[5A-75]

적응 에너지 문턱 값을 적용한 DTW 처리시간 단축에 관한 연구

서지호, 최남대, 배명진
숭실대학교
jihoseo@lycos.co.kr

A Study on the Improvement of Processing Time of DTW Using the Adaptive Energy Threshold

Jiho Seo, Namdae Choi, Myungjin Bae
Soongsil University

요 약

화자인식은 음성의 특징에 의해 화자의 신원을 확인하는 기술이다. 이러한 기술은 화자식별과 화자검증으로 분류된다. 첫 번째 방법은 이미 등록된 그룹으로부터 화자 구분과 단어를 인식하고, 두 번째 방법은 식별을 주장하는 화자를 검증한다. 음성으로부터 화자 정보를 추출해서 개별적으로 신원을 확인하는 이 방법은 전화 공중망을 통해 서비스되는 가장 효과적인 기술중의 하나가 될 것이다. 그러나, 실제 적용을 위해서는 다음과 같은 몇 가지 문제가 해결되어야 한다. 첫째는 오직 미리 등록된 고객을 위해 식별이 확인 되지 않은 사칭자를 거부하는 안전에 관한 사항이다. 둘째는, 시간이 지남에 따라 음성의 특징이 변한다는 사실이다. 이러한, 사실은 인식률의 심한 저하와 말하는 문장의 수의 증가에 따라 오류를 유발한다는 사실이다. 마지막으로 화자들 중에서 일반적인 특성이 잘못된 인식 결과를 유발한다는 사실이다. 묵음 구간이 음성 내에 포함되어 식별률이 감소한다. 이 논문에서는 식별 알고리즘이 수행되기 전에 묵음 구간을 제거할 뿐만 아니라 주변 잡음에 대한 에너지 여기를 주어 높은 에너지 레벨의 음성구간에 대해서 알고리즘을 수행한다.

음성의 시작점과 끝점의 에너지 신호 추출에 의해 DTW(Dynamic Time Warping) 알고리즘을 수행한다. 실험결과 제안된 방법이 기존의 방법과 비교해서 인식율의 향상을 얻을 수 있었다.

I. 서론

현대가 정보화 사회로 급속히 진행됨에 따라 대규모의 데이타베이스에 등록되어있는 개인이나 단체의 수많은 정보의 접근, 갱신, 수정이 빈번해지고 있다. 따라서 이에 따른 정보의 보안 문제가 심각해지고, 특정 지역의 출입 통제를 위한 보안 시스템이나 특정시스템을 사용할 때 사용자의 신분에 대한 확인 수단이 필수적이다. 그러나 종래의 개인 신분 확인 수단인 도장, 신분증, 카드 등은 도난, 분실, 위조 등의 위험을 수반한다. 또한 전화나 통신망을 이용해서 정보 접근을 할 경우에 개인 확인이 더욱 어려워진다. 이에 반해 음성을 이용한 화자 식별 시스템은 음성에 포함되어 있는 개개인 마다 화자정보를 추출하여 개인을 확인하는 기술로서 사칭자에 대한 처리, 처리시간, 원격자 확인 등 시스템 사용의 간편하고, 여러 가지 측면에서 가장 효과적인 기술이고 응용분야도 다양하다는 장점이 있다[1][2]. 그러나 기존의 DTW 를 이용한 화자 식별 시스템에서는 많은 화자를 처리할 경우 처리량이 증가하여 인식결과를 얻기 위해서는 많은 시간이 소요된다는 단점을 수반하고 사칭자의 경우에 잘못된 인식을 수행한다는 단점을 수반하게 된다. 화자 식별율은 화자수에 비례하여 정확도가 감소하므로 화자확인에 비해 어려우며, 실제 응용에서는 비협조적인 화자를 대상으로 하는 경우가 많으므로 화자의 정확한 판단에 어려움이 있다. 기존의 방법은 음성의 시작점과 끝점만을 검출하여 인식 알고리즘을 수행하는 방법을 택하고 있다. 이렇게 되면 비교할 음성데이터 중간에 포함 되어 있는 묵음구간이 인식률을 저하시키는 요인으로 작용하게 된다. 본 논문에서는 이를 개선하기 위하여 인식알고리즘을 수행하기 전에 묵음 구간을 제거함으로써 인식률을 개선하는 방법을 제안하였다.

II. 화자식별 시스템

화자인식은 인식대상에 따라 화자식별(speaker identification)과 화자확인(speaker verification)으로 나눌 수 있다. 화자 식별은 입력된 미지의 음성이 이미 등록된 여러 명의 화자중 어떤 화자에 의해 발성된 음성인지를 판정하는 것을 말하고, 화자확인 방법은 신분 확인 및 음성인식 기술과 조합하여 본인 여부를 가려내는 것이다. 그리고 화자인식은 인식 방법에 따라서 다음과 같이 4 가지로 구분할 수가 있다. 그 중 첫 번째는 입력패턴을 미리 정해진 기준 패턴(reference pattern)과 비교하여 최적화된 유사성을 판단하는 방법으로 패턴정합법(Pattern Matching)인 동적정합법(dynamic time warping, DTW), 각 화자별로 신경 회로망을 구성하고 화자간의 변별력을 갖도록 학습을 수행하도록하여 인식하는 신경회로망이 있다[9]. 그러나 이 방법은 새로운 화자의 추가 시 다시 학습시켜야 한다는