주요부에서 CRC 및 콘벌루션부호화된 데이터의 매우 중요한 부분, 즉 인코드처리측은 복호화처리부(33)에 의해 복호화되어서 배드프레임 마스크처리부(34)로 보낸진다. 나머지부분 즉, 덜 중요한 부분은 곧장 배드프레임 마스크처리부(34)로 보내진다. 복호화처리부(33)는 소위 비터비복호화처리 또는 CRC코드를 이용한 에러검출처리를 수행하는 동작을 행한다. 배드프레임 마스크처리부(34)는 보간처리를 통해 고빈도의 에러프레임의 파라미터를 구하는 동작을 행하여, 각각 피치데이터, 유성음

/무성음데이터 및 벡터양자화 진폭데이터를 따로따로 취한다.
배드프레임 마스크처리부(34)로부터의 벡터양자화 진폭데이터는 데이터가 역 양자화되는 역벡터양자화부(35)로 보내진다. 그런다음, 데이터는 데이터가 역변환되는 데이터수역변환부(36)로 보내진다. 데이터수역변환부(36)는 도 1에 나타낸 바와같은 데이터수변환부(19)의 동작에 반하는 역변환동작을 수행한다. 역변환된 진폭데이터는 유성음합성부(37) 및 무성음합성부(38)로 보내진다. 마스크처리부(34)로부터의 피치데이터도 유성음합성부(37) 및 무성음합성부(37)로 보내진다. 마스크처리부(34)로부터의 유성음/무성음판별에 대한 데이터도 또한 유성음합성부(37) 및 무성음합성부(38)로 보내진다. 또한, 마스크처리부(34)로부터의 유성음/무성음판별에 대한 데이터는 유성음/무성음프레임 검출회로(39)로도 마찬가지로 보내진다.
유성음합성부(37)는 예를들어 코사인합성처리를 통해 시간축상의 유성음파형을 합성하는 동작을 행한다. 무성음합성부(38)에 있어서는, 화이트노이즈가 대역통과필터에 의해 필터링처리되어 시간축상의 무성음파형을 합성하게 된다. 유성음합성파형 및 무성음합성파형은 가산부(41)에서 가산되어 합성된 다음, 출력단자(42)로 출력된다. 이 경우, 진폭데이터와 피치데이터 및 유성음/무성음판별에 대한 데이터가 각각의 1프레임(=L샘플, 예를들어 160샘플)에서 상기 분석시 갱신된다. 인접하는 프레임들 사이의 연속성을 강화하기 위해서, 즉 프레임들간의 접속을 평활화 하기 위해서, 각각의 진폭데이터와 피치데이터의 값은 각각 예를들어 1프레임의 중심위치의 데이터값으로 설정된다. 현재프레임의 중심과 다음프레임의 중심간(예를들어 파형을 합성할 때, 예를들어 분석프레임의 중심에서 다음의 분석프레임의 중심까지의 주어진 1프레임을 의미)의 각 데이터값은 보간처리를 통해 구해진다. 즉, 파형을 합성시 주어진 1프레임에서, 선단샘플점에서의 각 데이터값과 (다음의 합성프레임의 선단인) 종단샘플점에서의 각 데이터값이 주어져서, 이러한 샘플점들간의 각 데이터값을 보간처리를 통해 구하게 된다.
유성음/무성음판별에 대한 데이터에 따르면, 모든 그러한 대역들은 하나의 구분된 위치에서 유성음영역과 무성음영역으로 분리되게 된다. 그런다음, 이러한 분리에 따라서, 유성음/무성음판별에 대한 데이터가 각 대역에 대해 얻어질 수 있다. 상술한 바와같이, 이러한 구분위치가 조정될 수 있어, 저역측의 유성음대역이 고역측으로 확대될 수 있다. 분석측(엔코드처리측)이 일정수(예를들어 약 12개)의 대역으로 이미 감소된(제거된) 대역을 가지는 경우에는, 디코드처리측은 이러한 대역의 감소를 본래의 피치에 위치된 가변개수의 대역으로 복원하여야 한다.
이하, 유성음합성부(37)에서 실행될 합성처리에 대한 설명을 기술한다.
유성음이라 판별된 m번째 대역(m번째 고조파의 대역)에서 시간축상에 있는 (예를들어 160샘플의 L샘플로 이루어진) 하나의 합성프레임의 유성음(Vm(n))은 다음과 같이 표현된다.

