(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.[6]
G10L 3/00

(45) 공고일자 1999년06월15일
(11) 등록번호 10-0192854
(24) 등록일자 1999년01월30일

(21) 출원번호 10-1991-0701482
(22) 출원일자 1991년10월28일
번역문제출일자 1991년10월28일
(86) 국제출원번호 PCT/US 91/01333
(86) 국제출원일자 1991년02월25일
(81) 지정국 EP 유럽특허 : 오스트리아 벨기에 스위스 리히텐슈타인 독일 덴마크 스페인 프랑스 영국 이탈리아 룩셈부르크 네덜란드 스웨덴 그리스
국내특허 : 오스트레일리아 캐나다 일본 대한민국

(65) 공개번호 특1992-0701942
(43) 공개일자 1992년08월12일
(87) 국제공개번호 WO 91/13430
(87) 국제공개일자 1991년09월05일

(30) 우선권주장 486,462 1990년02월28일 미국(US)
(73) 특허권자 에스알아이 인터내셔널 도널드 엘. 앤드루소
미합중국 캘리포오니아 94025 멘로 파아크 레벤스우드 애비뉴 333
(72) 발명자 아도람 에렐
이스라엘 라마트 아비브 69978 텔아비브 대학 엘렉트로닉 시스템즈 공학부
미첼 와인트라우브
미합중국 캘리포니아 94536 프리몬트 코로나도 드라이브 36360
(74) 대리인 황주명

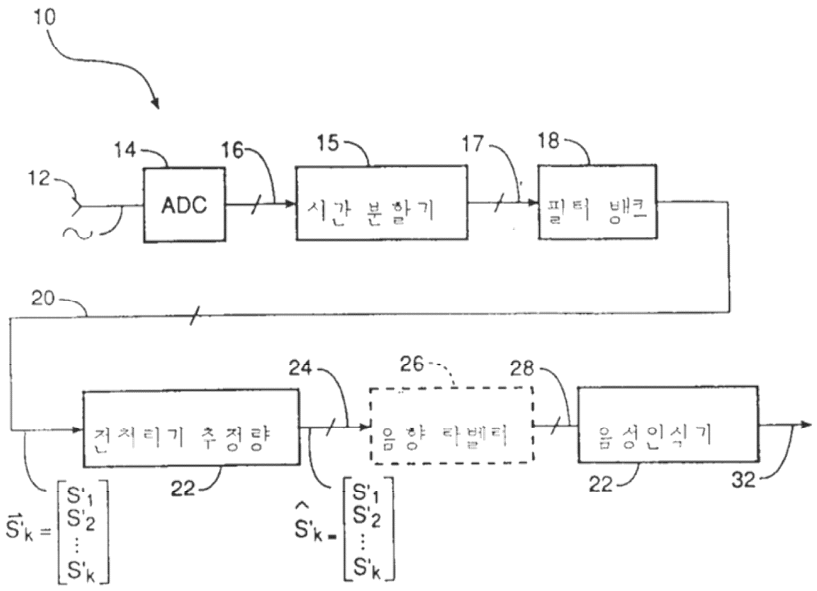
심사관 : 유병철

(54) 음성인식의 잡음강도를 개선하기 위한 스텍트랄 추정 방법

요약

음성인식기에 사용하기 위하여 추정의 오차의 유사성을 최소화하기 위한 노이즈 음성을 전처리하는데 사용되는 방법이 공지되었다. 이하의, 혼합 모델과 마르코프 모델을을 사용하는 최소-평균-로그-스펙트랄-거리(MMLSD) 추정이라 불리우는 계산가능한 기술에는 단일의 시간 프레임에 상당하는 노이즈 존재하의 음성의 각 벡터를 계산하는 단계들, 깨끗한 음성의 추정이 포함되고, 여기서 추정 방법의 가정들은 깨끗한 음성의 확률분포가 서로다른 주파수 채널들이 각 클래스내에서는 상호관련이 없다고 가정하여 서로다른 음성 클래스를 각기 나타내는 콤포넌트들의 혼합에 의해 모델화 될 수 있다는 것과, 서로다른 주파수 채널들에서 노이즈는 상호 무관하다는 것이다. 본 발명의 다른 실시예에서, 그 방법에는 일련의 시간 프레임들에 상당하는 노이즈 존재하의 음성의 일련의 벡터들을 계산하는 단계와 깨끗한 음성의 추정 단계를 포함하고, 여기서, 추정 방법의 기본적 가정들은 깨끗한 음성의 확률분포가 서로다른 주파수 채널들은 마르코프 프로세스의 각 상태내에서 상호무관하다고 가정하는 마르코프 프로세스에 의해 모델화 될 수 있다는 것과 서로다른 주파수 채널들에서 노이즈는 상호 무관하다는 것이다.

대표도



명세서

[발명의 명칭]

음성인식의 잡음강도를 개선하기 위한 스펙트랄 추정방법

[발명의 상세한 설명]

[저작권 경고]

본 특허명세서 내용의 일부는 저작권으로 보호되고 있는 내용을 포함하고 있다. 저작권 소유자는 본 특허서류 또는 특허명세서가 특허 상표국 파일 또는 기록에 나타나 있는한 어느 누군가가 팩시밀리 재생하는 것을 막지 않으나 그외의 경우는 무엇이든지 저작권에 저촉된다.

[본 발명의 배경]

본 발명은 국립과학재단(National Science Foundation)(IRI 8720403)에서 일부지원하고 캘리포니아, 멘로파크의 에스알아이 인터내셔널(SRI International)에서 일부 지원되었다.

본 발명은 잡음 존재하에서의 음성인식에 관한 것으로, 특히 음성인식 시스템과 관련하여 사용하기 위한 음성전처리(preprocessing)를 위한 방법에 관한 것이다.

음성인식 시스템들은 잡음이 없는 훈련조건(training condition)과 잡음이 존재하는 동작조건간의 차이에 매우 민감하다. 특히, 음성인식 시스템은 무잡음 상태하의 특정 음성 패턴을 인식하도록 훈련되고나서 고음질 음성을 훈련할 수 있게 되어 있다. 하지만 그러한 시스템은 잡음 상태하에서는 현저하게 (성능이) 저하된다.

이러한 문제점을 처리하는 방법이 몇가지 있는데 그중 하나는 통계측정기를 음성인식기에 제공하여 음향전처리를 보충하는 방법이다. 여기 사용된 통계측정기는 음성정보를 명확히 할 수 있는 입력치 또는 입력신호를 음성인식기에 공급하게 되어 있다.

음성인식을 위한 통계측정기의 설계의 과제는 음성인식기와 매치될 최적화 표준(optimality criterion)을 한정하는 것과 이 최적화 표준을 기본으로 하는 통계측정기를 연산하기 위한 알고리즘을 추측하는 것이다. 최적화 표준을 한정하는 것은 청취자를 위한 음성 확대(speech enhancement)에 대한 것보다 음성인식에 대한 것이 더욱 용이하다. 왜냐하면 신호처리기술은 전자에 대해서는 알려져 있으나 후자에 대해서는 알려져 있지 않기 때문이다. 거리 측정 규준(distance metric)을 기본으로 하는 인식 시스템에 대해서는 그것이 템플레이트 매칭(template matching)이든지 또는 벡터 양자화(vector quantization)이든지 최적화 표준이 거리 측정 규준에 의해 측정된 바와 같이 평균 왜곡(average distortion)을 최소화하는 것을 가정하는 것이 타당하다. 최적화 표준을 성취하는 것을 종종 계산적으로 실행 불가능하다.

