(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0098100
(43) 공개일자 2010년09월06일

(51) Int. Cl.
G10L 11/00 (2006.01) *G10L 11/02* (2006.01)
(21) 출원번호 10-2009-0017109
(22) 출원일자 2009년02월27일
심사청구일자 2009년02월27일

(71) 출원인
고려대학교 산학협력단
서울 성북구 안암동5가 1
(72) 발명자
육동석
서울 용산구 서빙고동 신동아아파트 1동 102호
양경철
경기도 군포시 대야미동 3B 1,2,3 센트럴아이파크 106동 603호
(74) 대리인
특허법인충현

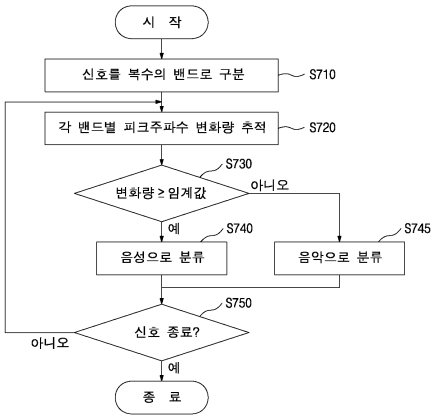
전체 청구항 수 : 총 13 항

(54) 음성과 음악을 구분하는 방법 및 장치

(57) 요 약

본 발명은 음성과 음악을 구분하는 방법에 관한 것으로, 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법은 입력되는 신호에서 일정 구간 마다 피크 주파수 변화량의 평균을 산출하는 단계; 및 상기 피크 주파수 변화량이 임계값 이상이면 상기 신호를 음성으로 분류하고, 상기 피크 주파수 변화량이 임계값 미만이면 상기 신호를 음악으로 분류하는 단계를 포함한다. 본 발명에 의하면, 음성과 음악을 구분함에 있어서 빠른 응답을 유지하면서도 높은 성능을 구현할 수 있는 효과가 있다.

대 표 도 - 도7



특허청구의 범위

청구항 1

입력되는 신호에서 일정 구간 마다 피크 주파수 변화량의 평균을 산출하는 단계; 및

상기 피크 주파수 변화량이 임계값 이상이면 상기 신호를 음성으로 분류하고, 상기 피크 주파수 변화량이 임계값 미만이면 상기 신호를 음악으로 분류하는 단계

를 포함하는, 음성과 음악을 구분하는 방법.

청구항 2

제1항에 있어서,

상기 분류된 결과에 따라, 상기 신호로부터 음성으로부터 분류된 신호를 추출하는 단계를 더 포함하는 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 3

제1항에 있어서,

상기 피크 주파수 변화량은,

$d(t,b)$는 시간 t에서 밴드 b의 주파수 변화량이며, $f(t,b)$는 시간 t에 밴드 b의 피크 주파수일 때, 수학식 $d(t,b)=f(t,b)-f(t-1,b)$에 따라 연산되는 순간 주파수 변화량을 이용하여 산출되는 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 4

제3항에 있어서,

상기 순간 주파수 변화량은,

f_{max}가 변화 제한 폭일 때, $|f(t,b)-f(t-1,b)|>f_{max}$이면, 0의 값을 갖는 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 5

음성과 음악이 혼합된 신호를 복수의 밴드로 구분하는 단계;

상기 각 밴드마다 일정 구간 동안의 피크 주파수 변화량의 평균을 산출하는 단계; 및

상기 피크 주파수 변화량이 임계값 이상인 밴드의 신호를 음성으로 분류하고, 상기 피크 주파수 변화량이 임계값 미만인 밴드의 신호를 음악으로 분류하는 단계

를 포함하는, 음성과 음악을 구분하는 방법.

청구항 6

제5항에 있어서,

상기 복수의 밴드는,

주파수 축 상에서 125Hz 간격으로 분할된 주파수 대역인 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 7

제5항에 있어서,

상기 복수의 밴드는,

270Hz부터 3010Hz 사이의 주파수 대역을 일정 간격으로 분할한 주파수 대역인 것을 특징으로 하는, 음성과 음악

을 구분하는 방법.

청구항 8

제5항에 있어서,

상기 피크 주파수 변화량은,

$d(t,b)$는 시간 t에서 밴드 b의 주파수 변화량이며, $f(t,b)$는 시간 t에 밴드 b의 피크 주파수일 때, 수학식 $d(t,b)=f(t,b)-f(t-1,b)$에 따라 연산되는 순간 주파수 변화량을 이용하여 산출되는 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 9

제8항에 있어서,

상기 순간 주파수 변화량은,

f_{max}가 변화 제한 폭일 때, $|f(t,b)-f(t-1,b)|>f_{max}$이면, 0의 값을 갖는 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 10

제5항에 있어서,

상기 피크 주파수 변화량의 평균을 산출하는 단계는,

각 밴드마다 평균 에너지를 산출하는 단계; 및

상기 산출된 평균 에너지가 상기 신호의 전체 평균 에너지의 일정 비율 이하인 밴드의 순간 주파수 변화량을 0으로 계산하는 단계를 포함하는 것을 특징으로 하는, 음성과 음악을 구분하는 방법.

청구항 11

제 1 항 내지 제 10 항 중 어느 한 항에 따른 음성과 음악을 구분하는 방법을 컴퓨터 시스템에서 실행하기 위한 프로그램이 기록된 컴퓨터 시스템이 판독할 수 있는 기록매체.

청구항 12

음성과 음악이 혼합된 신호를 복수의 밴드로 구분하고, 상기 각 밴드마다 일정 구간 동안의 피크 주파수 변화량의 평균을 산출하는 하모닉스 변화 추적부; 및

상기 피크 주파수 변화량이 임계값 이상인 밴드의 신호를 음성으로 분류하고, 상기 피크 주파수 변화량이 임계값 미만인 밴드의 신호를 음악으로 분류하는 과정을 각 밴드별로 수행하는 음성 음악 구별부

를 포함하는, 음성과 음악을 구분하는 장치.

