

(19) 대한민국특허청(KR)
(12) 특허공보(B1)

(51) Int. Cl. ⁶ G11B 20/02	(45) 공고일자 1997년05월 15일	(11) 공고번호 특1997-0007661
(21) 출원번호 특1993-0003050	(24) 등록일자 1997년05월 15일	(65) 공개번호 특1993-0020409
(22) 출원일자 1993년02월27일	(43) 공개일자 1993년10월19일	
(30) 우선권주장 844,804 1992년03월02일 미국(US)		

(73) 특허권자	아메리칸 텔리폰 앤드 텔레그라프 캄파니	윌리엄 라이언
(72) 발명자	미합중국, 뉴욕 10013-2412 뉴욕, 애비뉴 오브 디 아메리카즈 32 제임스 데이빗 존스톤	
(74) 대리인	미합중국, 뉴저지 07059, 워렌, 밸리 뷰 로드 8 이병호, 최달용	

심사관 : 이상찬 (책자공보 제5008호)

(54) 스테레오포닉 오디오 신호의 입력세트 코딩방법

요약

내용없음.

대표도

도1

명세서

[발명의 명칭]

스테레오포닉 오디오 신호의 입력세트 코딩방법

[도면의 간단한 설명]

제1도는 본 발명을 응용할 수 있는 종류인 종래 기술의 예시적인 무선 통신/기억 시스템 및 그 향상 및 확장도.

제2도는 본 발명의 진보성과 가름침을 응용할 수 있는 예시적인 지각 오디오 코더(PAC)와 그 향상 및 확장도.

제3도는 임계값 계산에 사용된 유용한 마스킹 수준차 항목의 표현도.

제4도는 본 발명의 일면에 따른 예시적인 분석 필터 बैं크.

제5a도 내지 제5e도는 다양한 윈도우 함수의 동작도.

제6도는 윈도우 스위칭 기능을 예시하는 흐름도.

제7도는 출력 비트 스트림을 유도하기 위한 총체적인 입력 신호 처리를 예시하는 블록/흐름도.

제8도는 임의의 임계값 변화도.

제9도는 임의의 비트 할당 기능에 대한 흐름도 표현.

제10도는 비트 스트림 구성도.

제11a도 내지 제11c도는 임의의 호프만 코딩동작도.

제12도는 인코더 동작에 상보적인 디코더에서의 동작도.

제13도는 본 발명의 일면에 따른 임의의 양자화 동작을 예시하는 흐름도.

제14a도 내지 제14g도는 제4도의 필터 बैं크와 사용하기 위한 예시적인 윈도우 도면.

* 도면의 주요부분에 대한 부호의 설명

102 : 전처리기	104 : 지각 오디오 코더
106 : 통신채널/기억매체	108 : 지각 오디오 디코더
110 : 후처리기	202 : 분석 필터 बैं크
204 : 지각 모델 처리기	206 : 양자화기/비율 루프 처리기

208 : 엔트로피 인코더	302 : 버퍼
304 : 윈도우 증배기	306 : 윈도우 메모리
308 : FFT 처리기	310 : MDCT 처리기
311 : 연결기	312 : 지연 메모리
314 : 데이터 선택기	

[발명의 상세한 설명]

본 발명은 정보신호의 처리, 특히 기억시키거나 또는 전송하기 위해 음성과 음악정보를 나타내는 신호를 포함하는 모노포닉 및 스테레오포닉(monophonic and stereophonic) 오디오 신호의 효율적인 인코딩 및 디코딩에 관한 것이다.

높은 질의 오디오 신호를 기억시키고, 처리하며 그리고 통신하기 위한 소비자, 산업, 스튜디오 및 실험용 제품의 수요가 크다. 예를 들어, 소위 음악용 콤팩트디스크('CD')와 디지털 오디오 테이프('DAT') 레코딩이 오랫동안 인기를 끌던 축음기와 카세트 테이프를 전면적으로 교체시키고 있다. 게다가 또한, 최근에 이용 가능하게 된 디지털 오디오 테이프('DAT') 레코딩은 높은 질의 오디오 신호를 위해 보다 큰 유연성과 보다 높은 기억 밀도를 제공한다. 역시 Tan과 Vermeulen의 '데이터 기억용 디지털 오디오 테이프' IEEE 스펙트럼, PP.34~38(1989년 10월)을 참조한다. 게다가 CD와 유사한 질을 제공하는 디지털 기술을 방송에 응용할 요구가 일어나고 있다.

이와 같이 부상하는 디지털 기술이 높은 질의 신호를 발생시킬 수 있지만, 종종 이와 같은 성능은 상당한 데이터 기억용량 또는 전송 대역을 희생시킴으로서만 달성된다. 따라서, 기억과 전송을 위해 높은 질의 오디오 신호를 압축시키려는 시도에 많은 연구가 이루어지고 있다.

전송 및 기억용으로 신호를 압축시키려던 종래의 대부분의 연구는, 신호의 소오스를 신호에 두는 중복을 감소시키는데 중점되고 있다. 그리하여, 1984년 Prentice-Hall 출판사, N.S. Jayant와 P.Noll의 '파형의 디지털 코딩'에 설명된 ADPCM, 서브 밴드 코딩 및 변화 코딩과 같은 기술은, 소오스 신호에 달리 존재하게 될 중복(redundancies)을 제거하는데 역점이 되고 있다.

다른 해결책에서, 소오스 신호에 존재하는 무관한 정보는 인간의 지각 체계 모델에 근거한 기술을 사용하여 제거시키는데 역점이 되고 있다. 이와 같은 기술은, 1987년 11월 Consumer Electronics, Vol. CE-33, No.4에서 E.F. Schroeder와 J.J. Platte의 'MSC' : Stereo Audio Coding with CD-Quality and 256 kBIT/SEC, '와'의 IEEE J.S.C.A Vol.6, No.2에서 Johnston의 Transform Coding of Audio Signals Using Noise Criteria에 설명되어 있다.

상술된 바와 같이, 즉 Johnston 논문의 지각 코딩(Perceptual Coding)은, 소정의 비트율(또는 이용가능한 비트의 재활당) 혹은 오디오 신호를 나타내는 총 비트수를 낮추는 기술에 관한 것이다. 이와 같은 코딩 형식에서, 원치않는 신호에 대한 마스킹 임계값(a masking threshold)은 소정의 신호에 대한 주파수 함수로서 식별되고 있다. 이때, 특히 소정의 신호의 신호성분을 나타내는데 사용된 양자화의 거칠기(the coarseness of quantizing)는, 코딩에 의해 유입된 양자화 노이즈가 임계값에 매우 근접할지라도 노이즈 임계값 이상으로 증가하지 않도록 선택된다. 그래서 유입된 노이즈는 지각 프로세스(the perception process)에서 마스킹된다. 이렇게 지각적으로 코딩된 신호에 대한 종래의 신호대 잡음비가 비교적 낮지만, 그럼에도 불구하고 인간에 의해 인지된 것과 같이 디코딩하게 되면 이와 같은 신호의 질은 높다.

1991년 8월 13일 등록된 Brandenburg등에 의한 미국특허 5,040,217은 이와 같은 인지를 고려하여 사용한 높은 질의 오디오 신호를 효율적으로 코딩하고 디코딩하는 시스템을 설명한다. 특히, 입력 신호의 '노이즈와 같은' 혹은 '톤과 같은' 질에 대해서 측정된 것을 사용하여, 상기 후자의 시스템에 설명된 실시예는 모노포닉 오디오 신호에 대해 매우 효율적인 코딩을 제공한다.

물론, 오디오 신호를 압축시키는데 사용된 코딩 기술이 자체적으로 바람직하지 않은 성분이나 혹은 인공물(artifacts)을 유입시키지 않는 것이 중요하다. 이것은 특히, 하나의 스테레오 채널에 상당하는 코드화 정보가 재생시키기 위해 디코딩 될 때 다른 스테레오 채널에 상당하는 코딩 정보와 간섭하거나 혹은 상호 작용하는 스테레오포닉 오디오 정보를 코딩할 때 특히 중요하다. 두 개의 스테레오 채널을 코딩하기 위한 구현상의 선택에는 소위 고정된 비트율로 동작하는 두 개의 독립적인 코더를 사용하는 '듀얼 모노(sual mono)' 코더가 있다. 이와 대조적으로, '조인트 모노(joint mono)' 코더는 두 개의 모노포닉 코더를 사용하나 결합된 하나의 비트율, 즉 두 코더의 비트율이 고정 비트율보다 작거나 혹은 같은 비트율을 공유하지만 그러나, 각 코더의 비트율 사이에 절충이 이루어질 수 있다. '조인트 스테레오' 코더는 부수적인 코딩 이득을 실현하기 위한 스테레오 쌍(the stereo pair)의 채널간 특성을 사용하려고 시도하는 코더이다.

특히 낮은 비트율로 스테레오 쌍의 두 채널을 독립적으로 코딩하게 되면 바람직하지 않은 많은 수의 심리음향학적 인공물(psychoacoustic artifacts)이 일어날 수 있는 것으로 밝혀지고 있다. 그 중에는 동적으로 이미지화된 신호의 국소화와 부합하지 않는 코딩 노이즈의 국소화와 관련된 것이 있다. 그리하여 인간의 스테레오포닉 지각처리는, 만약 이와 같이 부정확한 국소화를 피하려고 한다면 인코딩 처리에 제한을 가하는 것으로 보인다. 이와 같은 발견은, 적어도 낮은 주파수에 대해 존재하는 것으로 보이는 입체음향의 마스킹 수준차(binaural masking-level differences)에 대한 보고서와 일치하므로, 노이즈가 공간적으로 격리될 수도 있다. 이와 같은 입체음향 마스킹 수준차는 모노포닉 시스템에서 마스킹 될 수 있는 노이즈 성분을 탈마스킹(unmask)시키는 것이 고려된다. 예를 들어, 1982년, 플로리다, 오를랜드, 아카데미 출판사, B.C.J.Morre의 '감청의 심리소개 제2판', 특히 제5장을 참조한다.

스테레오포닉 콘텍스트내 심리음향학적 인공물을 감소시키는 한가지 기술은 ISO-WG11-MPEG-오디오 심리음향학 11[ISO] 모델을 사용한다. 상기 모델에서, 신호대 잡음비('SNR')의 두번째 한계값이 심리음향학 모델 내부의 신호대 잡음비에 인가된다. 그러나, 이와 같이 부수적인 SNR 제약성은 전형적으로, 저주파수에서 역시 코딩의 모노포닉 성능을 저하시키지만 부수적인 채널용량 또는(기억장치 응용에서) 부수적인 기

역 장치의 사용을 필요로 한다.

본 발명에 따라 높은 질을 갖는 오디오 채널의 스테레오 쌍을 코딩하는 방법과 장치에서 종래 기술의 한계가 극복되며 기술적 진보가 이루어진다. 채널간 중복과 무관성(irrelevancy)은, 디코딩 후 높은 질의 재생을 유지하면서 보다 낮은 비트율을 달성하는데 사용된다. 특히 스테레오포닉 코딩과 디코딩에 적합하지만, 본 발명의 장점은 또한 종래의 듀얼 모노포닉 스테레오 코더에서 실현될 수도 있다.

본 발명의 예시적인 실시예는 변형 이산 코사인 변환(MDCT)을 사용하는 필터 뱅크 아키텍처(a filter bank architecture)를 사용한다. 시스템에 들어올 수 있는 모든 범위의 신호를 코딩하기 위해, 예시적인 실시예는, 신호에 따른 형식으로 주파수와 시간 모두가 스위칭 된 L/R 레프트와 라이트(Left and Right)와 M/S(합/차) 모두를 양호하게 사용한다. 새로운 스테레오포닉 노이즈 마스킹 모델은 코딩된 스테레오포닉 신호내 인공물을 양호하게 검출하고 회피한다. 채널간 중복은 오디오 질을 저하시키지 않고 압축의 향상을 제공하는데 사용된다.

라이트와 레프트 오디오 채널 모두의 시간동작은 양호하고 정확하게 모니터되며 그 결과는 코딩처리의 잠정적인 해를 제어하는데 사용된다. 그래서, 한 면에서, 본 발명의 예시적인 실시예는 정상적인 MDCT 윈도우 또는 신호상태가 보다 짧은 윈도우를 표시할 때 중의 어느 하나로서 입력신호 처리를 제공한다. 게다가, RIGHT/LEFT 또는 SUM/DIFFERENCE 코딩 모델간의 동적 스위칭이, 시간과 주파수 모두에 제공되어 원치 않는 임체음향 노이즈 국소화를 제어함으로써 SUM/DIFFERENCE 신호의 과잉 코딩의 필요성을 방지하고 전역 코딩 이득을 극대화한다.

코더 출력을 형성하는데 유용한 유연성을 제공하는 전형적인 비트스트림 정의 및 비율 제어 루프가 설명된다. 채널간 무관성이 양호하게 제거되며 스테레오포닉 노이즈 마스킹이 향상됨으로써 합동으로 코딩된 스테레오포닉 쌍으로 재생된 오디오 질을 향상시킨다. 예시적인 실시예에서 사용된 비율 제어 방식(The rate control method)은 절대 임계값과 코더비율 한계 이하의 신호에 대한 마스킹 임계값 사시의 보간과 비율이 제한된 조건하에서 임계값 평가 방법을 사용한다.

본 발명의 전체적인 코더/디코더 시스템에 따르면, 채널 비트율 요구조건 혹은 기억 장치에 사용하기 위한 기억용량을 더욱 감소시키기 위해 향상된 호프만 유형의 엔트로피 코더/디코더를 사용하는 것이 바람직한 것으로 증명된다. 예시적으로 사용된 노이즈 없는 압축 방법은, 인지 임계값(the perceptual threshold)에 의해 표시될 수 있는 것과 같이 L, R, M과 S에 대한 주파수 샘플을 효율적으로 코딩하기 위해 주파수-분할 방법과 함께 호프만 코딩을 사용한다.

본 발명은, 종래 기술과는 다른 접근법을 사용하지만 종래 기술의 양자화기/비율-루프의 많은 제약과 비용을 피하면서 오디오 신호(즉, 분석 필터 뱅크에서 나오는 MDCT계수)를 양자화하는데 사용될 척도인수(the scale factors)를 결정하는 메카니즘을 제공한다. 본 발명에 따라 양자화된 오디오 신호는 종래 기술보다 노이즈가 보다 적고 보다 적은 비트로 인코딩시킨다.

상기 결과는 본 발명의 예시적인 실시예에서 달성되며, 그것에 의해 양자화될 각 스펙트럼 계수의 주파수에 해당하는 주파수를 히어(hearing at)하는 산출된 임계값으로부터 유도된 척도인수와 허용가능한 한계 내에서 양자화된 스펙트럼 계수가 인코딩될 때까지 상기 주파수를 히어하는 절대 임계값으로부터 유도된 척도 인수사이를 보간함으로써 사용된 척도인수가 반복적으로 유도된다.

