(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.[6]
G10L 21/00

(45) 공고일자 2003년03월15일
(11) 등록번호 10-0365300
(24) 등록일자 2002년12월05일

(21) 출원번호	10-1997-0705131	(65) 공개번호	특1998-0701735
(22) 출원일자	1997년07월28일	(43) 공개일자	1998년06월25일
번역문제출일자	1997년07월28일		
(86) 국제출원번호	PCT/SE1996/00024	(87) 국제공개번호	WO 1996/24128
(86) 국제출원일자	1996년01월12일	(87) 국제공개일자	1996년08월08일
(81) 지정국	국내특허 : 오스트레일리아 바베이도스 불가리아 브라질 캐나다 중국 체코 에스토니아 그루지야 헝가리 아이슬란드 일본 북한 대한민국 스리랑카 라이베리아 리투아니아 라트비아 마다가스카르 몽고 멕시코 노르웨이 뉴질랜드 폴란드 루마니아 AP ARIPO특허 : 케냐 말라위 수단 EA 유라시아특허 : 아르메니아 벨라루스 키르기즈 카자흐스탄 몰도바 러시아 EP 유럽특허 : 오스트리아 스위스 리히텐슈타인 독일 덴마크 스페인 핀랜드 영국 룩셈부르크 포르투칼 스웨덴 OA OAPI특허 : 부르키나파소 베넹 중앙아프리카 콩고 코트디브와르 카메룬 가봉 기네 말리 모리타니 니제르 세네갈 차드 토고		

(30) 우선권주장 9500321-6 1995년01월30일 스웨덴(SE)

(73) 특허권자 텔레폰아크티에볼라게트 엘엠 에릭슨
스웨덴,스톡홀름, 에스-126 25
(72) 발명자 피터 핸델
스웨덴왕국 엎살라 에스-753 36 베프나르가탄 15비
(74) 대리인 김기종, 권동용, 최재철

심사관 : 남인호

(54) 스펙트럼감산잡음억제방법

명세서

기술분야

<1> 본 발명은 디지털 프레임을 기반으로 하는 통신시스템에서 잡음억제에 관한 것으로서, 특히 그러한 시스템에서 스펙트럼적 감산 잡음억제방법에 관한 것이다.

배경기술

<2> 음성신호 프로세싱에서 공통적인 문제는 음성신호의 잡음 측정으로 음성신호를 강화시키는 것이다. 단일 채널(송화기)측정을 기초로 음성을 강화시키기 위한 한 접근법은 스펙트럼 감산기술(spectral subtraction technique)을 적용하여 주파수영역을 필터링하는 것이다[1], [2] 배경잡음이 (음성에 비해) 장시간 변동이 없다는 가정 하에서, 음성행위가 없는 시간주기 동안에 배경잡음의 모델을 추정한다. 그런 다음, 음성행위가 있는 데이터프레임 동안에, 이 추정된 잡음모델은 음성을 강화시키기 위하여 잡음이 있는 음성의 추정된 모델과 함께 사용된다. 스펙트럼적 감산기술에 있어서, 이들 모델들은 전형적인 FFT방법을 사용하여 추정되는, 전력 스펙트럼밀도(Power Spectral Density : PSD)로서 주어진다.

<3> 상기 언급된 기술 중 어느 것도 이동전화응용에서 만족스리운 가청품질을 가지는 출력신호를 기본적으로 출력할 수 없다. 즉,

<4> 1. 왜곡되지 않은 음성출력

<5> 2. 잡음레벨의 충분한 감소

<6> 3. 인위적인 음에 영향을 미치지 않는 잔여잡음을 가지는 출력신호를 출력할 수 없다.

<7> 특히, 스펙트럼적 감산방법은 상기 2를 충족하면 상기 1을 위배하거나 또는 상기 1을 충족하면 상기 2를 위배하는 것으로 알려져 있다. 이외에도, 대부분의 경우에 있어서, 스펙트럼적 감산 방법이 소위 음악적인 잡음을 도입하기 때문에 3은 더 또는 덜 위배된다.

<8> 스펙트럼적 감산방법의 상기 단점들은 공지되었고, 그리고 문헌에, 소정의 스피치-인-노이즈(speech-in-noise;잡음하의 음성) 시나리오를 위한 기초 알고리즘의, 여러 개의 임시 변통적 변형이 나타나 있다. 그러나, 상기 1~3을 충족하는 스펙트럼 감산방법을 어떻게 설계할지의 문제가 아직 해결되지 않았다.

<9> 잡음이 있는 데이터로부터 음성을 강화시키는데 있어서의 곤란함에 집중하기위하여, 스펙트럼적 감산방법은 입력되는 데이터의 추정된 모델을 사용하는 필터링을 기초로 한다는 것을 명심해야 한다. 만일 이들 추정된 모델들이 근원적인 "참(true)"모델에 가깝다면, 이는 훌륭한 작업 접근법이다. 그러나, 이동전화제품 주위의 물리적인 실체(8000Hz 샘플링 주파수, 0.5~3.0초의 잡음의 정상성(定常性)(stationarity))뿐만 아니라 음성의 짧은 정상성(10~40ms) 때문에, 추정된 모델들은 근원적인 실체와 상당히 달라질 수 있어서, 따라서 낮은 가청품질을 가지는 필터출력이 된다.

<10> EP, A1, 0 588 526호는 스펙트럼분석이 고속 퓨리에 변환(Fast Fourier Transformation : FFT) 또는 선형 예측 코딩(Linear Predictive Coding : LPC)으로 각각 수행되는 방법을 기술한다.

발명의 상세한 설명

<11> 본 발명의 목적은 가청품질을 손상시키는 일이 없이 더 나은 잡음 감소를 제공하는 스펙트럼적 감산 잡음억제방법을 제공하는 것이다.

<12> 이 목적은 청구항 1의 특징부로 해결된다.

도면의 간단한 설명

<13> 도 1은 본 발명의 방법을 실행하기에 적합한 스펙트럼적 감산 잡음억제시스템의 블록도.

<14> 도 2는 도 1의 시스템에 사용될 수 있는 음성행위 검출기(VAD)의 상태도.

<15> 도 3은 두 개의 상이한, 음성프레임의 전력스펙트럼 밀도 추정의 도면.

<16> 도 4는 음성과 배경잡음을 포함하는 샘플링 된 음성신호의 시간도.

<17> 도 5는 선행기술에 따른 스펙트럼적 잡음감산 후 도 3의 신호의 시간도.

<18> 도 6은 본 발명에 따른 스펙트럼적 잡음감산 후 도 3의 신호의 시간도.

<19> 도 7은 본 발명의 방법을 설명하는 흐름도.

실시예

<20> 스펙트럼적 감산기술

<21> 부가적인 잡음에 의해 품질이 떨어진 음성의 프레임을 고려하자.

$$x(k) = s(k) + v(k) \qquad k=1, \ldots, N \qquad (1)$$

<23> 이때 x(k), s(k) 및 v(k)는 각각 음성의 잡음 측정치, 음성 및 부가적인 잡음을 나타내고, 그리고 N는 프레임에서 샘플의 수를 나타낸다.

<24> 음성은 프레임 전체에 걸쳐 불변(정상)으로 가정되는 한편, 잡음은 장시간 불변(정상), 즉 수 개의 프레임에 걸쳐 불변이라고 가정된다. v(k)가 불변인 프레임의 수는 $\tau >> 1$로 표시된다. 또한, 음성행위가 충분히 낮아서, 비음성행위 동안에 잡음의 모델을 정확히 추정할 수 있다고 가정한다.

<25> 측정치, 음성 및 잡음의 전력 스펙트럼밀도(PSD)를 $\Phi_x(\omega)$, $\Phi_s(\omega)$, $\Phi_v(\omega)$로 각각 표시한다.

$$\varphi_x(\omega) = \varphi_s(\omega) + \varphi_v(\omega) \qquad (2)$$

<27> $\Phi_x(\omega))$와 $\Phi_v(\omega)$를 알면, 량 $\Phi_s(\omega))$와 s(k)가 아래에서 간략히 기술되는, 표준 스펙트럼적 감산방법([2]를 참조하라)을 사용하여 추정될 수 있다.

<28> $\hat{S}(k)$가 s(k)의 추정을 표시하도록 하자. 그러면,

$$\hat{s}(k) = \mathcal{F}^{-1}(H(\omega)X(\omega))$$

$$X(\omega) = \mathcal{F}(x(k)) \qquad (3)$$

<30> 이때, $F(\cdot)$는 몇몇 선형변환, 예컨대 이산 퓨리에 변환(DFT)을 나타내고, 그리고 $H(\omega)$는 $\omega \in (0,2\pi)$에서 실수치의 우함수를 나타내는 것으로 $0 \le H(\omega) \le 1$이 된다. 함수 $H(\omega)$는 $\Phi_x(\omega)$와 $\Phi_v(\omega)$에 의존한다. $H(\omega)$가 실수치이기 때문에, $\hat{S}(\omega)=H(\omega)X(\omega)$의 위상은 품질이 저하된 음성의 위상과 동일하다. 실수치의 $H(\omega)$의 사용은 위상 왜곡에 대해 민감하지 않은 인간의 귀에 의한 것이다.

<31> 일시적으로, $\Phi_x(\omega)$와 $\Phi_v(\omega)$는 알 수 없고 그리고 $H(\omega)$에서 추정된 량 $\hat{\Phi}_x(\omega)$와 $\hat{\Phi}_v(\omega)$로 대체되게 된다. 음성의 비-불변성 때문에, $\Phi_x(\omega)$는 데이터의 단일 프레임으로부터 추정되는 반면, $\Phi_v(\omega)$는 τ음성이 없는 프레임의 데이터를 사용하여 추정된다. 간략하게 하기 위하여, 잡음이 있는 음성을 포함하는 프레임과 잡음만을 포함하는 프레임을 구별하기 위하여 음성행위검출기(VAD)를 이용할 수 있다고 가정한다. 예컨대, 다음 식을 사용하여 여러 개의 프레임을 평균함으로써 비-음성행위 동안의 $\Phi_v(\omega)$를 추정한다고 가정한다.

$$\hat{\Phi}_v(\omega)^{\ell} = \rho\,\hat{\Phi}_v(\omega)^{\ell-1} + (1-\rho)\,\overline{\Phi}_v(\omega) \qquad (4)$$

<33> 식 (4)에서, $\hat{\Phi}_v(\omega)^{\ell}$는 프레임 수(ℓ)을 포함하는 지금까지의 데이터를 토대로 한 (연속)평균된 PSD추정이고, 그리고 $\overline{\Phi}_v(\omega)$는 현 프레임을 토대로 한 추정이다. 스칼라 $\rho \in (0,1)$는 v(k)의 추정된 불변성에 관하여 조정된다. τ프레임에 대한 평균은 다음에 의해 함축적으로 주어진 ρ에 개략적으로 대응한다.

$$\frac{2}{(1-\rho)} = \tau \qquad (5)$$

<35> 적절한 PSD추정(배경잡음의 스펙트럼적 형태에 선천적인 추정이 없다고 가정한다)은 다음으로 주어진다.

$$\overline{\Phi}_v(\omega) = \frac{1}{N} V(\omega)V^*(\omega) \qquad (6)$$

<37> 이때 "*"는 복소 공액이고 그리고 $V(\omega)=F(v(k))$이다. $F(\cdot)=FFT(\cdot)$(고속퓨리에 변환)으로, $\overline{\Phi}_v(\omega)$는 페리오다이그램(Periodigram)이고 그리고 식(4)에서 $\overline{\Phi}_v(\omega)$는 평균된 페리오다이그램이고, 둘다는 근사 분산으로 점근선(漸近線)적으로 (N>>1) 편향리지 않은 PSD추정이 된다.

$$\mathrm{Var}(\overline{\Phi}_v(\omega)) \approx \Phi_v^2(\omega)$$

$$\mathrm{Var}(\hat{\Phi}_v(\omega)) \approx \frac{1}{\tau}\Phi_v^2(\omega) \qquad (7)$$

<40> (7)과 비슷한 식도 음성행위(식(7)의 $\Phi_v^2(\omega)$를 $\Phi_x^2(\omega)$로 대체함) 동안 $\hat{\Phi}_x(\omega)$에 대해 유효하다.

<41> 본 발명의 방법을 실행하기에 적합한 스펙트럼적 감산 잡음억제시스템이 도 1에 블록형태로 설명된다. 송화기(10)로부터 음성신호(x(t))가 A/D변환기(12)로 전달된다. A/D변환기(12)는 디지털화 된 음성 샘플을 프레임형태 {x(k)}로 변환블럭(14), 예컨대 FFT(고속 퓨리에 변환)블럭으로 보내고, 변환블럭은 각 프레임을 상응하는 주파수 변환된 프레임 {X(ω)}로 변환시킨다. 변환된 프레임은 블록(16)에서 $\hat{H}(\omega)$에 의해 필터된다. 이 단계는 실제 스펙트럼적 감산을 수행한다. 최종신호 $\{\hat{S}(\omega)\}$는 역변환블럭(18)에 의해 시간영역으로 역변환된다. 결과는 잡음이 억제된 프레임{ $\hat{s}(k)$}가 된다. 이 프레임은 반향 소거기(20)로 전달되고, 다음에 음성부호기(22)에 전달될 수 있다. 음성 부호화된 신호는 송신을 위해 채널부호기와 복조기(이들 장치들은 도시되지 않았음)로 전달된다.