청구항 13

제12항에 있어서,

상기 복수의 밴드는,

주파수 축 상에서 125Hz 간격으로 분할된 주파수 대역인 것을 특징으로 하는, 음성과 음악을 구분하는 장치.

명 세 서

발명의 상세한 설명

기 술 분 야

[0001] 본 발명은 음성인식 시스템의 전처리 과정에 관한 것으로, 특히, 음성과 음악을 구분하는 방법 및 장치에 관한

것이다.

배 경 기 술

[0002] 개인용 컴퓨터의 고성능화, 대용량 저장 장치의 보편화 및 월드와이드웹(World Wide Web; WWW)으로 대변되는 컴퓨터 네트워크의 발전에 따라 디지털로 표현되는 멀티미디어 정보의 생성, 전송, 가공이 매우 용이해졌다. 빠르게 진보하는 정보환경 속에서 엄청난 속도로 증가하는 멀티미디어 정보 중에서 사용자가 필요로 하는 내용의 정보를 찾기 위해 기존의 기반 검색은 효과적이지 않다. 따라서, 사용자가 원하는 정보를 내용에 기반하여 검색할 수 있는 방법이 요구되고 있다.

[0003] 최근 각종 응용에서 멀티미디어 데이터 중에서 오디오 신호의 음성과 음악을 구분하는 연구가 계속 진행되고 있다. 특히, 음성 인식 시스템의 응용 분야가 넓어지면서, 실제 생활 환경에서도 좋은 성능을 얻기 위한 전처리 방법이 많은 관심을 받고 있다. 음성 인식을 위한 전처리 응용 분야도 다시 세분화 되면서 방송과 같은 음악 환경에서 음성을 음악으로부터 구분해 내는 방법에 대한 연구가 진행되고 있다.

[0004] 기존의 SMD (Speech and Music Discrimination) 방법을 살펴보면 음악의 주요 특성이라고 생각될 수 있는 시간에 따라 변하는 리듬을 이용하여 음성과 음악을 구분하는 방법들이 제안되었다. 이러한 방법들은 대체로 음악은 음성의 변화에 비해 상대적으로 느리며 비교적 일정한 간격으로 변한다는 원리를 사용하였기 때문에, 음악의 종류에 따라 템포가 빨라지거나 사용하는 악기가 변화하면 그 성능이 크게 변할 수밖에 없다.

[0005] 2007년에 발표된 논문 "Speech/music discrimination for robust speech recognition in robots" (M. Y. Choi, H. J. Song, and H. S. Kim, *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 118-121)에서는 일정한 프레임 사이의 켑스트럼 거리의 최소값의 평균 (Mean of Minimum Cepstral Distance; MMCD)이 작으면 음성으로 분류하고 크면 음악으로 구분하였다. 또한, 스펙트럼 플럭스 (Spectral Flux)의 경우 프레임 사이의 스펙트럼 에너지 차이를 구하여 음성과 음악을 구분하기도 하였다. 이러한 방법들은 비교적 좋은 성능을 보였지만, 빠른 응답을 얻지 못하는 단점을 갖고 있다.

발명의 내용

해결 하고자하는 과제

[0006] 따라서, 본 발명이 이루고자 하는 첫 번째 기술적 과제는 빠른 응답을 유지하면서 높은 성능을 구현할 수 있는 음성과 음악을 구분하는 방법을 제공하는 데 있다.

[0007] 본 발명이 이루고자 하는 두 번째 기술적 과제는 빠른 응답을 유지하면서 높은 성능을 구현할 수 있는 음성과 음악을 구분하는 장치를 제공하는 데 있다.

과제 해결수단

[0008] 상기의 첫 번째 기술적 과제를 이루기 위하여, 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법은 입력되는 신호에서 일정 구간 마다 피크 주파수 변화량의 평균을 산출하는 단계; 및 상기 피크 주파수 변화량이 임계값 이상이면 상기 신호를 음성으로 분류하고, 상기 피크 주파수 변화량이 임계값 미만이면 상기 신호를 음악으로 분류하는 단계를 포함한다.

[0009] 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법은 상기 분류된 결과에 따라, 상기 신호로부터 음성으로부터 분류된 신호를 추출하는 단계를 더 포함할 수 있다.

[0010] 바람직하게는, 상기 피크 주파수 변화량은 $d(t,b)$는 시간 t에서 밴드 b의 주파수 변화량이며, $f(t,b)$는 시간 t에 밴드 b의 피크 주파수일 때, 수학식 $d(t,b)=f(t,b)-f(t-1,b)$에 따라 연산되는 순간 주파수 변화량을 이용하여 산출될 수 있다. 여기서, 상기 순간 주파수 변화량은 f_{max}가 변화 제한 폭일 때, $|f(t,b)-f(t-1,b)|>f_{max}$이면, 0의 값을 갖도록 할 수 있다.

[0011] 상기의 첫 번째 기술적 과제를 이루기 위하여, 본 발명의 다른 실시 예에 따른 음성과 음악을 구분하는 방법은 음성과 음악이 혼합된 신호를 복수의 밴드로 구분하는 단계; 상기 각 밴드마다 일정 구간 동안의 피크 주파수 변화량의 평균을 산출하는 단계; 및 상기 피크 주파수 변화량이 임계값 이상인 밴드의 신호를 음성으로 분류하

고, 상기 피크 주파수 변화량이 임계값 미만인 밴드의 신호를 음악으로 분류하는 단계를 포함한다.

[0012] 바람직하게는, 상기 복수의 밴드는 주파수 축 상에서 125Hz 간격으로 분할된 주파수 대역일 수 있다.

[0013] 바람직하게는, 상기 복수의 밴드는 270Hz부터 3010Hz 사이의 주파수 대역을 일정 간격으로 분할한 주파수 대역일 수 있다.

[0014] 바람직하게는, 상기 피크 주파수 변화량은 $d(t,b)$는 시간 t에서 밴드 b의 주파수 변화량이며, $f(t,b)$는 시간 t에 밴드 b의 피크 주파수일 때, 수학식 $d(t,b)=f(t,b)-f(t-1,b)$에 따라 연산되는 순간 주파수 변화량을 이용하여 산출될 수 있다.