[1. 개관]

본 발명의 설명을 간단히 하기 위해, 상세히 설명된 바와 같이 다음의 특허, 특허출원 및 출판물들 : K. Brandenburg 등에 의해 1991년 8월 31일 등록된 미국 특허 5,040,217; 1988년 12월 30일 출원된 발명의 명칭이 '오디오 신호의 지각 코딩'인 미국 특허출원 Serial No. 07/292,598 ; J.D Johnston의 '지각 노이즈 기준을 사용한 오디오 신호의 변환 코딩' 통신에서 선정된 분야에 대한 IEEE 저널 Vol. 6, No.2(1988. 2); 1988년 3월 10일 출원된 국제 특허출원(PCT) WO 88/01811; Brandenburg 등에 의해 1990년 3월 9일 출원된 발명의 명칭이 '하이브리드 지각 코딩'인 미국 특허출원 Serial No. 07/491,373; AES 90회 회의(1991)의 Aspect : 높은 질의 음악 신호에 대한 적응적 스펙트럼 엔트로피 코딩 : Johnston J.,의 노이즈 마스킹 기준을 사용한 지각 엔트로피의 평가, ICASSP(1998); J.D. Johnston의 광대역 스테레오 신호의 지각 변환 코딩, ICASSP(1989); 1987년 11월 Consumer Electronics에 대한 IEEE Trans, Vol. CE-33, NO.4에서 E.F. Schroeder와 J.J. Platte의 'MSC' : CD-품질과 256kBIT/SEC를 갖는 스테레오 오디오 코딩; 그리고 IEEE J.S.C.A(1988. 2) Vol. 6, No.2에서 Johnston의 노이즈 기준을 사용한 오디오 신호의 변환 코딩은 본 발명의 공개에서 참고로서 일체화된다.

설명을 간략히 하기 위해, 본 발명의 예시적인 실시예는 개별적인 기능블록('프로세서'로 명명된 기능 블록을 포함)을 구비하는 것으로서 제시되어 있다. 이러한 블록이 나타내는 기능은, 공유 혹은 전용 하드웨어의 사용을 포함하지만 여기에 한정되지 않으며, 소프트웨어를 실행할 수 있는 하드웨어의 사용을 통해 제공된다('프로세서'라는 용어의 사용은 소프트웨어를 실행할 수 있는 하드웨어에 전적으로 의미하는 것으로 해석되어서는 안된다). 예시적인 실시예는 AT&T DSP 16 혹은 DSP 32C와 같은 디지털 신호 프로세서(DSP) 하드웨어와 이하에 논의된 동작을 수행하는 소프트웨어를 구비할 수 있다. 하이브리드 DSP/VLSI 실시예는 물론 본 발명의 대규모 집적(VLSI) 하드웨어의 실시예도 역시 제공될 수 있다.

제1도는 본 발명의 예시적인 실시예를 일체화하는데 유용한 시스템의 전체 블록도이다. 도시된 수준에서, 제1도의 시스템은 종래 기술에 공지된 시스템을 예시하지만, 여기에 서술된 그것의 변형과 확장은 본 발명의 향상을 명확하게 할 것이다. 제1도에서, 아날로그 음성신호(101)는 그것이 샘플링(전형적인 48KHz에서)되는 전처리(102)로 공급되며 표준형으로 디지털 펄스 코드 변조('PCM') 신호(103)(전형적으로 16비트)로 변환된다. PCM 신호(103)는, PCM 신호를 압축시켜서 압축된 PAC 신호를 통신 채널/기억 매체(105)로 출력시키는 지각 오디오 코더(104)('PAC')에 공급된다. 통신 채널/기억 매체로부터 압축된 PAC 신호는 지각 오디오 디코더(107)로 공급되며, 이것은 압축된 PAC 신호를 압축해서(decompresses)하여 압축된 PAC 신호를 나타내는 PCM 신호(108)를 출력한다. 지각 오디오 디코더로부터, PCM 신호(108)는 PCM 신호(108)에 대한 아날로그 표현을 만드는 후처리(109)로 공급된다.

지각 오디오 코더(104)의 예시적인 실시예는 제2도의 블록도에 도시되어 있다. 제1도에 예시된 시스템의 경우에서와 같이, 제2도의 시스템은 임의의 종래 기술 시스템, 즉 Brandenburg등의 미국 특허 5,040,217에 공개된 시스템은 똑같이 설명할 수 있다. 그러나, 이하에 설명된 그 확장과 변형으로서 중요하고 새로운 결과가 얻어진다. 제2도의 지각 오디오 코더는 분석 필터 뱅크(102), 지각 모델 프로세서(204), 양자화기/비율-루프 프로세서(206) 및 엔트로피 코더(208)를 구비하는 것으로서 양호하게 간주될 수 있다.

제2도의 필터 뱅크(202)는, 신호처리 이득(즉, 중복추출)과 인간의 지각체계에 비추어 의미 있는 방식으로 필터 뱅크 입력의 맵핑 모두에 대한 일부 측정을 제공하는 방법과 같이 시간/주파수에서 입력 오디오 신호를 양호하게 변환한다. 양호하게 1986년 10월 IEEE Trans. ASSP, Vol. 34, No.5에서 J.P. Princen과 A.B. Bradley에 의한 '시간도메인 얼라이징 소거에 기초한 분석/합성 필터 뱅크 설계'에 설명된 공지된 변형 이산코사인 변환(MDCT)은 입력 신호를 변환시키는데 적용될 수도 있다.

본 발명에 유용하게 될 MDCT의 특성에는, 임계 샘플링 특성 즉 필터 뱅크로의 모든 n 샘플에 대해 필터 뱅크로부터 n 샘플이 얻어진다는 것이 있다. 추가적으로, MDCT는 하프-오버랩(half-overlap), 즉 변환 길이가 필터 뱅크로 시프트된 샘플수(n) 길이의 정확히 2배라는 것이다. 하프 오버랩은 우수한 분석 윈도우 주파수 응답을 제공함은 물론 각 필터탭(tap)으로 독립적으로 주사(injected)된 노이즈의 제어로 양호하게 다루는 방법을 제공한다. 게다가, 양자화없이, MDCT는 단지 샘플의 적분수를 지연시킬 뿐, 입력 샘플에 대한 정확한 재구성을 제공한다.

매우 효율적인 스테레오포닉 오디오 코더와 연결하여 사용하기 위해 MDCT가 양호하게 변형되는 한 면은, 정선된(critically) 샘플과 정확한 재구성 특성을 유지하는 식으로 강한 비-정적 성분(strongly non-stationary components)를 갖는 신호부분에 대한 분석 윈도우의 길이를 스위칭하는 능력을 제공한다는 것이다. 본 출원과 함께 출원된 Ferreira와 Johnston에 의한 '오디오 신호의 지각 코딩 방법 및 장치'(이하의 '필터 뱅크 출원'이라고 칭함) 미국 특허 출원은 제2도의 소자(202)의 기능을 수행하기 적합한 필터 뱅크를 설명한다.

제2도에 도시된 지각 모델 처리기(204)는 지각 중요도의 평가, 노이즈 마스킹 특성 또는 분석 뱅크내 다양한 신호성분에 대한 매우 두드러진 노이즈 플로어(noise floor)를 산출한다. 이때 이와 같은 양들을 나타내는 신호가 다른 시스템 소자에 제공되어 채널 또는 기억 매체에 보내질 데이터에 대한 필터링 동작 및 구성의 향상된 제어를 제공한다. 1988년 2월, 통신의 선정된 분야에 대한 IEEE J.에서 J. D. Johnston에 의한 '지각 노이즈 기준을 사용하는 오디오 신호의 변형 코딩'에 설명된 임계 밴드 분석에 의한 임계 밴드를 사용하기보다는, 본 발명의 예시적인 실시예는 임계값(thresholds)의 산출에 보다 정밀한 주파수 해를 양호하게 사용한다. 그래서 마지막에 인용된 Johnston 논문에서와 같이 전체 음조 메트릭(an overall tonality metric)을 사용하는 대신에, AES 89회 회의, 1990년 K. Brandenburg와 J. D. Johnston의 '세대 지각 오디오 코딩 : 하이브리드 코더,'에서 언급된 것에 음조 방법은 주파수에 따라 변함으로서 복잡한 신호에 보다 적합한 음조 평가를 제공한다.

지각 모델 처리기(204)에서 수행된 심리음향학적 분석은, 정상적인 MDCT 윈도우와 보다 짧은 윈도우 모두에 적합할 수 있는 바와 같이 L(레프트), R(라이트), M(합)과 S(차)에 대한 노이즈 임계값을 제공한다.

동작에서, 지각 모델 처리기(204)의 예시적인 실시예는 THR_L과 THR_R로 명명한 좌측과 우측 채널에 대한 임계값을 평가한다. 다음에 윈도우 임계값은 예시적인 35코더 주파수 분할(윈도우가 스위칭된 활성 블록의 경우에 있는 56분할) 각각에서 비교된다. 일정양 이하 전형적으로 2dB 만큼 우측과 좌측 사이에서 두 개의 임계값이 변하는 각 분할에서, 코더는 M/S 모드로 스위칭된다. 즉, 주파수의 상부 밴드에 대한 좌측 신호는 $M=(L+B)/2$ 로 교체되고, 우측 신호는 $A=(L-R)/2$ 로 교체된다. 마지막에 언급된 치환을 야기하는 실제 차이값은 비트를 제약과 다른 시스템 변수에 따라 변할 것이다.

임계값이 실제 M과 S 신호상에서 산출된 것으로서 L과 R에 대해 사용된 동일한 임계값 계산이 역시 M과 S 임계값을 위해 사용된다. 먼저, BTHR_m과 MLDs로 명명된 기본 임계값이 산출된다. 다음에, M과 S 신호의 스테레오 마스킹 공여(the stereo masking contribution)를 산출하기 위해 다음 단계가 사용된다.

1. 각 M과 S 임계값에 대해 부수적인 인수가 산출된다. MLD_m과 MLDs라고 하는 이 인수는, 제3도에 예시적으로 도시된 마스크 수준차 인수만큼 스프레드 신호 에너지(the spread signal energy)를 곱함으로써 (1988년 2월, 통신에서 선정된 분야에 대한 IEEE J.에서 J. D. Johnston의 '지각 노이즈 기준을 사용한 오디오 신호의 변형 코딩', 1990년 AES 89회 회의에서 K. Braudengurg와 J. D. Johnston의 '제2세대 지각 오디오 코딩 : 하이브리드 코더', Brandenburg등의 미국 특허 5,040,217에 유도된 바와 같이) 산출된다. 이것은 여러 가지 소오스에 도시된 마스킹 수준차에 근거하여 M과 S 채널에서 주파수 전체의 제2노이즈 검출 가능성을 산출한다.

2. M(THR_m)에 대한 실제 임계값은 $THR_m = \max(BTHR_m, \min(BTHR_s, MLD_s))$ 로서 산출되고 임계값 $m = \max(BTHR_m, \min(BTHR_s, MLD_s))$ 와 S에 대한 임계값은 $THR_s = \max(BTHR_s, \min(BTHR_m, MLD_m))$ 로서 산출된다.

사실, MLD 신호는 스테레오 탈마스킹(unmasking)의 기회가 존재하는 경우에 BTHR 신호를 치환한다. L과 R 임계값이 같기 때문에, 같지 않은 L과 R 임계값으로 인하여 M과 S 임계값 억제(depression) 문제를 고려할 필요가 없다.

제2도의 예시적인 코더에 사용된 양자화기와 비율 제어 처리기(206)는 분석 뱅크와 지각 모델로부터 출력을 얻어 비트, 노이즈를 할당하며, 응용하는데 필요한 비트율을 충족시키기 위해 다른 시스템 변수를 제한한다. 일부의 코더에서 이것은, 비트율을 고려할(명시적으로)필요없이 지각 모델의 두드러진 차(the just noticeable difference)가 결코 초과되지 않도록 양자화로만 구성될 수 있고, 일부 코더에서 이것은 비트율과 코딩노이즈간에 균형을 이루기 위해 왜곡과 비트율을 조정하는 복합적인 순환 루프 세트일 수도 있다. 특히 유용한 양자화기와 비트율 처리기는, 본 발명과 동시에 출원된 J. D. Johnston에 의한 발명의 명칭이 '지각 인코더/디코더용 비율 루프 처리기' (이하에 '비율 루프'라고 칭함)인 미국 특허 출원에 일체화되어 설명되어 있다. 또한 비율 루프 처리기(206)에 의해 바람직하게 수행되며 비율 루프 처리기에 설명되어 있는 것은, 양자화된 분석 신호에서 나오는 정보와 임의의 필수적인 측면정보를 수신하고 동기화를 삽입시켜 정보를 프레임화시키는 기능이다. 다시, 이것과 동일한 기능은 Brandenburg등에 의한 미국

특히 5,040,217에 포괄적으로 설명되어 있다.

엔트로피 코더(208)는 비율 제어 처리기(206)와 협동하여 노이즈 없는 부가적인 압축을 달성하는데 사용된다. 특히, 본 발명의 다른 면에 따라, 엔트로피 코더(208)는 양자화기/비율-루프(206)로부터 나오는 양자화된 오디오 신호 출력을 포함하는 입력을 양호하게 수신하고, 양자화된 오디오 신호에 대해 손실없는 인코딩을 수행하여 통신 채널/기억매체(106)로 압축된 오디오 신호를 출력한다.

예시적인 엔트로피 코더(208)는 양자화된 오디오 신호 각각을 인코딩하기 위해 최소-중복 호프만 코딩 기법의 독창적인 변형으로 구성된다. 호프만 코드는, Proc. IRE, 40 : 1098~1101(1952)에 D. A. Huffman에 의한 '최소 중복 코드 구성방법'과 T. M. Cover와 J. a. Thomas의 정보이론 요소, PP. 92~101(1991)에 설명되어 있다. 제2도 코더의 콘텍스트에 양호하게 사용된 호프만 코드의 유용한 적용은, 본 발명의 출원과 동시에 출원되어 본 출원의 양도인에게 양도된 J. D. Johnston과 J. Reeds(이하에 '엔트로피 코더 출원'이라고 칭함)에 의한 미국 특허 출원에 보다 상세히 설명되어 있다. 데이터 통신기술 분야에 숙달된 숙련자는, 공지된 Lempi-Ziv 압축방법을 포함하는 기타 노이즈 없는 데이터 압축기술을 사용하여 엔트로피 코더(208)의 선택적인 실시예를 구현하는 방법을 쉽게 알 수 있는 것이다.

제2도에 도시된 각 소자의 사용은 전체 시스템 기능의 콘텍스트에서 보다 상세히 설명될 것이다. 세부동작은 지각 모델 처리기(204)를 위해 제공될 것이다.

[2. 1 분석 필터 뱅크]

지각 오디오 코더(104)의 분석 필터 뱅크는 펄스 코드 변조('PCM') 디지털 오디오 신호(전형적으로 48KHz에서 샘플링된 16비트 신호)를 입력으로 수신하여 입력 신호의 개별적인 주파수 성분을 식별하는 입력 신호의 표현을 출력한다. 특히, 분석 필터 뱅크(202)의 출력은 입력 신호의 변형 이산 코사인 변환('MDC T')을 구비한다. IEEE ICASSP PP. 2161~2164(1984)에서 J. Princen 등에 의한 '시간도메인 얼라이징 소거에 기초한 필터 뱅크 설계를 사용하는 서브-밴드 변환 코딩'을 참조한다.

본 발명의 일면에 따른 예시적인 분석 필터 뱅크(202)는 제4도에 제시되어 있다. 분석 필터 뱅크(202)는 입력신호 버퍼(302), 윈도우 증배기(304), 윈도우 메모리(306), FET 처리기(308), MDCT 처리기(310), 연결기(311), 지연 메모리(312) 및 데이터 선택기(132)를 구비한다.

분석 필터 뱅크(202)는 프레임마다 동작한다. 프레임은 입력신호 버퍼(302)에 의해 보유된 2N PCM 입력 오디오 신호 샘플로서 편리하게 선정된다. 상술된 바와 같이, 각각의 PCM 입력 오디오 신호 샘플은 M 비트로 표현된다. 예시적으로, N=512와 M=16이다.