<42> 블록(16)에서 $\hat{H}(\omega)$의 실제 형태는 PDS추정기(24)에서 형성되는 추정 $\hat{\Phi}_x(\omega)$, $\hat{\Phi}_v(\omega)$와 사용되게 되는 이들 추정의 해석적인 식에 따라 다르다. 다른 식들의 예가 다음 장의 표 2에 주어져 있다. 다음 설명의 대부분은 입력프레임 {x(k)}으로부터 추정 $\hat{\Phi}_x(\omega)$, $\hat{\Phi}_v(\omega)$를 형성하는 다른 방법에 집중된다.

<43> PSD추정기(24)는 프레임이 음성(S) 또는 배경잡음(B)을 포함하는지를 판단하기 위하여 입력프레임 {x(k)}을 사용하는 음성행위검출기(VAD(26))에 의해 제어된다. 적절한 VAD가 [5],[6]에 설명되어 있다. VAD는 도 2에 설명된 4상태를 가지는 상태기계로서 구현될 수 있다. 최종 제어신호(S/B)는 PSD추정기(24)로 간다. VAD(26)가 음성(S), 상태(21 및 22)를 나타내면, PSD추정기(24)는 $\hat{\Phi}_x(\omega)$를 형성하게 된다. 한편, VAD(26)가 비-음성행위(B), 상태(20)를 나타내면, PSD추정기(24)는 $\hat{\Phi}_v(\omega)$를 형성하게 된다. 후자의 추정은 다음 음성프레임 시퀀스 동안(이 시퀀스의 프레임들 각각의 $\hat{\Phi}_x(\omega)$와 함께) $\hat{H}(\omega)$를 형성하는데 사용된다.

<44> 또한 신호(S/B)는 스펙트럼 감산블럭(16)으로 가게 된다. 이 방법에서, 블록(16)은 음성 및 비-음성프레임 동안 다른 필터를 적용할 수 있다. 음성프레임 동안에 $\hat{H}(\omega)$는 상기 언급된 $\hat{\Phi}_x(\omega)$, $\hat{\Phi}_v(\omega)$의 식이다. 한편, 비-음성프레임 동안, $\hat{H}(\omega)$는 잡음억제 후에 음성프레임에 남는 배경음 레벨과 동일한 레벨로 배경음 레벨을 감소시키는 상수 $H(0 \le H \le 1)$일 수 있다. 이 방식에 있어서, 감지된 잡음레벨은 음성 및 비-음성프레임 동안 동일하게 된다.

<45> 바람직한 실시예에 있어서, (3)의 출력신호($\hat{s}(k)$))가 계산기기 전에, $\hat{H}(\omega)$는 다음에 따라 뒤에 필터될 수 있다.

$$H_p(\omega)=\max(0.1,W(\omega)\,\overline{H}(\omega)) \qquad \forall\omega \qquad (8)$$

표 1 : 후필터링 함수

상태(st)	$\overline{H}(\omega)$	주해
0	$1(\forall\omega)$	$\hat{s}(k)=x(k)$
20	$0.316(\forall\omega)$	묵음-10dB
21	$0.7\hat{H}(\omega)$	cautios 필터링(-3dB)
22	$\hat{H}(\omega)$	

<47> 이때 $\overline{H}(\omega)$는 표 1에 따라 계산된다. 스칼라 0.1은 잡음 기저(noise floor)가 -20dB라는 것을 의미한다.

<48> 게다가, 신호(S/B)는 또한 음성부호기(22)로 간다. 이는 음성과 배경음의 상이한 부호화를 가능하게 한다.

<49> PSD 에러분석

<50> s(k) 및 v(k)에 부과된 불변성(정상성) 가정은 잡음이 없는 음성신호(s(k))와 비교해볼 때 추정($\hat{s}(k)$)이 얼마나 정확한가에 대한 경계의 근원이라는 것은 명확하다. 이 장에서, 스펙트럼적 감산방법에 대한 분석기술이 도입된다. 이는 도입한 편차의 정확성을 위한 근사(0차 근사)식과 함께, PSD추정 $\hat{\Phi}_x(\omega)$ 및 $\hat{\Phi}_v(\omega)$의 일차 근사치를 기초로 한다(아래 (11)를 보라). 명확하게, 다음에서 사용된 방법(변환함수 H(ω)의 선택)과 그리고 사용된 PSD추정기의 정확성으로 때문에, 추정된 신호($\hat{s}(k)$)의 주파수 영역 에러에 대한 식이 유도된다. 인간의 귀가 위상왜곡에 대해 민간하지 않기 때문에, 이는 다음으로 규정되는 PSD에러를 고려하는 것이 적절하다.

$$\tilde{\Phi}_s(\omega) = \hat{\Phi}_s(\omega) - \Phi_s(\omega) \qquad (9)$$

이때

$$\hat{\Phi}_s(\omega) = \hat{H}^2(\omega)\Phi_x(\omega) \qquad (10)$$

<52> $\tilde{\Phi}_s(\omega)$의 구성은 필터된 잡음 측정치의 크기와 음성의 크기 간의 차이(주파수 영역에서)를 나타내는 에러항이라는 것을 명심해야 한다. 따라서, $\tilde{\Phi}_s(\omega)$는 양과 음의 값 둘 다를 가질 수 있고 그리고 어떠한 시간영역 신호의 PSD가 아니다. 식 (10)에서, $\hat{H}(\omega)$는 $\hat{\Phi}_x(\omega)$ 및 $\hat{\Phi}_v(\omega)$를 기초로 한 H(ω)의 추정을 나타낸다. 이 장에서, 분석은 전력감산(Power Subtraction : PS)의 경우로 한정된다([2]) 다른 종류들의 $\hat{H}(\omega)$도 비슷한 방식으로 분석될 수 있다(부록 A~C를 참조하라). 이외에도, 새로운 종류의 $\hat{H}(\omega)$가 도입되어 분석된다(부록 D~G를 보라). 적절한, 상이한 종류들의 $\hat{H}(\omega)$의 요약이 표 2에 나타나 있다.

<53> 표 2. 다른 스펙트럼 감산방법의 예 : 전력감산(PS)(δ=1에 대한 표준 PS, $\hat{H}_{ps}(\omega)$, 크기감산(MS), 위너필터링(WF)을 기초로 한 스픽트럼 감산방법 및 최대 가망성(ML)방법론 및 본 발명의 바람직한 실시예에 따른 개선된 전력감산(IPS).

$$\hat{H}(\omega)$$

$$\hat{H}_{\delta PS}(\omega)=\sqrt{1-\delta\hat{\Phi}_v(\omega)/\hat{\Phi}_x(\omega)}$$

$$\hat{H}_{MS}(\omega)=1-\sqrt{\hat{\Phi}_v(\omega)/\hat{\Phi}_x(\omega)}$$

$$\hat{H}_{WF}(\omega)=\hat{H}^2_{PS}(\omega)$$

$$\hat{H}_{ML}(\omega)=\frac{1}{2}(1+\hat{H}_{PS}(\omega))$$

$$\hat{H}_{IPS}(\omega)=\sqrt{\hat{G}(\omega)}\,\hat{H}_{PS}(\omega)$$

<55> 정의에 의해, H(ω)는 간격 0≤H(ω≤1에 속하고, 이는 표 2의 상응하는 추정된 량에 들어맞을 필요는 없어서, 따라서 실제로 반파 또는 전파정류([1])가 사용된다.

<56> 분석을 실행하기 위하여, $\hat{\Phi}_x(\omega)$와 $\hat{\Phi}_v(\omega)$가 거의 편향되지 않도록 프레임 길이(N)가 충분히 크다고(N>>1) 가정한다. 1차 편차를 도입한다.

$$\hat{\Phi}_x(\omega) = \Phi_x(\omega) + \Delta_x(\omega)$$

$$\hat{\Phi}_v(\omega) = \Phi_v(\omega) + \Delta_v(\omega) \qquad (11)$$

이때, $\Delta_x(\omega)$와 $\Delta_v(\omega)$는 $E[\Delta_x(\omega)/\Phi_x(\omega)]^2 << 1$ 및 $E[\Delta_v(\omega)/\Phi_v(\omega)]^2$

<58> << 1가 되도록 제로-평균 확률론적 변수이다. 여기서 결국, 기호 E[·]는 통계적 기대값을 나타낸다. 또한, 잡음의 상관시간이 프레임 길이에 비해 짧다면, ℓ ≠k에 대해 $E[(\overline{\Phi}_v(\omega)^{\ell}-\Phi_v(\omega))(\overline{\Phi}_v(\omega)^{k}-\Phi_x(\omega))] \approx 0$이고, 이때 $\overline{\Phi}_v(\omega)^{\ell}$는 ℓ번째 프레임의 데이터를 기초로 한 추정이다. 이는 $\Delta_x(\omega)$와 $\Delta_v(\omega)$는 거의 독립적이라는 것을 나타낸다. 그렇지 않고, 만일 잡음이 강하게 상관되었다면, $\Phi_v(\omega)$는 주파수($\omega_1,\ldots,\omega_n$)에 위치된, 제한된 수 (<<N)의 (강한)피크들을 가진다고 가정한다. 그러면, $E[(\overline{\Phi}_v(\omega)^{\ell}-\Phi_v(\omega))(\overline{\Phi}_v(\omega)^{k}-\Phi_v(\omega))]\approx 0$는 $\omega \neq \omega_j$ j=1,...,n 및 ℓ ≠ k에 들어맞고 그리고 분석은 여전히 ω≠ωj j=1,...,n에 들어맞게 된다.

<59> 식(11)은 페리오드그램 또는 평균된 페리오드그램과 같은 점근적으로 (N>>1) 편향되지 않은 PSD가 사용된다는 것을 의미한다. 그러나, 블랙맨-터키 PSD추정기와 같은, 점근적으로 편향된 PSD추정기를 사용하면, 식(11)을 다음과 같은 것으로 대체하여도 유사한 분석이 들어맞는다.

$$\hat{\Phi}_x(\omega) = \Phi_x(\omega) + \Delta_x(\omega) + B_x(\omega)$$

및

$$\hat{\Phi}_v(\omega) = \Phi(\omega) + \Delta_v(\omega) + B_v(\omega)$$

<61> 이때, $B_x(\omega)$와 $B_v(\omega)$는 PSD추정기에서 점근적인 편향을 설명하는 결정론적인 항이다.

<62> 또한, 식(11)은 식 (9)의 $\tilde{\Phi}_s(\omega)$가 $\Delta_x(\omega)$와 $\Delta_v(\omega)$에서 (1차 근사)선형 함수라는 것을 의미한다. 다음에서, 편향에러($E[\tilde{\Phi}_s(\omega)]$)와 에러변동($Var(\tilde{\Phi}_s(\omega))$)의 항으로 상이한 방법의 실행이 고려된다. 다음 장에서 완벽한 유도가 $\hat{H}_{ps}(\omega)$에 대해 주어지게 된다. 표 1의 다른 스펙트럼 감산방법의 비슷한 유도가 부록 A~G에 주어져 있다.

$\hat{H}_{PS}(\omega)$의 분석(δ=1일 때 $\hat{H}_{\delta PS}(\omega)$)

<64> 식 (10)과 표 2의 $\hat{H}_{ps}(\omega)$를 식 (9)에 삽입하고, 테일러 급수 전개식($(1+x)^{-1}\simeq 1-x$)을 사용하고 그리고 1차 편차보다 높은 것을 무시함으로써, 다음의 간단한 계산이 구해진다.

$$\tilde{\Phi}_s(\omega) \simeq \frac{\Phi_v(\omega)}{\Phi_x(\omega)} \Delta_x(\omega) - \Delta_v(\omega) \quad (12)$$

<66> 이때 "≃ "는 우세한 항들만이 존속되는 근사 등식을 나타내는데 사용된다. 량 $\Delta_x(\omega)$와 $\Delta_v(\omega)$는 제로-평균 통계론적인 변수이다. 따라서,

$$E[\tilde{\Phi}_s(\omega)] \simeq 0 \quad (13)$$

및

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \frac{\Phi_v^2(\omega)}{\Phi_x^2(\omega)} \mathrm{Var}(\hat{\Phi}_x(\omega)) + \mathrm{Var}(\hat{\Phi}_v(\omega)) \quad (14)$$

<68> 식을 계속하기 위해, 점근적으로 편향되지 않은 스펙트럼추정기($\hat{\Phi}(\omega)$)에 대한 식의 일반적인 결과를 사용하여((7)참조), 몇몇(주파수 종속) 변수 $\gamma(\omega)$에 대한 식이 만들어진다.

$$\mathrm{Var}(\hat{\Phi}(\omega)) \simeq \gamma(\omega)\Phi^2(\omega) \quad (15)$$

<70> 예컨대, 페리오드그램은 $\gamma(\omega) \approx 1+(\sin\omega N/N\sin\omega)^2$에 대응하고, N>>1에 대해 $\gamma \approx 1$로 감소한다. 식(14)와 (15)를 결합하면,

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \gamma \Phi_v^2(\omega) \quad (16)$$

<72> 이 된다.

