



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0085725
(43) 공개일자 2013년07월30일

(51) 국제특허분류(Int. Cl.)

G10L 25/93 (2013.01) G10L 19/02 (2006.01)

(21) 출원번호 10-2012-0006735

(22) 출원일자 2012년01월20일

심사청구일자 2012년01월20일

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

김무영

서울특별시 강남구 도곡2동 도곡렉슬아파트 210동 402호

이명석

서울특별시 구로구 오류동 68-21 미래빌라 1동 209호

(74) 대리인

특허법인태백

전체 청구항 수 : 총 6 항

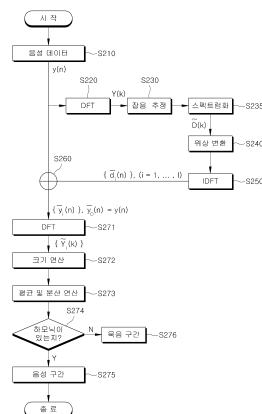
(54) 발명의 명칭 위상 변환된 잡음 신호를 이용한 음성 영역 검출 방법 및 그 장치

(57) 요약

본 발명은 음성 영역 검출 방법 및 장치에 관한 것이다. 본 발명에 따르면, 음성 데이터를 입력받는 단계와, 상기 음성 데이터를 주파수 영역의 신호로 변환하는 단계와, 상기 변환된 음성 데이터로부터 잡음을 추정하는 단계와, 상기 추정된 잡음에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성하는 단계와, 위상이 다른 상기 I개의 잡음 신호들을 각각 시간 영역의 신호로 변환하는 단계와, 상기 변환된 I개의 잡음 신호들 및 상기 입력받은 음성 데이터를 각각의 잡음 신호 별로 연산 처리하여 I개의 합산 신호들을 획득하는 단계, 및 상기 I개의 합산 신호들을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 단계를 포함하는 음성 영역 검출 방법을 제공한다.

상기 음성 영역 검출 방법 및 그 장치에 따르면, 입력받은 음성 데이터로부터 잡음을 추정하고 추정된 잡음에 대해 위상 변환을 수행한 다음, 위상 변환된 잡음 신호들과 상기 입력받은 음성 데이터를 서로 연산 처리한 결과를 바탕으로 잡음 섞인 환경에서 음성 영역과 묵음 영역을 정확히 검출할 수 있는 이점이 있다.

대표도 - 도3



특허청구의 범위

청구항 1

음성 데이터를 입력받는 단계;
 상기 음성 데이터를 주파수 영역의 신호로 변환하는 단계;
 상기 변환된 음성 데이터로부터 잡음을 추정하는 단계;
 상기 추정된 잡음에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성하는 단계;
 위상이 다른 상기 I개의 잡음 신호들을 각각 시간 영역의 신호로 변환하는 단계;
 상기 변환된 I개의 잡음 신호들 및 상기 입력받은 음성 데이터를 각각의 잡음 신호 별로 연산 처리하여 I개의 합산 신호들을 획득하는 단계; 및
 상기 I개의 합산 신호들을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 단계를 포함하는 음성 영역 검출 방법.

청구항 2

청구항 1에 있어서,
 상기 음성 구간 및 묵음 구간을 판별하는 단계는,
 상기 I개의 합산 신호들을 각각 주파수 영역의 신호로 변환하는 단계;
 상기 변환된 I개의 합산 신호들의 크기를 각각 연산하는 단계; 및
 상기 I개의 합산 신호들의 크기에 대한 평균 값 또는 분산 값의 연산 결과를 이용하여 상기 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 음성 영역 검출 방법.

청구항 3

청구항 2에 있어서,
 상기 평균 값 또는 분산 값의 연산 결과를 이용하여 상기 음성 구간 및 묵음 구간을 판별하는 단계는,
 상기 평균 값 또는 분산 값의 연산 결과로부터 하모닉 구간이 있는지를 판단하는 단계; 및
 상기 하모닉 구간이 있는 경우, 상기 음성 구간으로 판단하고, 상기 하모닉 구간이 없는 경우, 상기 묵음 구간으로 판단하는 단계를 더 포함하는 음성 영역 검출 방법.