수학식 9

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>Vm(n)=Am(n)cos(θ m(n)) 0≤ n < L</FONTID>

여기서 n은 합성프레임의 내측에 있는 시간인덱스(샘플수)를 나타낸다. 유성음이라 판별된 모든 대역의 유성음이 합해져서(ΣVm(n)) 최종의 유성음(V(n))을 합성하게 된다.
수학식 9의 Am(n)은 합성프레임의 선단에서 종단까지의 범위에서 보간처리된 m번째의 고조파의 진폭을 나타낸다. 가장 단순한 수단은 1프레임단위로 갱신된 진폭데이터의 m번째 고조파의 값을 선형보간처리하는 것이다. 즉, 합성프레임의 선단(n=0)에서 m번째 고조파의 진폭값을 A_{0m}, 합성프레임(n=L : 다음 합성프레임의 선단)에서 m번째 고조파의 진폭값을 A_{Lm} 이라 가정하면, Am(n)은 다음식으로 계산될 수 있다.

수학식 10

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>Am(n) = (L - n) A_{0m} / L + nA_{Lm} / L</FONTID>

이어서, 수학식 9의 위상(θm(n))은 다음식에 의해 구해질 수 있다.

수학식 11

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>θ m(n) = mω 01n + n²m (ω L1-ω 01) / 2L + ??0m + ??ω n</FONTID>

여기서, ??0m은 합성프레임의 선단(n=0)에서의 m번째 고조파의 위상(프레임의 초기위상)을 나타내고, ω01은 합성프레임의 선단(n=0)에서의 기본각주파수를 나타내며, ωL1은 합성프레임의 종단(n=L : 다음의 합성프레임의 선단)에서의 기본각주파수를 나타낸다. 수학식 11의 ??ω는 n = L일때 θm(L)와 같은 위상(fLm)을 만드는 최소의 ??ω로 설정된다.
임의의 m번째의 대역에 있어서, 프레임의 시점은 n = 0이고 프레임의 종점은 n = L이다. 프레임의 종점이 n = L인 경우, 주어지는 위상(psi(L)m)은 다음과 같이 계산된다.

수학식 12

<FONTIDKOR=6ENG=6HANJA=2JAPAN=3SPECIAL=2>psi(L)m = mod2π (psi(0)m + mL(ω 0 + ω L)/2)</FONTID>

여기서 psi(0)m은 프레임의 시점이 n = 0일때 주어지는 위상을 나타내고, ω0은 피치주파수를 나타내며, ωL은 프레임의 종점이 n = L일때 주어지는 피치주파수를 나타내고, mod2π(x)는 -π +π의 범위로 x의 기본값을 되돌리는 함수이다. 예를들어, x=1.3π일때는 mod2π(x) = 0.3π가 주어지고, x = -1.3π일때는 mod2π(x) = 0.7π가 주어