<73> $\hat{\mathbf{H}}_{MS}(\omega)$에 대한 결과

<74> $\hat{\mathbf{H}}_{MS}(\omega)$에 대한 유사한 계산은(상세한 것은 부록A에 주어져 있다.)

$$E[\tilde{\Phi}_s(\omega)] \simeq 2\Phi_v(\omega)\left(1 - \sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right)$$

및

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \left(1 - \sqrt{1 + \frac{\Phi_s(\omega)}{\Phi_v(\omega)}}\right)^2 \gamma \Phi_v^2(\omega)$$

<76> 이다.

<77> $\hat{\mathbf{H}}_{WF}(\omega)$에 대한 결과

<78> $\hat{\mathbf{H}}_{WF}(\omega)$에 대한 계산은(상세한 것은 부록B에 주어져 있다.),

$$E[\tilde{\Phi}_s(\omega)] \simeq -\left(1 - \frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)\Phi_v(\omega)$$

및

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq 4\left(1 - \frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)^2 \gamma \Phi_v^2(\omega)$$

<81> 가 된다.

<82> $\hat{\mathbf{H}}_{ML}(\omega)$에 대한 결과

<83> $\hat{\mathbf{H}}_{ML}(\omega)$에 대한 계산(상세한 것은 부록 C에 주어져 있다)은,

$$E[\tilde{\Phi}_s(\omega)] \simeq \frac{1}{2}\Phi_v(\omega) - \frac{1}{4}\left(\sqrt{\Phi_x(\omega)} - \sqrt{\Phi_s(\omega)}\right)^2$$

및

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \frac{1}{16}\left(1+\sqrt{\frac{\Phi_x(\omega)}{\Phi_s(\omega)}}\right)^2 \gamma\ \Phi_v^2(\omega)$$

<85> 가 된다.

<86> $\hat{\mathbf{H}}_{IPS}(\omega)$에 대한 결과

<87> $\hat{\mathbf{H}}_{IPS}(\omega)$에 대한 계산은($\hat{\mathbf{H}}_{IPS}(\omega)$는 부록 D에 유도되어 있고 그리고 부록 E에 분석되어 있다.),

$$E[\tilde{\Phi}_s(\omega)] \simeq (\overline{G}(\omega)-1)\Phi_s(\omega)$$

및

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \overline{G}^2(\omega)$$

$$\times \left(\overline{G}(\omega) + \gamma\,\Phi_v(\omega)\frac{\Phi_v(\omega)+2\Phi_x(\omega)}{\Phi_s^2(\omega)+\gamma\,\Phi_v^2(\omega)}\right)^2 \gamma\,\Phi_v^2(\omega)$$

<89> 가 된다.

<90> 공통특징

<91> 고려한 방법에 있어서, 편향에러는 $\hat{\mathbf{H}}(\omega)$의 선택에만 의존하는 반면, 에러변화는 $\hat{\mathbf{H}}(\omega)$의 선택과 사용된 PSD추정기의 변화 둘 다에 의존한다는 것을 명심해야 한다. 예컨대, $\Phi_v(\omega)$의 평균된 페리오드그램 추정에 대해, 식(7)로부터 $\gamma_v \approx 1/\tau$을 가진다. 한편, $\Phi_x(\omega)$의 추정에 대해 단일 프레임 페리오드그램을 사용하여, $\gamma_x \approx 1$을 가진다. $\tau >> 1$에 대해, 상기 변화식에 나타나는 $\gamma = \gamma_x + \gamma_v$에서 우세한 항은 γ_x이고, 따라서 주 에러원은 잡음이 있는 음성을 기초로 한 단일 프레임 PSD추정이다.

<92> 상기의 설명으로부터, 스펙트럼적 감산기술을 개선하기 위하여, γ_x의 값을 감소시키고(가능한 훌륭한 실행을 가지는, 근사적으로 편향되지 않은 추정기인 적절한 PSD주정기를 선택하고) 그리고 "훌륭한" 스펙트럼적 감산기순을 선택($\hat{\mathbf{H}}(\omega)$를 선택)하는 것이 바람직하다는 것을 알아야 한다. 본 발명의 해결방법은 γ_x의 값이 물리적인 음성영역의 모델링을 사용하여(자유도의 수를 N(한 프레임에서 샘플의 수)에서 N보다 작은 값으로 감소시킴으로써) 감소될 수 있다. s(k)가 자기회귀(AR) 모델(전형적으로 계수 $p \approx 10$)로 정확히 기술될 수 있다는 것을 잘 공지되어 있다. 이는 다음 두 장의 주제이다.

<93> 이외에도, $\tilde{\Phi}_s(\omega)$의 정확성(및 함축적으로, $\hat{S}(k)$의 정확성)은 $\hat{\mathbf{H}}(\omega)$의 선택에 의존한다. 새로운, 바람직한 $\hat{\mathbf{H}}(\omega)$의 선택이 부록 D~G에서 유도되어 분석된다.

<94> 음성 AR 모델링

<95> 본 발명의 바람직한 실시예에서 s(k)는 자기회귀(AR)프로세스로서 모델된다.

$$s(k) = \frac{1}{A(q^{-1})}\,\omega(k) \qquad k=1, \ldots, N \qquad (17)$$

<97> 이때 $A(q^{-1})$은 역변위 연산자($q^{-1}w(k)=w(k-1)$등)에서 모닉(최고차계수는 1과 같다) p번째 차 다항식이다.

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_p q^{-p} \qquad (18)$$

<99> 그리고 w(k)는 변이량 σ_w^2를 가지는 백색 제로-평균잡음이다. 언뜻보아, AR모델만 고려하는 것으로 제한되게 보일 수 있다. 그러나, 음성모델링에 대해 AR모델을 사용하는 것은 추정된 모델의 정확성에 대한 음성영역의 물리적 모델링과, 여기서 보다 중요한, 잡음이 있는 음성으로부터의 물리적 제한으로부터 동기를 부여받는다.

<100> 음성신호 프로세싱에서, 프레임 길이(N)는 변이를 감소시키고 그리고 PSD추정기의 비편향을 유지하기 위하여 프레임 내에 평균기술의 적용이 충분히 이루어지도록 크지 않을 수 있다. 따라서, 예컨대 식(12)에서 제1항의 영향을 감소시키기 위하여, 음성영역의 물리적 모델링이 사용되어야 한다. AR구조(17)가 s(k)에 부과된다. 명백히,

$$\Phi_x(\omega) = \frac{\sigma_\omega^2}{|A(e^{i\omega})|^2} + \Phi_v(\omega) \qquad (19)$$

<102> 이외에도, Φv(ω)는 파라메터에 관한 모델로서 기술될 수 있다.

$$\Phi_v(\omega) = \sigma_v^2 \frac{|B(e^{i\omega})|^2}{|C(e^{i\omega})|^2} \qquad (20)$$

<104> 이때, $B(q^{-1})$ 및 $C(q^{-1})$는 (18)의 $A(q^{-1})$와 비슷하게 규정된, 각각 q 및 r차 다항식이다. 간략하게 하기 위하여, 식 (20)의 파라메터에 관한 잡음모델은 파레메터에 관한 모델의 차수가 추정되는 아래의 설명에 사용된다. 그러나, 배경잡음의 다른 모델이 또한 가능하다는 것을 알 수 있다. 식 (19)와 식 (20)을 결합하면,

$$x(k) = \frac{D(q^{-1})}{A(q^{-1})C(q^{-1})}\eta(k) \qquad k=1, \ldots, N \qquad (21)$$

<106> 가 될 수 있다. 이때, η(k)는 변이량 σ_η^2를 가지는 제로평균 배색잡음이고 그리고 $D(q^{-1})$는 다음 항등식으로 주어진다.

$$\sigma_\eta^2 |D(e^{i\omega})|^2 = \sigma_\omega^2 |C(e^{i\omega})|^2 + \sigma_v^2 |B(e^{i\omega})|^2 |A(e^{i\omega})|^2 \qquad (22)$$

<108> 음성파라메터 추정

<109> 아무런 잡음이 존재하지 않으면 식 (17)~(18)의 파라메터를 추정하는 것은 간단하다. 잡음이 없는 경우에, 식 (22)의 오른쪽의 제2항은 사라져, 따라서 식(21)은 폴-제로(pole-zero) 소거후 식 (17)로 감소된다.

<110> 여기서, 자기상관 방법을 기초로 한 PSD추정기가 모색된다. 이에 대한 동기는 네 가지다.

<111> ● 자기상관 방법은 잘 공지되어 있다. 특히, 추정된 파라메터들은 최소 위상이어서, 최종 필터의 안정성을 보장한다.

<112> ● 레빈슨(Levinson) 알고리즘을 사용. 이 방법은 용이하게 구현되고 그리고 낮음 계산의 복잡도를 가진다.

<113> ● 최적의 절차는 비선형 최적화를 포함하고, 명백하게 몇몇 초기화를 필요로 한다. 자기상관 방법은 아무 것도 필요로 하지 않는다.

<114> ● 실제적인 관점에서 보아, 만일 동일한 추정절차를 사용할 수 있다면, 동일한 추정절차는 품질이 떨어진 음성과 깨끗한 음성에 사용될 수 있는 것이 바람직하다. 달리 말하면, 추정방법은 실제 동작시나리오에 관계 없어야 한다. 이는 음성-대-잡음비에 관계없다.

<115> ARMA모델(식 (21)과 같은)이 무한차 AR프로세스로 모델될 수 있다는 것은 잘 공지되어 있다. 유한수의 데이터들이 파라메터추정을 위해 사용될 수 있으면, 무한차 AR모델은 끝수가 버려져야 한다. 여기서 사용된 모델은,

$$x(k) = \frac{1}{F(1^{-1})}\eta(k) \qquad (23)$$

<117> 이다. 이때, $F(q^{-1})$은 $\bar{p}$차이다. 적절한 모델차수는 아래 설명에 주어진다. 근사모델(23)은 만일 그들의 PSD들이 근사적으로 동일하다면, 잡음프로세스에서 음성에 가깝다. 즉

$$\frac{|D(e^{i\omega})|^2}{|A(e^{i\omega})|^2|C(e^{i\omega})|^2} \approx \frac{1}{|F(e^{i\omega})|^2} \qquad (24)$$

<119> 이다. 음성영역의 물리적 모델링을 기초로, $p=\deg(A(q^{-1}))=10$을 고려하는 것이 일반적이다. 식(24)로부터, $\bar{p}=\deg(F(q^{-1}))>>\deg(A(q^{-1}))+\deg(C(q^{-1}))=p+r$ 이 일어난다. 이때 p+r은 $\Phi_x(\omega)$에

서 피크의 수와 거의 같다. 한편, AR모델을 사용하여 잡음이 있는 협소한 대역프로세스를 모델링하는 것은 신뢰성이 있는 PSD추정을 보장하기 위하여 $\bar{p}$<<N을 필요로 한다. 요약하면,

$$p+r << \bar{p} << N$$

<121> 적절한 경험법칙은 $\bar{p} \sim \sqrt{N}$으로 주어진다. 상기 설명으로부터 N >> 100일때 파라메터에 관한 접근법이 유리하다는 것을 예기할 수 있다. 또한 식 (22)로부터 잡음스페트럼이 평탄하면 할수록 N의 보다 작은 값이 허용된다는 것을 결론내릴수 있다. $\bar{p}$가 충분히 크지 않은 경우에도, 파라메터에 관한 접근법이 적당한 결과를 제공할 것으로 기대된다. 이 이유는, 에러 변이량의 항으로, 파라메터에 관한 접근법이 페리오드그램을 기초로 한 접근법(전형적인 예에서, 변이량 간의 비는 1:8과 같다 ; 아래를 보라) 보다 상당히 보다 정확한 PSD추정을 제공하기 때문이다. 이는 출력에서 음성잡음과 같은 인위음을 상당히 저감시킨다.

<122> 파라메터에 관한 PSD추정기는 다음과 같이 요약된다. 식 (23)에서 AR파라메터 $\{\hat{f}_1, \ldots, \hat{f}_p\}$와 잡음 변이량($\sigma_\eta^2$)을 계산하기 위하여 자기상관 방법과 고 차수 AR모델(모델 차수 $\bar{p}$ >> p 및 $p \sim \sqrt{N}$)을 사용한다. 추정된 AR모델로부터 다음 식에 따라 $\hat{\Phi}_x(\omega)$를(식 (3)의 X(ω)의 주파수빈(frequency bin)에 대응하는 N불연속점에서) 계산한다.

$$\hat{\Phi}_x(\omega) = \frac{\hat{\sigma}_{\eta^2}}{|\hat{F}(e^{i\omega})|^2} \qquad (25)$$

<124> 그런 다음, 표 2에서 고려한 스펙트럼 감산기술 중 하나가 음성(s(k))을 향상시키기 위하여 사용된다.