입력신호 버퍼(302)는 두 부분을 구비하는데, 첫 부분은 버퍼위치 1 내지 N 샘플을 구비하고, 두번째 부분은 버퍼위치 N+1 내지 2N에서 N 샘플을 구비한다. 지각 오디오 코더(104)에 의해 코딩될 각 프레임은 입력 오디오 신호의 연속하는 N 샘플을 입력신호 버퍼(302)를 시프트시킴으로서 정의된다. 보다 오래된 샘플은 새로운 샘플보다 보다 높은 버퍼 위치에 위치된다.

소정 시간에, 입력신호 버퍼(302)가 2N 오디오 신호 샘플의 프레임을 포함한다고 가정하면, 연속하는(the succeeding) 프레임은, (1) 버퍼위치 1내지 N에서 N 오디오 신호 샘플을 버퍼위치 N+1에서 2N으로 각각 시프트시키고 (위치N+1 내지 2N에 있는 이전의 오디오 신호 샘플은 중복기재되거나 혹은 삭제될 수 있다) (2) 버퍼위치 1 내지 N에서, 전처리기(102)에서 나오는 N 새로운 오디오 신호 샘플을 입력신호 버퍼(302)로 시프트시킴으로써 얻어진다. 그래서, 연속하는 프레임은 공통적으로 N 샘플을 포함하는데, 연속하는 프레임의 첫번째는 버퍼위치 1 내지 N에서 공통 샘플을 가지며, 연속하는 프레임의 두번째는 버퍼위치 N+1 내지 2N에서 공통 샘플을 갖는다는 것을 알 수 있다. 분석 필터 뱅크(202)는 임계적으로 샘플링된 시스템(a critically sampled system) (즉, 입력신호 버퍼(302)에 의해 수신된 모든 N 오디오 신호 샘플에 대해)으로 양자화기/비율-루프(206)에 N 스케일러 벡터(a vector of N scalars)를 출력한다.

입력 오디오 신호의 각 프레임은, 윈도우 증배기(304)가 7 구별되는 데이터 윈도우를 프레임에 인가할 수 있도록 입력신호 버퍼(302)에 의해 윈도우 증배기(304)에 제공된다. 각 데이터 윈도우는 '계수'라고 하는 스케일러의 벡터이다. 7모든 데이터 윈도우가 2N 계수(즉, 프레임내 오디오 신호 샘플이 존재하는 것과 같은 수)를 갖지만, 단지 7중 네 개만이 N/2비-제조 계수(즉, 프레임내 오디오 신호 샘플수의 1/4)를 갖는다. 이하에 설명된 바와 같이, 데이터 윈도우 계수는 MDCT 처리기(310)의 출력에 대한 지각 엔트로피(the perceptual entropy)를 감소시키는데 양호하게 선택될 수 있다.

데이터 윈도우 계수에 대한 정보는 윈도우 메모리(306)에 기억되어 있다. 윈도우 메모리(306)는 예시적으로, 액세스 메모리('RAM'), 판독전용 메모리('ROM') 혹은 기타 자기 혹은 광학 매체를 구비할 수 있다. 윈도우 증배기(304)에 의해 인가된 예시적인 7데이터 윈도우에 대한 도면은 제4도에 제시되어 있다. 제4도에 제시된 7데이터 윈도우 각각에 대한 전형적인 계수의 벡터는 부록 A에 제시되어 있다. 제4도와 부록 A 모두에서 알 수 있는 바와 같이, 일부의 데이터 윈도우 계수는 0과 같을 수도 있다.

데이터 윈도우가 2N 스케일러의 벡터이며 오디오 신호 프레임도 역시 2N 스케일러의 벡터라는 것을 염두 해두고, 데이터 윈도우 계수는 1대 1곱셈(즉, 첫번째 오디오 신호 프레임 스케일러는 첫번째 데이터 윈도우 계수에 의해 곱해지고, 두번째 오디오 신호 프레임 스케일러는 두번째 데이터 윈도우 계수에 의해 곱해진다. ...등)을 거쳐 오디오 신호 프레임 스케일러에 인가된다. 그래서 윈도우 증배기(304)는 7데이터 윈도우 중 하나를 입력신호 버퍼(302)에 의해 보유된 오디오 신호 프레임에 인가하기 위해 각각 2N 곱셈을 수행하며 병렬로 동작하는 7마이크로 프로세서를 구비할 수 있다. 윈도우 증배기(304)의 출력을 '윈도우된 프레임 벡터'라고 불리울 2N 스케일러의 7벡터이다.

상기 7윈도우된 프레임 벡터는 윈도우 증배기(304)에 의해 FFT 처리기(308)에 제공된다. 상기 FFT 처리기(308)는 각각의 7윈도우된 프레임 벡터상에서 기수-주파수 FFT를 실행한다. 상기 기수 주파수 FFT는 주파수에서 판정된 이산프리에 변환이다.

$$\frac{kf_H}{2N}$$

여기서, $K=1, 3, 5, \dots, 2N$ 이고, f_H =샘플링 비율의 1/2과 같다. 예시적인 FFT 처리기(308)는, 각각 윈도우 된 다른 프레임 벡터에서 동작하며 병렬로 동작하는 종래의 7디시메이션-인-타임(decimation-in time) FFT처리기를 구비할 수 있다. FFT 처리기(308)의 출력은, 집합적으로 'FFT 벡터'로 불리울 2N 복합항목의 7벡터이다.

FFT 처리기(308)는 지각 모델 처리기(204)와 MDCT 처리기(310) 모두에 7 FFT 벡터를 공급한다. 지각 모델 처리기(204)는 FFT 벡터를 사용하여 데이터 선택기(314)와 양자화기/비율-루프 처리기(206)의 동작을 지시한다. 데이터 선택기(314)와 지각 모델 처리기(204) 동작에 관한 세부사항은 이하에 제시되어 있다.

MDCT 처리기(310)는 FFT 처리기(308)로부터 수신된 7 FFT 벡터 각각에 대한 실수성분을 기초하여 MDCT를 수행한다. P MDCT 처리기(310)는 병렬로 동작하는 7마이크로프로세서를 구비할 수 있다. 상기 각 마이크로프로세서는 7 각 FFT 벡터중 하나에 기초하여 N 실수 스케일러의 7 'MDCT 벡터' 중 하나를 결정한다. 각각의 FFT 벡터(F(k))에 대해, 그 결과 MDCT 벡터(X(k))는 다음과 같이 형성된다.

$$X(k) = \text{Re}\{F(k)\} \cos\left[\frac{\pi(2k+1)(1+N)}{4N}\right] \quad | \quad 1 \leq k \leq N.$$

결과에서 중복 때문에, 2N가 아니라 단지 N에 대해서만 k를 실행할 절차 필요하다. 간략히, $N < k < 2N$ 에 대해 $X(k) = -X(2N-k)$. MDCT 처리기(310)는 연결기(311)와 지연 메모리(312)에 7 MDCT 벡터를 공급한다.

윈도우 증배기(304)와 관련하여 상술된 바와 같이, 7데이터 윈도우 중 4는 N/2 비-제로 계수(제4C 내지 4F로 참조)를 갖는다. 이것은 윈도우된 프레임 벡터 중 네 개는 단지 N/2 비-제로값만을 포함할 수 있다는 것을 의미한다. 그래서, 상기 4벡터의 비-제로값은 MDCT 처리기(310)로부터 출력되면 연결기(311)에 의해 길이 2N의 유일한 벡터로 연결될 수도 있다. 상기 벡터의 결과로 나타나는 연결은 다음 목적을 위해 하나의 벡터로서 취급된다. 그리하여, 지연 메모리(312)는 7보다는 4 MDCT 벡터로서 제시되어 있다.

지연 메모리(312)는, 일시적으로 저장하기 위해 MDCT 처리기(314)와 연결기(311)로부터 4 MDCT 벡터를 수신한다. 지연 메모리(312)는 필터 बैं크(202)를 거쳐 4 MDCT 벡터의 흐름상으로 한개의 오디오 신호 프레임의 지연(입력신호 버퍼(302)에 의해 정의된 바대로)을 제공한다. 지연은, (i) 연속하는 오디오 신호 프레임을 나타내는 MDCT 벡터의 가장 최근에 연속하는 2세트를 저장하고, (ii) 입력으로서 데이터 선택기(314)에 연속하는 벡터 세트의 오래된 벡터 세트를 전해증으로서 지연된다. 지연 메모리(312)는 크기를 갖는 랜덤 액세스 메모리(RAM)를 구비할 수 있다.

$M \times 2 \times 4 \times N$ 여기서 2는 연속하는 벡터 세트의 수이고, 4는 세트내 벡터수이며, N는 MDCT 벡터내 항목수이며 그리고 M은 MDCT 벡터항목을 표현하는데 사용된 비트수이다.

데이터 선택기(314)는, 필터 बैं크(202)로부터 양자화기/비율-루프(206)로 출력되며 메모리(312)에 의해 공급된 4 MDCT 벡터 중 하나를 선택한다. 상술된 것과 같이, 지각 모델 처리기(204)는 FFT 처리기(308)에 의해 공급된 FFT 벡터에 근거하여 데이터 선택기(314)의 동작을 지시한다. 지연메모리(312)의 동작으로 인해, 지각 모델 처리기(204)에 공급된 7 FFT 벡터와 데이터 선택기(314)에 동시에 공급된 4 MDCT 벡터는 동일한 오디오 입력 프레임에 기초한 것이 아니라, 그 보다는 2 연속하는 입력신호 프레임에 기초하고 있는데, 여기서 두개의 프레임은 프레임의 초기에 기초한 MDCT 벡터와 프레임의 나중에 기초한 FFT 벡터가 있다. 그리하여, 특정한 MDCT 벡터의 선택은, 다음의 연속하는 오디오 신호 프레임에 포함된 정보에 기초한다. 지각 모델 처리기(204)가 MDCT 벡터의 선택을 지시하는 기준은 이하 2. 2절에 설명되어 있다.

예시적인 스테레오 실시예를 위해, 상기 분석 필터 बैं크(202)가 각각의 좌측과 우측 채널에 대해 제공된다.

[2. 2 지각 모델 처리기]

지각 코더는, 부분적으로 필터 बैं크(202)로부터의 MDCT 정보와 같은 신호를 띠고 있는 정보의 양자화와 연관된 노이즈를 유입시킴으로서 높은 질의 오디오 신호를 정확하게 표현하는데 필요한 비트수를 성공적으로 감소시킨다. 물론, 그 목적은 상기 노이즈를 감지하기 어렵거나 혹은 유연한 방법으로 유입하는 것이다. 이와 같은 노이즈 셰이퍼(noise shaping)은 주로 주파수 분석 측정이며, 그래서 신호를 스펙트럼 표현(즉, 필터 बैं크(202)에 의해서 공급된 MDCT 벡터)으로 변환하며, 상기 신호에 의해 표시될 노이즈의 형태와 양을 계산하며 그것은 스펙트럼 값을 양자화시킴으로서 주입된다. 상기 및 다른 기본적인 동작은 제2도에 도시된 지각 코더의 구조에 제시되어 있다.

지각 오디오 커더(104)의 지각 모델 처리기(204)는, 예시적으로 그 입력을 연속적인 프레임에 대해 작용하는 분석 필터 बैं크(202)로부터 수신한다. 다음에 지각 모델 처리기 입력은 전형적으로 분석 필터 बैं크(202)로부터의 고속 푸리에 변환(FFT)을 구비한다. 이것은, 각각 윈도우된 프레임 벡터 중 하나에 대응하는 2N 복합항목(complex elements)의 7벡터 형태로 FFT 처리기(308)의 출력이다.

신호로서 양자화 노이즈를 마스킹하기 위해, 신호의 스펙트럼 내용과 신호의 특정한 스펙트럼 패턴의 지속기간을 반드시 고려하여야 한다. 이와 같은 두 가지 면은, 신호와 노이즈가 거의 일정한 상태-만약 히어링 시스템의 총 주기(the integration period)가 주어진다면-인 주파수 도메인에서의 마스킹과 신호와 노이즈가 와우각(cochlear) 필터 처리되는 시간 도메인에서의 마스킹과 관련되어 있다. 이와 같은 필터의 형태와 길이는 주파수에 의존적이다.

주파수 도메인에서의 마스킹은 동시 마스킹의 개념에 의해 서술된다. 시간 도메인에서의 마스킹은 전마스킹과 후마스킹(premasking and postmasking)의 개념에 특징이 있다. 상기 개념은 다음 문헌에 포괄적으로 설명되어 있는데, 예를 들어, 1990년 Springer-Verlag 출판사, E. Zwicker와 H. Fastl의 '심리음향학, 사실학, 그리고 모델'을 참조한다. 이러한 개념을 지각 코딩에 유용하도록, 이것들은 다른 방법으로 실시된

다.

동시 마스크링(Simultaneous masking)은 각각 노이즈 웨이핑 모델을 사용하여 평가된다. 만약 신호의 스펙트럼 내용과 그 설명이 노이즈와 같거나 혹은 톤과 같은 동작으로서 주어진다면, 이와 같은 모델은 각 스펙트럼 성분의 양자화 수준을 규정하는 가설적인 마스크링 임계값을 산출한다. 이 노이즈 웨이핑은 어떠한 지각차(any perceptual difference)를 유발하지 않고 본래의 신호에 유입될 수 있는 최대 노이즈 양을 표현한다. 지각 엔트로피(PE)라고 하는 측정법은 이와 같은 가설적 마스크링 임계값을 사용하여 트랜스퍼런트 인코딩을 위한 비트율의 이론적 하부한계(the theoretical lower bound)를 평가한다. 1989년, ICASSP에서 J. D. Johnston의 '노이즈 마스크링 기준을 사용한 지각 엔트로피의 평가'를 참조한다.

전마스크링(premasking)은, 노이즈보다 시끄러운 마스크 신호(the masker signal) 이전의 시간에서 시작하는 노이즈의 (불)가청가능성(the (in)audibility)에 특징이 있다. 노이즈 진폭은 지연이 증가함에 따라 반드시 보다 더 감쇄되어야 한다. 이러한 감쇄 수준도 역시 주파수에 의존한다. 만약 노이즈가 합성 원도우의 첫번째 절반만큼 감쇄된 양자화 노이즈라면, 실험에 따르면 수용가능한 최대지연은 각 1밀리초이다.

상기 문제는 매우 중요하며 양호한 코딩 이득을 달성하는데 직접적인 장애가 될 수 있다. 정적조건(stationary conditions)-가설인-인 경우, 코딩 이득은 보다 큰 변환에 대해 보다 크지만, 양자화 에러는 재구성된 시간 세그먼트가 시작될 때까지 전파한다(spreads). 그래서, 만약 1024 포인트의 변환 길이가 사용된다면, 디지털 신호는 48000Hz의 비율로 샘플되어, 신호 이전의 대부분 21밀리초에 노이즈가 나타나게 될 것이다. 이것은, 신호가 흔히 '공격(attack)'으로 알려진 시간도메인에 일시적으로 머무는(transient) 날카로운 모양(the form of a sharp)을 띠 때 특히 중요하다. 이 경우 양자화 노이즈는 공격 전에 가청가능하다. 이러한 효과는 사전에코(preecho)로 알려져 있다.

그리하여, 고정된 길이의 필터 बैं크는 우수한 지각 해결법이 아니며 더군다나 시간의 비-정적 영역에 대한 신호처리 해결법도 아니다. 이러한 문제점을 극복하기 위한 한가지 가능한 방법은, 분석/합성 원도우 길이를 감소시킴으로서 코더의 잠정적 해결을 향상시키는 것이라는 것을 다음에 나타나게 될 것이다. 이것은, 공격의 조건이 검출될 때 원도우 스위칭 메카니즘으로서 실시된다. 이 방법에서, 긴 분석/합성 원도우를 사용하여 얻어진 코딩 이득은, 보다 짧은 분석/합성 원도우로 스위칭할 필요와 함께 이러한 검출이 일어날 때만이 영향을 받게 될 것이다.