이산 푸리에 변형(DFT:Discrete Fourier Transform), 즉 필터 뱅크 베이스 시스템(filter-bank based system)으로서 전형적으로 사용되는 거리 측정 방법은 필터 베이스 시스템의 출력 에너지의 로그(logarithm)의 코사인 변형(cosine transform)상의 유클리드 거리에 비중을 두고, 종종 리프터드 켑스트랄 거리(liftered cepstral distance)로 언급된다(필터 뱅크 시스템의 켑스트럼은 필터 에너지의 변형으로 정의된다). 이 거리 측정 방법을 사용한 추정 표준획득(방법)은 추가의 노이즈로 인하여 계산적으로 매우 어렵다. 필터-뱅크 시스템에 응용되어 왔던 공지된 추정 알고리즘은 최소제곱 평균오차(MMSE:minimum mean square error) 알고리즘 및 스펙트럼 감법 알고리즘(spectral subtraction algorithm)이고 이것들은 이산 푸리에 변형(DTE) 계수 또는 필터-뱅크 출력 에너지에 각기 적용된다(이하에서 논의한 포터(porter)등 및 반 콤퍼놀 1 2(Van Compernolle 1 2) 참조). 다차원 켑스트랄 거리 최적화 표준 및 단일 주파수 채널 최소제곱 평균오차 거리표준은 켑스트랄 거리가 특징 벡터(feature vector)의 결합 추정(joint estimation)을 포함하므로써 MMSE 거리는 스칼라량의 독립 추정을 포함하게 하는 것이다. 다른 주파수들에서 음성 스펙트랄 에너지들은 사실상 상호 관련이 있기 때문에 각 주파수 채널들의 독립 추정치의 사용은 서브옵티말 추정(suboptimal estimation)을 초래한다.

이 기술은 통계학과 마르코프 과정(Markov process)과 기본적인 관련이 있음을 추정할 수 있고 마찬가지로 히든 마르코프 모델(Hidden Markov Model)을 사용한 음성인식 시스템에서의 공지 기술과 관련이 있다. 다음의 특허 및 공고된 공지기술을 참고할 수 있다. 이것들은 본 발명과 관련하여 발명자들의 주의를 끌어온 것들이다. 이 참고문헌들은 모두(본 발명과) 관련된 문헌들은 아니라는 생각이 든다.

발명자	미국특허번호	등록일
Bahl외	4,817,156	03/28/89
Levinson 외	4,587,670	05/06/86
Juang 외	4,783,804	11/08/88
Bahl 외	4,741,036	04/26/88

발명자	외국특허번호	공고일
Sedgwick	EP 240,330	10/07/87

논문

Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," "Proc. IEEE, Vol. 77, No. 2, 1989년 2월.

Nadas 외., "Speech Recognition Using noise-adaptive prototypes," IEEE Trans. on ASSP, Vol. 37, No. 10, 1989년 10월.

Stern 외., "Acoustical pre-processor for robust speech recognition," Proc. DARPA Speech and Natural Language Workshop, Session, 1989년 10월.

Ephraim 외., "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Estimator," IEEE Trans. ASSP, Vol. 32, pp. 1109-1112(1984년 12월).

Ephraim 외., "Speech Enhancement Using a Miniumu Mean-Square Error Log-Spectral Amplitude Estimator," IEEE Trans. ASSP, Vo. 33, pp. 443-447(1985년 4월).

Porter 외., "Optimal Estimators for Spectral Restoration of Noisy Speech," Proc. ICASSP, Vol. 2, pp. 18A2.1-2.4 (1984).

Van Compernolle, "Noise Adaptation in a Hidden Markov Model Speech Recognition System," Computer Speech and Language, Vol. 3, pp. 151-167, 1989.

Van Compernolle, "Spectral Estimation Using a Log-Distance Error Criterion Applied to Speech Recognition," Proc. ICASSP, Vol. 1, pp. 258-261(1989).

Gray, "Vector Quantizatiopn," The ASSP Magazine, Vol. 1, No. 2, pp. 3-29(1984년 4월).

래비너(Rabiner) 논문은 본 발명의 이해에 기초가 되는 음성인식에 히든(hidden) 마르코프 모델 응용에 관하여 개관(survey)하는 논문이다. 이 논문은 노이즈 문제를 언급하지 않았다. 이 논문은 여기에 참고 문헌으로 채택되었다.

나머지 특허들과 공고들에는 일반적인 분야에서의 다른 연구들이 기술되었다.

Nadas의 논문은 노이즈 문제를 언급하고 있다. 그러나 전처리 방법은 아니다.

Stern의 논문은 스펙트럼감법(spectral subtraction)을 기본으로 하는 전처리기를 사용하여 노이즈 문제를 해결하는 방법을 기술하고 있다.

Van Compernolle의 논문(1)은 스펙트럼감법 타입의 전처리를 언급하고 있다.

Van Compernolle의 논문(2)은 신호채널의 최소제곱 평균오차 독립 추정을 사용한 전처리기를 기술하고 있다.

Porter 및 Ephraim의 논문들은 DFT진폭의 로그를 포함하는 디지탈 푸리에 변형(DFT) 계수들의 여러기능들의 최소제곱 평균오차를 사용하는 음성인식 기술을 사용하는 전처리기를 기술하고 있다. 이 기술들은 단일 DFT 계수만을 다루고 있다.

Sedgwick의 논문은 음성인식 시스템을 위한 노이즈 보상기에 관한 국립자원개발회사(National Resource Development Corporation)의 연구를 기술한 것으로 주파수 스펙트럼 영역의 레벨에 상당하는 입력신호들이 파생되고 노이즈 입력 셀들을 위한 (유클리드) 거리는 인식과 훈련중의 노이즈 레벨을 이용하여 결정된다. 인식과 훈련처리에서 마이크로폰에 도착하는 신호들은 디지탈 처리되고, 주파수 채널내에서 분리되어지도록 필터 뱅크(fklter bank)를 통과한다. 훈련처리에서, 노이즈 추정량과 마스커(masker)는 인식되어지는 워드들의 마르코프 모델들을 부분적으로 한정하는 각 채널에 대하여 확률밀도함수(PDFs:probability density functions)들을 준비하고 저장하기 위해 인식기와 함께 사용된다. PDFs는 입력신호로부터 노이즈 레벨들 이상인 때에만 추론되어진다. 그러나 추론은 각 PDF전체가 나타나도록 한다. 인식에서, 인식의 기본이 되는 거리(distance) 측정은 각 채널에 대해서 추론된다. 만일 한 채널에서의 신호가 노이즈 레벨 이상이라면, 거리는 PDF의 네가티브 로그(negative logarithm)로부터 인식기에 의해 결정되어진다. 만일 채널 신호가 노이즈 레벨 이하이면, 거리는 노이즈 레벨에 대하여 PDF의 축적거리(cumulative distance)의 네가티브 로그로부터 결정되어진다. 이 공고공보는 노이즈 보상을 가진 인식 시스템을 기술한다. 그러나 노이즈 문제를 언급하는 전처리를 언급하지 않았다.

Bahl의 '156특허는 다음 화자(subsequent speaker)에 대한 마르코프 모델 음성인식기의 통계를 훈련하기 위한 장치 및 방법에 관한 IBM연구를 기술하고 있다. 거기에서, 그 발명은 희박한 훈련 데이타가 있는 다음 화자에 상당하는 마르코프 모델에서의 전이(transition)에 레이블 출력 확률(label output probability)을 결정한다. 이 특허는 노이즈 문제를 언급하지 않았다.

Levinson의 특허는 다수의 저장되고 한정된 히든 마르코프 모델 참조 템플레이트(reference templates)와 다수의 참조 패턴들의 전술한 음향 특징의 일련의 대표적인 저장된 신호들을 포함하는 음성인식기에 관한 AT T벨 연구소의 연구이다. 이 특허는 노이즈 문제를 언급하지 않았다.