<125> 디음, 파라메터에 관한 PSD추정기의 변이량에 대한 저 차수의 근사치(고려한 비파라메터에 관한 방법에 대한 식 (7)과 비슷한)와, s(k)의 퓨리에 급수 전개식이 잡음이 백색이라는 가정 하에 사용된다. 그런 다음, $\hat{\Phi}_x(\omega)$의 점근적(데이타의 수(N >> 1)와 모델 차수($\bar{p}$ >> 1)둘 다에 대한) 변이량이 다음에 의해 주어진다.

$$Var(\hat{\Phi}_x(\omega)) \simeq \frac{2p}{N}\Phi_x^2(\omega) \qquad (26)$$

<127> 상기 식은 또한 순수(고-차수) AR프로세스에 들어맞는다. 식 (26)으로부터, $\gamma_x \approx 2\bar{p}/N$이 상기 언급한 경험법칙에 따라 $\gamma_x \simeq 2/\sqrt{N}$과 거의 동일하게 되고, 페리오드그램된다.

<128> 예로서, 이동전화의 핸드프리 환경에서, 잡음은 약 0.5초(8000Hz의 샘플링속도와 프레임길이 N=256에서) 동안 정적이어서, $\tau \approx 15s$ 가 되고, 따라서 $\gamma_v \simeq 1/15$가 된다는 것을 추정할 수 있다. 또한, $\bar{p}=\sqrt{N}$에서, $\gamma_x 1/80$이다.

<129> 도 3은 전형적인 음성프레임에 대한 페리오드그램 추정과 본 발명에 따른 파라메터에 관한 PSD추정 간의 차이점을 설명한다. 이 예에서 N=256(256샘플)이고, 그리고 10개의 파라메터를 가지는 AR모델이 사용되었다. 파라메터에 관한 PSD추정($\hat{\Phi}_x(\omega)$)은 상응하는 페리오드그램 PSD추정 보다 훨씬 더 매끄럽다는 것을 명심해야 한다.

<130> 도 4는 잡음이 있는 배경에서 음성을 포함하는, 5초 동안 샘플된 음성신호를 설명한다. 도 5는 높은 가청품질에 우선권을 주는 페리오드그램 PSD추정을 기초로한 스펙트럼 감산 후 도 4의 신호를 설명한다. 도 6은 본 발명에 따른 파라메터에 관한 PSD추정을 기초로 한 스펙트럼 감산 후 도 4의 신호를 설명한다.

<131> 도 5와 도 6의 비교는, 상당한 잡음억제(10dB정도의)가 본 발명에 따른 방법으로 이루어진다는 것을 보여준다(도 1의 설명과 관련해 상기에서 설명된 바와 같이, 감소된 잡음레벨은 음성 및 비음성 프레임 둘 다에서 동일하다). 도 6에서 명확하지 않은 다른 차이점은 최종 음성신호가 도 5의 음성신호보다 덜 왜곡된다는 것이다.

<132> PSD에러의 편향과 에러 변이량의 항으로, 고려된 모든 방법에 대한 이론적인 결과가 표 3에 요약되어 있다.

<133> 상이한 방법들을 분류할 수 있다. 적절한 방법을 어떻게 선택하는 가에 대한 두 개의 기준을 분류할 수 있다.

<134> 먼저, 낮은 순간 SNR에 대해, 방법은 $\hat{S}(k)$에서 음색치 인위를 피하기 위하여 낮은 변이를 가지는 것이 바람직하다. 이는 편향의 증가없이 불가능하고, 그리고 이 편향항은 낮은 순간 SNR을 가지는 주파수영역을 억제하기 위해(및 증폭시키지 않기 위해) 음의 부호를 가져야 한다(따라서, (9)의 $\hat{\Phi}_s(\omega)$가 0이 되게 한다)이 기준을 충족시키는 후보군은 각각 MS, IPS 및 WF이다.

<135> 둘째로, 높은 순간 SNR에 대해, 낮은 비율의 음성왜곡이 바람직하다. 또한, 만일 편향항이 우세하다면, 이는 양의 부호를 가져야 한다. ML, $\overline{\delta}$-PS, PS, IPS, 및 (가능하다면) WF가 제1 선언문을 충족시킨다. 편향항은 ML 및 WF에 대한 MSE식에서 우세하고 편향항의 부호는 ML에 대해서는 양이고, WF에 대해서는 음이다. 따라서, ML, $\overline{\delta}$ PS, PS, 및 IPS는 이 기준을 충족시킨다.

<136> 표 3 : 전력감산(PS)(표준 PS, δ=1에 대한 $\hat{H}_{PS}(\omega)2$, 크기감산(MS), 개선된 전력감산(IPS) 및 위너필터링(WF) 및 최대 가망성(ML)방법론을 기초로 한 스펙트럼 감산방법, 순간 SNS은 $SNR=\Phi_s(\omega))/\Phi_v(\omega))$로 규정된다. PS에 대해 최적 감산인자($\overline{\delta}$)는 식 (58)로 주어지고 그리고 IPS에 대해, $\hat{G}(\omega)$는 식 (45)에 의해 주어지고 $\Phi_x(\omega)$와 $\Phi_v(\omega)$는 각각 $\widehat{\Phi}_x(\omega)$와 $\widehat{\Phi}_v(\omega)$로 대체된다.

$\hat{H}(\omega)$	편향	변이량
	$E[\tilde{\Phi}_s(\omega)]/\Phi_v(\omega)$	$Var(\tilde{\Phi}_s(\omega))/\gamma\Phi_v^2(\omega)$
δPS	$1-\delta$	δ^2
MS	$-2(\sqrt{1+SNR}-1)$	$(\sqrt{1+SNR}-1)^2$
IPS	$-\frac{\gamma SNR}{\gamma+SNR^2}$	$\left(\frac{SNR^2}{SNR^2+\gamma}\right)^2\left(1+2\gamma\frac{1+SNR}{SNR^2+\gamma}\right)^2$
WF	$-\frac{SNR}{SNR+1}$	$4\left(\frac{SNR}{SNR+1}\right)^2$
ML	$\frac{1}{2}-\frac{1}{4}(\sqrt{SNR+1}-\sqrt{SNR})^2$	$\frac{1}{16}\left(1+\sqrt{1+\frac{1}{SNR}}\right)^2$

<138> 알고리즘 관점

<139> 이 장에서 본 발명에 따른 스펙트럼 감산방법의 바람직한 실시예가 (7)을 참조해 설명된다.

1. 입력 : $X=\{x(k) \mid k=1,\ldots,N\}$

2. 설계변수

$\overline{p}$ 잡음하의 음성 모델차수

ρ $\widehat{\Phi}_v(\omega)$에 대한 실행 평균갱신인자

<141> 3. 입력데이타의 각 프레임에 대해 하는 일 :

<142> (a) 음성검출(단계110)

<143> VAD출력이 st=21 또는 st=22와 동일하다면 가용 음성은 참(true)으로 설정된다. 만일 st=20이면 음성은 거짓(false)으로 설정된다. 만일 VAD출력이 st=0과 같다면, 알고리즘은 재초기화된다.

<144> (b) 스펙트럼 추정

<145> 만일 음성 추정($\widehat{\Phi}_x(\omega)$)이면 :

<146> i. 제로평균 조정된 입력데이타{s(k)}에 적용된 자기상관방법을 사용하여 모든-폴 모델(22)의 계수(다항식 계수{ $\hat{f}_1,\ldots,\hat{f}_p$ }와 변이량($\widehat{\sigma}_\eta^2$))를 추정한다(단계120).

<147> ii. 식 (25)에 따라 $\widehat{\Phi}_x(\omega)$를 계산한다(단계130).

<148> 그렇지 않다면 $\widehat{\Phi}_v(\omega)$를 추정(단계140)

<149> i. 식 (4)를 사용하여 배경잡음 스펙트럼모델 $\hat{\Phi}_v(\omega)$를 갱신한다. 이때 $\hat{\Phi}_v(\omega)$는 제로평균 조정되고 그리고 해닝/해밍 윈도우된(Hanning/Hamming windowed) 입력데이타(x)를 기초로 한 페리오드그램이다. 여기서 윈도우된 데이터가 사용되는 한편, $\hat{\Phi}_x(\omega)$는 윈도우되지 않은 데이터를 기초로 하기 때문에, $\hat{\Phi}_v(\omega)$는 적절히 정규화되게 된다. $\hat{\Phi}_v(\omega)$의 적절한 초기값은 인자 0.25로 스케일된 제 1 프레임의 페리오드그램의 (주파수빈에 대한)평균으로 주어지고, 이는 선천적인 백색잡음 가정이 배경잡음에 부가된다는 것을 의미한다.

<150> (c) 스펙트럼 감산(단계150)

<151> i. 표 1에 따른 주파수 가중함수($\hat{H}(\omega)$)를 계산한다.

<152> ii. 가능한 후필터링, 무팅 및 잡음기저 조정.

<153> iii. 식 (3)과 제로-평균 조정된 데이터{x(k)}를 사용하여 출력을 계산한다. 데이터{x(k)}는 실제 프레임오버랩에 따라 윈도우되거나 또는 되지 않을 수 있다(장방형 윈도우가 비-오버랩핑 프레임에 대해 사용되는 반면, 해닝윈도가 50%오버랩으로 사용된다).

<154> 상기 설명으로부터, 본 발명은 가청품질을 손상시킴이 없이 상당히 잡음을 감소시킬 수 있다는 것을 명확히 알 수 있다. 이러한 개선은 음성 및 비음성 프레임에 대해 사용된 개별적인 전력 스펙트럼 추정방법으로 설명할 수 있다. 이들 방법들은 각 전력 스펙트럼추정의 변이량을 최소화하기 위하며 음성 및 비음성(배경잡음)의 상이한 특성을 이용한다.

<155> ● 비음성 프레임에 대해, $\hat{\Phi}_v(\omega)$는 각 프레임의 모든 N샘플을 사용하는 비-파라메터에 관한 전력 스펙트럼 추정방법, 예컨대 페리오드그램추정을 기초로한 FFT로 추정된다. 비음성 프레임의 모든 N 자유도를 유지함으로써, 다양한 종류의 배경잡음을 모델할 수 있다. 배경잡음이 수 개의 프레임에 걸쳐 불변인 것으로 가정하기 때문에, $\hat{\Phi}_v(\omega)$의 변이량의 감소는 수 개의 비음성 프레임에 대한 전력 스펙트럼 추정을 평균함으로써 이루어질 수 있다.

<156> ● 음성 프레임에 대해, $\hat{\Phi}_x(\omega)$는 음성의 파라메터에 관한 모델을 기초로한 파라메터에 관한 전력 스펙트럼 추정방법으로 추정할 수 있다. 이 경우, 음성의 특별한 특성을 음성프레임의 (파라메터에 관한 모델에서 파라메터의 수에 대한) 자유도 수를 감소시키기 위하여 사용한다. 보다 소수의 파라메터를 기초로 한 모델은 전력 스펙트럼추정의 변이량을 감소시킨다. 이러한 접근법은, 음성이 프레임 동안에만 불변되는 것으로 추정되기 때문에 음성 프레임에 바람직하다.

<157> 본 기술분야의 당업자라면 첨부된 특허청구의 범위를 벗어남이 없이 본 발명에 대해 다양한 수정과 변경이 이루어질 수 있다는 것을 알 수 있을 것이다.

<158> 부록 A

<159> $\hat{H}_{MS}(\omega)$의 분석

<160> $\hat{H}_{MS}(\omega)$에 대한 계산을 패러렐링(paralleling)하면,

$$\tilde{\Phi}_s(\omega) = \left(1-\sqrt{\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}}\right)^2 \Phi_x(\omega) - \Phi_s(\omega)$$

$$\simeq \left(1-\sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right)\left(2\Phi_v(\omega) - \frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)+\Delta_v(\omega)\right) \qquad (27)$$

<162> 이때, 두 번째 등식에서, 또한 테일러 급수 전개식 $\sqrt{1+x} \simeq 1+x/2$가 사용된다. 식 (27)로부터, $\tilde{\Phi}_s(\omega)$의 기대값이 0이 아니라는 것을 알 수 있고 기대값은 다음 식으로 주어진다.

$$E[\ \widetilde{\Phi}_s(\omega)] \simeq 2\Phi_v(\omega)\left(1-\sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right) \qquad (28)$$

및,

$$\mathrm{Var}(\ \widetilde{\Phi}_s(\omega)) \simeq$$

$$\left(1-\sqrt{\frac{\Phi_x(\omega)}{\Phi_v(\omega)}}\right)^2\left(\frac{\Phi_v^2(\omega)}{\Phi_x^2(\omega)}\ \mathrm{Var}(\ \widehat{\Phi}_x(\omega)) + \mathrm{Var}(\ \widehat{\Phi}_v(\omega))\right) \qquad (29)$$

<164> 식 (29)와 식 (15)를 결합하면,

$$\mathrm{Var}(\ \widetilde{\Phi}_s(\omega)) \simeq \left(1-\sqrt{1+\frac{\Phi_s(\omega)}{\Phi_v(\omega)}}\right)^2 \forall \Phi_v^2(\omega) \qquad (30)$$

<166> 부록 B

<167> $\hat{H}_{WF}(\omega)$의 분석

<168> 이 부록에서, PSD에러가 위너필터링을 기초로 한 음성강화([2])를 위해 유도되었다. 이 경우, $\hat{H}(\omega)$는 다음과 같이 주어진다.

$$\hat{H}_{WF}(\omega) = \frac{\widehat{\Phi}_s(\omega)}{\widehat{\Phi}_s(\omega) + \widehat{\Phi}_v(\omega)} = \hat{H}_{PS}^2(\omega) \qquad (31)$$

<170> 여기서, $\widetilde{\Phi}_s(\omega)$는 $\Phi_s(\omega)$의 추정이고 그리고 두 번째 등식은 $\widetilde{\Phi}_s(\omega)=\widetilde{\Phi}_x(\ \omega)-\ \widetilde{\Phi}_v(\omega)$로부터 도출된다.

$$\hat{H}_{WF}^2(\omega) \simeq \frac{\Phi_s(\omega)}{\Phi_x^2(\omega)}\left(\Phi_s(\omega) + 2\left\{\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega)\right\}\right) \qquad (32)$$

<172> 를 명심하라.

