

C# 언어를 이용한 음성 분석 시스템의 구현

김윤중(국립 한밭대학교 정보통신공학부 교수)
오지영(국립 한밭대학교 컴퓨터공학과 석사과정)
오신영(국립 한밭대학교 컴퓨터공학과 석사과정)

Dept. of Computer Engineering Hanbat national University
대전광역시 유성구 덕명동 산16-1 국립 한밭대학교
Tel +82-42-821-1143

Keyword : c#, 음성인식, cepstral 계수

요 약

c#은 MS사에서 개발하여 발표된 언어로서 다른 언어들(C++, Visual Basic, Java)의 장점을 채택하고 단점을 보완하여 개발되었으며, MSIL이라는 중간언어 방식을 가지고 있다. 또한 다중플랫폼을 지원한다.

음성인식 기술은 일부 실용화되는 단계에 있다. 웹 환경에서도 웹브라우저에 적용하고자 하는 요구가 증대되고 있다. 그러므로 웹 환경에서 용이하게 사용될 수 있는 인식 시스템이 요구된다. 즉, c#을 이용하여 MSIL 중간언어의 방식으로 음성인식엔진이 개발되면 Visual Basic, c++, c#등의 언어로 응용프로그램을 개발하는 것이 용이하게 된다. 웹 환경에서는 클라이언트가 음성을 분석하여 결과를 서버로 전송하고 서버에서 인식을 수행하는 방식이 필요하다.

따라서 본 연구에서는 음성 인식 과정 중 음성 분석 과정을 c#으로 구현하고자 한다. 이와 같이 c#으로 구현되면 다음과 같은 장점을 얻게 된다.

- ① 개발환경과 사용환경이 같아야 하는 제약 극복
- ② 다중 플랫폼에서 사용 가능
- ③ 웹 환경에서 클라이언트가 분석하고 결과를 서버에 전송하는 시스템의 개발이 용이
- ③ 여러 가지 언어로 응용 애플리케이션 개발이 용이

최근 음성과 자연언어의 기본적인 성질의 이해에 관한 관심도 높아지고 있고 통신판매, 물류처리, 제품홍보 등의 분야에의 적용이 활발해지고 있다.

이렇듯 많은 분야에서 관심을 끌고 있는 음성인식의 중요 과정인 음성 분석 과정을 본 연구에서는 c# 언어로 구현을 하였다.

- ① c# 언어는 MS사에서 출시한 .NET 플랫폼의 기본언어 중 하나로서 c++과 Visual Basic, Java의 장점을 채택하고 단점을 보완하여 만들어진 언어이다.
- ② 또한 .NET이 분산된 데스크 탑이나 모바일 애플리케이션을 만드는 새로운 방법으로 대두 되었다.

- ③ c# 언어는 공통 언어 런타임, 공통 타입 시스템, 공통 언어 명세를 가지고 소스코드를 중간언어(Microsoft Intermediate Language)로 만든다. 이는 언어의 통합을 가져올 수 있고 또한 다중플랫폼(Windows, UNIX외) 지원이 가능하게 해준다.

기존의 프로그램과 c# 언어로 구현한 프로그램의 차이점은 기존의 프로그램은 “지금 항상 실행 될 수 있는 상태인가?” 라는 물음에 플랫폼이 지원되지 않으면 그 답이 확실하지 않았다. 그에 반해 c# 언어로 구현했을 경우 어떠한 플랫폼이라도 지금 실행이 가능하다는 장점이 있다.

음성인식 시스템은 음성신호로부터 음성특성을 분석하는 과정부터 시작된다.

음성인식의 방법 중 패턴 인식적 방법은 먼저 n 개의 샘플링 된 음성 신호($S(n)$)를 분석한다. 테스트 패턴을 규정하기 위하여 입력 신호로 만들어진 특징 측정한다. 음성 신호들을 위해, 특징 측정은 대체로 필터 뱅크 분석기(Filter Bank), LPC 분석, 또는 DFT 분석과 같은 분석 기법으로 진행된다.