진다.
위상을 연속하여 유지시키기 위해서, 현재프레임의 종단에서의 위상(psi(L)m)의 값을 다음 프레임의 시점에서의 위상(psi(0)m)의 값으로 사용할 수도 있다.
유성음프레임이 연속할 때는, 각 프레임의 초기위상이 연속적으로 결정된다. 모든 대역이 무성음인 프레임은 피치주파수(ω)의 값을 불안정하게 하여, 그 결과 상기 법칙이 전체의 대역에 대해서 작용하지 않게 된다. 그러한 피치주파수(ω)에 적당한 상수를 사용하여도 어느정도의 예측은 가능하다. 그렇지만, 그러한 예측된 위상은 본래의 위상에서 점차적으로 변환된다.
따라서, 하나의 프레임내의 모든 대역이 무성음인 경우에는, 0 또는 π/2의 주어진 초기값은 프레임의 종단이 n = L 일때는 위상(psi(L)m)으로 대치된다. 이러한 대치는 정현파 또는 코사인파를 합성가능하게 해준다.
유성음/무성음판별에 대한 데이터를 기초로 하여, 무성음프레임 검출회로(39)는 모든 대역이 무성음인 두개 이상의 연속하는 프레임이 존재하는지 의 여부를 검출한다. 두개이상의 연속하는 프레임이 존재하는 경우에는, 위상초기화 제어신호가 유성음합성회로(37)로 보내지며, 여기서 위상이 무성음프레임으로 초기화된다. 위상초기화는 연속하는 무성음프레임의 간격에서 일정하게 수행된다. 연속적인 무성음프레임중 최종의 프레임이 유성음프레임으로 변환될 때, 정현파의 합성이 초기화위상에서부터 개시된다.
이는 연속하는 무성음프레임의 간격에서 위상의 엇갈림으로 인한 음질의 열화를 방지할 수 있도록 해준다. 3개의 연속하는 무성음프레임이 존재하는 경우, 피치정보대신에 다른 종류의 정보를 보내는 방식에 있어서는, 연속적인 위상예측이 곤란하게 된다. 따라서, 상술한 바와같이, 무성음프레임으로 위상을 초기화하는 것은 굉장히 효과적이다.
이어서, 무성음합성부(38)에서 수행되는 무성음 합성처리에 대한 설명을 기술한다.
화이트노이즈 발생부(43)는 시간축의 화이트노이즈 신호파형을 윈도우처리부(14)로 보낸다. 그 파형은 소정의 길이(예를들어, 256샘플)로 윈도우처리된다. 그러한 윈도우처리는 적당한 윈도우함수(예를들어 허밍윈도우)에 의해 수행된다. 그러한 윈도우파형은 STFT처리부(45)로 보내져서 그 파형에 대한 STFT(단기간푸리에변환)처리가 행해진다. 그 결과데이터는 화이트노이즈의 시간축 파워스펙트럼이 된다. 파워스펙트럼은 STFT처리부(45)에서 대역진폭처리부(46)로 보내진다. 대역진폭처리부(46)에서, 진폭(｜Am｜uv)은 유성음대역에 의해 승산되고 다른 유성음대역의 진폭은 0으로 초기화된다. 대역진폭처리부(46)는 진폭데이터와, 피치데이터 및 유성음/무성음 판별에 대한 데이터를 수신한다.
대역진폭처리부(46)로부터의 출력은 ISTFT처리부(47)로 보내진다. ISTFT처리부(47)에서는, 위상이 역STFT처리를 통해서 시간축의 신호로 변환된다. 역STFT처리는 본래의 화이트노이즈위상을 사용한다. ISTFT처리부(47)로부터의 출력은 오버랩 및 가산부(48)로 보내지며, 여기서 시간축상의 데이터에 적당한 무게를 가함으로써 오버랩 및 가산을 반복하여서 본래의 연속적인 노이즈파형을 복원하게 된다. 오버랩 및 가산의 반복은 시간축상의 연속적인 파형을 합성시키는 결과로 귀결된다. 오버랩 및 가산부(48)로부터의 출력신호는 가산부(41)로 보내진다.
합성부(37 및 38)에서 시간축으로 합성되어 되돌려진 유성음 또는 무성음신호는 적당한 고정혼합비로 가산부(41)에서 가산된다. 재생된 음성신호는 출력단자(42)로 출력된다.
본 발명은 상기 실시예에 한정되지 않는다. 예를들어, 도 1에 나타낸 음성합성측(엔코드측)의 구성과 도 6에 나타낸 음성합성측(디코드측)의 구성은 하드웨어의 관점에서 설명되었다. 대신에, 이러한 구성을 소프트웨어프로그램, 구체적으로, 소위 디지털신호프로세서로 실현할 수도 있다. 각 고조파에 대한 대역의 주어진 수의 대역으로의 수집(제거)을 반드시 행할 필요는 없다. 이는 필요할때 행하여도 좋다. 주어진 대역의 수가 12개로 한정되지는 않는다. 또한, 주어진 구분위치에서 모든 대역의 저역유성음영역과 고역무성음영역으로의 분할을 반드시 실행할 필요가 없다. 더구나, 본 발명의 적용은 멀티밴드 음성분석/합성방법에 제한되지 않는다. 대신에, 본 발명은 정현파합성을 통해서 실행되는 여러 종류의 음성분석/합성방법에도 용이하게 적용될 수 있다. 예를들어, 그러한 방법은 각 프레임의 모든 대역을 유성음 또는 무성음으로 절환하고 CELP(Code-Excited Linear Prediction;부호여기선형예측)부호화시스템등의 또다른 부호화시스템을 무성음으로 판별된 프레임에 적용하도록 구성하는 방법이다. 또는, 그러한 방법은 각종 부호화시스템을 LPC(선형예측부호화)잔차신호에 적용하는 방법이다. 또한, 사용방법으로써, 본 발명은 신호의 전송과 기록 및 재생과, 피치변환과, 음성변환 및 잡음억제등의 여러종류의 사용방식으로 적용될 수도 있다.
본 발명의 진의 및 범위를 벗어나지 않는 한, 큰폭의 서로 다른 본 발명의 실시예들이 구성될 수도 있다. 본 발명은 첨부된 청구범위에 규정된 바를 제외하고는, 명세서에 설명된 특정한 실시예들에 한정되지 않는다는 것을 알아야 할 것이다.