<173> 간단한 계산으로,

$$\widetilde{\Phi}_s(\omega) \simeq \left(1-\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)$$

$$\times\left(-\Phi_v(\omega) + 2\left\{\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega)\right\}\right) \qquad (33)$$

<175> 가 된다. 식 (33)으로부터,

$$E[\ \widetilde{\Phi}_s(\omega)] \simeq -\left(1-\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)\Phi_v(\omega) \qquad (34)$$

및

$$\mathrm{Var}(\ \widetilde{\Phi}_s(\omega)) \simeq 4\left(1-\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\right)^2 \forall \Phi_v^2(\omega) \qquad (35)$$

<177> 가 도출된다.

<178> 부록 C

<179> $\hat{H}_{ML}(\omega)$의 분석

<180> 알려지지 않은 진폭 및 위상의 결정론적인 파형으로 음성을 특징지움으로써, 최대 가망성(ML)스펙트럼 감산방법은 다음으로 규정된다.

$$\hat{H}_{ML}(\omega) = \frac{1}{2}\left(1+\sqrt{1-\frac{\hat{\Phi}_v(\omega)}{\hat{\Phi}_x(\omega)}}\right) \qquad (36)$$
$$= \frac{1}{2}(1+\hat{H}_{PS}(\omega)$$

<182> 식 (36)에 식 (11)을 삽입함으로써, 직접적인 계산은,

$$\hat{H}_{ML}(\omega) = \frac{1}{2}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\left(1-\frac{\Delta_v(\omega)}{\Phi_s(\omega)}+\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\frac{\Delta_x(\omega)}{\Phi_s(\omega)}\right)^{\frac{1}{2}}\right)$$

$$\simeq \frac{1}{2}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\right) \qquad (37)$$

$$+\frac{1}{4}\frac{1}{\sqrt{\Phi_x(\omega)\Phi_s(\omega)}}\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right)$$

<184> 이때, 첫 번째 등식에서는 테일러 급수 전개식 $(1+x)^{-1} \simeq 1-x$ 및 두 번째 등식에서는 $\sqrt{1+x} \simeq 1+x/2$가 사용된다. 이제 PSD에러를 계산하기가 간단해진다. 식(9)~(10)에 식 (37)을 삽입하고, $\hat{H}^2_{ML}(\omega)$의 전개식에서 1차 편차보다 큰 것을 무시하면 ,

$$\tilde{\Phi}_s(\omega) \simeq \frac{1}{4}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\right)^2\Phi_x(\omega)-\Phi_s(\omega)$$

$$+\left(1+\sqrt{\frac{\Phi_x(\omega)}{\Phi_s(\omega)}}\right)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)-\Delta_v(\omega)\right) \qquad (38)$$

<186> 가 된다. 식 (38)로부터,

$$E[\tilde{\Phi}_s(\omega)] \simeq \frac{1}{4}\left(1+\sqrt{\frac{\Phi_s(\omega)}{\Phi_x(\omega)}}\right)^2\Phi_x(\omega)-\Phi_s(\omega)$$

$$\frac{1}{2}\Phi_v(\omega)-\frac{1}{4}\left(\sqrt{\Phi_x(\omega)}-\sqrt{\Phi_s(\omega)}\right)^2 \qquad (39)$$

<188> 가 도출된다. 이때, 두 번째 등식에서 식 (2)가 사용된다. 또한

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \frac{1}{16}\left(1+\sqrt{\frac{\Phi_x(\omega)}{\Phi_s(\omega)}}\right)^2 \forall \Phi_v^2(\omega) \qquad (40)$$

<190> 가 도출된다.

<191> 부록 D

<192> $\hat{H}_{IPS}(\omega)$의 유도

<193> $\hat{\Phi}_s(\omega)$ 및 $\hat{\Phi}_v(\omega)$를 정확히 알면, $\Phi_x(\omega)$ 및 $\Phi_v(\omega)$를 대체한 $\hat{\Phi}_x(\omega)$ 및 $\hat{\Phi}_v(\omega)$를 가지는 $\hat{H}_{PS}(\omega)$인 제곱된 PSD에러는 $\hat{H}_{PS}(\omega)$로 최소화 된다. 이 사실은 식 (9)와 식(10)으로부터 바로 도출된다. 즉, $\tilde{\Phi}_s(\omega)=[H^2(\omega)\Phi_x(\omega)-\Phi_s(\omega)]^2=0$, 이때 식 (2)가 마지막 항에 사용된다. 이 경우,

H(ω)는 결정론적인 량인 반면, $\hat{H}(\omega)$는 통계적인 량이라는 것을 명심해야 한다. PSD추정의 불확실성을 고려하면, 일반적으로 이 사실은 더 이상 들어맞지 않아, 이 장에서 데이터에 상관없는 가중함수가 $\hat{H}_{PS}(\omega)$의 실행을 향상시키기 위하여 유도된다. 이를 위하여, 다음 형태의 변이량 전개식이 고려된다.

$$\mathrm{Var}(\tilde{\Phi}_s(\omega)) \simeq \xi \nu \Phi_v^2(\omega) \qquad (41)$$

<195> (PS에 대해 ξ=1이고 그리고 MS와 $\gamma = \gamma_x + \gamma_v$ 에 대해 $\xi = (1-\sqrt{1+SNR})^2$ 임). 변수 γ는 사용한 PSD추정방법에만 의존하고, 전이함수($\hat{H}(\omega)$)의 선택에 영향을 받지 않을 수 있다. 그러나, 첫 번째 인자 ξ는 $\hat{H}(\omega)$의 선택에 의존한다. 이장에서, 데이터에 상관없는 가중함수 $\bar{G}(\omega)$고 고려하면, $\hat{H}(\omega)=\sqrt{\bar{G}(\omega)}\hat{H}_{PS}(\omega)$는 제곱된 PSD에러의 기대값을 최소화 한다. 즉

$$\bar{G}(\omega) = \underset{G(\omega)}{\arg\min}\, E[\tilde{\Phi}_s(\omega)]^2$$

$$\tilde{\Phi}_s(\omega) = G(\omega)\hat{H}_{PS}^2(\omega)\Phi_{x(\omega)} - \Phi_s(\omega) \qquad (42)$$

<197> 식 (42)에서, G(ω)는 일반적인 가중함수이다. 계속 진행하기에 앞서, 만일 가중함수(G(ω))가 데이터에 종속되게 된다면, 많은 특수한 경우의, 공통적으로 사용한 방법과 같이, 예컨대 $G(\omega)=\hat{H}_{MS}^2(\omega)/\hat{H}_{PS}^2(\omega)$를 사용하는 크기 감산(Magnitude Subtraction)을 포함하는 일반적인 클래스의 스펙트럼 감산기술이 된다는 것을 명심해야 한다. 그러나, 이러한 결과는 데이터에 예속적인 G(ω)의 최적화가 G(ω)의 형태에 크게 의존하기 때문에 거의 관심이 없다. 따라서, 데이터에 예속적인 가중함수를 사용하는 방법은, 그러한 경우에 일반적인 결과가 유도될 수 없기 때문에 차례로 분석하여야만 한다.

<198> 식 (42)를 최소화하기 위하여, 간단한 계산은 다음과 같이 된다.

$$\tilde{\Phi}_s(\omega) \simeq (G(\omega)-1)\Phi_s(\omega)$$

$$+ G(\omega)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega)\right) \qquad (43)$$

<200> 제곱된 PSD에러의 기대치를 취하고 그리고 식 (41)를 사용하면 다음과 같이 된다.

$$E[\tilde{\Phi}_s(\omega)]^2 \simeq (G(\omega)-)^2\Phi_s^2(\omega) + G^2(\omega)\nu\Phi_v^2(\omega) \qquad (44)$$

<202> 식(44)는 G(ω)에서 이차식이고 그리고 분석적으로 최소화될 수 있다. 결과는

$$\bar{G}(\omega) = \frac{\Phi_s^2(\omega)}{\Phi_s^2(\omega)+\nu\Phi_v^2(\omega)}$$

$$= \frac{1}{1+\nu\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)-\Phi_v(\omega)}\right)^2} \qquad (45)$$

<204> 이때 두 번째 등식에서 식 (2)가 사용된다. $\bar{G}(\omega)$는 (알려지지 않은) PSD와 변수(γ)에 의존한다. 상기에서 설명된 바와 같이, 식 (45)의 알려지지 않은 PSD를 대응하는 추정으로 직접 대체할 수 없고 그리고 최종 수정된 PS방법은 최적화되고, 즉 식 (42)를 최소화 한다. 그러나, 설계절차에서 $\hat{\Phi}_x(\omega)$ 및 $\hat{\Phi}_v(\omega)$의 불확실성을 고려하면, 수정된 PS는 표준 PS보다 실행을 훨씬 더 "잘"하게 될 것이라는 것을 예상할 수 있다. 상기 이유 때문에, 이 수정된 PS방법은 개선된 전력 감산(IPS)으로 표시된다. IPS방법을 부록 E에서 분석하기 전에, 다음 의견사항이 정리된다.

<205> 높은 순간 SNR에 있어서(ω에 대해 $\Phi_s(\omega)/\Phi_v(\omega)\gg1$), 식 (45)로부터 $\bar{G}(\omega)\simeq$이 도출되고, 그리고 이 경우에 정규화된 에러 변이량 $\mathrm{Var}(\tilde{\Phi}_s(\omega))/\Phi_s^2(\omega)$(식 (41)을 보라)이 작기 때문에, IPS의 실행은 표준 PS의 실행에 (매우) 근사하다는 것을 결론내릴 수 있다. 한편, 낮은 순간 SNR에 있어서(ω에 대해 $\gamma\Phi_v^2(\omega) \gg \Phi_s^2(\omega)$), $G(\omega) \approx \Phi_s^2(\omega)/(\gamma\Phi_v^2(\omega))$는 다음과 같이 된다(식 (43)

참조).

$$E[\ \tilde{\Phi}_s(\omega)\ \approx\ -\Phi_s(\omega) \tag{46}$$

및

$$\mathrm{Var}(\ \tilde{\Phi}_s(\omega)\ \approx\ \frac{\Phi_s^4(\omega)}{\gamma\Phi_v^2(\omega)} \tag{47}$$

<207> 그러나, 낮은 SNR에서, 식 (45)의 $\bar{G}(\omega)$가 $\bar{G}(\omega)$로 대체되면, 즉 (45)의 $\Phi_x(\omega)$ 및 $\Phi_v(\omega)$를 그들의 추정된 $\hat{\Phi}_x(\omega)$ 및 $\hat{\Phi}_v(\omega)$로 대체하면 식 (46)~(47)은 거의 유효하다고 결론내릴 수 없다.