후마스크링은, 보다 강한 마스크 신호의 중지후에 남아 있는 노이즈의 (불)가청가능성에 그 특징이 있다. 이 경우 수용할 수 있는 지연은 거의 20밀리초이다. 만약 보다 크게 변환된 시간 세그먼트가 21밀리초(1024 샘플)동안 지속한다면, 이러한 상황을 다루는데 특별한 주의가 필요하지 않다.

[원도우 스위칭]

특정한 변환 세그먼트에 대한 지각 엔트로피(PE)의 추정은 그 세그먼트를 트랜스퍼런트하게 코드화시키기 위한 비트/샘플의 이론적 하부 단계를 제공한다. 전마스크링 보호와 관련되어 있는 그 메모리 특성으로 인하여, 이와 같은 측정은, 강한 신호의 비정적성질(즉, 공격)의 몇가지 상황이 존재할 때-이전의 세그먼트와 연관된-그 이전의 값보다 PE 값이 두드러진 증가를 나타낸다. 이러한 중요한 성질은, 사전에코를 줄이기 위해 원도우 스위칭 메카니즘을 활성화시키는데 사용된다. 이와 같은 원도우 스위칭 메카니즘은, 사용되어 오던, 즉 1990년, ISO/MPEG 오디오 코딩 보고서에 설명된 ASPEC 코더도 새로운 방법은 아니지만, 이면의 결정기술은 새로이 PE 정보를 사용하여 비-정적성질을 정확하게 국소화(to accurately localize the non-stationarity)시키고 스위칭을 동작시키기 위해 우측순간(the right moment)을 정의하는 것이다.

두개의 기본적인 원도우 길이, 즉 1024샘플과 256샘플이 사용된다. 전자는 약 21밀리초인 세그먼트 지속 기간에 상당하며 후자는 약 5밀리초의 세그먼트 지속기간에 상당한다. 짧은 원도우는, 큰 원도우(그러나 이것은 '다른' 잠정적 샘플 수를 표현한다) 만큼의 스펙트럼 데이터를 표현하기 위해 4의 세트로 연관되어 있다. 큰 것에서 짧은 원도우로 그리고 그 반대로 천이하기 위해서는 두 종류 이상의 원도우를 사용하는 것이 편리하다. 시작 원도우(A START Window)는 큰 것(규칙적)에서 짧은 원도우로 천이하며 정지(STOP) 원도우는 제5b도에 도시된 바와 같이, 그 반대로 천이한다. 본 주제에 유용한 정보인 상기 인용된 Princen 책자를 참고한다. 원도우 모두는 폭이 1024 샘플이다. 이것들은, 임계적으로 샘플된 시스템을 유지하며 또한 전이영역에서 시간 얼라이징 소거처리(the time aliasing cancellation process)를 보장하는데 유용하다.

채널간 중복과 무관성을 이용하기 위해, 동일한 종류의 원도우가 각 세그먼트의 우측과 좌측 채널에 사용된다.

신호의 정적동작(The stationarity behavior)은 두가지 수준에서 모니터된다. 첫번째는 규칙적인 큰 원도우에 의해, 그리고 필요하다면 짧은 원도우에 의해서이다. 따라서, 큰(규칙적) 원도우의 PE는 모든 세그먼트에 대해 계산되어 반면에 짧은 원도우의 PE는 단지 필요할 때만 계산된다. 그러나, 두 가지 종류에 대한 음조정보(the tonality information)는, 신호의 연속적인 변화를 따르기 위해 모든 세그먼트에 대해 갱신된다. 달리 언급되지 않는 한, 세그먼트에는 큰 규칙적 원도우의 길이인 1024 샘플이 존재한다.

제5a도는, 포인트 $\frac{N}{2}$ 으로부터 포인트 $\frac{3N}{2}$ 까지의 세그먼트가 분석될 때까지 모든 필터 가능성을 나타낸다. 도면에 연관되어 있는 것은 제6도의 흐름도로 이것은 모니터 시퀀스와 결정방법을 설명한다. 필요하다면 짧은 원도우의 시퀀스 이전에 시작 원도우를 삽입시킬 수 있도록 세그먼트의 세개 절반(three halves of a segment)을 버퍼에 유지시킬 필요가 있다. 제5a 내지 제5e도는 연속적인 세그먼트 사이의 50% 중복(overlap)을 명시적으로 고려하고 있다.

처리는, 512의 새로운 잠정적 샘플(나머지 512 샘플은 이전의 세그먼트에 속한다)로서 '새로운' 세그먼트를 분석함으로써 시작된다. 이와 같이 새로운 세그먼트의 PE와 이전의 세그먼트에 대한 차이 PE가 계산된다. 만약 후자값이 선정된 임계값에 도달한다면, 이때 현재의 세그먼트안에 비-정적의 존재가 선언되고 세부사항은 제5a도에 표현된 바와 같이 위치를 갖는 짧은 4원도우를 처리하여 얻어진다. 짧은 각 원도우의 PE 값은, 순서화된 시퀀스 즉 PE1, PE2, PE3와 PE4의 결과로 산출된다. 이 값으로부터, 강한 신호의

비-정적에 대한 정확한 시적이 유도된다. 단지 다섯개의 위치만이 가능하다. 이것은 제4a도에서와 같이 L1, L2, L3, L4 및 L5로서 식별된다. 앞으로 명확해지는 것과 같이, 만약 비-정적(non-stationarity)이

$\frac{N}{2}$ 부터 $\frac{15N}{16}$ 까지의 어딘가에서 발생하였다면, 그 상황은 이전의 세그먼트에서 감출되었을 것이다. PE1 값이 현재 세그먼트의 정적에 대한 유관한 정보를 포함하지 않는다는 것을 뜻한다. 짧은 윈도우의 평균 PE는 동일한 세그먼트의 큰 윈도우의 PE와 비교된다. 보다 작은 PE는, 보다 효율적인 코딩 상황을 나타낸다. 그리하여 만약 전자의 값이 후자의 값보다 작지 않다면, 이때 열화상황(a degenerate situation)에 직면하여 윈도우 스위칭 처리가 무산된다고 생각한다.

짧은 윈도우에 대해 정적에 대한 정보는, 이전의 윈도우의 PE 값에 대한 차에서 보다는 그 PE 값에 높이는 것으로 관찰되고 있다. 따라서, 선정된 임계값 보다 큰 PE 값을 갖는 첫번째 윈도우가 감출된다. PE2는 위치 L1과 동일시되며, PE3는 L2와 PE4는 위치 L3과 동일시된다. 어느 경우에서, 시작 윈도우는, 짧은 윈도우로 코드화될 현재의 세그먼트 이전에 놓인다. 처리를 완료하는데 정지 윈도우가 필요하다. 그러나, 두 가지 가능성이 존재한다. 만약 신호의 강한 비-정적이 시작하는 확인된 위치가 L1 혹은 L2이라면, 이때 이것은 짧은 윈도우 시퀀스 안쪽 깊숙이 존재하므로, 아무런 코딩 인공물이 발생하지 않으며 코딩 시퀀스는 제5b도에 서술되어 있다. 만약 L4, 이때 최악의 경우에 위치하면, 비-정적은 마지막 짧은 윈도우의 우측 가장자리에 매우 근접하여 시작할 수도 있다. 이전의 결과는 정지 윈도우를 놓는 것-코딩 상태에서-이 상황에서 상기 스위칭 포인트내 신호의 재구성을 두드러지게 저하시키는 것을 시종일관 보여주고 있다. 이와 같은 이유로, 정지 윈도우 이전에 다른 세트의 짧은 4윈도우가 놓인다. 결과로 나타나는 코딩 시퀀스는 제5e도에 제시되어 있다.

만약 짧은 PE 중 어느 것도 임계값 이상이 아니라면, 나머지 가능성은 L4 또는 L5이다. 이 경우, 짧은 윈도우 시퀀스의 범위 앞에 문제가 놓이며 버퍼내 첫번째 세그먼트는 규칙적인 큰 윈도우를 사용하여 즉시 코드화될 수 있다.

올바른 위치를 식별하기 위해, 다른 짧은 윈도우가 처리되어야 한다. 이것은 제5a도에 점선으로된 곡선으로 표시되어 있으며 그 PE 값, PE_{1+n1}도 역시 계산된다. 쉽게 알 수 있는 것과 같이, 이와 같이 짧은 윈도우는 이미 다음 세그먼트에 속한다. 만약 PE_{1+n1}이 임계값 이상이라면, 이때 위치는 L4이며, 제5c도에 도시된 바와 같이, 시작 윈도우 다음에는 정지 윈도우가 올 수도 있다. 이 경우에 양자화 노이즈의 전파는 짧은 윈도우 길이에 한정되지 않을 것이며 보다 우수한 코딩 이득이 얻어진다. 위치가 L5인 드문 상황에, 이때 코딩은 제5d도의 시퀀스에 따라 이루어진다. 이 경우 올바른 해인 것을 증명하는 방법은 PE_{2+n1}이 임계 이상일 것이라는 것을 확인함으로이다. PE_{2+n1}은 PE_{1+n1}로 확인된 윈도우가 바로 다음에 오는 짧은 윈도우(제5a 내지 5e도에 표시되어 있지 않다)의 PE이다.

각 세그먼트에 대해 이전에 언급한 것과 같이, 우측과 좌측 채널은 동일한 종류의 분석/합성 윈도우를 사용한다. 이것은, 적어도 한개의 채널이 스위치를 필요로 할때 채널 모두에 대해 스위치된다는 것을 뜻한다.

양호한 국소 심리음향학적 해를 표현할지라도, 낮은 비트율을 적용하는 경우 제5c도의 해는, 다음 세그먼트의 코딩 품질에 역효과를 일으킬 수 있는 터무니없이 많은 비트수를 요구한다는 것이 관찰되어 있다. 이와 같은 이유로, 상기 코딩해는 궁극적으로 억제될 수도 있다.

또한 짧은 윈도우가 사용될 때 재구성된 신호의 세부 사항은, 단지 규칙적인 큰 윈도우가 사용될때 보다 본래의 신호에 보다 가깝다는 것은 명확하다. 이것은, 공격(the attack)이 기본적으로 광대역 신호이며 단지 매우 짧은 시간 주기 동안에 대해서 정적으로 간주될 수 있기 때문이다. 짧은 윈도우가 큰 윈도우보다 보다 큰 잠정적해를 갖기 때문에, 이것은 변하는 스펙트럼 패턴이 보다 충실하게 따르고 재생할 수 있다. 다른 말로, 이것은 보다 정밀한 신호의 국보(시간에서) 양자화와 신호의 전역(주파수에서) 양자화간의 차이이다.

스테레오포닉 코더의 최종 마스킹 임계값은 모노포닉과 스테레오포닉 임계값의 조합을 사용하여 산출된다. 모노포닉 임계값이 각 채널에 대해 독립적으로 계산되지만, 스테레오 포닉임계값은 두 채널 모두를 고려한다.

우측 혹은 좌측 채널에 대한 독립적인 마스킹 임계값은, 톤 마스킹 노이즈와 노이즈 마스킹 톤에 대한 식을 포함하는 심리 음향학적 모델을 사용하여 계산된다. 후자는 노이즈마스킹 노이즈식에 대한 보존적인 개라근사 값으로서 사용된다. 모노포닉 임계값은 이전과 같이 동일한 절차를 사용하여 계산된다. 특히 음조측정(atonality measure)은, 신호를 보다 톤과 같거나 또는 노이즈와 같은 것으로서 식별하기 위해 마지막 세개의 세그먼트 전체에서 각 주파수 계수의 위상과 제곱의 전개를 고려한다. 따라서, 각각의 심리 음향학적 식은 다른 것보다 다소 부가가 되어 있다. 문헌에서 찾을 수 있는 상기 식은 보다 우수한 성능을 내기 위해 다음과 같이 갱신된다.

$$TMN_{ab} = 19.5 + \text{바크} \frac{18.0}{26.0}$$

$$NMT_{ab} = 6.56 - \text{바크} \frac{3.06}{26.0}$$

여기서 바크(bark) 바크 척도(Bark scale)의 주파수이다. 이 척도는, 수위 카클레이 필터(cochlear filter) 혹은 임계 밴드(critical bands)라고 부를 수 있는 것에 관련된 것으로, 교대로 베실라 멤브레인(basilar membrane)의 일정한 길이를 갖는 세그먼트와 동일시된다. 최종 임계값은 마스킹의 절대 임계값을 고려하고 또한 부분적인 전마스킹 보호를 고려하여 조정된다.

1. 첫째, 분할되어 구성될 수 있다. 각 분할은 연관된 하나의 바크 값을 갖는다. 상기 분할은 MDCT선 혹

은 임계 밴드의 1/3 어느 것 중 개략적으로 보다 넓게 해를 제공한다. 저주파수에서, 하나의 MDCT 선은 코더 분할을 구성할 것이다. 고주파수에서, 많은 선들이 하나의 코더 분할로 결합될 수도 있다. 이 경우, 연관된 바크값은 분할의 중간 바크포인트이다. 이와 같은 스펙트럼의 분할은 전파하는 함수에 대해 수용 가능한 해를 보장하는데 필요하다. 다음에 도시되는 것과 같이, 상기 함수는 이웃하는 임계밴드 가운데 마스킹 영향을 나타낸다.

2. 둘째, 스펙트럼이 밴드로 구성될 수도 있다. 밴드는 변수화일(a parameter file)로 정의된다. 각 밴드는, 최종 마스킹 임계값 벡터로부터 나오는 하나의 스케일 인수와 연관되어 있는 많은 스펙트럼선을 그룹 지운다.

3. 세째, 스펙트럼도 역시 부분(sections)으로 구성될 수도 있다. 부분에는 밴드의 정수가 있으며 동일한 호프만 코드북(code book)으로 코드화된 스펙트럼 영역을 표현한다는 것이 이하에 나타나게 될 것이다.

데이터 값에 대한 세개의 인덱스가 사용된다.

ω → MDCT 선 영역에서 주파수에 의해 계산이 인덱스된다는 것을 가르킨다.

b → 임계값 산출 분할 영역에서 계산이 인덱스된다는 것을 가르킨다. 그 영역에서 콘벌루션 혹은 합을 내는 경우에 합계 변수서 bb 가 사용될 수 있다.

n → 코더 밴드 영역에서 계산이 인덱스된다는 것을 가르킨다.

추가적으로 일부 기호도 또한 사용된다.