[0015] 여기서, 상기 순간 주파수 변화량은 f_{max}가 변화 제한 폭일 때, $|f(t,b)-f(t-1,b)|>f_{max}$이면, 0의 값을 갖도록 할 수 있다.

[0016] 바람직하게는, 상기 피크 주파수 변화량의 평균을 산출하는 단계에서, 각 밴드마다 평균 에너지를 산출하고, 상기 산출된 평균 에너지가 상기 신호의 전체 평균 에너지의 일정 비율 이하인 밴드의 순간 주파수 변화량을 0으로 계산할 수도 있다.

[0017] 상기의 두 번째 기술적 과제를 이루기 위하여, 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 장치는 음성과 음악이 혼합된 신호를 복수의 밴드로 구분하고, 상기 각 밴드마다 일정 구간 동안의 피크 주파수 변화량의 평균을 산출하는 하모닉스 변화 추적부; 및 상기 피크 주파수 변화량이 임계값 이상인 밴드의 신호를 음성으로 분류하고, 상기 피크 주파수 변화량이 임계값 미만인 밴드의 신호를 음악으로 분류하는 과정을 각 밴드별로 수행하는 음성 음악 구별부를 포함한다.

[0018] 바람직하게는, 상기 복수의 밴드는 주파수 축 상에서 125Hz 간격으로 분할된 주파수 대역일 수 있다.

효 과

[0019] 본 발명에 의하면, 음성과 음악을 구분함에 있어서 빠른 응답을 유지하면서도 높은 성능을 구현할 수 있는 효과가 있다.

발명의 실시를 위한 구체적인 내용

[0020] 이하에서는 도면을 참조하여 본 발명의 바람직한 실시 예를 설명하기로 한다. 그러나, 다음에 예시하는 본 발명의 실시 예는 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 다음에 상술하는 실시 예에 한정되는 것은 아니다.

[0021] 주파수 분석을 통해 음성과 음악의 특성을 살펴보면, 대부분 악기는 특정 주파수 소리를 지속적으로 내도록 고안되어 있다는 것을 알 수 있고, 음성은 조음 현상에 의해서 점차적인 주파수 변화가 발생하는 것을 알 수 있다.

[0022] 본 발명에서는 이러한 음성과 음악이 갖고 있는 주파수 변화 특성을 이용하여 음성과 음악을 구별한다. 즉, 음성과 음악을 구분하기 위한 특성 값으로서 주파수 변화율을 사용한다.

[0023] 이하에서는 음성과 음악의 스펙트로그램 상에서의 특성 차이를 비교하여 주파수 도메인에서 음성과 음악을 구분하기 위한 특성을 분석한다.

[0024] 도 1a 및 1b는 단모음 '아'(aa)와 '이'(iy)의 스펙트로그램에 각 밴드의 최고 에너지 값을 실선으로 표시한 그래프이다.

[0025] 각 밴드에서 최고 에너지 값을 스펙트럴 피크 (spectral peak)라고 한다. 도 1a 및 1b의 '아'와 '이'는 어느 정도 일정한 주파수 간격으로 스펙트럴 피크가 발생한다.

[0026] 도 1a의 '아'의 경우, 음소가 지속되는 중간 부분은 상대적으로 피크의 변화가 작고 음소가 시작되는 부분과 특히 끝나는 부분에서는 변화가 크게 발생하는 것을 볼 수 있다. 이 경우는 화자가 음소의 중간 부분은 일정한 주파수의 소리를 냈지만 발성하는 전후 과정에서 조음 기관을 움직여 다른 주파수로 전이하는 발성을 했기 때문이다.

[0027] 도 1b의 '이'의 경우에는 단모음이지만 음소가 발성되는 전 과정에서 피크의 변화가 발생한다. 많은 경우 화자

가 의도적으로 해당 주파수를 유지하지 않으면 단모음이 발성되는 동안에도 높은 주파수 또는 낮은 주파수로 점차적으로 움직여가는 주파수 변화가 쉽게 발생한다. 이와 같이 각각의 음소를 발성하기 위해 발성 기관을 움직이는 순간마다, 밴드 별 에너지가 최고인 지점의 주파수가 점차적으로 변화하는 것을 알 수 있다. 즉, 음성은 하모닉스의 변화가 매 순간 점차적으로 발생한다. 그 이유는 음성은 성도를 통해서 조음 기관이 변형될 때 주파수 변화가 발생하는데, 사람은 조음 기관을 움직이며 소리를 발성하므로 소리를 변화시킬 때마다 주파수 대역이 연속적으로 변화하기 때문이다. 이는 특히 악기가 기계적으로 단절된 주파수의 소리를 순간적으로 내는 것과는 구별된다.

[0028] 도 2는 연속 문장의 스펙트로그램에 밴드별 피크를 표시한 그래프이다.

[0029] 도 2에 사용된 연속 문장은 "She had your dark suit in greasy wash water all year" 이다. 연속 문장에서는 다양한 주파수 변화를 볼 수 있다. 즉, 현재 음소에서 다음 음소로 변화해 가면서 주파수 변화가 점차적으로 발생하고, 이미 살펴 본 단모음의 경우와 같이 음소가 시작하거나 끝나는 부분에서도 주파수 변화가 발생한다.

[0030] 도 3a는 기타 연주곡의 스펙트로그램에 스펙트럴 피크를 표시한 그래프이다.

[0031] 이 경우, 밴드 별 스펙트럴 피크가 일정하게 유지되다가 순간적으로 변화되는 것을 볼 수 있다. 일반적으로 악기들은 일정한 주파수의 소리를 내도록 고안되어 있으며, 연주된 소리가 특정 주파수를 일정한 시간 동안 지속하다가 새로운 음이 발생할 때 다른 주파수로 순간 변화하게 된다. 즉, 음악에서는 음성과 같이 점진적 주파수 변화는 발생하지 않는다.

[0032] 도 3b는 비교적 빠르고 강한음의 드럼 연주곡의 스펙트로그램에 스펙트럴 피크를 표시한 그래프이다.