청구항 4

음성 데이터를 입력받는 데이터 입력부;
 상기 음성 데이터를 주파수 영역의 신호로 변환하는 제1 신호 변환부;
 상기 변환된 음성 데이터로부터 잡음을 추정하는 잡음 추정부;
 상기 추정된 잡음에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성하는 위상 변환부;
 위상이 다른 상기 I개의 잡음 신호들을 각각 시간 영역의 신호로 변환하는 제2 신호 변환부;
 상기 변환된 I개의 잡음 신호들 및 상기 입력받은 음성 데이터를 각각의 잡음 신호 별로 연산 처리하여 I개의 합산 신호들을 획득하는 신호 연산부; 및
 상기 I개의 합산 신호들을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 구간 판별부를 포함하는 음성 영역 검출 장치.

청구항 5

청구항 4에 있어서,

상기 구간 판별부는,

상기 I개의 합산 신호들을 각각 주파수 영역의 신호로 변환한 후 상기 변환된 I개의 합산 신호들의 크기를 각각 연산한 다음, 상기 I개의 합산 신호들의 크기에 대한 평균 값 또는 분산 값의 연산 결과를 이용하여 상기 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 음성 영역 검출 장치.

청구항 6

청구항 5에 있어서,

상기 구간 판별부는,

상기 평균 값 또는 분산 값의 연산 결과를 이용하여 상기 음성 구간 및 묵음 구간을 판별 시에,

상기 평균 값 또는 분산 값의 연산 결과로부터 하모닉 구간이 있는지를 판단한 후, 상기 하모닉 구간이 있는 경우, 상기 음성 구간으로 판단하고, 상기 하모닉 구간이 없는 경우, 상기 묵음 구간으로 판단하는 음성 영역 검출 장치.

명세서

기술분야

[0001] 본 발명은 위상 변환된 잡음 신호를 이용한 음성 영역 검출 방법 및 그 장치에 관한 것으로서, 보다 상세하게는 입력된 음성 데이터로부터 음성 구간과 묵음 구간을 판별할 수 있는 위상 변환된 잡음 신호를 이용한 음성 영역 검출 방법 및 그 장치에 관한 것이다.

배경기술

[0002] 기존에 음성 데이터로부터 음성 영역을 검출하는 방법에 관하여 다양하게 개시되어 있다. 그 중에서 특허공개 제1999-0039422호는 G.729 음성 부호화기를 위한 음성 활성화 측정 방법에 관한 것이다. 이는 음성 프레임의 에너지를 문턱값과 비교하여 묵음 구간 및 실 음성 구간으로 구분하는 구성을 포함한다.

[0003] 그런데, 이러한 종래의 방법에 따르면, 단순히 문턱값을 이용하여 음성 영역을 검출하므로 에러가 발생할 수 있으며, 잡음이 섞이면서 특징 벡터들이 크게 변하여 정확한 음성 영역의 검출이 어려운 단점이 있다.

발명의 내용

해결하려는 과제

[0004] 본 발명은, 입력받은 음성 데이터로부터 잡음을 추정하고 추정된 잡음에 대해 위상 변환을 수행한 다음, 위상 변환된 잡음 신호들과 상기 입력받은 음성 데이터를 서로 연산 처리한 결과를 바탕으로 잡음 섞인 환경에서 음성 영역과 묵음 영역을 정확히 검출할 수 있는 위상 변환된 잡음 신호를 이용한 음성 영역 검출 방법 및 그 장치를 제공하는데 목적이 있다.