1. 산출 분할의 인덱스, b
2. 분할에서 최저 주파수 선, ω_{lowb}
3. 분할에서 최고 주파수 선 ω_{highb}
4. 분할의 중간 바크값, b_{valb}
5. 분할에 대한 톤 마스킹 노이즈(dB로) 값, TMN_b
6. 분할에 대한 노이즈 마스킹 톤(dB로) 값, NMT_b

다음 설명의 몇가지 점은 '확산 함수(spreading function)'를 일컫는다. 이것은 다음 방법에 의해 계산된다.

$$tmpx = 1.05(j-i),$$

여기서 i 는 확산(spread)인 신호의 바크값이며, j 는 확산하는(spread into) 밴드의 바크값이며, $tmpx$ 는 잠정적인 변수이다.

$$x = 8 \text{minimum}(tmpx-5)^2 - 2(tmpx-.5), 0)$$

여기서 x 는 잠정적인 변수이며, $\text{minimum}(a, b)$ 는 a 또는 b 중 더 음의 값을 복귀하는 함수이다.

$$tmpy = 15.811389 + 7.5(tmpx + .474) - 17.5(1 + (tmpx + .474)^2)^5$$

여기서 $tmpy$ 는 다른 잠정적인 변수이다.

$$\text{만약}(tmpy < -100) \text{이면 } \{ \text{sprdnf}^{(i, j)=0} \} \text{ 이고 그렇지 않으면 } \left\{ \text{sprdnf}(i, j) = 10^{\frac{(x + tmpy)}{10}} \right\} .c \text{ 이다.}$$

[임계값 산출 단계]

다음 단계는 코더에 사용된 SMRn을 산출하는 데 필요한 단계들이다.

1. 다른 1024 샘플세그먼트를 형성하기 위해 입력 신호의 512새로운 샘플을 연결(concatenate)한다. 제5a도 참조한다.
2. 2.0에 설명된 0-FFT를 사용하고 사인 윈도우를 사용하여 입력 신호의 복합 스펙트럼을 계산한다.
3. 예측된 r 과 ϕ 를 계산한다.

변환의 극성 표면이 계산된다. r_ω 과 ϕ_ω 는 변환된 세그먼트의 스펙트럼선에 대한 크기 및 위상 성분을 나타낸다.

예측된 크기(\hat{r}_ω)와 위상($\hat{\phi}_\omega$)은 선행하는 두개의 임계값 계산 블록(r 과 ϕ)으로부터 산출된다.

$$\hat{r}_\omega = 2r_\omega(t-1) - r_\omega(t-2)$$

$$\hat{\phi}_\omega = 2\phi_\omega(t-1) - \phi_\omega(t-2)$$

여기서 t 는 현재의 블록수를 표시하며, $t-1$ 은 이전의 블록 데이터를 인덱스하면, $t-2$ 는 이전에 임계값 산

출 블록으로부터의 데이터를 인덱스한다.

4. 예측 불가능 측정(C_ω)을 계산한다.

C_ω , 예측 불가능성 측정(the unpredictability measure)은 다음과 같다.

$$C_\omega = \frac{((r_\omega \cos \phi_\omega - \hat{r}_\omega \cos \hat{\phi}_\omega)^2 + (r_\omega \sin \phi_\omega - \hat{r}_\omega \sin \hat{\phi}_\omega)^2)^{.5}}{r_\omega + \text{abs}(\hat{r}_\omega)}$$

5. 임계값 계산 분할내 에너지 및 예측 불가능성을 계산한다.

각 분할내의 에너지(e_b)는 다음과 같다.

$$e_b = \sum_{\omega = \omega_{\text{low}_b}}^{\omega_{\text{high}_b}} r_\omega^2$$

그리고 가중된 예측 불가능성(c_b)는 다음과 같다.

$$c_b = \sum_{\omega = \omega_{\text{low}_b}}^{\omega_{\text{high}_b}} r_\omega^2 C_\omega$$

6. 분할된 에너지와 예측불가능성을 확산함수로서 컨벌브한다(convolve).

$$ecb_b = \sum_{bb=i}^{bmax} e_{bb} \text{sprdngf}(bval_{bb}, bval_b)$$

$$ct_b = \sum_{bb=i}^{bmax} c_{bb} \text{sprdngf}(bval_{bb}, bval_b)$$

ct_b 가 신호 에너지에 의해 가중되었기 때문에, cb_b 로 다시 노멀라이즈되어야만 한다.

$$cb_b = \frac{ct_b}{ecb_b}$$

동시에, 확산함수의 비-노멀화 성질로 인하여, ecb_b 가 다시 노멀화되어야 하며 노멀화된 에너지(en_b)가 계산된다.

$$en_b = \frac{ecb_b}{morm_b}$$

노멀화 계수($rnorm_b$)는 다음과 같다.

$$morm_b = \frac{1}{\sum_{bb=0}^{bmax} \text{sprdngf}(bval_{bb}, bval_b)}$$

7. cb_b 를 tb_b 로 변환한다.

$$tb_b = -.299 - .432 \log_e(cb_b)$$

각각의 tb_b 는 그 범위가 $0 \leq tb_b \leq 1$ 로 한정되어 있다.

8. 각 분할에 요구되는 SNR을 계산한다.

$$TMN_b = 19.5 + bval_b \frac{18.0}{26.0}$$

$$NMT_b = 6.56 - bval_b \frac{3.06}{26.0}$$

여기서 TMN_b 는 dB로 나타낸 톤 마스크링 노이즈이며 NMT_b 는 dB로 나타낸 노이즈마스크 톤이다.

요구되는 신호대 잡음비($sNnr_b$)는 다음과 같다.

$$SNR_b = tb_b TMN_b + (1 - tb_b) NMT_b$$

9. 제곱비(the power ration)를 계산한다. 제곱비(bc_b)는 다음과 같다.

$$bc_b = 10^{\frac{-SNR_b}{10}}$$

10. 실제 에너지 임계값(nbb)을 계산

$$nb_b = enb_b c_b$$

11. MDCT 선에 걸쳐 임계 에너지를 확산시켜 nb_ω 를 산출한다.

$$nb_\omega = \frac{nb_b}{\omega_{high_b} - \omega_{low_{b+1}}}$$

12. 절대임계값을 포함시켜 가청가능성(thr_ω)에 대한 최종 에너지 임계값을 산출한다.

$$\max(nb_\omega, abstr_\omega).$$

'절대임계표(Absolute Threshold Tables)'에 도시된 $abstr$ 의 dB 값은, $\pm 1/2|sb$ 의 사이파가 임계값 계산을 위해 사용된 MDCT에 갖는 수준과 관계된다. dB 값은, 실제로 사용된 MDCT 노말화를 고려한 후 에너지 영역으로 반드시 변환되어야 한다.

13. 사전 에코 제어

14. 마스크 비율에 대한 신호(SMR_n)를 계산한다. '코더의 밴드'에 대한 표는 다음을 나타낸다.

1. 밴드의 인덱스(n)

2. 밴드(n)의 상위 인덱스(ω_{high_n}), 하위 인덱스(ω_{low_n}) $\omega_{high_{n-1}+1}$ 과 같이 이전 밴드로부터 계산된다.

각 밴드를 더욱 분류하기 위해, 다른 변수가 만들어진다. 폭 인덱스($width_n$)는, 만일 n 이 지각적으로 협대역이라면 값 $width_n=1$ 로, 만약 n 이 지각적으로 광대역이라면 값 $width_n=0$ 으로 가정한다.

$$bval_{\omega_{high_b}} - bval_{\omega_{low_b}} < bandlength$$

$bandlength$ 는 초기화 루틴에 설정된 변수이다. 그렇지 않으면 후자의 경우가 가정된다. 이때, 만약 ($width_n=1$)이면, 코더밴드의 노이즈 수준($nband_n$)은 다음과 같이 계산된다.

$$nband_n = \frac{\sum_{\omega=\omega_{low_n}}^{\omega_{high_n}} thr_\omega}{\omega_{high_n} - \omega_{low_{n+1}}}$$

그렇지 않으면

$$nband_n = \text{minimum}(thr_{\omega_{low_a}}, \dots, thr_{\omega_{high_a}})$$

이 경우, 여기서 $\text{minimum}(a, \dots, z)$ 는 인수($a \dots z$)의 가장 부정적인 인수나 혹은 가장 작은 양의 인수를 복귀하는 함수이다.

디코더로 보내질 비율(SMR_n)은 다음과 같이 계산된다.

$$SMR_n = 10 \cdot \log_{10} \left(\frac{[12.0 * nband_n]^{0.5}}{\text{minimum}(abstr)} \right)$$

음조측정은 스펙트럼 분석 처리의 출력이기 때문에, 분석윈도우가 크거나 짧은 인수의 모든 경우에 대해 사인 형태를 갖는다는 것이 중요하다. 특히, 인수가 시작 혹은 정지 윈도우로 코드화되도록 선정될 때, 그것의 음조정보는, 사인 윈도우인 나머지 연산 즉 임계값 계산과 계수의 양자화를 고려하여 얻어지며, 적절한 윈도우로 얻어진 스펙트럼을 고려한다.

[스테레오포닉 임계값]

스테레오포닉 임계값은 몇가지 목표가 있다. 대부분의 시간에 두 개의 채널은 '유사'한 것으로 알려져 있다. 그리하여, 코딩 이득에서 변환될 수 있는 몇가지 상호관계가 존재한다. 두 채널의 장정적인 표현을 돌이켜보면, 이 상호 관계는 명확하지 않다. 그러나, 스펙트럼 표현은 양호하게 사용될 수 있는 흥미있는 많은 특징이 있다. 사실, 매우 실용적이며 유용한 가능성은 두 채널을 표현하기 위해 새로운 기준을 만드는 것이다. 이 기준에는 두개의 직교 벡터가 있는데, 이것은 다음의 선형 조합에 의해 정의된 벡터 합(SUM)과 벡터차(DIFFERENCE)이다.

$$\begin{bmatrix} \text{합} \\ \text{차} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \text{우측} \\ \text{좌측} \end{bmatrix}$$

사용되는 윈도우 길이를 갖는 이 벡터들은, 변환 처리가 선형 연산에 의해 정의되기 때문에 주파수내에서 발생된다. 이것은 연산상의 부하를 단순하게 하는 장점이 있다.

첫번째 목표는, 두 신호가 보다 상관성이 없는 표현을 갖게 하는 것이다. 대부분의 에너지를 상기 새로운 채널 중 하나에 집중시키는 것은 우측과 좌측 채널 사이에 존재하는 중복의 결과로서 이것은 항상 코딩 이들을 낳게 된다.

두번째 목표는, 우측과 좌측 채널의 양자화 노이즈를 상호 관련시켜 노이즈의 국소화 또는 언마스킹 효과(the unmasking effect)를 제어하는 것이다. 이 문제는, 만약 우측과 좌측 채널이 양자화되어 독립적으로 코드화된다면 발생한다. 이 개념은 다음과 같은 내용으로서 예시되는 데, 특정한 신호에 대한 마스킹 임계값이 계산되었다고 가정하면 두 가지 상황이 발생할 수도 있다. 첫째 임계값에 상당하는 노이즈의 양을 신호에 부가한다. 만약 동일한 노이즈를 갖는 동일한 신호를 두귀에 제시한다면 이때 노이즈가 마스크된다. 그러나, 만약 임계값에 상당하는 노이즈양을 신호에 부가하여 이 조합을 하나의 귀(one ear)에 제시한다면, 다른 귀에 대해 똑같은 동작을 하지만 이전의 노이즈와는 상관성이 없으며 이때 상기 노이즈는 마스크되지 않는다. 다시 마스킹을 이루기 위해, 양쪽 귀에 있는 노이즈는 마스킹 수준차(MLD)에 의해 주어진 수준으로서 감소되어야만 된다.

언마스킹 문제는 다음 형식으로 일반화될 수도 있는데, 만약 마스킹 신호의 국소화를 따르지 않는다면 양자화 노이즈는 마스크되지 않는다. 여기서, 특히, 두 가지 한정적인 경우가 있는데, 그것은 언마스킹이 청취자의 쪽에서 보다 눈치챌 수 있는 신호의 중앙 국소화와 언마스킹이 중심선에서 보다 눈치챌 수 있는 신호의 측면국소화이다.

새로운 벡터 합과 차는, 청취자 중앙에 그리고 또한 양쪽에 국소화된 신호를 표현하기 때문에 매우 편리하다. 또한, 이것으로서 중앙과 측면이미지를 갖는 양자화 노이즈를 제어하는 것이 가능케 된다. 그리하여, 언마스킹 문제는, 상기 벡터를 거치는 MLD에 대한 보호수준을 제어함으로써 해결된다. 몇가지 심리음 방향적 정보와 다른 실험 그리고 그 결과에 기초하여, MLD 보호는, 특히 양 3KHz인 매우 낮은 주파수에 대해 특히 중요하다. 그 음조 특성이 아니라 신호력에만 의존하는 것처럼 보인다. MLD에 대한 다음 식은 좋은 결과를 산출하는 것으로 증명되었다.

$$MLD_{dB}(i) = 25.5 \left[\cos \frac{\pi b(i)}{32.0} \right]^2$$

여기서 i 는 스펙트럼의 분할인덱스([7]을 참조)이며, $b(i)$ 는 분할(i) 중심의 바크주파수이다. 상기 식은, 단지 $b(i) < 16.0$ 즉 3KHz 이하의 주파수에 대해서만 유효하다. MLD 임계값에 대한 식은 다음과 같다.

$$THR_{MLD}(i) = C(i) 10^{-\frac{MLD_{dB}(i)}{10}}$$

$C(i)$ 는 분할(i)에만 대응하는 베실라 멤브레인에 대한 확산 신호 에너지이다.

세번째 그리고 마지막 목표는, 이미지에 의해 마스크된 신호의 방향(directions)으로부터 무관성(irrelevance)을 추출하기 위해 특정한 스테레오포닉 신호 이미지의 장점을 이용하는 것이다. 이론적으로, 이것은 풍부한 스테레오 신호를 저하시키지 않기 위해, 스테레오 이미지가 한 방향으로 세계 정의될 때에만 이루어진다. 벡터 합과 차에 기초하여, 상기 목표는 다음의 두 가지 이중 원칙을 가정함으로써 실행된다.

1. 만약 청취자 양쪽에 신호의(그리고 노이즈의) 강한 표현이 존재한다면, 이때 중간선(중심 이미지)상의 노이즈 증가는 시각적으로 억제된다. 상부한계는 측면노이즈이다.
2. 만약 중심선에 강한 신호의 노이즈화(그리고 노이즈의)가 존재한다면, 이때 양쪽상의(상호 연관된)노이즈 증가는 시각적으로 억제된다. 상부한계는 중앙 노이즈이다.

그런, 노이즈 수준의 임의의 증가는 MLD 임계값에 의해 반드시 교정되어야 한다.

상기 목표에 따라, 최종 스테레오포닉 임계값은 다음과 같이 계산된다. 첫째, 채널 합과 차에 대한 임계값은 노이즈-마스킹-톤 및 톤-마스킹-노이즈에 대한 모노포닉을 모델을 사용하여 산출된다. 상기 절차는, 단계 10까지 3. 2에 제시된 것과 똑같다. 이때 채널 모두에 대해 단위 밴드당 실제 에너지 임계값(nb_n)을 갖는다. 편의상, 이것을 각각 채널 합과 채널 차에 대한 THR_{nSUM} 과 THR_{nDIF} 라고 부른다.

두번째, 양 채널에 대한 MDL 임계값, 즉 $THR_{nMLD, SUM}$ 과 $THR_{nMLD, DIF}$ 도 역시 다음에 의해 산출된다.

$$THR_{nMLD, SUM} = en_{b, SUM} 10^{-\frac{MLD_{n, dB}}{10}}$$

$$THR_{nMLD, DIF} = en_{b, DIF} 10^{-\frac{MLD_{n, dB}}{10}}$$

MLD 보호와 스테레오 무관성은 계산함으로써 침착된다.

$$nthr_{SUM} = \text{MAX}[THR_{nSUM}, \text{MIN}(THR_{nDIF}, THR_{nMLD, DIF})]$$

$$nthr_{DIF} = \text{MAX}[THR_{nDIF}, \text{MIN}(THR_{nSUM}, THR_{nMLD, SUM})]$$

이것을 연산한 다음 3. 2에 제시된 바와 같이 11번째 후의 나머지 단계도 역시 두 채널 모두를 위해 택해진다. 본질적으로 상기 마지막 임계값은 절대임계값 그리고 또한 부분적인 전마스킹 보호를 고려하여 더욱 조정된다. 상기 전마스킹 보호는 모노포닉의 경우로부터 간단하게 채택되었다는 것에 주목하여야 한다. 이것은 약 2밀리초인 모노오럴(monoaural)시간 해를 고려한다. 그러나, 바이오럴(binaural)시간 해는 6밀리초만큼 정확하다. 유관한 스테레오 이미지를 채널간 시간차에 기초하여 코드 스테레오 신호를 편리하게 코드화하는 것은 더욱 연구되어야 할 주제이다.