<208> 부록 E

<209> $\hat{H}_{IPS}(\omega)$의 분석

<210> 이 부록에서, IPS방법이 분석된다. 식 (45)를 고려하여, $\hat{G}(\omega)$를 (45)로 정의하고, $\Phi_v(\omega)$ 및 $\Phi_x(\omega)$가 상응하는 추정된 량으로 대체된다. 이는 식 (43)과 비교할 수 있는 다음 식으로 나타날 수 있다.

$$\tilde{\Phi}_s(\omega)\ \simeq\ (\bar{G}(\omega)-1)\Phi_s(\omega)$$

$$+\ \bar{G}(\omega)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega)\ -\ \Delta_v(\omega)\right) \tag{48}$$

$$\times\left(\bar{G}(\omega)+\gamma\Phi_v(\omega)\frac{\Phi_v(\omega)+2\Phi_x(\omega)}{\Phi_s^2(\omega)+\gamma\Phi_v^2(\omega)}\right)$$

명백히,

$$E[\ \tilde{\Phi}(\omega)]\ \simeq\ (\bar{G}(\omega)-1)\Phi_s(\omega) \tag{49}$$

및

$$\mathrm{Var}(\ \tilde{\Phi}_s(\omega))\ \simeq\ \bar{G}^2(\omega)$$

$$\times\left(\bar{G}(\omega)+\gamma\Phi_v(\omega)\frac{\Phi_v(\omega)+2\Phi_x(\omega)}{\Phi_s^2(\omega)+\gamma\Phi_v^2(\omega)}\right)^2\gamma\Phi_v^2(\omega) \tag{50}$$

<212> 가 된다. $\Phi_s(\omega)/\Phi_v(\omega)>>1$ 인 높은 SNR에 대해, 몇몇 직관이 식 (49)~(50)에서 구해질 수 있다 이 경우,

$$E[\ \tilde{\Phi}_s(\omega)]\ \simeq\ 0 \tag{51}$$

및

$$\mathrm{Var}(\ \tilde{\Phi}_s(\omega))\ \simeq\ \left(1+4\gamma\frac{\Phi_v(\omega)}{\Phi_s(\omega)}\right)\gamma\Phi_v^2(\omega) \tag{52}$$

<215> 를 알 수 있다. 식 (51)과 (52)에서 무시된 항은 차수 $O((\Phi_v(\omega))/\Phi_s(\omega))^2$이다. 따라서, 앞서 주장한 바와 같이, IPS의 실행은 높은 SNR에서 PS의 실행과 비슷하다. 한편, 낮은 SNR에 대해(ω에 대해 $\Phi_s^2(\omega)/(\gamma\Phi_v^2(\omega))<<1$), $\overline{G}(\omega) \simeq \Phi_s^2(\omega)/(\gamma\Phi_v^2(\omega))$, 및

$$E[\widetilde{\Phi}_s(\omega)] \simeq -\Phi_s(\omega) \qquad (53)$$

및

$$\mathrm{Var}(\widetilde{\Phi}_s(\omega)) \simeq 9\frac{\Phi_s^4(\omega)}{\gamma\Phi_v^2(\omega)} \qquad (54)$$

기 된다. 식 (53)~(54)를 상응하는 PS결과(식 (13) 및 (16))와 비교하면, 낮은 순간 SNR에 대해, IPS방법은 식 (9)의 $\widehat{\Phi}_s(\omega)$를 제로로 만듬으로써 표준 PS방법과 비교해 $\widetilde{\Phi}_s(\omega)$의 변이량을 상당히 감소시킨다는 것을 알 수 있다. 명백히, IPS와 PS변이량 사이의 비는 $O(\Phi_s^4(\omega)/\Phi_v^4(\omega))$ 정도이다. 또한 식 (53)~(54)를 근사적인 전개식(47)과 비교하면, 이들 사이의 비가 9와 같다는 것을 알 수 있다.

<217> 부록 F

<218> 최적 감산인자 $\overline{\delta}$를 가지는 PS

<219> 종종 고려되는 전력 감산방법의 변형은 다음과 같다.

$$\widehat{H}_{\delta PS}(\omega) = \sqrt{1-\delta(\omega)\frac{\widehat{\Phi}_v(\omega)}{\widehat{\Phi}_x(\omega)}} \qquad (55)$$

<221> 이때, $\delta(\omega)$는 주파수 예속함수이다. 특히, 몇몇 상수 $\delta>1$에 대한 $\delta(\omega)=\delta$ 때문에, 본 방법은 종종 과잉감산이 있는 전력감산으로 언급된다. 이러한 변형은 잡음레벨을 상당히 감소시키고 그리고 음성적인 인위를 저감시킨다. 이외에도, 이는 음성을 상당히 왜곡시키는데, 이러한 변형은 고품질 음성향상에 무용하게 된다. $\delta>>1$일 때, 이러한 사실은 식 (55)로부터 쉽게 알 수 있다. 따라서, 중간정도 및 낮은 잡음 대 음성비(ω-영역에서)에 대해 루트부호 아래의 전개식은 자주 음이되고 그리고 따라서 정류장치는 이를 제로에 설정하게 된다(반파 정류). 이는 SNR이 높은 주파수 대역만이 식 (3)의 출력신호(s(k))에 나타나게 된다는 것을 의미한다. 비선형 정류장치 때문에, 본 분석기술은 이 경우에 직접 적용될 수 없고, 그리고 $\delta>1$이 열악한 가청품질을 가진 출력을 만들기 때문에, 이 변형은 더 이상 연구되지 않는다.

<222> 그러나, $(\delta)\leq 1$일 때 다음의 발견적인 토의에서 잘 알 수 있듯이 흥미로운 경우가 된다. 앞서 설명한 바와 같이, $\Phi_x(\omega)$와 $\Phi_v(\omega)$를 정확히 알면, $\delta(\omega)=1$로 식 (55)가 제곱된 PSD에러를 최소화하는데 있어서 최적이 된다. 반면, $\Phi_x(\omega)$와 $\Phi_v(\omega)$를 완전히 모르면, 즉 이들의 추정이 이용될 수 없으면, 최선은, $\delta=0$로 식(55)를 사용하는 것에 상응하는 잡음측정 그 자체, 즉 $\hat{s}(k)=x(k)$로 음성을 추정하는 것이다. 상기 두 개의 외항 때문에, 알려지지 않은 $\Phi_x(\omega)$와 $\Phi_v(\omega)$가 각각 $\widehat{\Phi}_x(\omega)$와 $\widehat{\Phi}_v(\omega)$로 대체되면, 에러 $E[\widetilde{\Phi}_s(\omega)]^2$는 간격 $0<\delta(\omega)<1$에 있는 몇몇 $\delta(\omega)$에 대해 최소화되게 된다는 것을 예상할 수 있다.

<223> 이외에도, 경험적인 양에서, PSD에러와 비슷한 평균된 스펙트럼 왜곡개선이 MS에 대한 감산인자에 대해 실험적으로 연구되었다. 여러실험을 기초로, 최적 감산인자는 0.5에서 0.9에 이르는 간격 내에 있어야만 한다고 결론되었다.

<224> 명백히, 이 경우에서 PSD에러를 계산하면,

$$\widetilde{\Phi}_s(\omega) \simeq (1-\delta(\omega))\Phi_v(\omega) + \delta(\omega)\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega)\right) \qquad (56)$$

<226> 가 된다. 제곱된 PSD에러의 기대치를 취하면,

$$E[\widetilde{\Phi}_s(\omega)]^2 \simeq (1-\delta(\omega))^2\Phi_v^2(\omega) + \delta^2\gamma\Phi_v^2(\omega) \qquad (57)$$

<228> 이때, 식 (41)이 사용된다. 식 (57)은 $\delta(\omega)$에서 이차방정식이고 그리고 분석적으로 최소화될 수 있다.

<229> $\overline{\delta}$ 로 최적값을 나타내면, 결과는

$$\overline{\delta} = \frac{1}{1+\gamma} < 1 \qquad (58)$$

<231> 이다. 식 (58)의 γ가 거의 주파수에 독립적이기 때문에(적어도 N>>1에 대해), δ 또한 주파수에 독립적이다. 특히, δ는 $\Phi_x(\omega)$와 $\Phi_v(\omega)$에 독립적인데, 이는 $\widetilde{\Phi}_s(\omega)$의 변이량과 편향이 식 (57)로부터 직접 도출된다는 것을 의미한다.

<232> $\overline{\delta}$의 값은 몇몇 (현실적인)경우에 1보다 작은 것으로 여질 수 있다. 예컨대, 다시 $\gamma_v=1/\tau$ 및 $\gamma_x=1$을 고려하면, $\overline{\delta}$는

$$\overline{\delta} \simeq \frac{1}{2}\,\frac{1}{1+1/2\tau}$$

<234> 로 주어지고, 전체에 대해 τ는 0.5보다 작다. 이 경우, $\overline{\delta}$<<1이라는 사실은 PSD추정기에서의 불확실성(및 특히, $\widehat{\Phi}_x(\omega)$에서의 불확실성)이 출력의 품질(PSD에러의 항으로)에 커다란 영향을 미친다는 것을 나타낸다. 특히, $\overline{\delta}$<<1의 사용은 입력에서 출력신호로 잡음 대 음성비 개선이 작다는 것을 의미한다.

<235> 발생하는 문제점은, 만일 부록 D의 IPS방법에 대한 가중함수와 유사한 데이터에 독립적인 가중함수($\bar{G}(\omega)$)가 존재하느냐는 것이다. 부록 G에서, 그러한 방법이 유도된다(δIPS로 표시된다).

<236> 부록 G

<237> $\hat{H}_{\delta IPS}(\omega)$의 유도

<238> 이 부록에서, 몇몇 상수 δ(0≤δ≤1)에 대한 $\hat{H}(\omega)=\sqrt{\overline{G}(\omega)}\hat{H}_{\delta IPS}(\omega)$가 제곱된 PSD에러의 기대치를 최소화시키는 데이터에 독립적인 가중함수($\bar{G}(\omega)$)를 찾는다(식 (42)참조). 간단한 계산은,

$$\widetilde{\Phi}_s(\omega) = (G(\omega)-1)\Phi_s(\omega) + G(\omega)(1-\delta)\Phi_v(\omega)$$

$$G(\omega)\delta\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)}\Delta_x(\omega) - \Delta_v(\omega)\right) \qquad (59)$$

<240> 가 된다. 제곱된 PSD에러의 기대치는 다음과 같이 주어진다.

$$E[\widetilde{\Phi}_s(\omega)] = (G(\omega)-)^2\Phi_s^2(\omega) + G^2(\omega)(1-\delta)^2\Phi_v^2(\omega)$$

$$2(G(\omega)-1)\Phi_s(\omega)G(\omega)(1-\delta)\Phi_v(\omega) + G^2(\omega)\delta^2\gamma\Phi_v^2(\omega) \qquad (60)$$

<242> 식 (60)의 우측편은 G(ω)에서 이차방정식이고 그리고 분석적으로 최소화될 수 있다. 결과 $\bar{G}(\omega)$는 다음과 같이 주어진다.

$$\overline{G}(\omega) =$$

$$\frac{\Phi_s^2(\omega) + \Phi_s(\omega)\Phi_v(\omega)(1-\delta)}{\Phi_s^2(\omega) + 2\Phi_s(\omega)\Phi_v(\omega)(1-\delta) + (1-\delta)^2\Phi_v^2(\omega) + \delta^2\gamma\Phi_v^2(\omega)} \qquad (61)$$

$$= \frac{1}{1+\beta\left(\frac{\Phi_v(\omega)}{\Phi_x(\omega)-\Phi v(\omega)}\right)^2}$$

<244> 이때 두 번째 등식에서 β는 다음과 같이 주어진다.

$$\beta = \frac{(1-\delta)^2 + \delta^2\gamma + (1-\delta)\Phi_s(\omega)/\Phi_v(\omega)}{1 + (1-\delta)\Phi_v(\omega)/\Phi_s(\omega)} \qquad (62)$$

<246> δ=1에 대해, 상기 식 (61)~(62)는 IPS방법(식 (45))으로 감소되고 그리고 δ=0에 대해, 표준 PS가 되게 된다. 식 (61)~(62)의 $\Phi_s(\omega)$와 $\Phi_v(\omega)$를 그들의 상응하는 추정된 량 $\hat{\Phi}_x(\omega) - \hat{\Phi}_v(\omega)$ 및 $\hat{\Phi}_v(\omega)$로 각각 대체하면, IPS방법을 고려하면 δIPS로 표시되는 방법이 되게 된다. δIPS방법의 분석은 IPS방법의 분석과 비슷하지만, 상당한 노력과 지루한 계산을 필요로 하기 때문에 이는 생략된다.

<247> 참조문헌

[1] S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, April 1979, pp. 113~120.

[2] J.S. Lim and A.V. Oppenheim,"Enhancement and Bandwidth Compression of Noisy Speech", Proceedings of the IEEE, Vol 67, No 12, December 1979, pp. 1586~1604.

[3]J.D. Gibson, B.Koo and S.D. Gray,"Filtering of Colored Noise for Speech Enhancement and Coding", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol ASSP~39, NO. 8, August 1991, pp. 1732~1742.

[4] J.H.L Hansen and M.A. Clements,"Constrained Iterative Speech Enhancement with Application to Speech Recognition", IEEE Transactions on Signal Processing, Vol. 39, No. 4, April 1991, pp. 795~805.

[5] D.K. Freeman,G. Cosier, C.B. Southcott and I. Boid,"The Voice Activity Detector for the Pan-European Digital Cellular Mobile Telephone Service", 1989 IEEE International Conference Acoustics, Speech and Signal Processing, Glasgow, Scotland, 23~26 March 1986, pp. 369~372.

[6] PCT application WO 89/08910, British Telecommunications PLC.