발명의 효과

이상으로 살펴본 바와같이, 본 발명에 관계되는 음성합성방법 및 장치에 따르면, 무성음으로 판별된 프레임에서는, 정현파합성을 위한 기본파 및 그 고조파의 위상을 초기화하기 때문에, 무성음프레임에서 위상의 엇갈림에 따른 음질열화를 미연에 방지할 수 있다. 또한, 무성음프레임이 2프레임이상 연속하는 경우, 위상의 초기화를 행함으로써, 피치검출미스 등에 의한 유성음이 되어야 할 프레임이 무성음으로 판별되는 등의 오동작을 방지할 수 있다.

(57) 청구의 범위

청구항 1.

음성신호에서 얻어진 입력신호를 프레임단위로 구분하고, 각 구분된 프레임에 대하여 피치를 구하고, 유성음을 함유하는지 무성음을 함유하는지가 결정된 데이터에서 음성을 합성하는 단계들을 취하도록 구성된 음성합성방법에 있어

서,
상기 프레임이 유성음을 함유하는 것으로 판별되는 경우에는 유성음을 피치의 기본파 및 그 고조파와 합성하는 단계와,
상기 프레임이 무성음을 함유하는 것으로 판별되는 경우에는 상기 기본파 및 그 고조파의 위상을 주어진 값으로 초기화하는 단계와,
를 포함하여 구성된 것을 특징으로 하는 음성합성방법.

청구항 2.
제 1항에 있어서,
무성음을 함유하는 것으로 판별되는 프레임을 유성음을 함유하는 것으로 판별되는 프레임으로 변환하는 시점에서 기본파 및 그 고조파의 위상을 초기화하는 것을 특징으로 하는 음성합성방법.

청구항 3.
제 1항에 있어서,
무성음을 함유하는 것으로 판별되는 두개이상의 연속하는 프레임이 존재하는 경우, 기본파 및 그 고조파의 위상을 초기화하는 것을 특징으로 하는 음성합성방법.

청구항 4.
제 1항에 있어서,
상기 입력신호는 음성신호에 대해서 선형예측부호화동작을 수행함으로써 얻어진 선형예측부호화 잔차신호인 것을 특징으로 하는 음성합성방법.