[0033] 드럼과 같은 타악기의 경우 주파수 변화 현상은 거의 없다는 것을 알 수 있다.

[0034] 이러한 분석을 통해 음악은 특정 주파수에서 시작하여 일정한 시간 동안 같은 주파수를 유지하는 반면, 음성의 경우 발성하는 매 순간마다 지속적으로 조금씩 변한다는 사실을 확인할 수 있다. 이하에서는 이러한 특성을 이용하여 음성과 음악을 구분하는 방법을 설명한다.

[0035] 본 발명은 주파수 변화율을 이용하여 음성과 음악을 구분하는 STR (Spectral Transition Rate) 특징 기반 SMD 알고리즘을 제공한다.

[0036] 피크 주파수의 변화량은 아래 수학식 1과 같이 계산한다.

수학식 1

[0037]
$$d(t,b)=\begin{cases}0, & \text{if } |f(t,b)-f(t-1,b)|>f_{\max}\\ f(t,b)-f(t-1,b), & otherwise\end{cases}$$

[0038] 여기서 $d(t,b)$는 시간 t에서 밴드 b의 주파수 변화량이며, $f(t,b)$는 시간 t에 밴드 b의 피크 주파수이고, f_{max}는 변화 제한 폭이다. 순간 주파수 변화량은 시간 t와 t-1 사이의 주파수 변화량이다. 이때, 주파수 변화량이 f_{max} 이상이면 새로운 소리가 다른 주파수에 발생한 것으로 간주하고 $d(t,b)$를 0으로 계산하는 것이 바람직하다. 또한, 밴드의 평균 에너지가 전체의 평균 에너지에 비해 일정한 비율 이하인 경우에도 순간 변화량 $d(t,b)$를 0으로 계산하여 상대적으로 낮은 에너지를 갖는 주파수 대역의 변화를 제외 하도록 할 수도 있다.

[0039] 수학식 2는 일정한 기간 동안의 피크 주파수 변화량이다.

수학식 2

[0040]
$$STR(t)=\sum_{b=start}^{end}\left(\sum_{\tau=0}^{T}d(t+\tau,b)\right)^2$$

[0041] 여기서 $STR(t)$은 입력된 소리가 T 시간 동안 점차적으로 높은 주파수 대역으로 또는 낮은 주파수 대역으로 움직여가는 변화량이다. 순간 주파수 변화량인 $d(t,b)$를 각 밴드 별로 T까지 더한 후 그 제곱 값을 유효 밴드까지 합한 값이다. 여기서 $start$부터 end까지의 대역이 주파수 변화 현상을 관찰하는 유효 밴드 대역이다. 음성과 음악을 구분하는 특징 값을 만들기 위해서, 하모닉스의 변화를 일정한 밴드로 나누어 그 추이를 추적한다. 예를 들어, 평균 270Hz부터 3,010Hz 사이에 포만트 (formant) 주파수와 평균 피치를 고려하여, 주파수 대역을 125Hz

간격으로 나누어 스펙트럴 피크 값을 추적할 수 있다.

[0042] 음성의 $STR(t)$은 소리가 변화하는 순간에 그 크기가 크게 나타나지만 소리가 유지되는 구간에서는 작게 나타날 수 있다. 이런 경우를 보상하기 위해 일정한 구간의 평균값을 사용한다. 수학식 3은 SMD 알고리즘에서 사용하는 최종 STR 값을 구하기 위한 식이다.

수학식 3

[0043] $$MSTR(t)=\left(\sum_{t=0}^{W} STR(t)\right)/W$$

[0044] 여기서 W는 평균을 구하는 윈도우의 크기이다.

[0045] 도 4a 및 4b는 각각 도 1a 및 1b의 단모음 '아'(aa)와 '이'(iy)에 대한 STR 값을 도시한 그래프이다.

[0046] 단모음 '아'의 경우에는 주파수 변화 현상이 발성 시작 부분과 끝 부분에서 크게 발생하므로 시작하는 시점과 끝나는 시점에서 STR 값이 크게 나타난다. 단모음 '이'의 경우는 발성하는 과정에서 점차적으로 낮은 주파수로 변화하는 경우다. 주파수 변화 현상은 발성 초기에 크게 나타나므로 STR 값도 전반부에서 크게 나타난다.

[0047] 도 5는 연속 문장의 STR 값을 도시한 그래프이다.

[0048] 연속 문장에 대한 음성의 경우 다양한 주파수 변이 현상으로 인해 STR 값이 매 순간 크게 나타난다.

[0049] 도 6a 및 6b는 각각 도 3a의 기타 연주곡 및 3b의 드럼 연주곡의 STR 값을 도시한 그래프이다.

[0050] 기타와 드럼 연주곡 모두 음성과 비교하면 주파수 변화 현상이 적게 발생하므로 STR 값이 음성에 비해 상대적으로 작게 나타난다.

[0051] 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법에서는 테스트 데이터의 STR 값이 임계값 보다 이상이면 음성으로 분류하고, 작으면 음악으로 분류한다.

[0052] 도 7은 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법의 흐름도이다.

[0053] 먼저, 입력되는 신호를 복수의 밴드로 구분한다(S710). 이때, 입력되는 신호는 음성과 음악이 혼합된 신호일 수 있다.

[0054] 다음, 각 밴드마다 일정 구간 동안의 피크 주파수 변화량의 평균을 산출한다(S720). 이 과정(S720)은 입력 신호가 종료될 때까지 계속될 수 있으므로, 피크 주파수 변화량을 추적하는 과정이다.

[0055] 다음, 산출되는 피크 주파수 변화량이 미리 정해진 임계값 이상인 밴드의 신호를 음성으로 분류한다(S730, S740). 한편, 산출되는 피크 주파수 변화량이 임계값 미만인 밴드의 신호는 음악으로 분류한다(S730, S745). 여기서, 비교 기준이 되는 임계값은 반복된 테스트 및 학습의 결과로부터 얻어진 학습 데이터를 이용하여 음성과 음악의 STR 분포를 구하고 음성과 음악의 구분 오차가 최소인 STR 값으로 정해질 수 있다.