[스테레오포닉 코더]

스테레오포닉 코더의 개략화된 구조는 제12도에 제시되어 있다. 분석되는 각 데이터 세그먼트에 대해, 두 신호 채널 모두의 독립적이며 상대 동작에 대한 상세한 정보는 크고 짧은 변환으로서 주어진 정보를 통해 이용 가능할 수도 있다. 이 정보는 특정 세그먼트를 코드화하는 데 필요한 필수적인 단계수에 따라 사용된다. 상기 단계에는, 필수적으로 분석 윈도우의 선택, 코딩모드(R/L 또는 S/D)의 밴드 기준에 대한 정의, 계수와 스케일 인수의 양자화 및 호프만 코딩 그리고 마지막으로 비트스트림 합성이 있다.

[코드 모드 선택]

새로운 세그먼트가 판독될 때, 크고 짧은 분석 윈도우에 대한 음조 갱신이 이루어진다. 모노포닉 임계값과 PE 값은 3. 1절에 설명된 기술에 따라 산출된다. 이것으로 두 채널에 사용될 윈도우 종류에 대해 첫번째로 결정된다.

일단 윈도우 시퀀스가 선정되면, 다음에 직교하는 코딩 결정이 고려된다. 이것에는, 채널의 독립적인 코딩 사이의 선택, 모드 우측/좌측(R/L) 혹은 합과 차 채널(S/D)을 사용한 조인트 고정이 있다. 이 결정은 코더의 밴드 기준에 대해 취해진다. 이것은, 바이오럴 지각이 양귀에서 동일한 임계 밴드의 출력 함수인 것에 근거한다. 만약 두 채널의 임계값이 매우 다르다면, 이때 MLD 보호의 필요성이 존재하지 않으면, 만약 채널 합과 차가 고려된다면 신호는 보다 상관관계가 없는 것이 아닐 것이다. 만약 신호가 스테레오 이미지를 발생한다면, 이때 MLD 보호는 반드시 활성화되어 부수적인 이득은 S/D 코딩 모드를 선택함으로써 이용할 수도 있다. 상기의 후자 상황을 검출하는 한 편리한 방법은, 우측과 좌측 채널 사이의 모노포닉 임계값을 비교하는 것이다. 만약 특정 밴드내의 임계값이 선정된 값, 즉 2dB 이상만큼 다르지 않다면, 이때 S/D 코딩 모드가 선택된다. 그렇지 않으면 독립적인 모드 R/L으로 가정된다. 각각의 밴드는 상기 밴드의 코딩 모드를 명시하는 하나의 비트 플래그이며 반드시 측면정보로서 디코더로 전송되어야만 된다. 지금부터 이것을 코딩모드 플래그라고 칭한다. 동일한 밴드에 대해 다음의 세그먼트가 다를 수 있기 때문에 코딩 모드 결정은 시간에 적응적이며 또한 동일한 세그먼트에 대해, 다음 밴드의 코딩 모드가 다를 수 있기 때문에 주파수에서 적응적이다. 코딩 결정의 예시는 제13도에 제시되어 있다. 상기 예시는 길고 또한 짧은 세그먼트에 대해 유효하다.

이때 윈도우 스위칭 메카니즘에는 다음 모노포닉 측정만이 관여하기 때문에 단위 세그먼트당 PE측정의 최대 수는 $10(2\text{채널} \times [1\text{큰 윈도우} + 4\text{짧은 윈도우}])$ 이라는 것은 명확하다. 그러나, 단위 세그먼트마다 계산하는데 필요할 수 있는 최대 임계값 수는 20이며 그래서 20음조측정은 항상 반드시 세그먼트마다 $(4\text{채널} \times [1\text{큰 윈도우} + 4\text{짧은 윈도우}])$ 갱신되어야 한다.

[비트율 조정]

윈도우 스위칭과 코딩 모드 선택에 대한 결정은, 그것이 서로 의존하지 않는다는 의미에서 직교한다고 이전에 서술하였다. 이 결정에 독립적인 것은 또한 양자화, 호프만 코딩 그리고 비트스트림 합성을 포함하는 코딩 처리의 최종 단계, 즉 피드백 경로가 존재하지 않는다. 이러한 사실은, 전체 코딩 지연을 최소값 $(1024/4800=21.3\text{밀리초})$ 으로 감소시키며 또한 직교하지 않는 코딩 상황으로 인한 불안정을 피하는 장점에 있다.

양자화 처리는 스펙트럼 계수와 스케일 인수 모두에 영향을 끼친다. 스펙트럼 계수는 밴드로 클러스터(clustered)되는데, 각 밴드는 똑같은 스텝크기 또는 스케일 인수를 갖는다. 각 스텝크기는 3. 2절 단계 4에 도시된 바와 같이 그 밴드에 상당하는 마스킹 임계값으로부터 직접 계산된다. 정수인 양자화된 값은 이때 가변 워드 길이 혹은 호프만 코드로 변환된다. 비트스트림의 부수적인 필드를 고려하여 세그먼트를 코드화하기 위한 총 비트수가 계산된다. 비트율은 반드시 일정하게 유지되어야 하기 때문에, 양자화 처리는 상기 비트수가 선정된 한계내에 있을 때까지 반드시 반복적으로 실시된다. 모든 세그먼트를 코드화시키는 데 필요한 비트수 이후에, 기본적인 마스킹 임계값을 고려하여, 조정의 정도 버퍼 제어 장치에 의해 지시된다. 상기 제어장치는 각각의 필요에 따라 공유한다.

비트율 조정 루틴 기술은 제9도의 흐름도에 의해 표시되어 있다. 현재의 세그먼트에 의해서 사용될 이용 가능한 비트의 총수가 산출된 다음, 순환 절차는, 만일 모든 초기의 임계값이 상기 인수만큼 곱해진다면 최종의 총 비트수가 보다 작고 그리고 이용 가능한 비트수의 에러(δ) 이내가 되도록 인수(α)를 구하려 시도한다. 비록 근사값의 곡선이 맞지 않아 α 가 순환의 최대수이내에서 발견되지 않더라도, 수용 가능한 한가지 해는 항상 이용 가능하다.

상기 루틴의 주요한 단계는 다음과 같다. 첫째, 해를 포함하는 간격이 구해진다. 다음에, 루프는 해로 급속히 집중하려고 한다. 각각의 순환에서, 최상의 해가 갱신된다.

크고 짧은 윈도우로 코드화된 세그먼트에 대해 동일한 절차를 사용하기 위해, 후자의 경우에, 4짧은 윈도우 계수는 유사한 밴드끼리 연결시킴으로서 클러스터 된다. 스케일 인수도 동일하게 클러스터 된다.

비트율 조정 루틴은, 호프만 코드화된 모든 워드(계수와 스케일 인수)를 표현하기 위해 전체 비트수를 계산하는 다른 루틴을 호출한다. 후자의 루틴은 계수의 진폭 분포에 따라 스펙트럼 분할을 실시한다. 그 목표는, 스펙트럼의 부분에 선정된 호프만 코드북을 지정하는 것이다. 각각의 부분은 밴드의 가변수들을 그룹지우고 그 계수는 편리한 북을 사용하여 호프만 코드화된다. 상기 코드북의 참조와 부분의 한계는 반드시 측면정보로서 디코더에 보내져야만 한다.

스펙트럼 분할은 최소비용 전략을 사용하여 실시된다. 주요한 단계는 다음과 같다. 첫째, 가능한 모든 부

분이 정의되는데, 한계는 단위밴드당 한 부분이며, 각 부분은 상기 부분내의 계수에 대한 진폭 분포에 가장 적합한 코드북을 갖는다. 전체 스펙트럼의 시작과 정지가 알려진 바와 같이, 만약 k 가 부분의 번호라면, 부분 사이에는 $k-1$ 격리자(separators)가 존재한다. 각 격리자를 제거하는 가격(The price)이 계산된다. 보다 낮은 가격을 갖는 격리자는 제거(초기가격이 음일 수 있다)된다. 다음 순환전에 다시 가격이 계산된다. 상기 처리는 허용 가능한 최대 부분이 얻어질 때까지 반복되며 다른 격리자를 제거하기 위한 최소가격은 선정된 값보다 크다.

지금부터 제2도의 양자화기/비율-루프(206)에 의해 달성된 처리의 면들이 제시될 것이다 종래 기술에서, 비율-루프(rate-loop) 메카니즘은 모노포닉의 경우와 관련된 가정을 포함하고 있다. 모노포닉으로부터 스테레오포닉 지각 모델로 쉬프트함에 따라, 비율-루프에 대한 요구가 증가된다.

제2도의 양자화기/비율-루프(206)에 들어가는 입력은 스펙트럼 계수에 상당하는 산출된 임계값을 포함하고 있는 분석 필터 뱅크(202)에 의해 유도된 스펙트럼 계수(즉, MDCT 계수)와 지각모델(204)의 출력을 포함한다.

양자화기/비율-루프(206)는 부분적으로, 히어링(hearing)의 절대 임계값과 계산된 임계값에 근거한 스펙트럼 정보를 양자화시키며, 그렇게 함으로써 엔트로피 코더(208)에 비트스트림을 공급한다. 비트스트림은 세 부분으로 나뉘어진 신호를 포함하는 데, 이것은 (1) 표준화된 측면정보를 포함하는 첫째 부분, (2) 35 또는 56에 대한 스케일 인수와 사용될 때(이 부분의 길이는 첫째 부분에 있는 정보에 따라 변할 수 있다) 소위 적응적-윈도우 스위칭을 위해 사용된 부수적인 측면 정보를 포함하는 두번째 부분과 (3) 양자화된 스펙트럼 계수를 구비하는 세번째 부분이 존재한다.

'사용된 스케일 인수'(Δ)는, 양자화된 스펙트럼 계수가 허용 가능한 한계 내에서 인코딩될 수 있을 때까지 양자화될 각 스펙트럼 계수의 주파수에 해당하는 주파수를 히어(hearing)하는 절대 임계값으로부터 유도된 스케일 인수와 계산된 스케일 인수 사이를 보간 함으로써 반복적으로 유도된다.

본 발명의 예시적인 실시예를 제W도에서 찾을 수 있다. W01에 도시된 바와 같이 양자화기/비율-루프는 스펙트럼 계수(C_i)와 상기 스펙트럼 계수에 해당하는 에너지 임계값(E)을 수신한다. '임계 스케일 인수'(Δ)는 다음에 의해 계산된다.

$$\Delta_0 = \sqrt{12E}$$

'절대 스케일 인수'(Δ_A)도 역시 히어링의 절대 임계값을 기초로 계산된다(즉, 스케일 인수에 해당하는 주파수가 들릴 수 있는 가장 조용한 음). 양호하게 보간 상수(α)와 보간 경계(α_{high} 와 α_{low})는 사용된 스케일 인수를 조정하는데 도움이 되도록 초기화된다.

$$\alpha_{high} = 1$$

$$\alpha_{low} = 0$$

$$\alpha = \alpha_{high}$$

다음에, W0.5에 도시된 바와 같이, 사용된 스케일 인수는 다음으로부터 결정된다.

$$\Delta = \Delta_0^\alpha \times \Delta_A^{(1-\alpha)}$$

다음에, W07에 도시된 바와 같이, 사용된 스케일 인수, 상기 계산된 바와 같이 사용된 스케일 인수가 단속적이지 않지만, 전송되고 사용될 때 양호하게 단속적이기 때문에 스스로 양자화된다.

$$\Delta = Q^{-1}(Q(\Delta))$$

다음에, W09에 도시된 바와 같이, 스펙트럼 계수는, '양자화된 스펙트럼 계수' $Q(C_i, \Delta)$ 를 만들기 위해 사용된 스케일 인수를 사용하여 양자화된다.

$$Q(C_i, \Delta) = \text{NINT}\left(\frac{C_i}{\Delta}\right)$$

여기서 'NINT'는 가장 가까운 정수함수이다. 양자화기/비율 루프(206)가 반드시 양자화된 스펙트럼 계수와 사용된 스케일 인수 모두를 전송하여야 하기 때문에 그것들 모두를 전송하는데 얼마나 많은 비트가 취해질 것인가와 연관되어 있는 비율(C)이 계산된다. 제W11도에 도시된 바와 같이 C 는 다음과 같다.

$$C = F00(Q(C_i, \Delta), Q(\Delta))$$

여기서 F00는, 특정한 실시예에 따라 데이터 통신의 기술 분야에 통상적인 기술을 지닌 사람들에게 의해 용이하게 결정될 수 있는 함수이다. W13에 도시된 바와 같이, 비용(C)이 검사되어 그것이 허용 가능한 범위(PR)내에 존재하는지를 결정한다. 비용이 허용가능한 범위내일 때, $Q(C_i, \Delta)$ 와 $Q(\Delta)$ 은 엔트로피 코더(208)로 전송된다.

양호하게 그리고 허용가능한 범위(PR)대 비용(C)의 관계에 따라서 보간 상수와 경계는, 사용된 스케일 인수가 허용 가능한 범위내의 비용을 갖는 양자화된 스펙트럼 계수를 산출할 때까지 조정된다. 예시적으로, 제W도의 W13에 도시된 바와 같이, 보간 경계는 2진 검색을 발생하도록 조작된다. 특히, $C > PR$ 일 때 $\alpha_{high} = \alpha$, 이며, 선택적으로, $C < PR$ 일 때, $\alpha_{low} = \alpha$ 이다.

어느 경우에서, 보간 상부는 다음에 의해 계산된다.

$$\alpha = \frac{\alpha_{\text{low}} + \alpha_{\text{high}}}{2}$$

다음 처리는, C가 허용 가능한 범위(PR) 이내가 될 때까지 W05에서 순환적으로 계속한다.

[스테레오포닉 디코더]

스테레오포닉 디코더는 매우 간단한 구조이다. 그것의 주요기능은 들어오는 비트스트림을 판독하고, 모든 데이터를 디코딩하며, 역 양자화시키고 우측 및 좌측 채널을 재구성하는 것이다. 상기 기술은 제12도에 표시되어 있다. 예시적인 실시예는 AT&T DSP 16 혹은 DSP32C와 같은 디지털 신호 처리기(DSP) 하드웨어와 이와 논의된 연산을 수행하는 소프트웨어를 구비할 수 있다. 하이브리드 DSP/VLSI 구현은 물론, 본 발명의 대규모 집적(VLSI) 하드웨어 구현도 역시 제공될 수 있다.