(57) 청구의 범위

청구항 1

각 프레임이 규정된 수(N)의 음성샘플을 포함하여 각 프레임에 N 자유도를 제공되며, 스펙트럼 감산함수 $\hat{H}(\omega)$는 비음성 프레임의 배경잡음의 전력 스펙트럼 밀도의 추정 $\hat{\Phi}_v(\omega)$와 음성프레임의 전

력 스펙트럼밀도의 추정 $\widehat{\Phi}_x(\omega)$을 기초로 하는, 프래임 기반 디지털 통신시스템의 스펙트럼적 감산 잡음억제방법에 있어서,

자유도의 수를 N보다 작게 감소시키는 파라메틱 모델로 각 음성 프레임을 근사시키는 단계 ; 및

근사적인 파라메트릭 모델을 기초로 한 파라메트릭 전력 스펙트럼 추정방법으로 각 음성프레임의 전력 스펙트럼밀도의 상기 추정 $\widehat{\Phi}_x(\omega)$을 추정하는 단계,

비-파라메트릭 전력 스펙트럼 수정방법으로 비-음성 프레임 각각의 전력 스펙터럼 밀도의 상기 추정 $\widehat{\Phi}_v(\omega)$를 추정하는 단계를 특징으로 하는 스펙트럼적 감산 잡음억제방법.

청구항 2

제1항에 있어서, 상기 근사적인 파라메트릭 모델은 자기회귀(AR)모델인 것을 특징으로 하는 방법.

청구항 3

제2항에 있어서, 상기 자기회귀(AR)모델은 $\sqrt{N}$ 차수인 것을 특징으로 하는 방법.

청구항 4

제3항에 있어서, 상기 자기회귀(AR)모델은 10 차수인 것을 특징으로 하는 방법.

청구항 5

제3항에 있어서, $\hat{G}(\omega)$가 가중함수이고 그리고 $\delta(\omega)$가 간산인자인 식 :

$$\hat{H}(\omega) = \sqrt{\hat{G}(\omega)\left(1 - \delta(\omega)\frac{\widehat{\Phi}_v(\omega)}{\widehat{\Phi}_x(\omega)}\right)}$$

에 따른 스펙트럼 감산함수 $\hat{H}(\omega)$를 특징으로 하는 방법.

청구항 6

제5항에 있어서, $\hat{G}(\omega)=1$인 것을 특징으로 하는 방법.

청구항 7

제5항 또는 제6항에 있어서, $\delta(\omega)$는 상수≤1 인 것을 특징으로 하는 방법.

청구항 8

제3항에 있어서, 식 :

$$\hat{H}(\omega)= 1-\sqrt{\frac{\widehat{\Phi}_v(\omega)}{\widehat{\Phi}_x(\omega)}}$$

에 따른 스펙트럼 감산함수 $\hat{H}(\omega)$를 특징으로 하는 방법.

청구항 9

제3항에 있어서, 식 :

$$\hat{H}(\omega)= \left(1-\frac{\widehat{\Phi}_v(\omega)}{\widehat{\Phi}_x(\omega)}\right)$$

에 따른 스펙트럼 감산함수 $\hat{H}(\omega)$를 특징으로 하는 방법.

청구항 10

제3항에 있어서, 식 :

$$\hat{H}(\omega)= \frac{1}{2}\left(1+\sqrt{\left(1-\frac{\widehat{\Phi}_v(\omega)}{\widehat{\Phi}_x(\omega)}\right)}\right)$$

에 따른 스펙트럼 감산함수 $\hat{H}(\omega)$를 특징으로 하는 방법.

요약

프레임기초 디지탈통신 시스템에서의 스펙트럼 감산 잡음억제방법이 설명된다. 각 프레임은 규정된 수(N)의 음성샘플을 포함하여. 각 프레임에 N자유도를 제공한다. 방법은 비음성 프레임의 배경잡음의

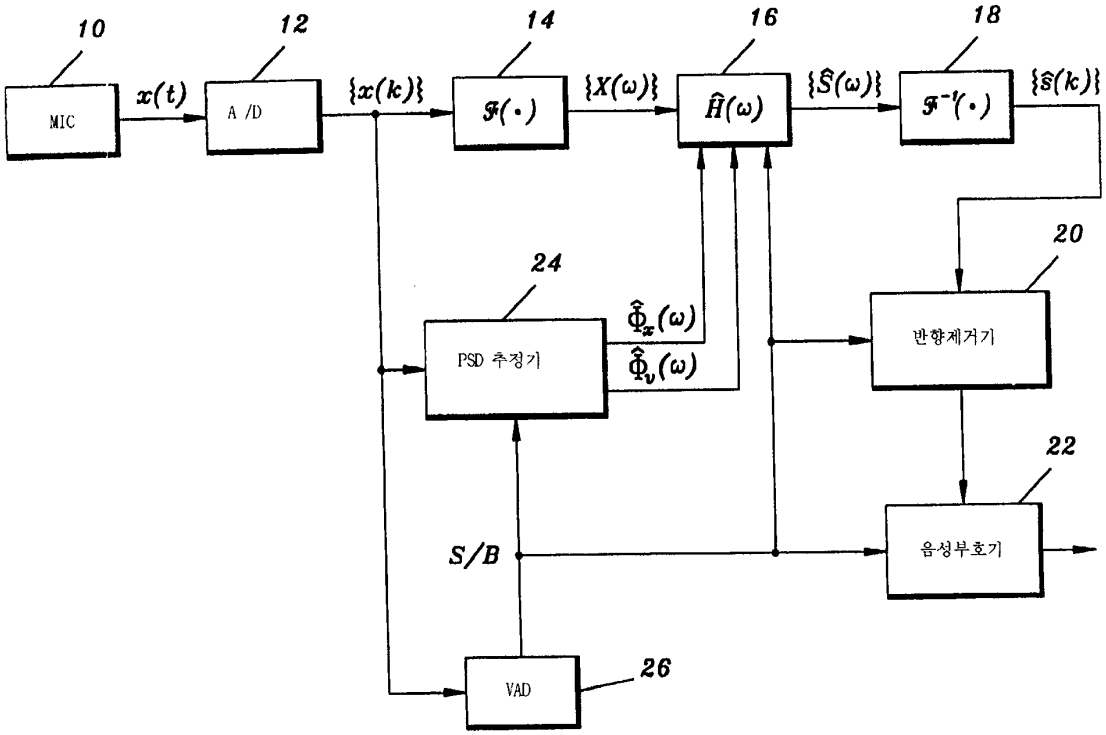
전력 스펙트럼밀도의 추정(140) $\hat{\Phi}_v(\omega)$과 음성 프레임의 전력 스펙트럼밀도의 추정(130) $\hat{\Phi}_x(\omega)$을 기초로 하는 스펙트럼 감산(150)함수 H(ω)로 실행된다. 각 음성 프레임은 자유도의 수를 N보다 작게 감소시키는 파라메터에 관한 모델에 의해 근사된다(120). 각 음성 프레임의 전력 스펙트럼밀도의 추정 $\hat{\Phi}_x(\omega)$은 근사적인 파라메터에 관한 모델로부터 추정된다(130).