[0056] 마지막으로, 신호의 입력이 종료되었는지 판단하고, 종료되지 않았으면, 피크 주파수 변화량을 추적하는 과정(S720)부터 시작하여 위의 과정을 반복한다.

[0057] 도 8은 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 장치의 블록도이다.

[0058] 하모닉스 변화 추적부(810)는 입력되는 신호를 복수의 밴드로 구분하고, 상기 각 밴드마다 일정 구간 동안의 피크 주파수 변화량의 평균을 산출한다. 여기서, 입력되는 신호는 음성과 음악이 혼합된 신호일 수 있다. 하모닉스 변화 추적부(810)는 피크 주파수 변화량을 주기적으로 산출하거나 일정한 패턴으로 주어진 시간에 피크 주파수 변화량을 산출할 수도 있다.

[0059] 음성 음악 구별부(820)는 하모닉스 변화 추적부(810)에서 산출되는 피크 주파수 변화량이 임계값 이상인 밴드의 신호를 음성으로 분류하고, 하모닉스 변화 추적부(810)에서 산출되는 피크 주파수 변화량이 임계값 미만인 밴드의 신호를 음악으로 분류한다.

[0060] 음성 음악 구별부(820)의 음성과 음악의 구분에 따라 음성 신호와 음악 신호를 생성할 수 있다. 음성 추출부(830)와 음악 추출부(840)는 당업자의 필요에 따라 추가 또는 생략될 수 있다. 음성 추출부(830)는 음성 음악 구별부(820)에서 분류된 밴드의 신호들을 이용하여 음성 신호를 생성한다. 또한, 음악 추출부(840)는 음성 음악

구별부(820)에서 분류된 밴드의 신호들을 이용하여 음악 신호를 생성한다.

[0061] 이하에서는 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 장치의 성능을 평가한다.

[0062] 실험을 위한 음성 데이터로 TIMIT 데이터 베이스를 사용하였으며, 음악 데이터로는 여러 장르의 음악을 사용하였다. 소리는 16,000Hz 모노로 녹음하였으며, 푸리에 변환 (Fourier Transform) 윈도우의 크기는 128ms로, 10ms 간격으로 전진하며 SMD를 수행하였다. STR 계산에 사용된 유효 음성 주파수 대역으로는 125Hz에서 2,000Hz까지로 하였다.

[0063] 수학식 1에서 f_{max}는 70.3Hz으로 하였으며, 수학식 2에서 주파수 변화 계산을 위한 시간 T는 실험을 통해 최적한 값인 200ms을 사용하였다. STR의 평균 윈도우 W에 따라서 임계값을 구하여 실험하였다. 또한, STR의 빠른 응답에 대한 성능을 알아보기 위해, 기존 MMCD의 켑스트럼 거리 (cepstral distance) 계산을 위한 시간 보다 작은 T=150ms, W=150ms으로도 실험을 진행하였다.

[0064] 도 9는 평균 윈도우 W가 250ms인 경우, MMCD와 본 발명의 일 실시 예에 따른 STR의 음성과 음악의 장르별 SMD 성능을 비교한 그래프이다.

[0065] MMCD는 음악의 종류에 따라서 성능의 변화가 크다. 반면, 본 발명의 일 실시 예에 따른 STR은 음악의 종류에 무관하게 상대적으로 안정된 성능을 보인다.

[0066] 도 10은 STR과 MMCD의 평균 SMD 성능을 도시한 그래프이다.

[0067] 본 발명의 일 실시 예에 따라 STR을 이용한 SMD는 평균 윈도우 W의 크기가 큰 경우 MMCD와 유사한 성능을 나타낸다. 윈도우 W가 크기가 작은 경우, 즉 빠른 응답에서도 MMCD에 비해 높은 성능을 보이는 것을 확인할 수 있다. 본 발명의 일 실시 예에 따른 알고리즘을 기존의 알고리즘과 비교할 때, 상대적으로 빠른 응답에서 좋은 성능을 보인다는 것이 확인된다.

[0068] 본 발명은 소프트웨어를 통해 실행될 수 있다. 바람직하게는, 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체에 기록하여 제공할 수 있다. 소프트웨어로 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 프로그램 또는 코드 세그먼트들은 프로세서 판독 가능 매체에 저장되거나 전송 매체 또는 통신망에서 반송파와 결합된 컴퓨터 데이터 신호에 의하여 전송될 수 있다.

[0069] 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, DVD±ROM, DVD-RAM, 자기 테이프, 플로피 디스크, 하드 디스크(hard disk), 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0070] 본 발명은 도면에 도시된 일 실시 예를 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시 예의 변형이 가능하다는 점을 이해할 것이다. 그리고, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

산업이용 가능성

[0071] 본 발명은 음성과 음악을 구분함에 있어서 빠른 응답을 유지하면서도 높은 성능을 구현할 수 있는 음성과 음악을 구분하는 방법 및 장치에 관한 것으로, 녹음 장치, 음향 편집 장치, 데이터 검색 방법, 음성인식 시스템의 전처리 장치 등에 적용될 수 있다.

도면의 간단한 설명

[0072] 도 1a 및 1b는 단모음 신호의 스펙트로그램에 각 밴드의 최고 에너지 값을 실선으로 표시한 그래프이다.

[0073] 도 2는 연속 문장의 스펙트로그램에 밴드별 피크를 표시한 그래프이다.

[0074] 도 3a는 기타 연주곡의 스펙트로그램에 스펙트럴 피크를 표시한 그래프이다.

[0075] 도 3b는 비교적 빠르고 강한음의 드럼 연주곡의 스펙트로그램에 스펙트럴 피크를 표시한 그래프이다.

[0076] 도 4a 및 4b는 각각 도 1a 및 1b의 단모음에 대한 STR 값을 도시한 그래프이다.

[0077] 도 5는 연속 문장의 STR 값을 도시한 그래프이다.