Juang의 특허는 히든 마르코프 모델 음성인식 설비에 관한 AT T벨 연구소 추가의 연구이다. 마르코프 모델 음성 패턴 템플레이트들은 음향 특징 신호들의 프레임 시퀀스(frame sequence)들을 발생하기 위한 확인된 음성패턴을 분석함으로써 형성된다.

Bahl의 '036특허는 음성인식 시스템의 IBM연구로서 유사하게 발음되는 단어들의 판별은 수화기의 참고 단어 시퀀스를 나타내는 마르코프 모델을 위하여 저장된 확률 벡터 데이타를 조작함으로써 개선되어진다. 조작 벡터(weigting vector)는 비터비 배열(Viterbi alignment)과 정확한 인식 다변량 분포와 부정확한 인식 다변량 분포 사이의 차이값을 최대로 하는 다변량 해석(multivariate analysis)을 사용하여 유사발음을 비교하므로써 각 참고단어에 대하여 파생된다. 이 특허는 노이즈 문제를 언급하지 않았다.

계산적으로 가능한 방법으로 음성인식 시스템에서의 노이즈 음성의 처리에 있어서 종래기술은 그 어떤것도 켑스트랄 거리 최적화 표준(cepstral distance optimality criterion)에 접근하는 방법을 제시하지 못한다.

[본 발명의 요약]

본 발명에 따르면, 음성인식기에 사용키 위한 추정 오차 발생 가능성을 최소화 하기 위한 노이즈 음성을 전처리하는데 사용하는 계산 가능한 방법을 제공한다. 여기서, 계산가능한 기술은 최소-평균-로그-스펙트랄-거리(Minimum-Mean-Log-Spectral-Distance)(MMLSD)라 불리우고, 단일 시간 프레임, 깨끗한 음성 추정에 상당하는 노이즈의 존재하의 각 음성 벡터에 대해서 계산하는 단계를 포함한다. 여기서 추정량의 방법에 대한 가정은 깨끗한 음성의 확률분포가 서로 다른 음성 클래스를 각기 나타내는 콤포넌트들을 혼합함으로써 모델화되어 질 수 있다는 것이다. 서로 다른 음성 클래스는 서로 다른 주파수 채널들은 각 클래스내에서 상호 상관관계가 없고, 서로 다른 주파수 채널들에서 그 노이즈는 아무관련이 없다는 것을 가정한 것이다(식 11 및 제2도).

본 발명의 다른 실시예에서는, 일련의 시간 프레임, 깨끗한 음성 추정에 상당하는 노이즈 존재하의 일련의 각 음성 벡터를 계산하는 단계를 포함한다. 여기서 추정치의 방법의 기본적인 가정은 깨끗한 음성의 확률분포가 마르코프 프로세스에 의해 모델화될 수 있다는 것인데 이 마르코프 프로세스는 서로 다른 주파수 채널은 마르코프 프로세스의 각 상태내에서 상호 관련이 없으며 서로 다른 주파수 채널에서의 노이즈는 상호 관련이 없다는 것을 가정한 것이다(식 21 및 제3도).

본 발명은 첨부한 도면을 참조하고, 이하 기술한 상세한 설명을 참고하면 이해에 더욱 도움이 될 것이다.

[도면의 간단한 설명]

제1도는 본 발명에 따른 방법을 채용한 전처리기를 채용한 음성인식 시스템의 블럭도이다.

제2도는 본 발명에 따른 제1방법에 따라 혼합 모델을 사용하는 단일 프레임 MMLSD 추정기의 블럭도이다.

제3도는 본 발명과 관련하여 사용되는 가우스 모델(Gaussians model)의 혼합의 계산을 도시한 플로우 챠

트이다.

제4도는 본 발명의 제2방법에 따라 마르코프 모델을 사용하는 일련의 시간 프레임에 대한 MMLSD 추정치의 블럭도이다.

[특정한 실시예들의 설명]

제1도는 본 발명에 따른 방법을 채용한 전처리 추정치(22)를 채택한 음성인식 시스템(10)의 블럭도이다. 도시한 음성인식 시스템(10)은 히든 마르코프 프로세스 인식기를 채용한 필터-뱅크-기본 시스템이다. 음성인식 시스템(10)은 입력(12)에서 노이즈가 존재하는 음성을 나타내는 아나로그 시간-정의역(도메인:domain) 신호를 받는다. 입력(12)은 디지탈 컨버터(ADC)(14)와 필터 뱅크(18)에 아나로그를 포함하는 특징 추출기(feature extractor)에 가해진다. ADC(14)는 아나로그 신호를 타임 세그멘터(15)에 디지탈 신호선(16)들상에 가해지는 디지탈화된 음성으로 변환한다. 타임 세그멘터(15)는 다음 처리를 위하여 시간 프레임들을 디지탈화된 신호로 쪼갠다. 선(17)들 상의 출력은 필터 뱅크(18)로 들어간다. 필터 뱅크(18)는 음성+노이즈(노이즈 음성) 콤포넌트를 각 시간 프레임에 대하여 콤포넌트 S'_k를 가지고 필터 로그 에너지들(또는 더욱 일반적으로는 스펙트랄 로그 에너지들) S'의 벡터로 분류한다. 각 콤포넌트는 음성정보에 대한 하나의 필터 채널을 나타낸다. 벡터 S'는 선(20)을 경유하여 전처리기(22)에 입력되고 전처리기(22)는 깨끗한 음성의 추정치로서 기능한다. 전처리기(22)의 출력은 깨끗한 음성의 추정형태, 벡터 $\hat{S}$ 이다.

벡터 $\hat{S}$ 는 선(24)상에서 음향 라벨러(26)에 선택적으로 입력되거나 또는 음성인식기(30)에 직접적으로 입력된다.

전처리기(22)는 다음 모든 처리가 마치 노이즈가 없는 것처럼 입력신호를 취급하도록 기능한다. 3가지 가능한 음성인식기는 본 발명을 따른 전처리기(22)와 함께 사용될 수 있다. 음성인식기(30)는 음향 라벨링을 위해 거리 측정을 사용하는 음향 라벨러와 함께 이산밀도 히든 마르코프 모델(HMM) 인식기로 될 수 있다. 다른 한편으로, 음성인식기(30)는 거리 측정을 사용치 않고 음향 라벨링을 위해 통계적 알고리즘을 사용하는 연속 밀도 HMM 인식기를 사용할 수도 있다. 또 다른 한편으로, 음성인식기(30)는 동적 시간 뒤틀림(dynamic time warping)과 같은 템플레이트 매칭을 사용하는데 동적시간 뒤틀림은 템플레이트 매칭을 위해 거리 측정을 사용한다. 선(32)들 상에 음성인식기(30)의 출력은 인식된 음성이다.

제2도에는 본 발명에 따른 전처리기(22)의 제1실시예를 도시하였다. 제2도의 전처리기(22)는 다음의 사항을 근거로 하여 각 시간 프레임에 대하여 노이즈 음성의 벡터 S'로부터 깨끗한 음성의 벡터 $\hat{S}$ 의 추정치를 계산한다.

1) 깨끗한 음성의 확률분포는 혼합 모델로서 콤포넌트의 혼합에 의해 모델화될 수 있고, 각 콤포넌트는 서로 다른 채널들은 각 클래스내에서 상호 연관이 없다는 가정하에 서로 다른 음성 클래스를 나타낸다는 가정. 이것은 다음을 근거로 한다.

2) 다른 주파수 채널들에서 노이즈는 상호 관련이 없다는 가정에 근거하는 깨끗한 음성의 벡터 주변에 노이즈 음성 벡터의 조건적 확률기능.