과제의 해결 수단

[0005] 본 발명은, 음성 데이터를 입력받는 단계와, 상기 음성 데이터를 주파수 영역의 신호로 변환하는 단계와, 상기 변환된 음성 데이터로부터 잡음을 추정하는 단계와, 상기 추정된 잡음에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성하는 단계와, 위상이 다른 상기 I개의 잡음 신호들을 각각 시간 영역의 신호로 변환하는 단계와, 상기 변환된 I개의 잡음 신호들 및 상기 입력받은 음성 데이터를 각각의 잡음 신호 별로 연산 처리하여 I개의 합산 신호들을 획득하는 단계, 및 상기 I개의 합산 신호들을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 단계를 포함하는 음성 영역 검출 방법을 제공한다.

[0006] 여기서, 상기 음성 구간 및 묵음 구간을 판별하는 단계는, 상기 I개의 합산 신호들을 각각 주파수 영역의 신호로 변환하는 단계와, 상기 변환된 I개의 합산 신호들의 크기를 각각 연산하는 단계, 및 상기 I개의 합산 신호들의 크기에 대한 평균 값 또는 분산 값의 연산 결과를 이용하여 상기 음성 데이터 내의 음성 구간 및 묵음 구간을 판별할 수 있다.

[0007] 또한, 상기 평균 값 또는 분산 값의 연산 결과를 이용하여 상기 음성 구간 및 묵음 구간을 판별하는 단계는, 상기 평균 값 또는 분산 값의 연산 결과로부터 하모닉 구간이 있는지를 판단하는 단계, 및 상기 하모닉 구간이 있는 경우, 상기 음성 구간으로 판단하고, 상기 하모닉 구간이 없는 경우, 상기 묵음 구간으로 판단하는 단계를 더 포함할 수 있다.

[0008] 그리고, 본 발명은, 음성 데이터를 입력받는 데이터 입력부와, 상기 음성 데이터를 주파수 영역의 신호로 변환하는 제1 신호 변환부와, 상기 변환된 음성 데이터로부터 잡음을 추정하는 잡음 추정부와, 상기 추정된 잡음에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성하는 위상 변환부와, 위상이 다른 상기 I개의 잡음 신호들을 각각 시간 영역의 신호로 변환하는 제2 신호 변환부와, 상기 변환된 I개의 잡음 신호들 및 상기 입력받은 음성 데이터를 각각의 잡음 신호 별로 연산 처리하여 I개의 합산 신호들을 획득하는 신호 연산부, 및 상기 I개의 합산 신호들을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별하는 구간 판별부를 포함하는 음성 영역 검출 장치를 제공한다.

발명의 효과

[0009] 본 발명에 따른 음성 영역 검출 방법 및 그 장치에 따르면, 입력받은 음성 데이터로부터 잡음을 추정하고 추정된 잡음에 대해 위상 변환을 수행한 다음, 위상 변환된 잡음 신호들과 상기 입력받은 음성 데이터를 서로 연산 처리한 결과를 바탕으로 잡음 섞인 환경에서 음성 영역과 묵음 영역을 정확히 검출할 수 있는 이점이 있다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 실시예에 따른 음성 영역 검출 장치의 구성도이다.

도 2는 도 1의 장치를 이용한 음성 영역 검출 방법을 나타내는 흐름도이다.

도 3은 도 2에 대응되는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0011] 그러면 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다.

[0012] 도 1은 본 발명의 실시예에 따른 음성 영역 검출 장치의 구성도이다. 상기 음성 영역 검출 장치(100)는 데이터 입력부(110), 제1 신호 변환부(120), 잡음 추정부(130), 위상 변환부(140), 제2 신호 변환부(150), 신호 연산부(160), 구간 판별부(170)를 포함한다.

[0013] 상기 데이터 입력부(110)는 음성 데이터를 입력받으며, 상기 제1 신호 변환부(120)는 상기 음성 데이터를 주파수 영역의 신호로 변환한다.

[0014] 상기 잡음 추정부(130)는 상기 변환된 음성 데이터로부터 잡음을 추정한다. 다음, 상기 위상 변환부(140)는 추정된 잡음에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성하고, 상기 제2 신호 변환부(150)는 위상이 다른 상기 I개의 잡음 신호들을 각각 시간 영역의 신호로 변환한다.