[0078] 도 6a 및 6b는 각각 도 3a의 기타 연주곡 및 3b의 드럼 연주곡의 STR 값을 도시한 그래프이다.

[0079] 도 7은 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 방법의 흐름도이다.

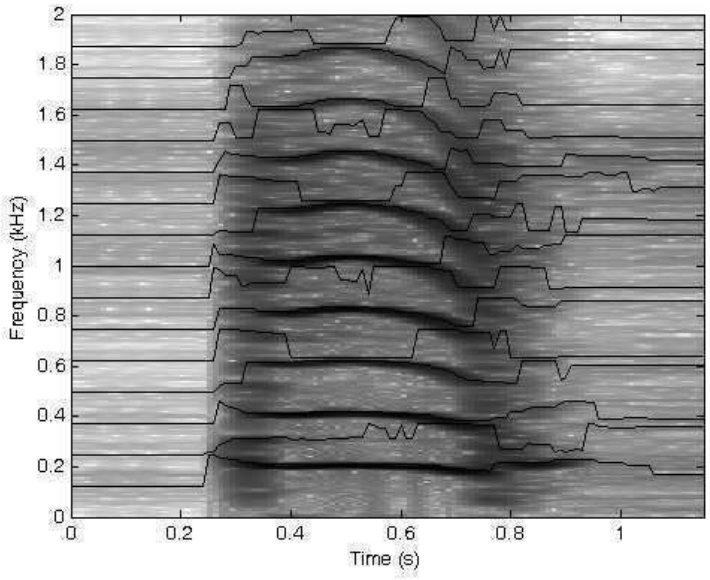
[0080] 도 8은 본 발명의 일 실시 예에 따른 음성과 음악을 구분하는 장치의 블록도이다.

[0081] 도 9는 평균 윈도우 폭가 250ms인 경우, MMCD와 본 발명의 일 실시 예에 따른 STR의 음성과 음악의 장르별 SMD 성능을 비교한 그래프이다.

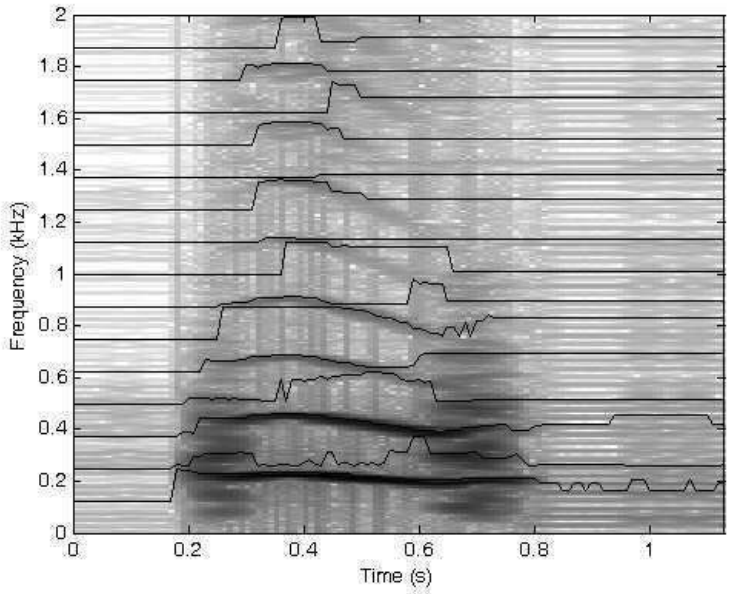
[0082] 도 10은 STR과 MMCD의 평균 SMD 성능을 도시한 그래프이다.