추정량은 벡터 S의 최소제곱 평균오차(MMSE) 추정이고, 여기서 제곱 평균오차는 유클리드 벡터이다. K필터 로그-에너지의 벡터 S상의 최소 유클리드 거리는 다음 벡터 추정치를 얻는다(여기서 볼드체는 벡터를 의미한다):

$$\hat{\mathbf{S}} = \int \mathbf{S}\, P(\mathbf{S}|\mathbf{S}')\, d\mathbf{S} \qquad (1)$$

베이즈 룰(Bayes' Rule)을 사용하면, 추정량은 다음 형태이다:

$$\hat{\mathbf{S}} = \frac{1}{P(\mathbf{S}')} \int \mathbf{S}\, P(\mathbf{S}'|\mathbf{S})\, P(\mathbf{S})\, d\mathbf{S} \qquad (2)$$

여기서 벡터 S'에 대한 확률은 다음과 같이 주어진다:

$$P(\mathbf{S}') = \int P(\mathbf{S}'|\mathbf{S})\, P(\mathbf{S})\, d\mathbf{S} \qquad (3)$$

이 추정량은 단일 채널형태 추정량보다 훨씬 더 복잡하다. 왜냐하면, 그것은 K-차원 확률분포 즉 25주파수 채널에 대한 25차원의 적분을 필요로 하기 때문이다. 본 발명에 따르면, 확률 S와 확률 S' 어라운드 S에 대한 근사 모델이 계산에 사용될 수 있다. 왜냐하면 노이즈가 추가되고 벡터 S는 필터-뱅크 로그-에너지의 벡터라고 가정하기 때문이다.

먼저, 벡터 S' 어라운드 벡터 S의 조건 확률 또는 P(S'|S)는 한계 확률의 산출로서 간단히 모델화 될 수 있기 때문이다. 또는

$$P(S'|S) = \prod_{k=1}^{K} P(S'_k|S_k) \qquad (4)$$

그것은 가우스 노이즈(Gaussian moise)는 주파수 정의역에서 상호 관련이 없다는 것을 가정하고, 주어진 노이즈 필터 S'_k의 에너지 값은 깨끗한 에너지 S_k 및 그 주파수에서의 노이즈의 레벨(즉 그 주파수에서의 필터의 통과대역내에서)에 관해서만 의존하기 때문이다. 이 모델은 단지 추정치일 뿐이지만, 필터의 통과대역들은 중복된다.

조건확률 $P(S'_k|S_k)$는 다음과 같이 모델화 될 수 있다.

추정되어지는 음성과 연관된 노이즈는 정상 ARMA(자동복귀 이동 평균; autoregressive moving average), 확률론적 처리(stochastic process)(즉, 선형여과로 채색된 백색 잡음(white noise)으로 보이는)에 의해 나타낼 수 있다. 그러므로 각 시간 프레임에서 노이즈에 대한 이산 푸리에 변형의 계수는 복잡한 임의의 가우스 변수와 상호 관련이 없다. 필터 출력 에너지는 M계수의 합에 의해 접근될 수 있다고 추가로 가정한다. 결국은 노이즈 스펙트랄 파워는 합의 범위내에서 일정하다고 가정한다. 노이즈만으로는(즉 음성이 없는) 그 합이 가우스인 제로평균 및 분산의 임의의 2M 변수들의 합을 넘는다:

$$\sigma^2 = (N_k)/(2M) \qquad (5)$$

여기서 N_k는 노이즈 필터 에너지의 기대값이다. 이런 조건들하에서, 분산-정규화 필터 에너지(variance mormalized filter energy)는 2M 등급 자유를 가진 카이(χ)제곱 확률분포를 따를 것이다.

음성과 노이즈의 존재하에서, 필터 에너지는 다음과 같이 주어진다.

$$E'_k = \sum_i |DFT_s(i) + DFT_n(i)|^2 \qquad (6)$$

여기서 DFTs는 음성계수이고 DFTn은 노이즈 계수이다.

임의의 변수는 식(5)를 식(4)로 나눔으로써 구해진다. 또한 $E'_k|\sigma^2$은 서로 다른 확률분포 즉 2M 등급 자유를 가진 비집중 카이제곱 확률분포 및 비집중변수(χ)에 따를 것이다.

그리하여 필터 에너지의 조건확률은 다음과 같이 주어진다:

$$P\left[\frac{E'_k}{N_k}\middle|\frac{E_k}{N_k}\right] \quad 2M.\chi^2 \quad N.C. \quad \left[\frac{2ME'_k}{N_k},\ 2M, \lambda\right] \qquad (7a)$$

여기서 $$\lambda = \sum_i \frac{|DFT_s(i)|^2}{\sigma^2} = \frac{2ME_k}{N_k} \qquad (7b)$$

정규화된 로그-에너지 변수는 다음과 같다.

$$S_k = 10 \log_{10}\left[\frac{E_k}{N_k}\right] ; \quad S'_k = 10 \log_{10}\left[\frac{E'_k}{N_k}\right] \qquad (8)$$

결과적으로 S'_k에 대한 조건확률은 다음과 같다.

$$P = (S'_k \mid S_k) = 0.23 \quad \frac{E'_k}{N_k} \quad P\left[\frac{E'_k}{N_k}\middle|\frac{E_k}{N_k}\right] \qquad (9)$$

S'_k값은 K^{th}필터에 대한 노이즈 존재하에 관찰된 필터 로그-에너지 값이고, S_k는 깨끗한 음성의 필터 로그-에너지값이다.

식(9)로 주어진 S_k 주위의 S'_k의 조건확률은 노이즈 존재하의 음성요소(S) 주위에 불확실성을 설명하기 위하여 확률분포내의 필요한 퍼지(fuzziness)를 제공한다.

모델 가정들로부터의 편차에 기인하여 실제 상황에서는 2M 등급의 자유이하가 될 것이다. 특히, 필터들은 전형적으로 (박스-카(box-car) 또는 장방형 주파수 윈도우들보다는 오히려) 사다리꼴 주파수 윈도우들로 한정되고 이산 푸리에 변형의 해밍 윈도우(hamming window)는 노이즈 DFT 계수들 사이의 상호관계들을 도입한다. 더우기, 매우 넓은 필터들의 노이즈 스펙트럼은 단일 필터의 영역내에서 납작해지지 않는다.

둘째, 깨끗한 음성 벡터 P(S)의 확률분포(PD)를 위한 모델이 요구되어진다. 깨끗한 음성의 확률분포는 한계확률의 산물로서 주파수 영역(domain)에 나타나지 않는다. 오히려, 이 확률분포는 다음식의 혼합 모델에 의해 모델화 될 수 있다.

$$P(S) = \sum_{n=1}^{N} C_n P_n(S), \quad P_n(S) = \prod_{k=1}^{K} P_n(S_k) \qquad (10)$$

여기서 C는 상수이고 N은 혼합 콤포넌트 또는 클래스의 수이다.