[0015] 그리고, 상기 신호 연산부(160)는 상기 변환된 I개의 잡음 신호들 및 상기 입력받은 음성 데이터를 각각의 잡음 신호 별로 연산 처리하여 I개의 합산 신호들을 획득한다. 이후, 구간 판별부(170)는 상기 I개의 합산 신호들을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별한다.

[0016] 도 2는 도 1의 장치를 이용한 음성 영역 검출 방법을 나타내는 흐름도이다. 도 3은 도 2에 대응되는 흐름도이다. 이하에서는 본 발명에 따른 음성 영역 검출 방법에 관하여 상세히 설명한다.

[0017] 먼저, 상기 데이터 입력부(110)에서는 음성 영역 검출 대상에 해당되는 테스트용 음성 데이터를 입력받는다 (S210). 입력받은 음성 데이터($y(n)$)는 잡음이 섞인 음성 신호로서, $y(n) = x(n) + d(n)$ 으로 나타낼 수 있다. 여기서, $x(n)$ 은 클린 음성 신호이고, $d(n)$ 은 잡음 신호이다.

[0018] 데이터 입력부(110)는 음성 수집모듈(미도시)로부터 미리 획득된 음성 데이터를 전송받아 입력받거나, 직접 음성 데이터를 수집하여 입력받을 수 있다. 음성 데이터는 시간 흐름에 따라 복수의 프레임들로 구분될 수 있다.

[0019] 이후, 상기 제1 신호 변환부(120)에서는 상기 입력받은 음성 데이터를 DTF(Discrete Fourier Transform)를 통

해 주파수 영역의 신호($Y(k)$)로 변환한다(S220).

[0020] 상기 잡음 추정부(130)에서는 상기 주파수 영역으로 변환된 음성 데이터로부터 잡음을 추정한다(S230). 그리고, 추정된 잡음($\tilde{D}(k)$)을 스펙트럼화한다(S235).

[0021] 여기서, 잡음을 추정하는 방식으로는 MS(Minimum statistics), WSA(Weighted spectral averaging), IMCRA(improve minima control recursive average), CDSTP(Codebook driven short-term parameter estimation) 등 공지된 다양한 방식이 적용될 수 있다.

[0022] 이후, 상기 위상 변환부(140)에서는 상기 추정된 잡음($\tilde{D}(k)$)에 대해 I개의 위상 변환을 수행하여 서로 다른 I개의 잡음 신호들을 생성한다(S240). 즉, 이러한 S240 단계는 추정된 잡음($\tilde{D}(k)$)에 대해 위상을 변화시켜서 서로 다른 I개의 잡음 신호를 생성한다.

[0023] 다음, 상기 제2 신호 변환부(150)에서는 위상이 다른 상기 I개의 잡음 신호들을 각각 IDFT(Inverse Discrete Fourier Transform)를 통해 시간 영역의 신호($\{\tilde{d}_i(n)\}, (i = 1, \dots, I)$)로 변환한다(S250).

[0024] 그러면, 신호 연산부(160)에서는 상기 S250 단계를 통해 시간 영역으로 변환된 I개의 잡음 신호들($\{\tilde{d}_i(n)\}$) 및 상기 S210 단계에서 입력받은 음성 데이터($y(n)$)를 각각의 잡음 신호($\{\tilde{d}_i(n)\}$)별로 합산 연산 처리하여, 총 I개의 합산 신호들($\{\tilde{y}_i(n)\}$)을 획득한다(S260).

[0025] 즉, S260 단계는 위상 변화를 통해 새롭게 생성된 I개의 잡음을 상기 입력된 음성 데이터에 각각 더해주는 과정에 해당된다. 이렇게 새로운 잡음을 음성 데이터에 더해줄 경우, 음성 구간에서의 하모닉은 크게 영향을 받지 않고 그대로 존재한다. 이에 반해, 묵음 구간에 대해서는 하모닉이 존재하지 않으므로, 이 묵음 구간에 다른 위상의 잡음이 섞일 경우 전혀 다른 스펙트럼을 보여주게 된다.