도면

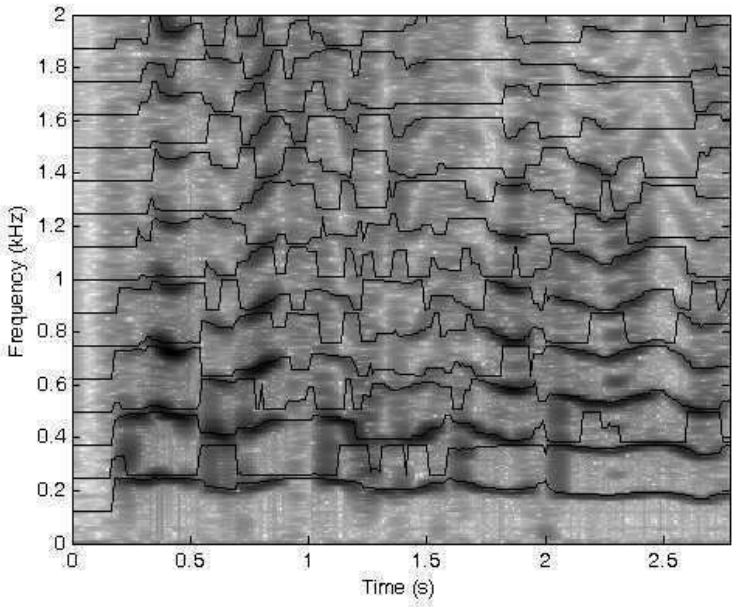
도면1a



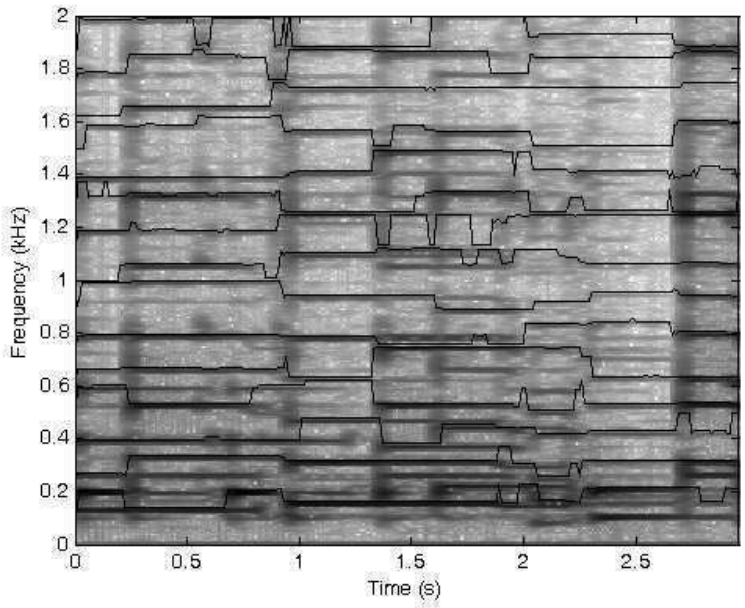
도면1b


Frequency (kHz)
Time (s)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

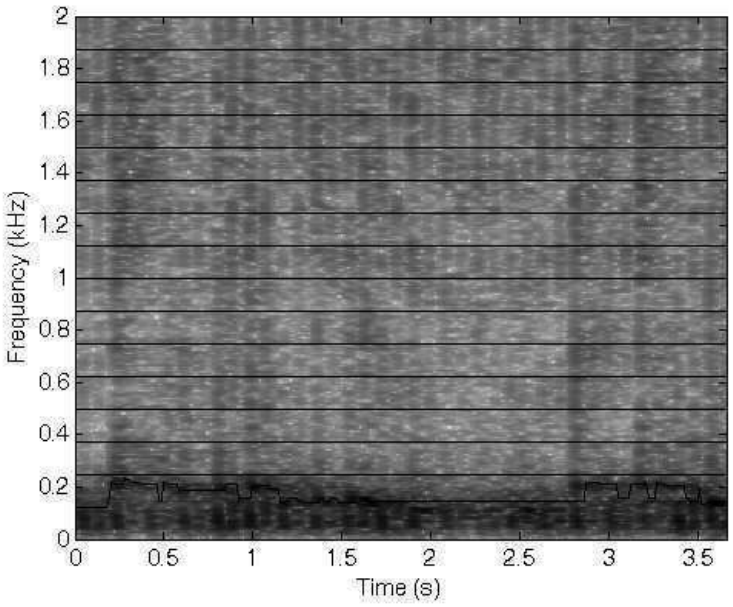
도면2


Frequency (kHz)
Time (s)
0
0.5
1
1.5
2
2.5
0.2
0.4
0.6
0.8
1.2
1.4
1.6
1.8

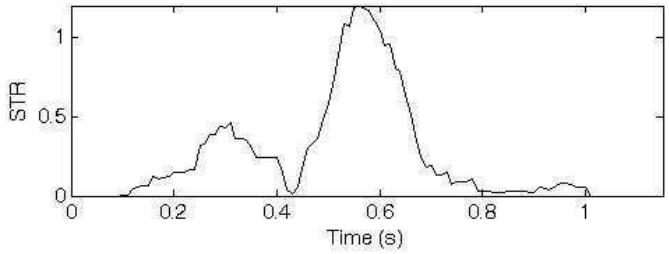
도면3a


Frequency (kHz)
Time (s)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0.5
1.5
2.5

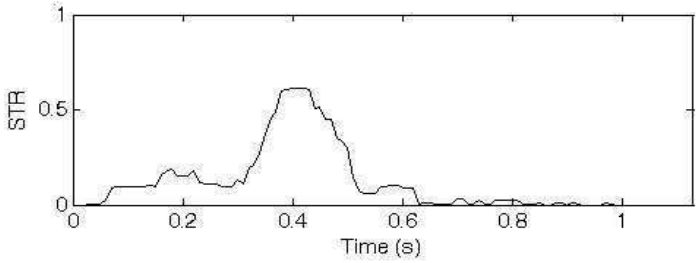
도면3b


Frequency (kHz)
Time (s)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0.5
1.5
2.5
3
3.5