(57) 청구의 범위

청구항 1

각각 좌측 및 우측 신호 채널 세트를 구비한 스테레오포닉 오디오 신호의 입력 세트 코딩 방법에 있어서, 첫번째 신호 세트가 복수개의 주파수 밴드 각각에 대한 진폭 및 위상 정보를 표현하는 신호를 구비하며, 입력 시퀀스의 주파수 내용을 표현하는 상기 좌측 및 우측 신호 채널 세트 각각의 형성단계, 상기 첫번째 신호 세트 각각의 상기 주파수 밴드 각각에 대한 랜덤 메트릭의 결정 단계, 상기 첫번째 신호 세트 각각에 대한 주파수를 구비한 전력 분산과 상기 첫번째 신호 세트 각각에 대한 상기 랜덤 메트릭스에 기초하여, 주파수 함으로서 음조함수를 형성하는 단계 및, 상기 랜덤 메트릭에 기초하여 상기 좌측 및 우측 신호 채널 세트 각각에 대한 첫번째 지각 임계값(a first perceptual threshold)을 결정하는 단계를 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 2

제1항에 있어서, 상기 첫번째 신호 세트의 형성 단계는, 복수개의 주파수 밴드 각각에 대한 상기 좌측 채널과 관계되는 진폭 및 위상 정보를 표현하는 신호와, 복수개의 주파수 밴드 각각에 대한 상기 우측 채널과 관계되는 진폭 및 위상 정보를 표현하는 신호를 형성하는 것을 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 3

제1항에 있어서, 상기 첫번째 신호 세트의 형성 단계는 상기 좌측 및 우측 신호 채널 세트 각각에 대한 단속적인 일시적 스펙트럼(discrete short-time spectra)을 유도하는 것을 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 4

제1항에 있어서, 양자화에 의해 유입된 노이즈가 상기 첫번째 신호세트 각각에 대해 상기 지각 임계값 이하 수준이 되도록 좌측 및 우측 채널 각각에 대한 상기 첫번째 신호 세트를 양자화하는 단계를 더 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 5

제3항에 있어서, 상기 좌측 및 우측 신호 채널 세트는 N 신호를 구비하며, 상기 단속적인 일시적 스펙트럼을 유도하는 것은 상기 N 좌측 및 N 우측 채널 신호 각각에 대한 N 스펙트럼 값을 유도하는 것을 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 6

제5항에 있어서, 상기 좌측 및 우측 신호 채널 세트는 선택적으로 N/k 신호를 구비하며, 상기 단속적인 일시적 스펙트럼을 유도하는 것은 선택적으로, 여기서 N/k는 정수, 상기 N/k 좌측 및 N/k 우측 채널 신호 각각에 대한 N/k 스펙트럼 값을 유도하는 것을 구비하며, 상기 방법은, 상기 좌측 및 우측 채널 신호 세트 각각에 대한 지각 엔트로피의 측정을 유도하는 단계와, 상기 선택적인 우측 및 좌측 채널 세트와 상기 지각엔트로피에 대한 값에 기초한 해당 스펙트럼 값 사이를 선택하는 단계를 더 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 7

제1항에 있어서, 상기 좌측 채널에 대한 첫번째 신호 세트와 상기 우측 채널에 대한 첫번째 신호 세트의 대응하는 신호세트의 합과 차에 각각 대응하는 합과 차의 채널 신호 세트를 형성하는 단계를 더 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 8

제7항에 있어서, 상기 주파수 밴드 각각에 대한 결정된 임계값에 기초하여 (i) 상기 합과 차의 채널 신호 또는 (ii) 상기 좌측 채널 및 우측 채널 신호 사이를 선택하는 것을 더 구비하는 것을 특징으로 하는 스테레오포닉 오디오 신호의 입력 세트 코딩 방법.

청구항 9

제8항에 있어서, 첫번째 신호가 선정된 양보다 작게 다른 상기 주파수 밴드에 대한 상기 합과 차의 채널

신호를 선택하는 것을 구비하는 것을 특징으로 하는 스테레오폰닉 오디오 신호의 입력 세트 코딩 방법.

청구항 10

제9항에 있어서, 상기 선정된 양의 거의 2dB와 같은 것을 특징으로 하는 스테레오폰닉 오디오 신호의 입력 세트 코딩 방법.

청구항 11

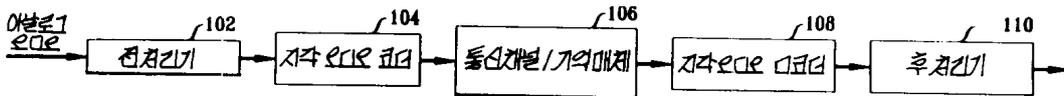
제9항에 있어서, L_i 는 상기 좌측채널에 대한 1번째의 첫번째 신호이며, R_i 는 상기 우측 채널에 대한 i 번째의 첫번째 신호이며 그리고 상기 채널 신호는 $M_i = (L_i + R_i) / 2$ 로 주어지는 것을 특징으로 하는 스테레오폰닉 오디오 신호의 입력 세트 코딩 방법.

청구항 12

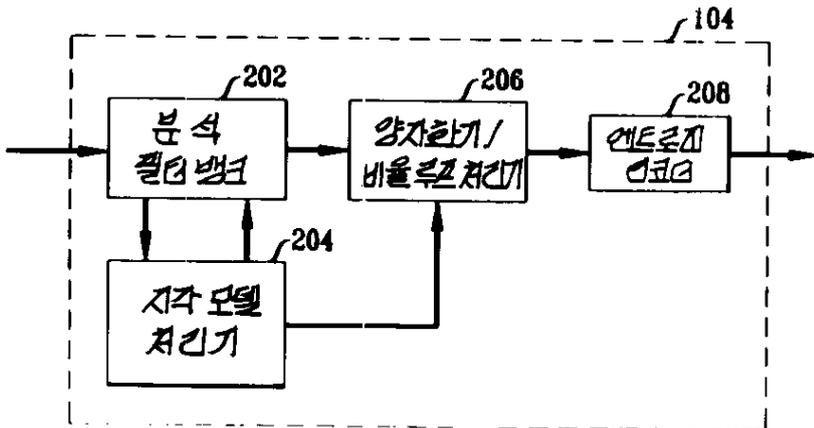
제6항에 있어서, 여기서 $k=4$ 인 것을 특징으로 하는 스테레오폰닉 오디오 신호의 입력 세트 코딩 방법.