대표도

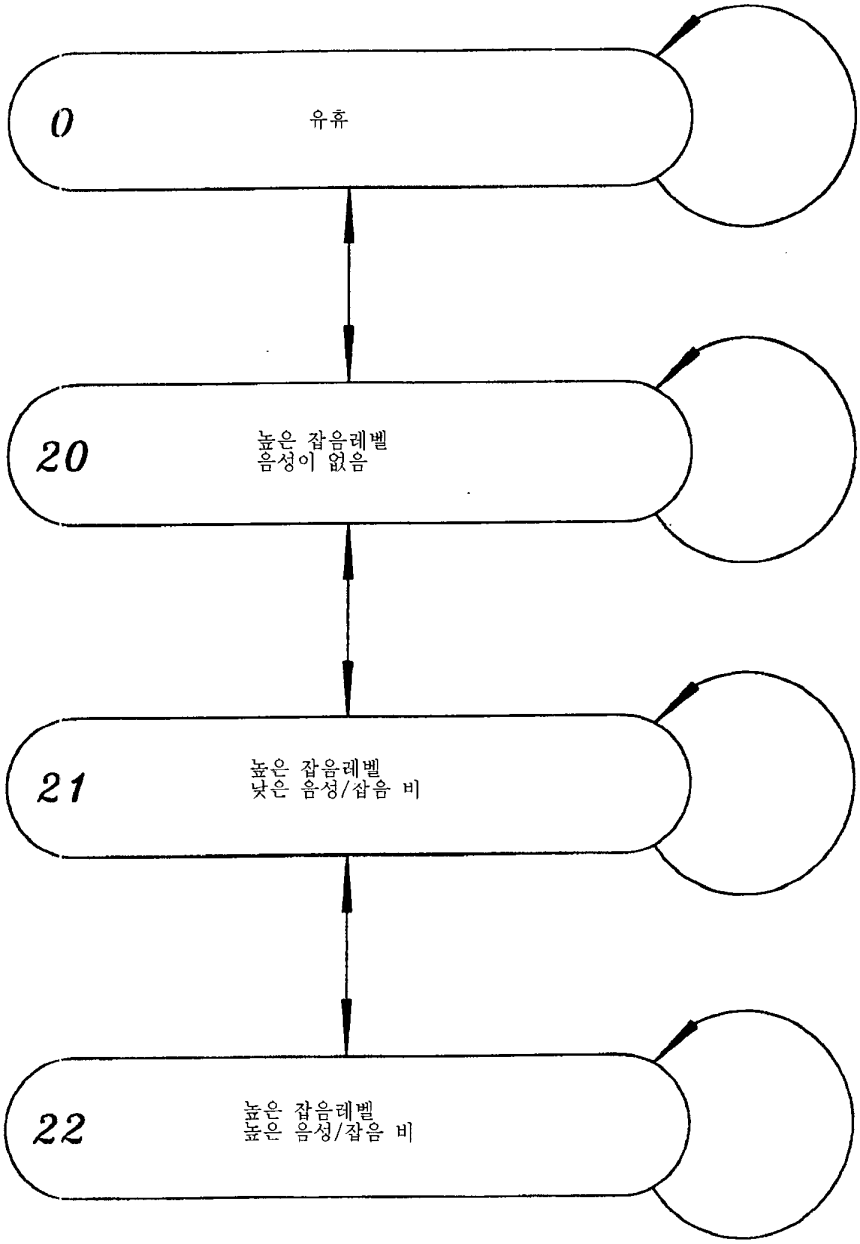
도1

도면

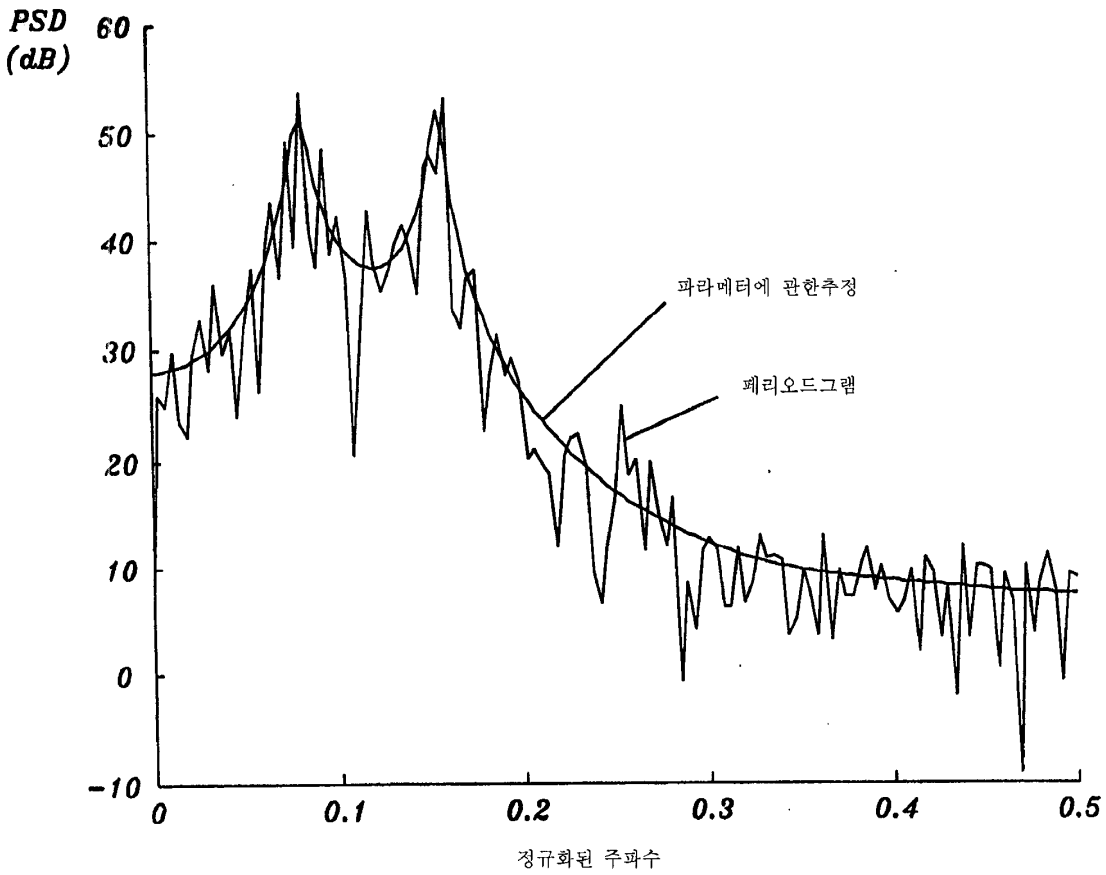
도면1



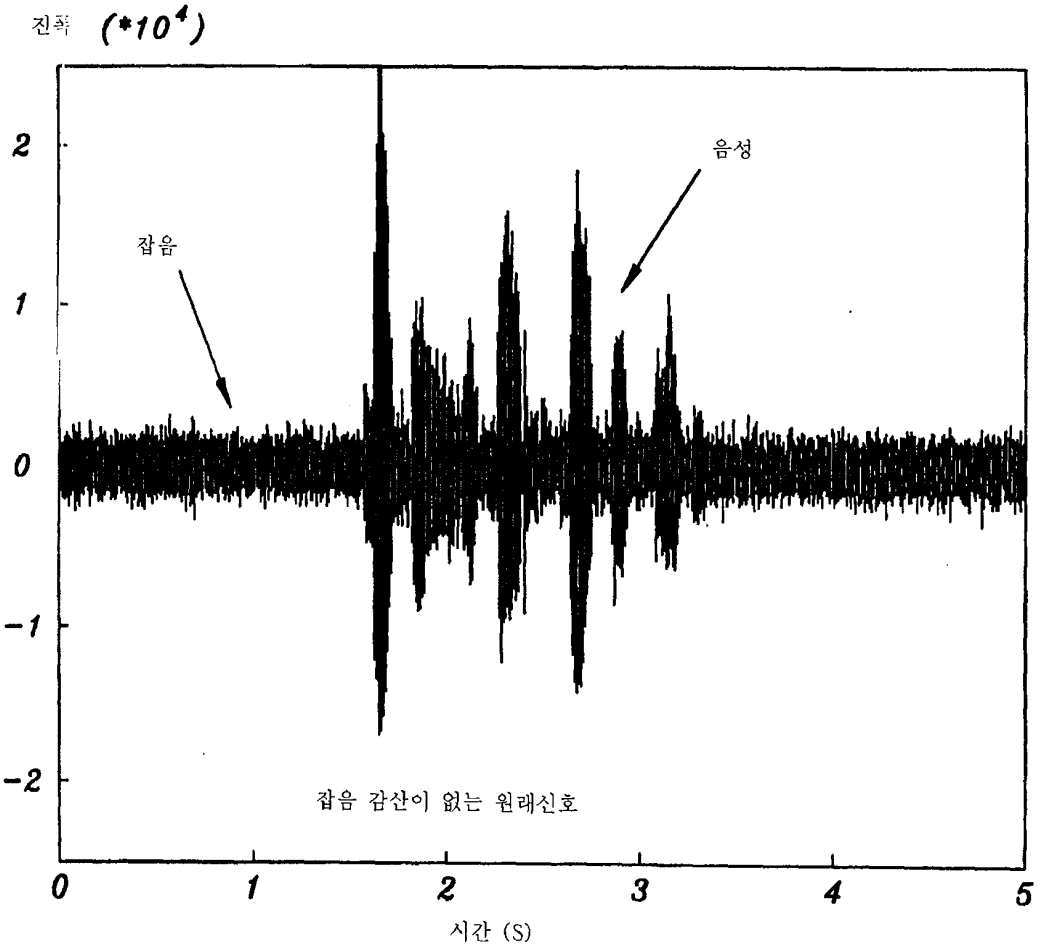
도면2



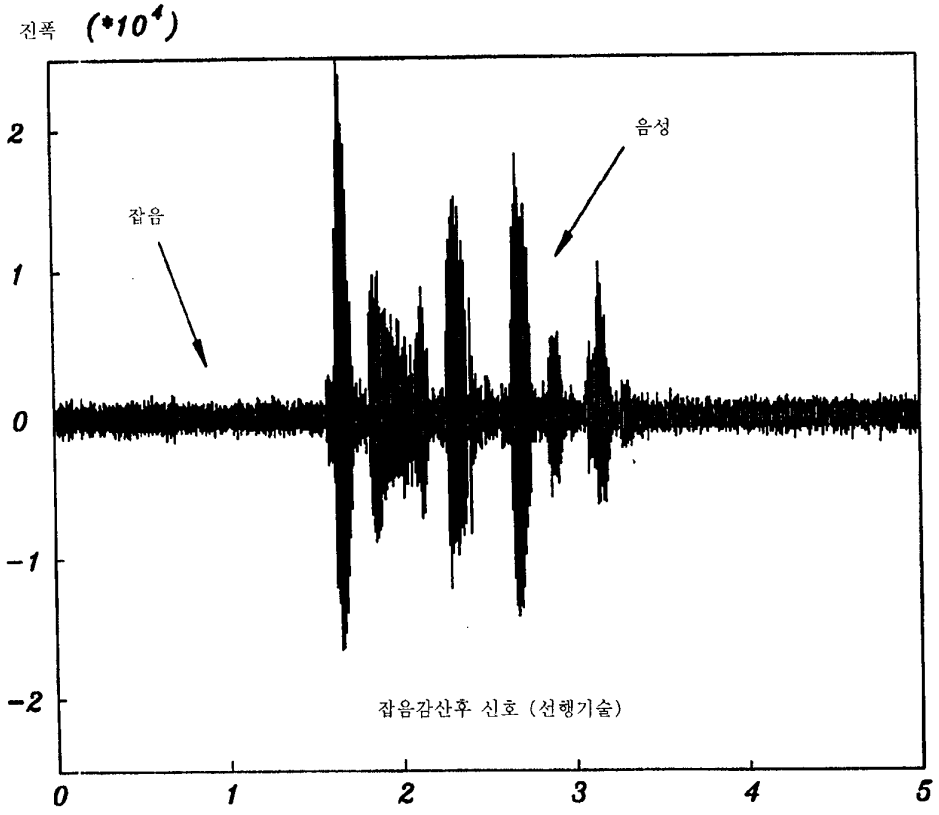
도면3


PSD
(dB)
60
50
40
30
20
10
0
-10
0
0.1
0.2
0.3
0.4
0.5
파라메터에 관한추정
페리오드그램
정규화된 주파수

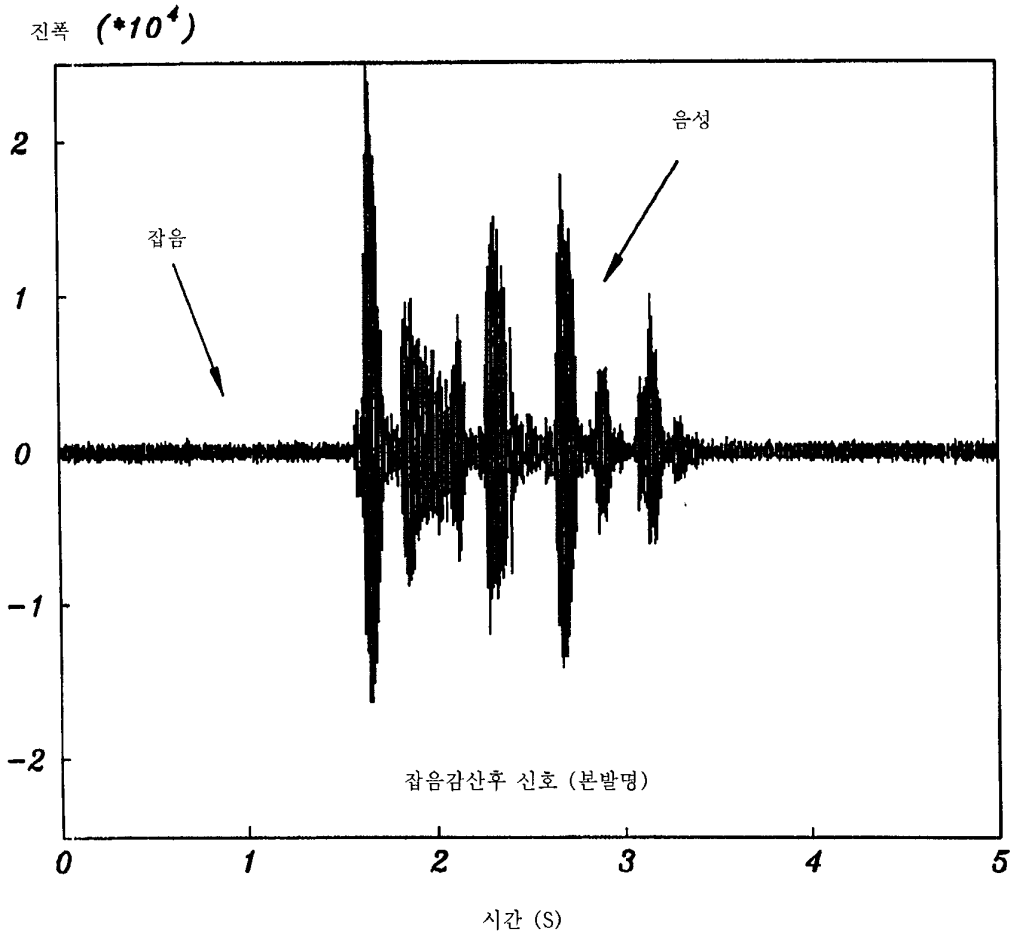
도면4



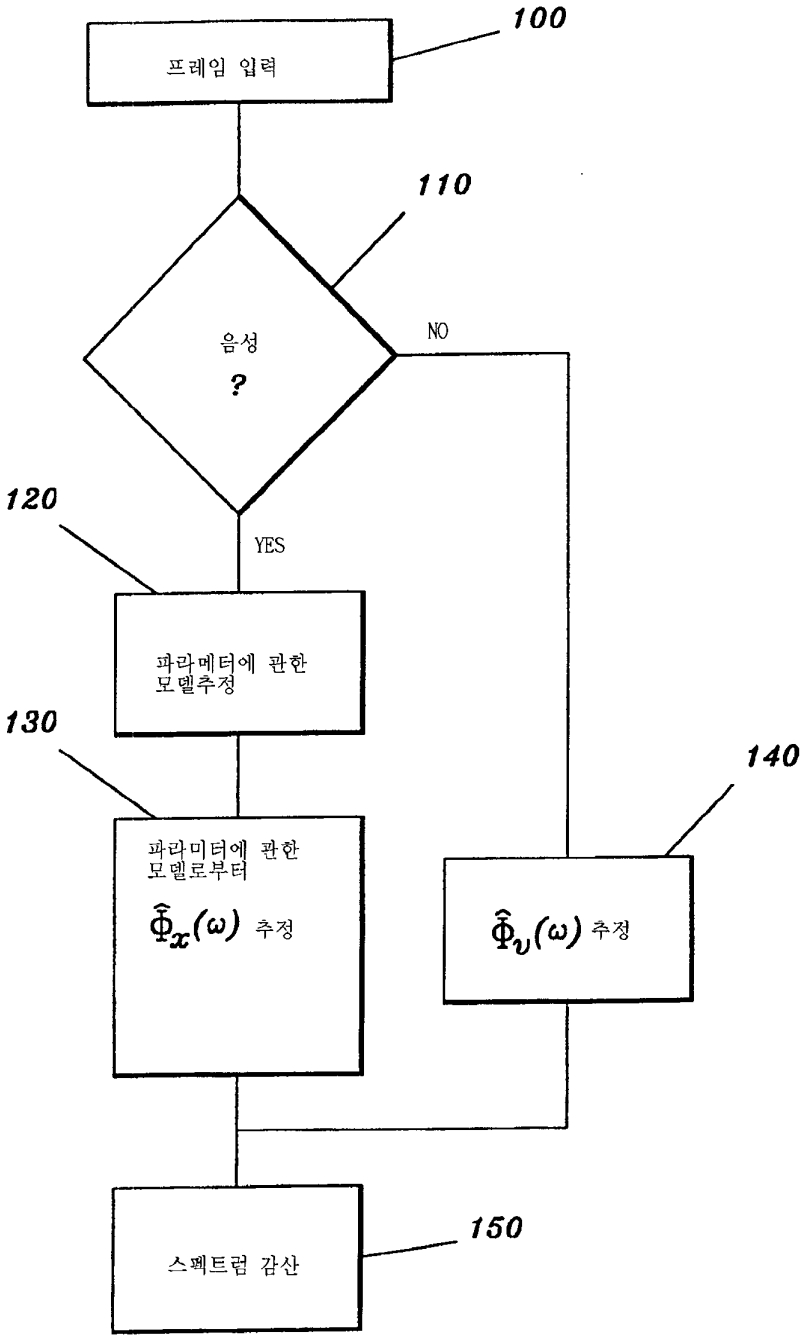
도면5



도면6



도면7


100
프레임 입력
110
음성
?
NO
YES
120
파라메터에 관한
모델추정
130
파라미터에 관한
모델로부터
$\hat{\Phi}_x(\omega)$ 추정
140
$\hat{\Phi}_v(\omega)$ 추정
150
스펙트럼 감산