[0026] 이후, 상기 구간 판별부(170)에서는 상기 I개의 합산 신호들($\{\tilde{y}_i(n)\}$)을 이용하여 음성 데이터 내의 음성 구간 및 묵음 구간을 판별한다(S270). 상기 구간 판별부(170)에서의 S270 단계의 과정을 보다 상세히 설명하면 다음과 같다.

[0027] 우선 상기 I개의 합산 신호들($\{\tilde{y}_i(n)\}$)을 각각 주파수 영역의 신호($\{\tilde{Y}_i(k)\}$)로 변환한다(S271). 이후, 상기 주파수 영역으로 변환된 I개의 합산 신호들의 크기를 각각 연산한다(S272). 이에 따라, 총 I개의 크기 값이 연산될 수 있다.

[0028] 이후에는, 상기 I개의 합산 신호들의 크기에 대한 평균 값과 분산 값을 연산한다(S273).

[0029] 여기서, 평균 값($\mu(k)$)과 분산 값($\sigma^2(k)$)은 각각 수학식 1과 수학식 2을 이용할 수 있다.

수학식 1

$$\mu(k) = \frac{1}{I+1} \sum_{i=0}^I |\tilde{Y}_i(k)|$$

[0030]

수학적식 2

$$\sigma^2(k) = \frac{1}{I+1} \sum_{i=0}^I \{ |\widetilde{Y}_i(k)| - \mu(k) \}^2$$

[0031]

[0032] 즉, 상기 S273 단계 이후에는, 상기 평균 값 또는 분산 값의 연산 결과로부터 하모닉 구간이 있는지를 판단한다(S274). 만약, 상기 하모닉 구간이 있는 경우, 상기 음성 구간으로 판단하고(S275), 상기 하모닉 구간이 없는 경우, 상기 묵음 구간으로 판단한다(S276).

[0033] 먼저, 평균 값을 이용하여 하모닉 구간의 여부를 판단하는 방법은 다음과 같다.

[0034] 즉, 수학적식 1과 같은 평균 값으로 얻어진 하나의 스펙트럼을 통하여 하모닉 구간의 존재 여부를 판단할 수 있고, 이를 통해 음성 구간인지 아닌지를 판단할 수 있다. 이때, 하모닉의 존재 여부를 판단하는 방법으로는 평균 값에 대하여 자기 상관 함수를 사용하여 주기성을 판단하는 방법을 이용할 수 있다. 이를 통해 주기성을 갖는 경우에는 하모닉이 존재하는 것이므로 음성 구간으로 결정을 하고, 주기성이 없는 경우에는 묵음 구간으로 판단한다.

[0035] 다음, 분산 값을 이용하여 하모닉 구간의 여부를 판단하는 방법은 다음과 같다. 일반적으로 잡음이 추가로 섞여도 음성 구간의 스펙트럼은 변화가 크지 않기 때문에 분산 값은 대체로 작게 나오게 된다. 따라서, 분산 값이 기준 값 미만이면 하모닉 구간이 있는 것으로서 음성 구간으로 판단하고, 분산 값이 기준 값 이상이면 하모닉 구간이 없는 것으로서 묵음 구간으로 판단한다. 여기서, 기준 값이란 묵음 구간과 음성 구간을 미리 알고 있는 음성 데이터들에 대하여 각 구간별 분산 값을 학습시켜서 얻어진 기준 값(또는 기준 범위)에 해당될 수 있다. 즉, 이와 같이 분산 값 자체를 특징 벡터로 사용하여 음성 구간과 묵음 구간을 판단할 수 있다.