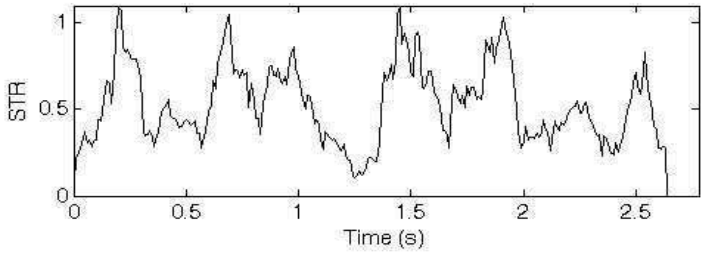
도면4a



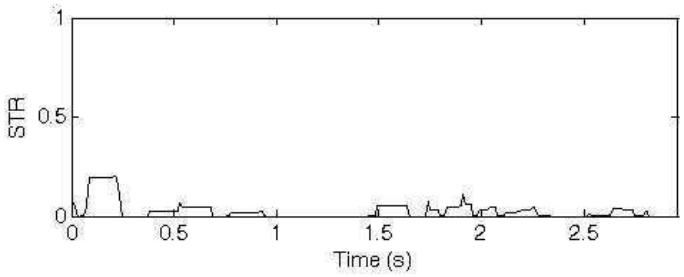
도면4b



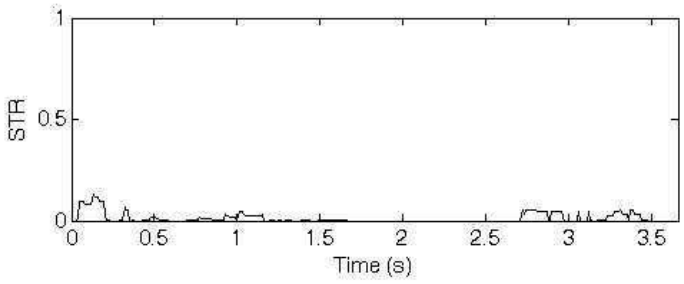
도면5



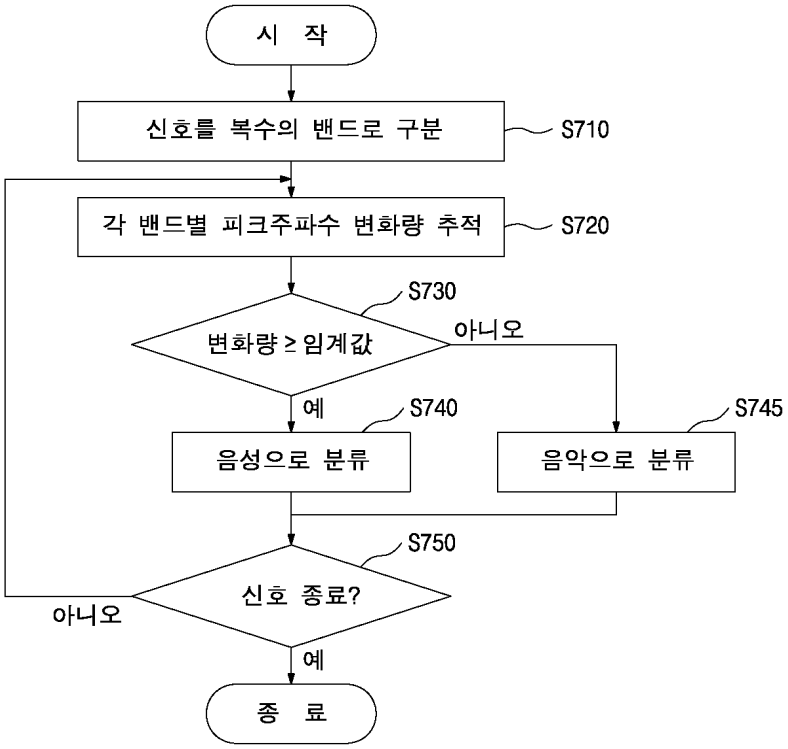
도면6a



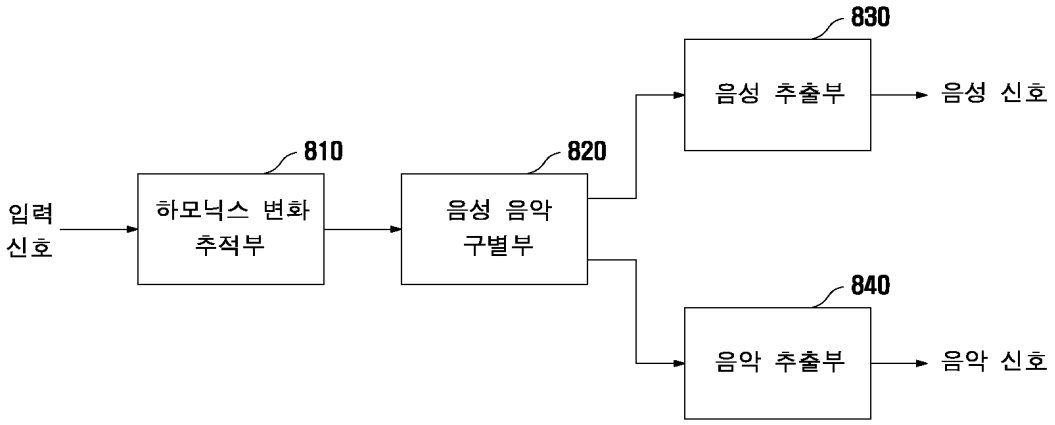
도면6b


STR
1
0.5
0
0
0.5
1
1.5
2
2.5
3
3.5
Time (s)

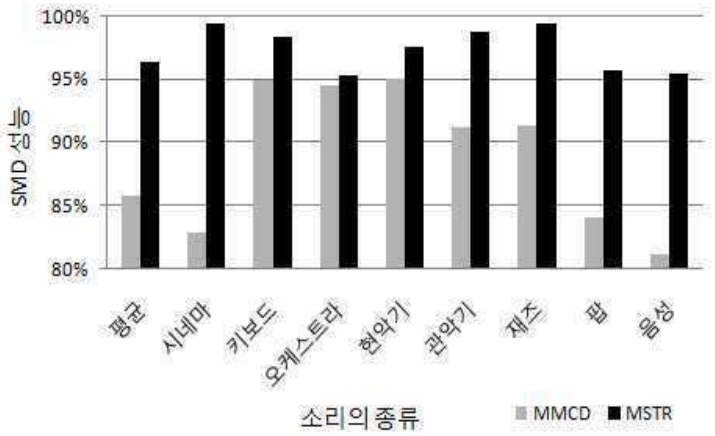
도면7


시 작
신호를 복수의 밴드로 구분
S710
각 밴드별 피크주파수 변화량 추적
S720
S730
변화량 ≥ 임계값
아니오
예
S740
음성으로 분류
S745
음악으로 분류
S750
신호 종료?
아니오
예
종 료

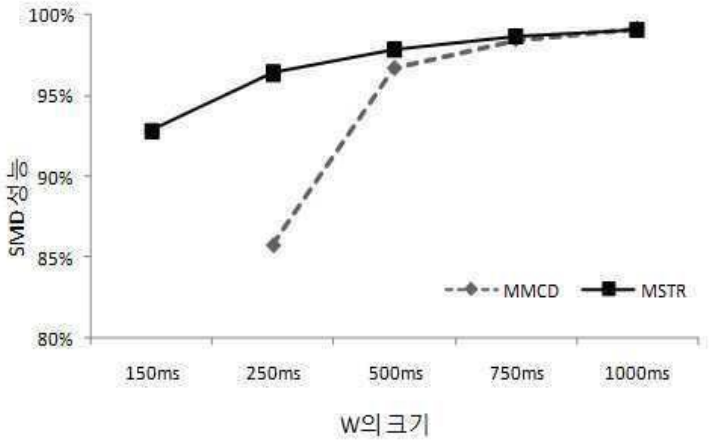
도면8


830
음성 추출부
음성 신호
810
820
입력
신호
하모닉스 변화
추적부
음성 음악
구별부
840
음악 추출부
음악 신호

도면9


100%
95%
90%
85%
80%
SMD 성능
평균
시네마
키보드
오케스트라
현악기
관악기
재즈
팝
음성
소리의 종류
MMCD
MSTR

도면10


100%
95%
90%
85%
80%
SMD 성능
MMCD
MSTR
150ms
250ms
500ms
750ms
1000ms
W의 크기