이 모델은 음향공간이 서로 다른 주파수 채널들 사이의 상호관계가 대체로 음향공간내에서보다 훨씬 더 작은 클래스들로 나뉘어질 수 있다는 생각에 근거를 두고 있다. 클래스들은 음향공간의 상호 배제 혹은 중복 영역을 나타낼 수 있다. 이하에 설명한 바와 같이, 깨끗한 음성에 대한 추정량은 다음과 같이 주어진다.

$$\hat{S}_k = \sum_{n=1}^{N} S_k|n \cdot P(n|S') \qquad (11)$$

여기서 인덱스 n은 클래스이고, 첫째항(n 어라운드 S_k의 추정)은 다음과 같은 계산 가능하게 주어진 n^{th} 클래스-조건의 MMSE 추정량이다.

$$\hat{S}_k|n = \frac{1}{P(S'_k|n)} \int S_k \, P(S'_k|S_k) \, P_n(S_k) \, dS_k \qquad (12)$$

여기서

$$P(S'_k|n) = \int P(S'_k|S_k) \, P_n(S_k) \, dS_k \qquad (13)$$

그리고 제2항(벡터 S'로 주어진 n의 다음 확률(posteriori probability) 즉, n^{th} 클래스에 속한 깨끗한 음성벡터)은 다음과 같다:

$$P(n|S') = \frac{C_n \, P(S'|n)}{\sum_{n=1}^{N} C_n \, P(S'|n)} \qquad (14)$$

여기서

$$P(\hat{S}'|n) = \prod_{k=1}^{K} P(S'_k|n). \qquad (15)$$

그러므로 추정량은 클라스-조건 MMSE 추정량들의 가중합(weighted sum)으로서 보여진다. 여기서 N=1이면 결과 추정량은 각 채널들의 MMSE 추정량과 동일하다.

본 발명에 따라 계산가능한 추정량을 실현하기 위하여, 식(10)에 의해 모델 P(S)에 대하여 혼합 모델을 채용하는 것이 바람직하다. 혼합 모델을 적용에 따라 몇가지로 구현하였다.

혼합 모델의 가장 단순한 구현은 벡터 양자화에 의한 것이다. 벡터 양자화 혼합 모델은 음향 공간의 고정-경계 구획들을 가진 클래스들을 동일화하고, 벡터 양자화에 의해 구획을 생성한다. 크기 N의 코드북(code book)은 로이드(Lloyd) 알고리즘을 사용하여 생성된다(로이드 알고리즘은 그레이Gray)의 문헌에서 주어진다). 코드북은 유클리드 거리에 의해 측정된 바와 같이 뒤틀림을 최소화하고 $P_n(S_k)$는 코드어(code word) n으로 양자화된 모든 음성 프레임들의 히소토그램(histogram)으로부터 추정된다.

계산가능한 방법에서 더욱 적합한 혼합 모델의 다른 구현예는 가우스 모델의 혼합이다. 이 모델은 주어진 관찰 데이타의 유사성을 최대화하도록 조정된 PDs(확률분포)에 대한 파라메트릭 모델이다.

가우스 모델의 혼합으로의 매개변수화(parameterization)는 확률 $P_n(S_k)$가 각기 평균값이 μ_{nk}이고, 표준편차가 σ_{nk}인 가우스 분포라는 가정으로 시작한다. 그러나 최대 가능성 문제는 연속 밀도를 가진 히든 마르코프 모델 변수들을 추정하는 문제가 된다. 여기서 그 모델은 다이아고날 코배리언스 매트릭스(diagonal covariance matrices)의 다변량 가우스 콤포넌트 N을 가진 단일 상태를 포함한다. 제3도에는 다음과 같은 반복 프로시져를 사용한 변수 추정단계를 도시하였다.

C, μ 및 σ(AA단계)를 초기 추정한다. 이들 시드(seed)값들에 대해서 합리적인 시작점은 각 클래스에서 벡터들의 상대적인 수들을 그들의 평균 및 그들의 표준편차값으로 사용하는 벡터 양자화에 의한 분류가 될 것이다.

그 다음은, 모든 음성 프레임에 관하여 루프를 수행하고 각 프레임 t에 대해서 다음식에 따라 확률 $\gamma_n(t)$를 계산한다.

$$\gamma_n(t) = \frac{C_n P_n(\hat{S}_t)}{\sum_n C_n P_n(\hat{S}_t)} \qquad (16)$$

여기서 $P_n(\hat{S}_t)$은 μ와 σ의 현재값을 사용하여 계산된다(AB단계). 이때 새로운 파라메타 추정은 다음 시간 평균에 의해 주어진다.

$$C_n = \langle \gamma n(t) \rangle \qquad (17)$$

$$\mu_{nk} = \gamma n(t) S_k(t) \qquad (18)$$

$$\sigma^2_{nk} = \gamma n(t)(S_k(t) - \mu_{nk})^2 \qquad (19)$$

결국, 전체 가능성(total likelihood)의 수렴은 다음식으로 주어진다:

$$L = \sum_t \log_{10} [\sum_n C_n P_n(S_t)] \qquad (20)$$

(AC단계). 만일 수렴이 획득되지 않으면, AB단계가 반복되고 수렴 테스트(AC단계)는 수렴이 획득될때까지 반복된다.

이상의 방법들은 실제로 잠재적 어려움이 있다. 첫째, 필터-뱅크-기본 시스템들은 전형적으로 중복되는 통과대역을 가지는 필터를 갖는다. 둘째,상기 모델에 대해서도 계산부하가 너부 과다하게 된다. 특히 대화형 또는 거의 실시간(real-time) 인식 시스템에서 그러하다. 중복 필터에 대해서 채널의 통계학적 독립성에 관한 가정은 부정확하다. 중복 필터들을 비중복 필터들로 교체할 수 있는 한편 계산부하를 줄이기 위하여 광대역(broadband) 혼합 모델을 적용하는 것도 또한 가능하다.

중복 필터에 대해서, 필터 로그-에너지의 K-차원 벡터는 비중복 통과대역이지만 광대역을 가지는 더 적은 수의 필터들을 가정하여 K차원 이하의 벡터로 나타낼 수 있다. 그리하여 VQ 혼합 모델을 사용하는 양자화는 단순하게 되고, 새롭고 더 차원적인 벡터가 다음과 같은 광대역 벡터 양자화 혼합 모델로서 도입

된다:

$$\hat{S}_k = \sum_{n=1}^{N} \hat{S}_k|n \cdot P(n|R') \qquad (21)$$

여기서 j는 광대역 채널이고, R_j는 채널 j내의 로그-에너지이며, 밴드들의 전체수는 J이고, 음성 프레임들의 분류는 사이즈 N의 코드북으로써 벡터 R을 양자화함으로써 진행된다. 식(10)의 혼합 콤포넌트 $P_n(S_k)$는 그 클래스들을 기초로 추정되고, 후(posteriori), 클러스터 확률은 식(11)에서와 같이 벡터 S'에 관해서라기보다 오히려 벡터 R'에 관해 필요하게 된다. 식(21)은 식(11)로 교체하고, 식(14) 및 (15)를 교체한 후 클러스터 확률은 다음과 같이 주어진다:

$$P(n|R') = \frac{C_n P(R'|n)}{\sum_{n=1}^{N} C_n P(R'|n)} \qquad (22)$$

여기서

$$P(R'|n) = \prod_{j=1}^{J} P(R'_j|n). \qquad (23)$$

$P(R'_j|\,)$는 식(13)에 유사하게 계산된다. 여기서 R_j 및 R'_j는 S_k 및 S'_k를 각기 대신한다. $P_n(R_j)$는 벡터 양자화를 사용하여 코드어 n으로 양자화된 음성 프레임들의 히스토그램들로부터 추정되거나 또는 가우스 혼합 모델링과 유사하게 가우스 방법에 의해 모델화 되어질 수 있다. $P(R'_j|R_j)$는 식(9)의 $P(S'_k|S_k)$ 다음에 유사하게 모델화되어 질 수 있다. 그러므로 이 방법의 잇점은 식(14)에 비하여 식(22)에서 곱해지고, 계산되어질 적분수가 적어진다는 것이다.

제4도를 참조하면, 본 발명에 따른 전처리기(22)의 제2실시예를 도시하였다. 제4도의 전처리기(22)는 다음을 근거로 하여 일련의 노이즈 음성 벡터 S'_0에서 $S'_{T'}$로부터 깨끗한 음성 벡터 $\hat{S}_0$에서 $\hat{S}_T$의 일련의 시간 프레임들에 대해 추정치를 계산한다:

1) 깨끗한 음성의 확률분포는 마르코프 모델에 의해서 모델화될 수 있다는 가정, 여기서 마르코프 모델의 각 상태는 다음을 근거로 하여 각 클래스내에서 다른 주파수 채널은 상호 연관되지 않는다고 가정하여 다른 음성 클래스를 나타낸다.