[0036] 이러한 본 발명에서는 잡음의 위상을 변화하여 다시 음성 데이터에 섞었을 때, 원래의 음성 구간에서의 하모닉에는 변화가 없기 때문에 이를 이용하여 음성 구간의 하모닉을 찾음으로써 음성 구간을 손쉽게 검출할 수 있다.

[0037] 이상과 같은 본 발명에 따른 음성 영역 검출 방법에 따르면, 입력받은 음성 데이터로부터 잡음을 추정하고 추정된 잡음에 대해 위상 변환을 수행한 다음, 위상 변환된 잡음 신호들과 상기 입력받은 음성 데이터를 서로 연산 처리한 결과를 바탕으로 잡음 섞인 환경에서 음성 영역과 묵음 영역을 정확히 검출할 수 있다.

[0038] 이러한 본 발명의 방법은 음성 데이터를 구성하는 프레임 단위로 수행할 수 있는데 본 발명이 반드시 이에 한정되는 것은 아니다.

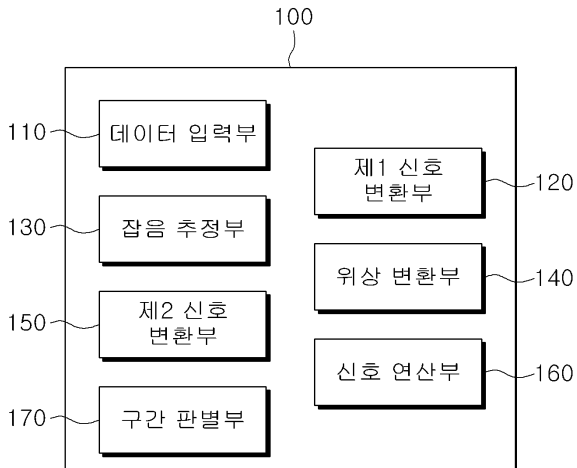
[0039] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

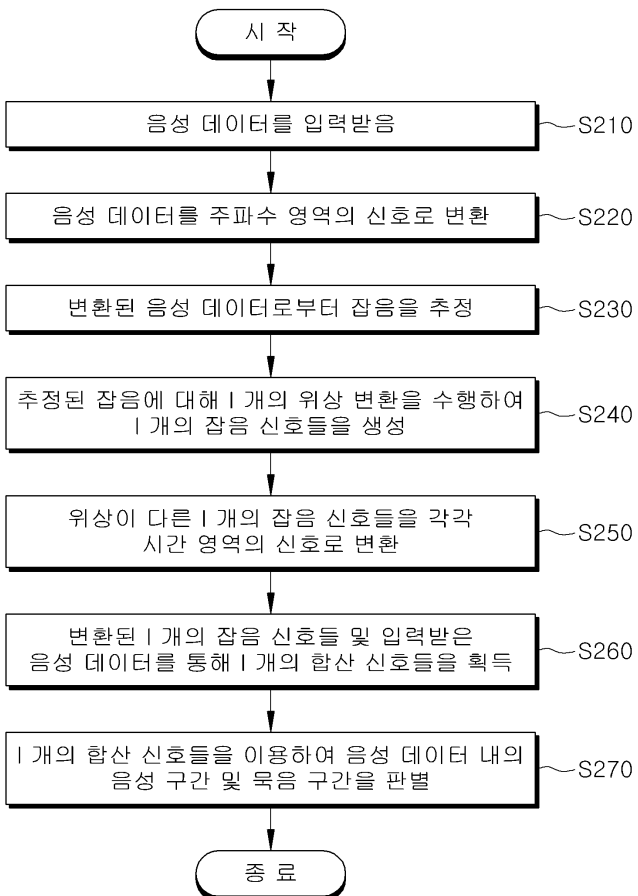
[0040]	100: 음성 영역 검출 장치	110: 데이터 입력부
	120: 제1 신호 변환부	130: 잡음 추정부
	140: 위상 변환부	150: 제2 신호 변환부
	160: 신호 연산부	170: 구간 판별부

도면

도면1



도면2



도면3