청구항 5.
제 1항에 있어서,
기본파 및 그 고조파의 위상을 0 또는 π/2로 초기화하는 것을 특징으로 하는 음성합성방법.

청구항 6.
음성신호에서 얻어진 입력신호를 프레임단위로 구분하고, 각 프레임에 대해 피치를 구하고, 유성음을 함유하는지 무성음을 함유하는지가 결정된 데이터에서 음성을 합성하도록 구성된 음성합성장치에 있어서,
상기 프레임이 유성음을 함유하는 것이라 판별되는 경우에는 유성음을 피치의 기본파 및 그 고조파와 합성하는 수단과,
상기 프레임이 무성음을 함유하는 것이라 판별되는 경우에는 상기 기본파 및 그 고조파의 위상을 주어진 값으로 초기화하는 수단과,
를 포함하여 구성된 것을 특징으로 하는 음성합성장치.

청구항 7.
제 6항에 있어서,
상기 초기화수단은 무성음을 함유하는 것이라 판별되는 프레임을 유성음을 함유하는 프레임으로 변환하는 시점에서 상기 기본파 및 그 고조파의 위상을 초기화하는 것을 특징으로 하는 음성합성장치.

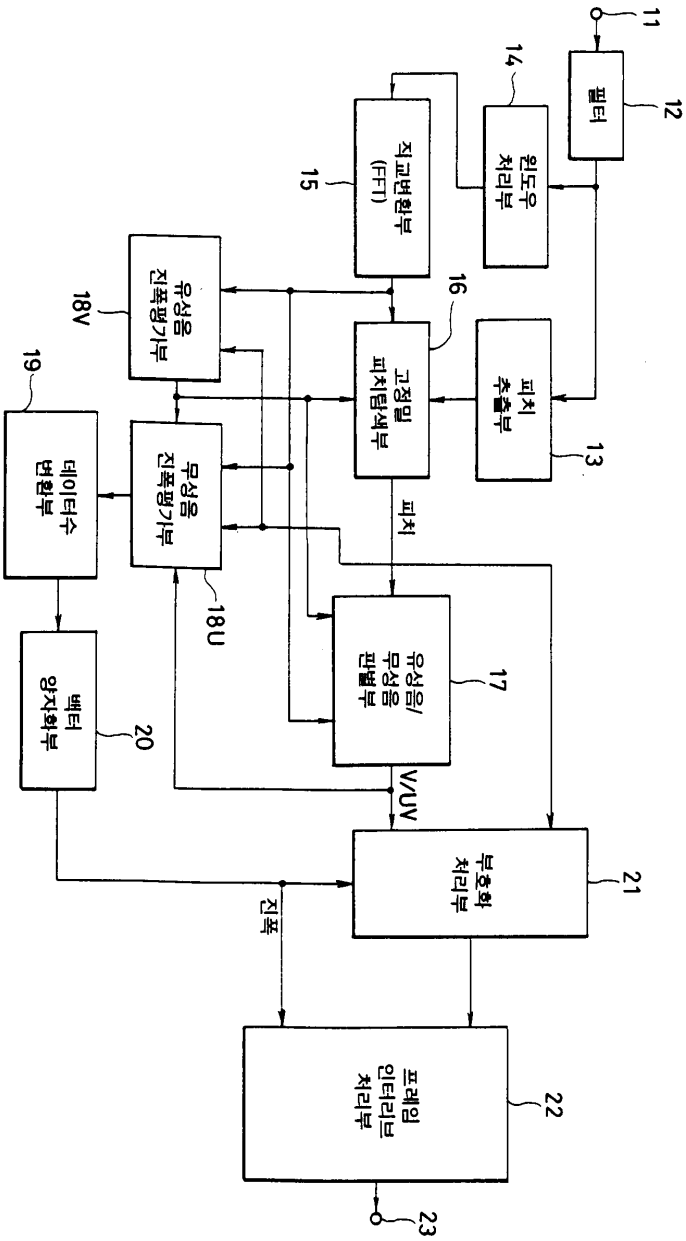
청구항 8.
제 6항에 있어서,
무성음을 함유하는 것이라 판별되는 두개이상의 프레임이 존재하는 경우, 상기 기본파 및 그 고조파의 위상을 초기화하는 것을 특징으로 하는 음성합성장치.