2) 다른 주파수 채널들에서 노이즈는 서로 관련이 없다는 가정에 기초하는 깨끗한 음성의 벡터에 대한 노이즈 음성 벡터의 조건 확률 함수

추정량은 일련의 벡터 S_0에서 $S_{T'}$의 최소제곱 평균오차(MMSE) 추정이다. 일련의 노이즈 벡터 S'_0에서 $S'_{T'}$에 주어진 K 필터 로그-에너지의 벡터 S의 최소 유클리드 거리 추정량은 다음 벡터 추정량을 얻는다:

$$\hat{S}_t = \int S_t \, P(S_t|S'_0,\ldots,S'_t\ldots S'_T) \, dS_t. \qquad (24)$$

이 벡터 추정량은 다음의 가정을 이용하여 계산된다. 음성은 임의의 시간 t에서 음성이 N개의 다른 상태들중 어느 하나에 있도록 제1차 마르코프 프로세스에 의해 모델화 될 수 있다. 상태 n에 대해 출력 확률분포(PD)는 다음 식으로 주어진다:

$$P_n(S) = \prod_{k=1}^{K} P_n(S_k). \qquad (25)$$

그 상태 출력 확률분포들과 상태들간의 전이 확률들은 음향공간의 구획들과 함께 상태들을 동일화 함으로써 계산되어진다. 이 구획들은 상술한 벡터 양자화 혼합 모델 또는 광대역 벡터 양자화 혼합 모델에 사용된 것들과 정확히 동일하다. 전이 확률들은 하나의 상태에서 다른 상태로 전이하는 수를 카운트함으로써 음성 데이타로부터 추정된다.

추가의 노이즈가 있는 경우, 노이즈 음성은 히든 마르코프 모델(HMM)에 의해 모델화되고, 그 상태는 깨

끗한 음성의 상태에 상당하고, 출력 확률분포들은 다음과 같이 주어진다:

$$P(S'|n) = \prod_{k=1}^{K} P(S'_k|n) \quad . \qquad (26)$$

여기서 $P(S'_k|n)$은 다음과 같다.

$$P(S'_k|n) = \int P(S'_k|S_k)\ P_n(S_k)\ dS_k \qquad (27)$$

또한 $P_n(S_k)$는 상태 n에 대해 깨끗한 음성 벡터 S의 K^{th} 콤포넌트의 출력 확률분포이다. S_k로 주어진 S'_k의 확률은 상술한 식(7) 내지 (9)를 사용하여 계산할 수 있다.

이상의 가정들을 모두 주면, 제4도의 추정량은 다음과 같이 된다:

$$S_k = \sum_n S_k|n \cdot P(n|S'_0,\ldots,S'_t\ldots S'_T). \qquad (28)$$

식(28)에서 일련의 노이즈 음성 벡터에 주어진 상태 n의 확률은 상술한 노이즈 음성의 HMM에 적용된 순방향-역방향 알고리즘(forward-backward algorithm)에 의해 계산된다(순방향-역방향 알고리즘은 한예로 본원의 한 부분으로 채택되고 참고된 상술한 논문들중 Rabiner(1989) 논문에 나타나 있다).

본 발명에 따른 6개의 다른 전처리기 추정량의 한 실시예의 예는 첨부한 부록 A에 나타나 있다. 부록 A는 프로그램언어 LISP으로 작성된 소스코드이다.

이제까지 본 발명은 특정된 실시예들을 참조하여 기술하였다. 그의 다른 실시예들은 당해 기술분야에 통상의 지식을 가진자들에게 자명한 것들이다. 그러므로 본 발명은 첨부한 특허청구범위에 제시된 것을 제외하고는 제한되지 않는다.

(57) 청구의 범위

청구항 1

디지탈화된 음성을 나타내는 신호들에 응답하기 위한 음성인식 시스템에서 노이즈 존재하의 음성을 전처리하는 방법에 있어서, 디지탈화된 음성과 노이즈를 노이즈 음성 벡터의 각 시간-불변 세그먼트와 노이즈 존재하의 음성정보의 한 주파수 채널을 나타내는 노이즈 음성 벡터의 각 엘리멘트로 분류하고, 상기한 노이즈는 주파수 정의역에 무관하고, 시간 불변하게 부가되어지는 것으로 가장되는 노이즈이며; 1) 깨끗한 음성의 확률분포가 혼합 모델로서 콤포넌트 혼합에 의하여 모델화되어 질 수 있고, 각 콤포넌트는 서로 다른 주파수 채널들이 2)항을 근거로 하여 각 클래스내에서 상호 관련이 없다고 가정하는 서로 다른 음성 클래스를 나타낸다는 가정 및 2) 추정된 깨끗한 음성 벡터를 획득하기 위하여 서로 다른 주파수 채널들이 상호 관련이 없다는 가정을 기본으로하는 깨끗한 음성 벡터 어라운드 노이즈 음성 벡터의 조건확률 함수를 근거로 하여 노이즈 음성 벡터로부터 상기한 각 엘리멘트에 대하여 음성값을 추정하는 단계를 포함하는 음성인식기의 전처리 방법.

청구항 2

제1항에 있어서,

$$\hat{S}_k = \sum_{n=1}^{N} S_k|n \cdot P(n \mid S')$$

여기서 인덱스 n은 클래스이고, S_k 어라운드 n의 추정은 다음과 같이 주어진 n^{th} 클래스-조건화된 MMSE 추정량이고,

$$\hat{S}_k|n = \frac{1}{P(S'_k|n)} \int S_k\ P(S'_k|S_k)\ P_n(S_k)\ dS_k$$

여기서

$$P(S'_k|n) = \int P(S'_k|S_k)\ P_n(S_k)\ dS_k$$

인 함수에 따라 다중-주파수 채널 함수에 관하여 필터 로그 스펙트랄 에너지를 추정하는 단계를 포함하는 음성인식기의 전처리 방법.

청구항 3

제2항에 있어서, 상기한 추정단계에 유클리드 거리의 추정을 획득하기 위하여 노이즈 음성의 벡터에 최소 평균-로그-스펙트랄-거리 추정을 적용하는 단계가 포함된 음성인식기의 전처리 방법.

청구항 4

제1항에 있어서, 깨끗한 음성 벡터 어라운드 노이즈 음성 벡터의 조건확률 $P(S'_k | S_k)$은 다음과 같이 모델화 되어진다:

1) 추정되어지는 음성과 연관된 노이즈는 각 시간 프레임에서 상기한 노이즈에 대하여 이산 푸리에 변형(DFT)의 계수가 상호 무관한 복소 가우스 임의 변수가 되도록 정상적이고 자동복귀 이동 평균 확률론적 처리로 가정된다; 2) 필터 출력 에너지는 M계수들의 합에 의해 접근되는 것으로 가정한다; 3) 노이즈 스펙트랄 파워는 필터들의 통과대역 범위내에서 균일하게 되도록 가정된다; 노이즈만 있는 경우, 합은 가우스인 임의의 2M 변수, 제로평균 및 편차의 합 이상으로 연장된다:

$$\sigma^2 = (N_k)/(2M)$$

여기서 N_k는 노이즈 필터 에너지의 기대값이고, 분산-정규화 필터 에너지는 2M 등급의 자유를 가진 카이제곱 확률분포(PD)를 따른다. 그리고 노이즈 및 음성이 존재하는 경우, 필터 에너지는 다음과 같이 주어진다:

$$E'_k = \sum_i |DFT_s(i) + DFT_n(i)|^2$$

여기서 DFTs는 음성계수이고, DFTn은 노이즈 계수이고, 분산 정규화 필터 에너지는 2M 등급자유를 가진 비집중 카이제곱 및 비집중 변수 λ의 확률분포를 따른다:

필터 에너지의 조건확률이 다음과 같이 주어지려면:

$$P\left[\frac{E'_k}{N_k} \left| \frac{E_k}{N_k}\right.\right] = 2M \cdot \chi^2_{N.C.}\left[\frac{2ME'_k}{N_k}, 2M, \lambda\right]$$

$$\text{여기서} \quad \lambda = \sum_i \frac{|DFT_s(i)|^2}{\sigma^2} = \frac{2ME_k}{N_k}$$

정규화 로그-에너지 변수는 다음과 같이 정의되고:

$$S_k = 10 \log_{10}\left[\frac{E_k}{N_k}\right]; \quad S'_k = 10 \log_{10}\left[\frac{E'_k}{N_k}\right]$$

S'_k에 대한 조건확률이

$$P = (S'_k | S_k) = 0.23 \frac{E'_k}{N_k} P\left[\frac{E'_k}{N_k} \left| \frac{E_k}{N_k}\right.\right]$$

인 음성인식기의 전처리 방법.

청구항 5

제1항에 있어서, 혼합 모델은 클래스들이 음향 공간의 구체적 경계 구획들과 동일시하고, 구획들은 벡터 양자화에 의해 생성되는 벡터 양자화 혼합 모델로서 실행되는 음성인식기의 전처리 방법.

청구항 6

제1항에 있어서, 혼합 모델은 확률 $P_n(S_k)$이 가우스 분포이고, 이들 가우스값들의 평균 및 표준편차들은 혼합 모델로 주어진 음성 데이타의 가능성을 최대화하도록 조정된다는 가정을 가지고 가우스 혼합 모델의 혼합으로서 실행되는 음성인식기의 전처리 방법.

청구항 7

제2항에 있어서, 혼합 모델의 벡터 S'는 벡터 R'와 교체되고, 여기서 벡터 R'는 주파수 광대역에서의 로그 스펙트랄 에너지의 보다 낮은 차원의 벡터이고, S'에 따라 달라지는 클래스 n의 확률은 새로운 클래

스 n 콤포넌트들이 벡터 R에 대해 혼합 모델의 클래스들로 참고하도록 R'에 관하여 달라지는 새로운 클래스 n의 확률로 교체되는 음성인식기의 전처리 방법.

청구항 8

제7항에 있어서, 벡터 R에 대한 혼합 모델이 벡터 양자화 혼합 모델로서 실행되고, 여기서 클래스들은 음향공간의 구체적 경제 구획들과 동일시되고, $\hat{S}_k$ 가 깨끗한 음성에 대한 추정량의 표현식에서

$$\hat{S}_k = \sum_{n=1}^{N} S_k|n \cdot P(n \mid R')$$

로 되어, 새로운 클래스 n에 대해 필요조건이 되는 음성인식기의 전처리 방법.

청구항 9

제7항에 있어서, 벡터 R에 대한 혼합 모델은 확률들 $P_n(R_j)$이 가우스 분포들이고, 가우스 평균 및 표준편차들이 혼합 모델로 주어진 음성 데이타의 유사성을 최대화하도록 조정된다는 가정하에 가우스 혼합 모델의 혼합으로서 실행되고, 깨끗한 음성에 대한 추정량에 대한 표현식에서 S_k는

$$\hat{S}_k = \sum_{n=1}^{N} S_k|n \cdot P(n \mid R')$$

되어 새로운 클래스 n에 대해 필요조건이 되는 음성인식기의 전처리 방법.

청구항 10

디지탈화된 음성을 나타내는 신호들에 응답하기 위한 음성인식 시스템에서 노이즈 존재하의 음성을 전처리하는 방법에 있어서, 디지탈화된 음성과 노이즈를 노이즈 음성 벡터의 각 시간-불변 세그먼트와, 노이즈 존재하의 음성정보의 한 채널을 나타내는 노이즈 음성 벡터의 각 엘리멘트로 분류하고, 상기한 노이즈는 주파수에 독립하고 시간에 불변한 것으로 가정되는 노이즈이며; 1) 깨끗한 음성의 확률분포가 마르코프 모델로서 모델화될 수 있고, 마르코프 모델의 각 상태는 다음 2)항을 근거로 각 클래스내에서 서로 다른 주파수 채널은 상호 관련이 없다고 가정하는 서로 다른 음성 클래스를 나타내고 2) 추정된 깨끗한 음성 벡터를 획득하기 위하여 서로 다른 주파수 채널들이 상호 관련이 없다는 가정을 기본으로하는 깨끗한 음성 벡터 어라운드 노이즈 음성 벡터의 조건확률 함수. 이상의 2가지 사항을 근거로 하여 일련의 노이즈 음성 벡터들 S'_0에서부터 S'_T까지 상기한 각 엘리멘트에 대하여 음성값을 추정하는 단계를 포함하는 음성인식기의 전처리 방법.

청구항 11

제10항에 있어서,

$$S_k = \sum_{n} \hat{S}_k|n \cdot P(n|S'_0, \ldots, S'_t \ldots S'_T)$$

에 따르는 다중-주파수 채널확률에 관해 필터 로그 스펙트랄 에너지들을 추정하는 단계를 포함하고, 여기서 인덱스 n은 상태이고, S_k 어라운드 n의 추정은 다음식으로 주어진 n^{th} MMSE 추정량이고:

$$\hat{S}_k|n = \frac{1}{P(S'_k|n)} \int S_k \, P(S'_k|S_k) \, P_n(S_k) \, dS_k$$

여기서

$$P(S'_k|n) = \int P(S'_k|S_k) \, P_n(S_k) \, dS_k$$

여기서 히든 마르코프 모델은 음성에 대하여 가정되고, 히든 마르코프 상태 출력 확률은 다음과 같고:

$$P(\hat{S}'|n) = \prod_{k=1}^{K} P(S'_k|n)$$

여기서 순방향-역방향 알고리즘이 $P(n \mid S'_0, \cdots S'_t \cdots S'_T)$를 계산하기 위해 적용되는 음성인식기의 전처리 방법.

청구항 12

제10항에 있어서, 깨끗한 음성 벡터 어라운드 노이즈 음성 벡터의 조건확률 $P(S'_k | S_k)$은 다음과 같이 모델화 되어진다:1) 추정되어지는 음성과 연관된 노이즈는 각 시간 프레임에서 상기한 노이즈에 대하여 이산 푸리에 변형(DFT)의 계수가 상호 무관한 복소 가우스 임의 변수가 되도록 정상적이고 자동복귀 이동평균 확률론적 처리로 가정된다; 2) 필터 출력 에너지는 M계수들의 합에 의해 접근되는 것으로 가정한다; 3) 노이즈 스펙트랄 파워는 필터들의 통과대역 범위내에서 균일하게 되도록 가정된다; 노이즈가 없는 경우, 합은 가우스인 임의의 2M 변수, 제로평균 및 편차의 합을 넘어 연장된다:

$$\sigma^2 = (N_k)/(2M)$$

여기서 N_k는 노이즈 필터 에너지의 기대값이고 분산-정규화 필터 에너지는 2M 등급 자유를 가진 카이제곱 확률분포(PD)를 따른다. 그리고 노이즈 및 음성이 존재하는 경우, 필터 에너지는 다음과 같이 주어진다:

$$E'_k = \sum_i |DFT_s(i) + DFT_n(i)|^2$$

여기서 DFTs는 음성계수이고, DFTn은 노이즈 계수이고, 분산 정규화 필터 에너지는 2M 등급자유를 가진 비집중 카이제곱 및 비집중 변수 λ의 확률분포를 따른다:

필터 에너지의 조건확률이 다음과 같이 주어지도록:

$$P\left[\frac{E'_k}{N_k} \mid \frac{E_k}{N_k}\right] = 2M.\chi^2_{N.C.}\left[\frac{2ME'_k}{N_k};\ 2M, \lambda\right]$$

여기서

$$\lambda = \sum_i \frac{|DFT_s(i)|^2}{\sigma^2} = \frac{2ME_k}{N_k}$$

정규화 로그-에너지 변수는 다음과 같이 정의되고:

$$S_k = 10 \log_{10}\left[\frac{E_k}{Nk}\right];\ S'_k = 10 \log_{10}\left[\frac{E'_k}{Nk}\right]$$

S'_k에 대한 조건확률이

$$P = (S'_k | S_k) = 0.23 \frac{E'_k}{N_k} P\left[\frac{E'_k}{N_k} \mid \frac{E_k}{N_k}\right]$$

인 음성인식기의 전처리 방법.

청구항 13

제10항에 있어서, 마르코프 모델이 벡터 양자화 마르코프 모델로서 실행되고 상태들은 음향공간의 견고한 경계구획과 동일하게 되고, 구획들은 벡터 양자화에 의해 생성되는 음성인식기의 전처리 방법.

청구항 14

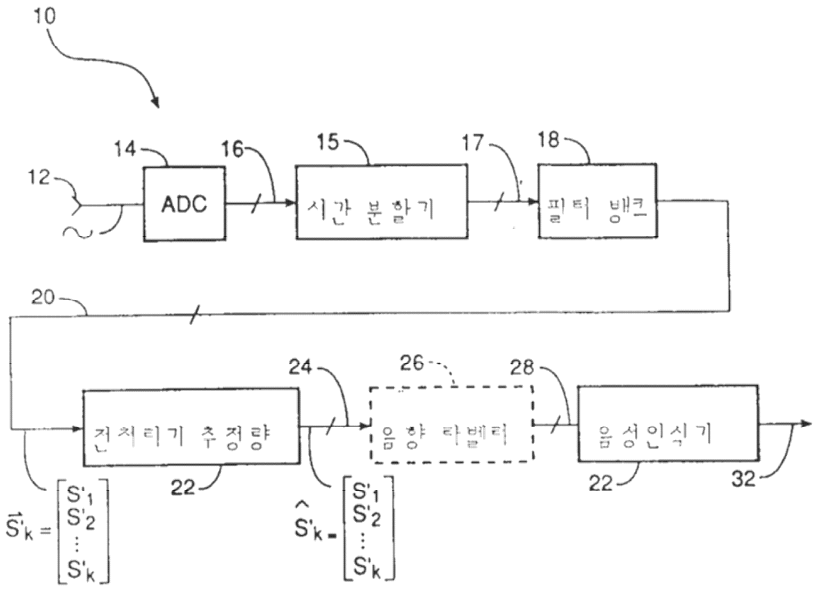
제11항에 있어서, 혼합 모델의 벡터 S'는 벡터 R'에 의해 교체되고 여기서, 벡터 R'는 주파수 광대역에서 로그 스펙트랄 에너지들의 더욱 낮은 차원의 벡터이고, t가 제로에서 T인 S'(t)의 순서에 따라 맞춰진 상태 n의 확률은 새로운 상태 n 콤포넌트가 벡터 R을 위한 마르코프 모델에서의 상태들을 언급하도록 t가 제로에서 T인 R'(t)의 순서에 관해 맞춰진 새로운 상태 n의 확률로 교체되는 음성인식기의 전처리 방법.

청구항 15

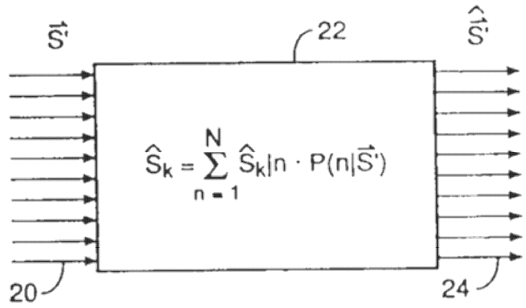
제14항에 있어서, 벡터 R에 대한 혼합 모델은 벡터 양자화 마르코프 모델로서 실행되고, 상태들은 음향공간의 견고한 경계구획들과 동일시되고, 구획들은 벡터 양자화에 의해 생성되고, $\hat{S}_k$는 새로운 상태 n에 따라 좌우되는 음성인식기의 전처리 방법.

도면

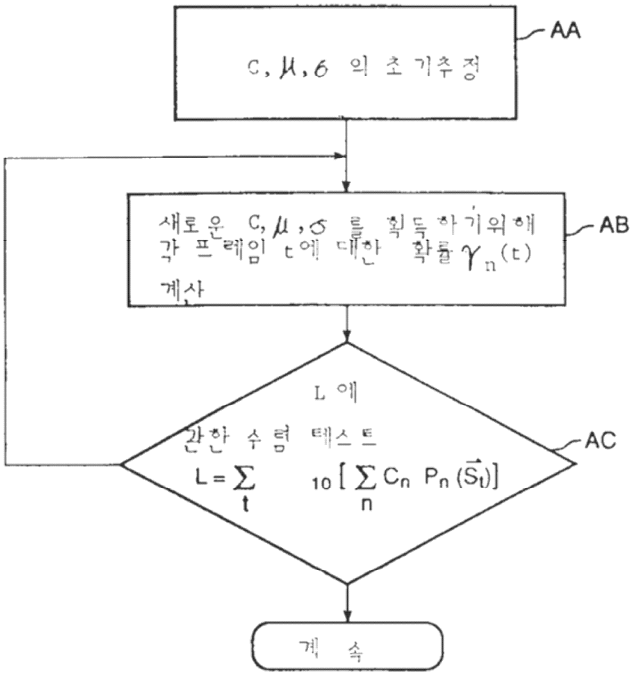
도면1


10
14
16
15
17
18
12
ADC
시간 분할기
필터 뱅크
20
24
26
28
전처리기 추정량
음향 라벨러
음성인식기
22
22
32

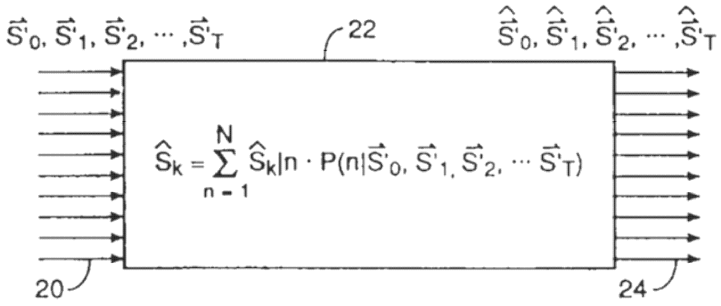
도면2


22
20
24

도면3


c, μ, σ 의 초기추정
AA
새로운 c, μ, σ 를 획득하기위해
각 프레임 t에 대한 확률 γn(t)
계산
AB
L 에
관한 수렴 테스트
L = Σt 10 [Σn Cn Pn (S⃗t)]
AC
계 속

도면4


S⃗'0, S⃗'1, S⃗'2, ··· ,S⃗'T
22
Ŝ⃗'0, Ŝ⃗'1, Ŝ⃗'2, ··· ,Ŝ⃗'T
Ŝk = Σ(n=1 to N) Ŝk|n · P(n|S⃗'0, S⃗'1, S⃗'2, ··· S⃗'T)
20
24