도면

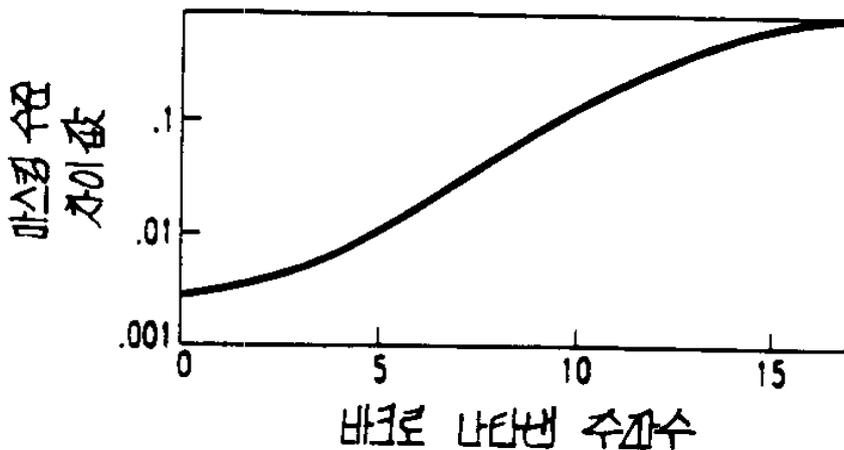
도면1



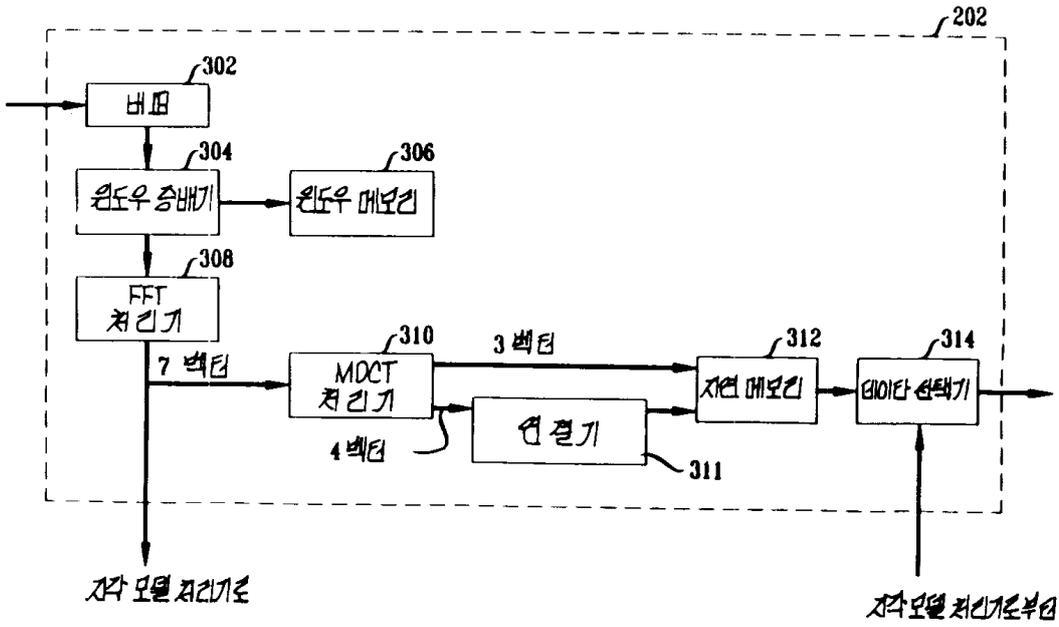
도면2



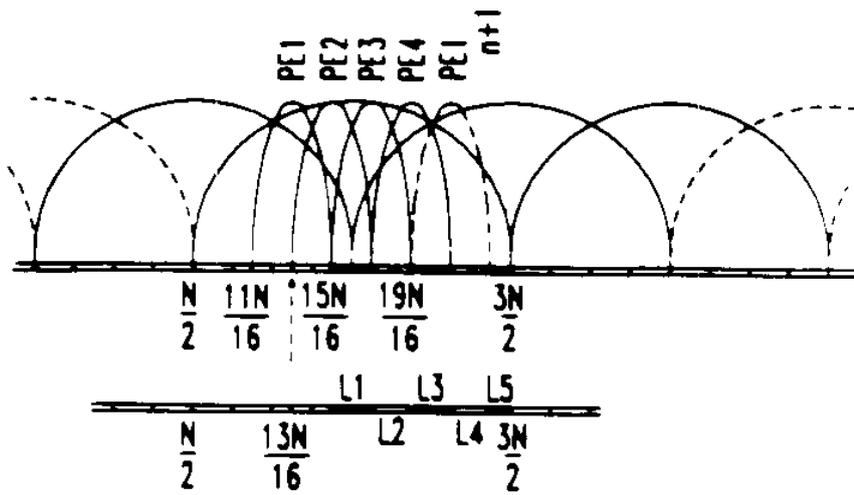
도면3



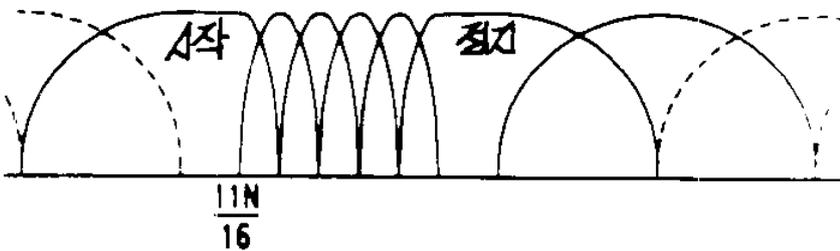
도면4



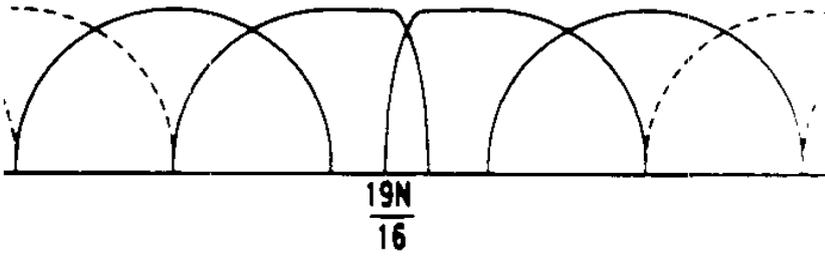
도면5a



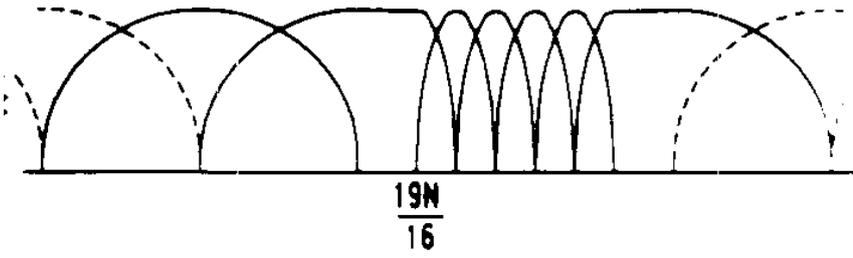
도면5b



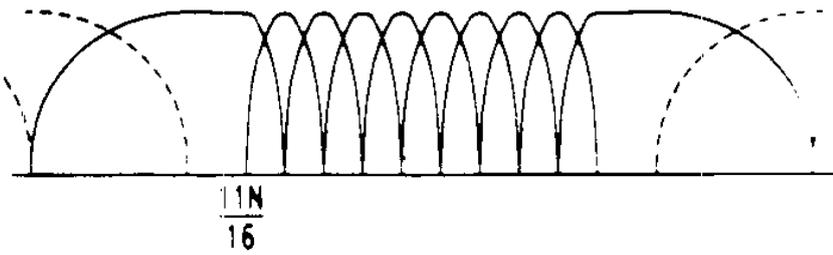
도면5c



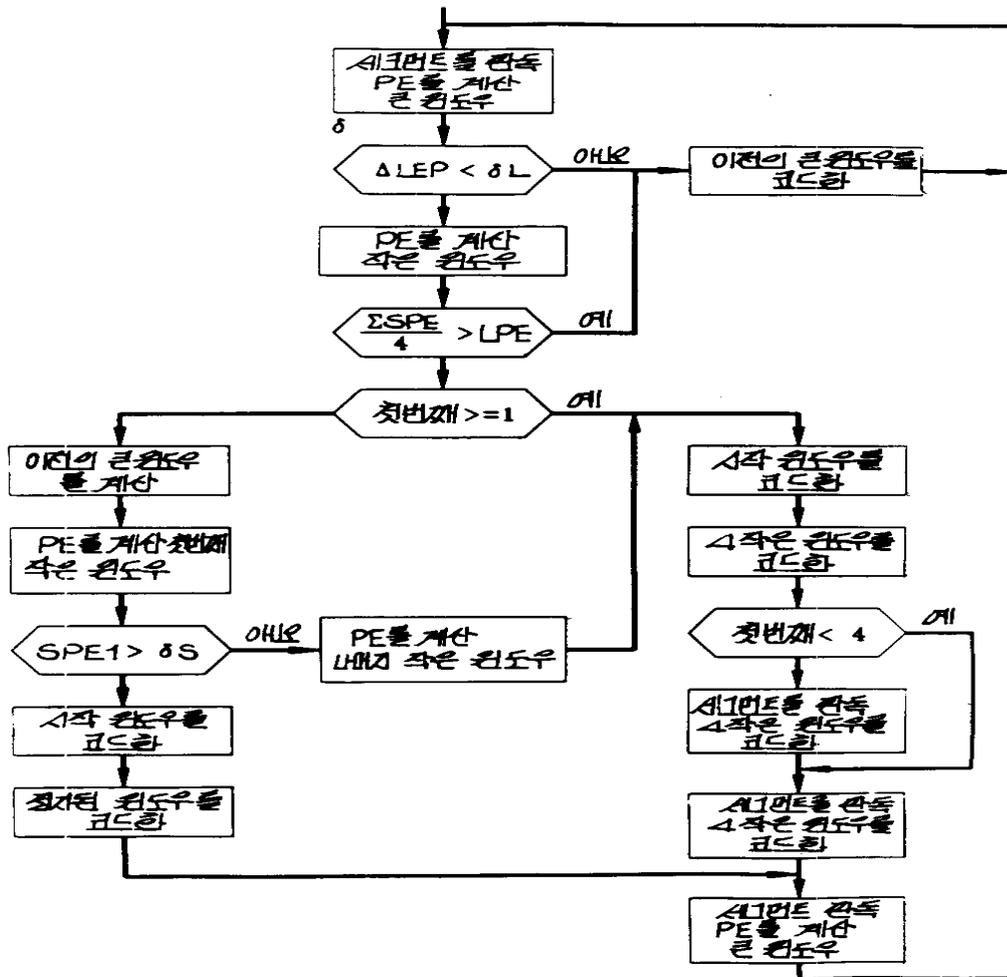
도면5d



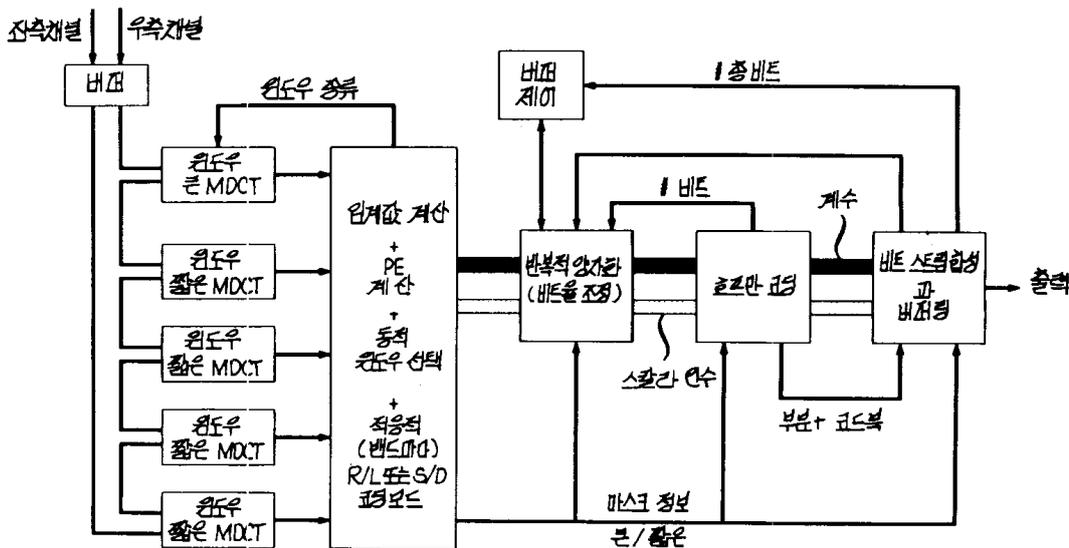
도면5e



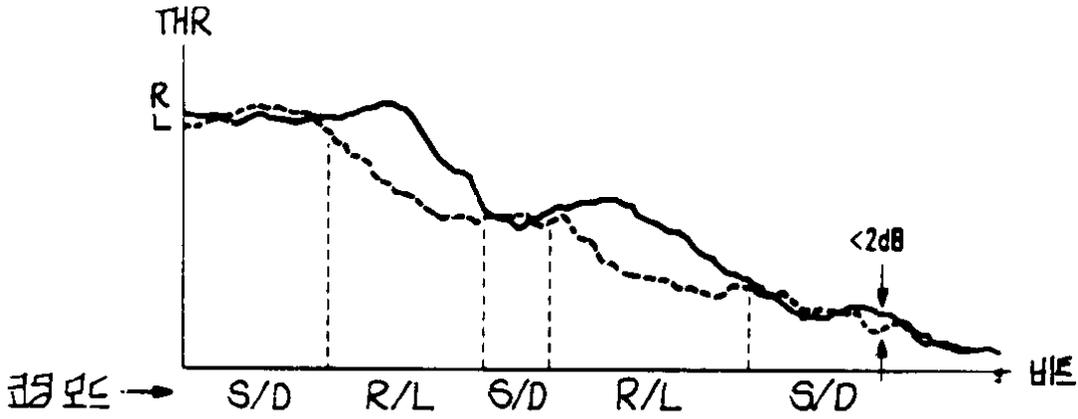
도면6



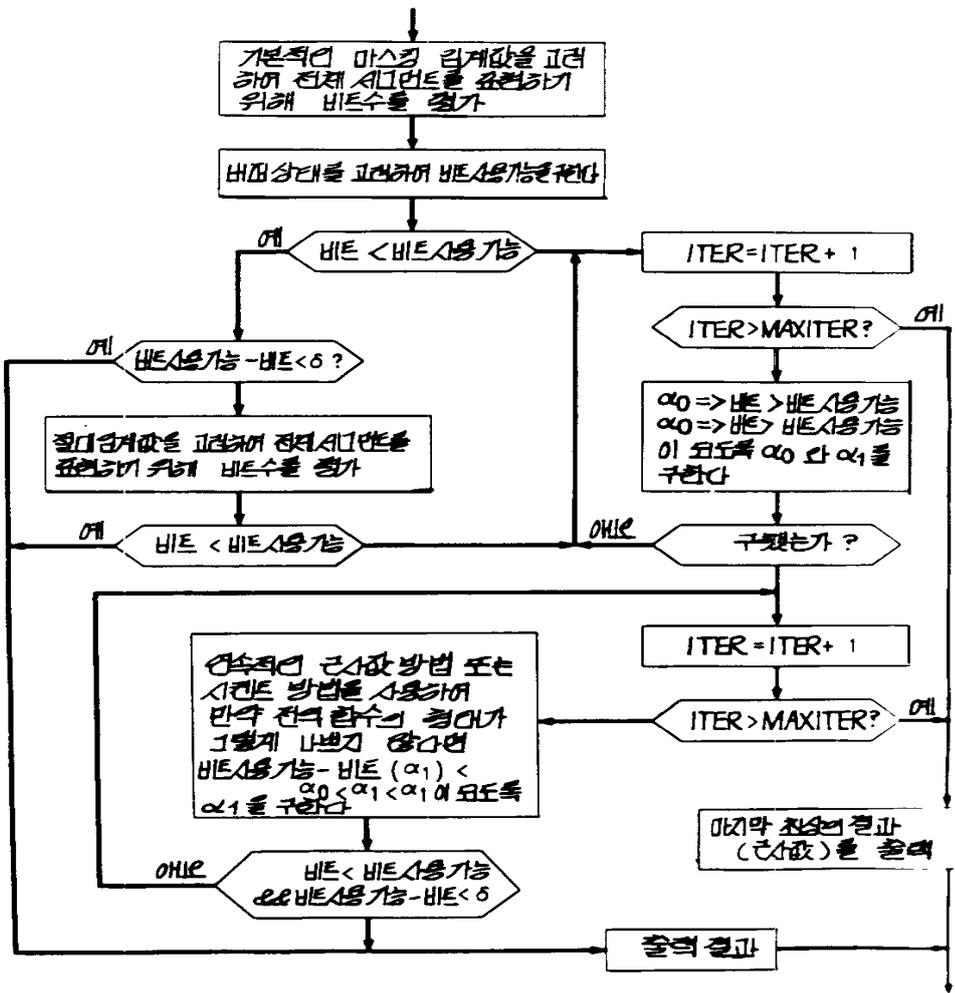
도면7



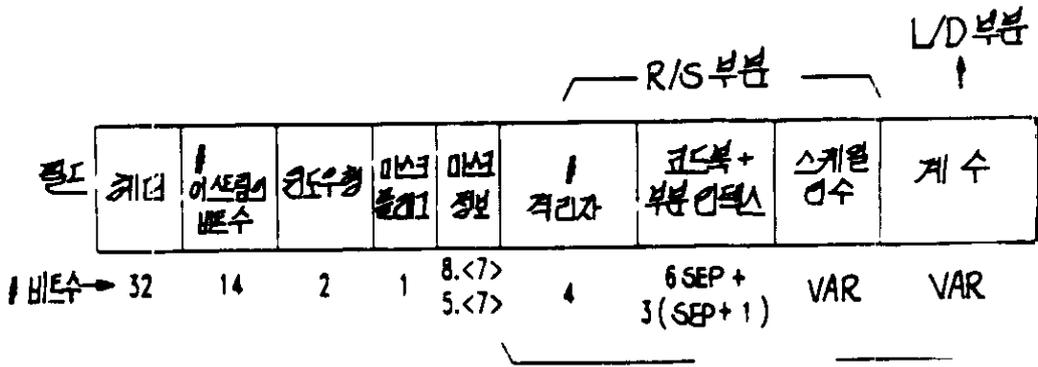
도면8



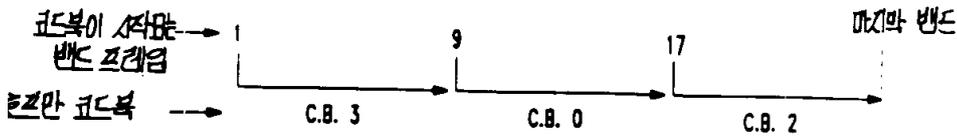
도면9



도면10



도면11a

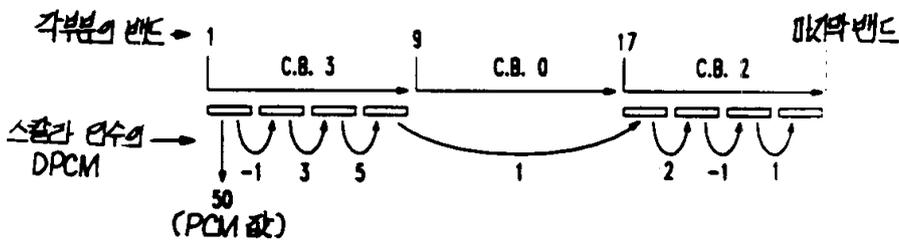


적외자 RS = K = 2

부분 인덱스	코드북	부분이 시작 되는 밴드
1	2	17
2	0	9
3	3	1

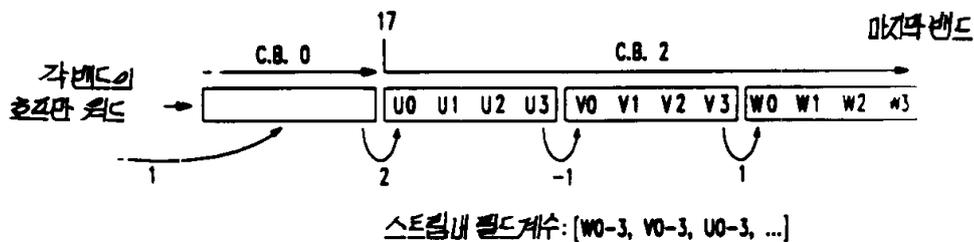
스트림내 필드 CB = 인덱스
: [2, 17, 0, 9, 3]

도면11b



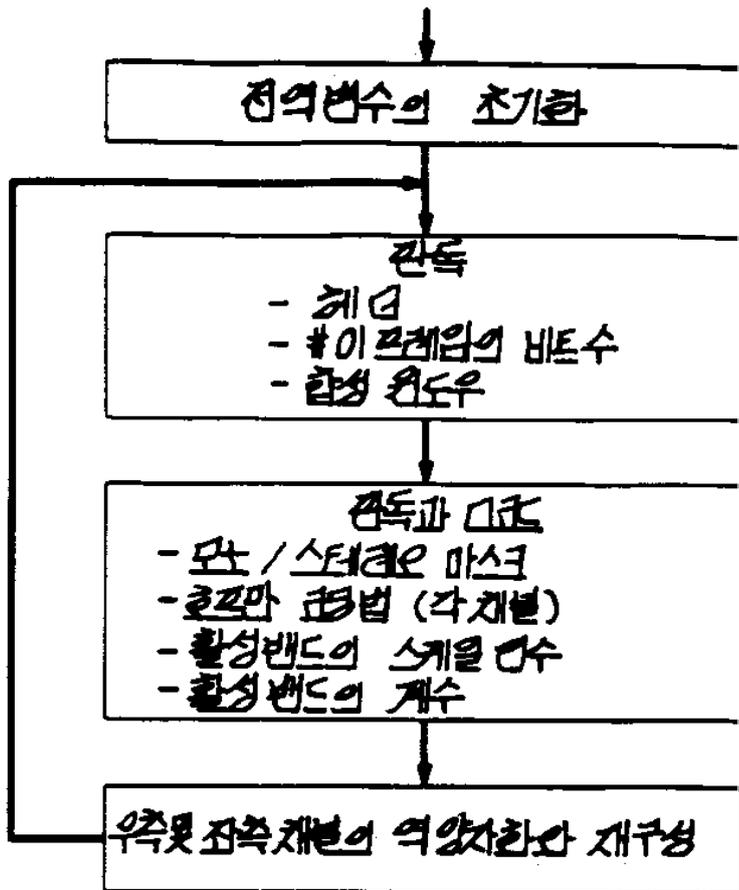
스트림내 필드 스케일 인수: [1, -1, 2, 1, 5, 3, -1, 50]

도면11c

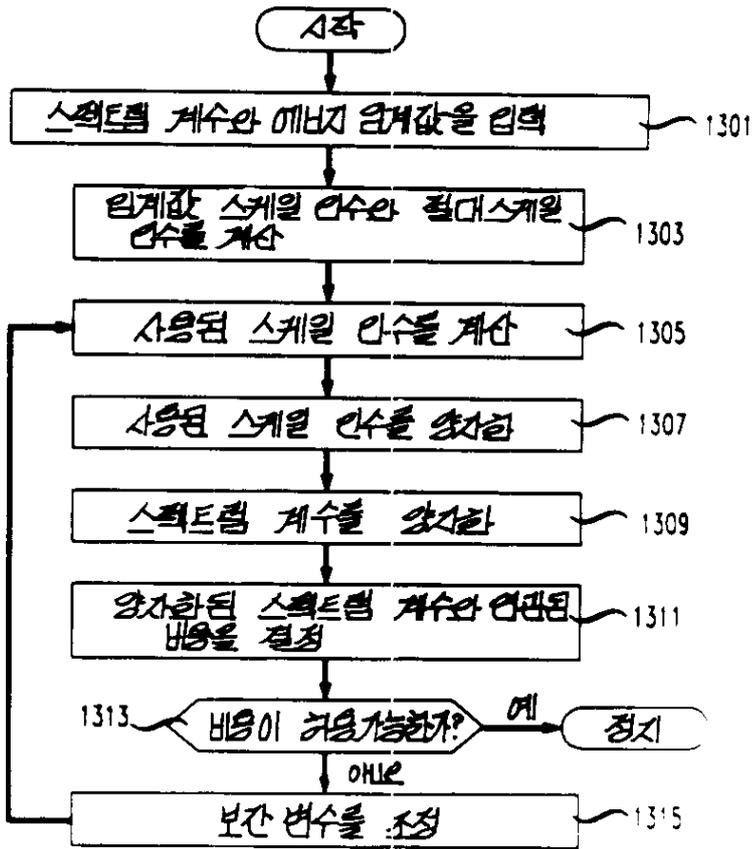


스트림내 필드 계수: [W0-3, V0-3, U0-3, ...]

도면 12



도면 13



도면 14