청구항 9.
제 6항에 있어서,
상기 초기화수단은 상기 기본파 및 그 고조파의 위상을 0 또는 π/2로 초기화하는 것을 특징으로 하는 음성합성장치.

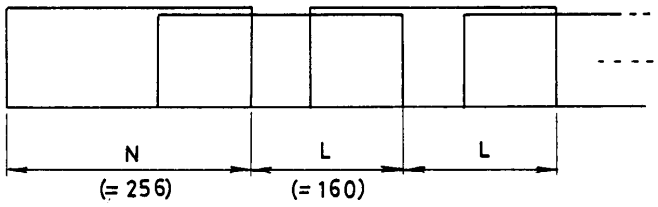
청구항 10.
제 6항에 있어서,
상기 입력신호는 음성신호에 대하여 선형예측부호화동작을 수행함으로써 얻어진 선형예측부호화 잔차신호인 것을 특징으로 하는 음성합성장치.

도면

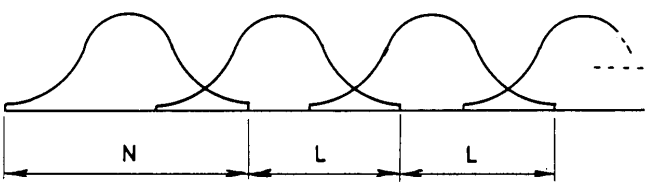
도면1



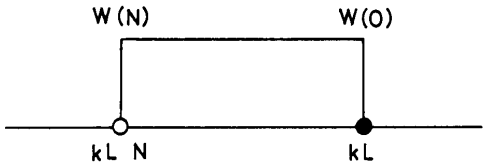
도면2a



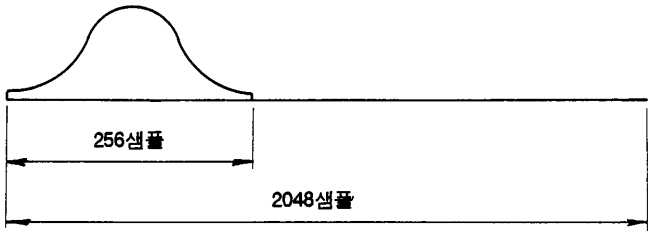
도면2b