같은 클래스의 언어음들에 부합하는 1 또는 더 많은 테스트 패턴들이 그 클래스의 특징들의 패턴 견본을 생성하도록 사용된 패턴 훈련을 한다. 결과로서 생긴 패턴을 참조하는 패턴이 본보기 또는 템플릿이 될 수 있다.

입력되는 테스트 패턴과 참조 패턴과의 유사점을 비교하여 어떤 참조 패턴이 최적의 결과물인가를 결정하도록 사용된다. 일반적으로 훈련이 더 있을 수록, 사실상 어떠한 작업을 위해 시스템의 성능이라도 더 높다

본 연구에서 다루는 음성 분석 시스템은 다음과 같은 단계를 거친다.

- ① 양자화 음성은 신호인 $S(n)$ 신호를 정밀한 영향에 반응이 있는 시스템을 통하여 고주파의 감쇄현상을 줄이기 위해 주파수를 평활화한다.
preemphasizer에서 사용된 디지털 시스템은 전송 조건, 소음 배경들을 평균하기 위하여 또는 심지어 신호 스펙트럼을 평균 위하여 사용된다.
- ② preemphasized 음성 신호는 N 개의 프레임으로 blocking되고 M 개의 샘플링 수만큼 시프트된다. M 은 N 보다 작을수록 좋다. 보통 $M=N/3=1$ 값이 이용된다.
- ③ 이는 너무 짧은 신호는 분석이 어렵기 때문에 신호의 길이를 조절하여 주파수의 특성을 추출하는 것이다.
- ④ 한 Frame 상의 신호는 처음과 끝에서 신호의 불연속성이 존재하므로 처음과 끝 부근의 신호를 작게 만들 필요가 있다. 이를 위하여 Hamming window를 사용한다.
- ⑤ 각 Frame에서 얻어진 p (LPC 분석기의 차수)+1개의 상관값을 이용하여 LPC 파라미터 SET을 LPC 분석기에서 구한 다음 LPC 계수, Cepstral 계수등을 얻어낸다.
- ⑥ 저차 cepstral 계수가 spectrum 전체의 slope에 영향을 주고 고차 계수가 noise에 영향을 주므로 이러한 영향을 제거하기 위하여 tapered window로 cepstral 계수를 weighting 할 필요가 있다.

본 연구에서는 다중플랫폼 지원이 가능한 c#을 이용하여 음성신호로부터 LPC 프로세서를 이용해서 분석하고 cepstral 계수를 추출하는 음성 신호 분석 시스템을 구현하고자 한다.

c# 언어를 이용하여 음성인식 시스템을 구현할 경우 얻을 수 있는 장점은 다음과 같다.

c# 언어의 고유 특징인 다중 플랫폼 지원으로 인해 기존의 프로그램이 가졌던 제약, 즉 개

발 플랫폼과 사용하는 플랫폼이 같아야 하는 제약을 받지 않는다.

또한 중간언어를 사용함으로써 c++, Visual Basic, Java등의 언어에 상관없이 애플리케이션 개발이 가능하다.

음성인식 기술은 매우 활발히 연구되고 일부 실용화되는 단계에 있다. 현재 인터넷 환경이 급속도로 발전함에 따라 웹 관련에서도 웹브라우저에 적용하고자 하는 요구가 증대되고 있다.

클라이언트가 음성 신호를 분석하고 결과를 서버에 전달하고 이것으로 서버에서 인식하여 클라이언트에 되돌려 주는 시스템의 개발이 요구된다. 그러므로 웹 환경에서 음성 신호 분석 과정이 중요하다.

향후 이 연구가 계속되어 웹 프로그램에서 마우스 대신 음성으로서 명령을 내릴 수 있는 시스템을 개발할 예정이다.