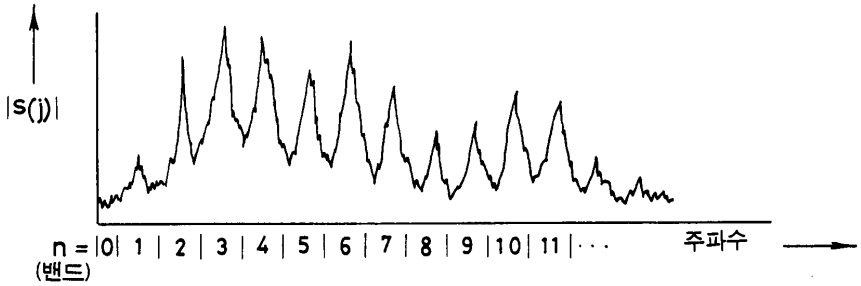
도면3



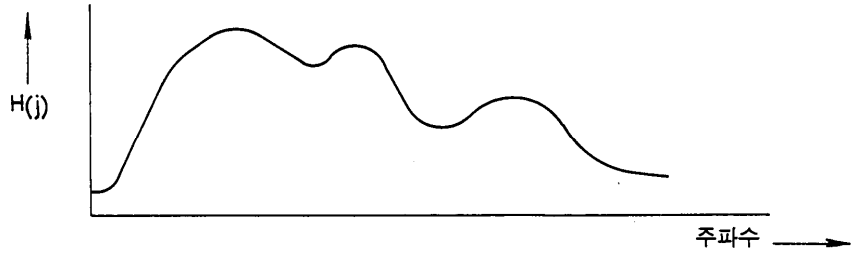
도면4



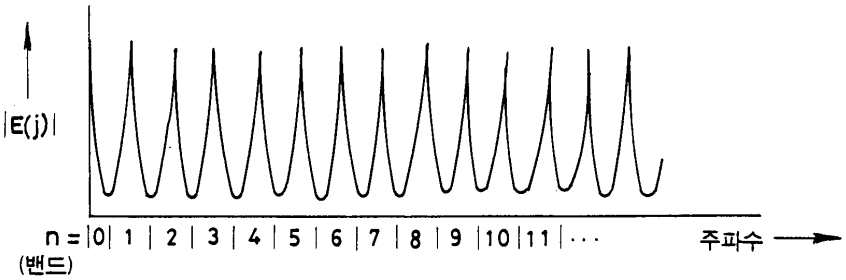
도면5a



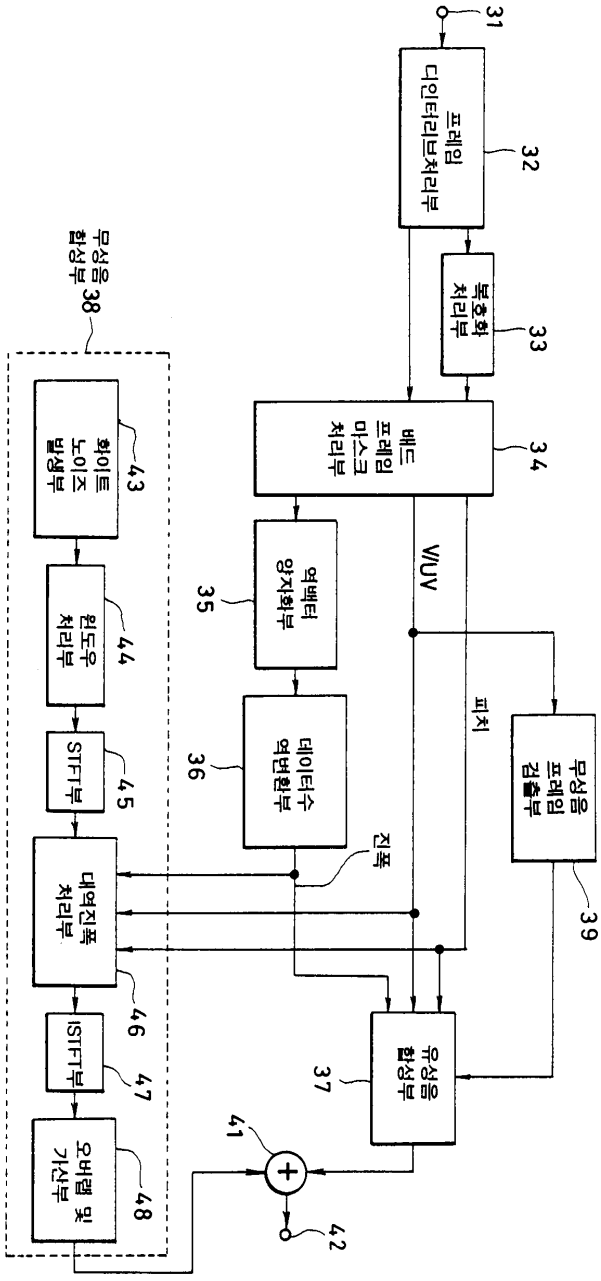
도면5b



도면5c



도면6


31
디인터리브처리부
32
복호화 처리부
33
배드 프레임 마스크 처리부
34
V/UV
역벡터 양자화부
35
데이터수 역변환부
36
피치
무성음 프레임 검출부
39
진폭
유성음 합성부
37
41
42
무성음 합성부 38
화이트 노이즈 발생부
43
윈도우 처리부
44
STFT부
45
대역진폭 처리부
46
ISTFT부
47
오버랩 및 가산부
48