아무도 인터넷이 인류에게 앞으로 어떠한 영향을 미칠 것인지 정확히 예측하지 못하고 있다.

휴대 정보 기기는 1980년대 말경에 PDA(Personal Digital Assistance)가 등장한 이래 아직까지는 보급이 크게 확대되지는 못하였으나, 최근 인터넷의 확산과 더불어 향후 5년 이내에 폭발적인 보급 확대가 이루어질 것으로 예상되고 있다. 이러한 이동 단말기들의 보급과 더불어 이동 통신기술의 발전으로 인터넷 정보검색, 파일공유, 메일서비스, 화상회의 서비스와 같은 기존의 인터넷 서비스를 그대로 제공받고자 하는 사용자들의 수요가 최근 급증하고 있는 추세다. 특히 이러한 추세와 더불어 이동 단말기의 사용자들에게 다자간 회의 서비스와 같은 복잡하고 제한된 환경에서의 서비스를 제공하기 위해 활발한 연구가 진행되고 있다.

기존의 회의 시스템에서는 PC와 같은 많은 데이터의 처리능력, 그리고 다양한 입력장치와 출력장치를 보유한 단말기를 통하여 음성, 화상, 데이터를 모두 포함하는 멀티미디어 서비스를 제공하고 있다. 이에 반해 이동 단말기는 소규모 데이터 처리능력, 제한된 입력장치와 출력장치로 인하여 기존 회의 서비스를 제공하기에는 많은 문제점을 가지고 있다. 특히 음성 서비스 위주의 이동 단말기들은 음성입력 이외에 DTMF(Dial Tone Message Frequency) 키 입력과 같이 단순한 입력만 가능하다. 또한 스피커를 통한 음성출력, 그리고 아주 제한적인 디스플레이를 통해서만 데이터의 출력이 가능하다. 뿐만 아니라 이동 단말기들은 기존 유선 망에 접속된 단말기들과는 근본적으로 네트워크를 통하여 데이터를 전송하는 능력에서 현저한 차이를 나타낸다.

이동 단말기들의 이러한 제한적인 기능들을 극복하고자 하는 노력들이 WML, VoiceXML, 음성인식이나 합성과 같은 기술들을 통하여 나타나고 있다. 그러나 이러한 기술들은 이동 단말기를 이용하여 아주 단순하고 기본적인 인터넷 서비스(문자전송, 메일확인, 게시판 등)를 제공하는 정도에 불과하다. 따라서 음성 통화중 파일전송 및 공유, 인터넷 자료검색과 같은 데이터 처리 기능을 가진 다자간 회의 서비스는 이동 단말기들이 제공하기에 어려움이 있다.

본 논문에서는 이러한 다양한 서비스 중에서 제약적인 기능을 가진 환경에서 이동 단말기를 통한 다자간 회의 서비스 제공이 가능한 구조 모델을 제시하고자 한다. 위에서 언급하였듯이 음성서비스 위주의 이동 단말기를 이용하여 회의 서비스 도중 데이터 처리를 위한 서비스를 제공하는 것은 매우 까다로운 부분이다. 이에 본 논문에서는 이러한 문제점들을 기존 회의 시스템에 WML과 VoiceXML이라는 새로운 기반기술의 접목과 발생하는 문제점의 극복을 통하여 보다 확장된 회의 서비스를 제공하는 새로운 형태의 회의 시스템 구조 모델을 제시하고 구조모델에 대한 구체적인 동작과 실험을 통하여 그 가능성을 타진한다.

82. 차분 파워 스펙트럼을 이용한 음성구간 추출 및 잡음제거

컴퓨터공학과 정 성 일
지도교수 신 옥 근

음성인식 기술은 지난 수십 년간의 연구결과 최근에는 괄목할만한 진보가 이루어졌으며, 음성인식 기술과 뿌리를 같이하는 화자인식 기술도 많은 진전이 있었다. 이러한 음성인식과 화자인식의 기술적인 진보는 그 기술이 실생활에 이용될 수 있으리라는 기대를 가지도록 하기에 충분했다.

그러나, 실생활에서는 화자의 발성방식이 조용한 환경에서의 발성방식과 다를 뿐 아니라, 학습시에 깨끗한 음성신호로 생성한 기준 패턴과는 다른 형태의 신호가 입력된다. 그 결과 훈련 환경과는 다른 환경에서 발성된 음성을 인식해야 하는 이른바 '불일치 조건(mismatch condition)' 현상이 발생하게 되며, 음성인식기와 음성부호기(vocoder) 등에 치명적인 성능 저하를 초래하는 것으로 알려져 있다[1]. 이러한 불일치 조건의 원인은 크게 음성신호에 첨가되어 입력되는 주변의 음향학적 잡음, 마이크로폰의 왜곡, 그리고 다양한 전송선로에서 발생하는 왜곡과 전자기적 잡음 등 세가지로 분류할 수 있다. 이런 불일치 조건 중 본 연구에서는 음성인식 시스템의 성능에 가장 큰 영향을 미치는 음향학적 잡음의 효과를 제거, 도는 감쇠시키기 위한 방법을 고려한다. 이러한 음향학적 잡음(이하 잡음)을 제거하기 위해 많은 연구가 이루어지고 있으며, 가장 효율적인 방법중의 하나가 필요한 음성성분을 제외한 나머지 잡음성분을 제거하는 것이다. 잡음을 제거하기 위해서는 정확한 잡음추정을 필요로 하며, 정확한 잡음추정은 정확한 음성구간 추출을 전제로 한다.

먼저, 음성구간 추출은 음성인식 시스템의 필수적인 전처리(preprocessing)과정으로 깨끗한 음성신호에서는 전력밀도[2]나 영교차율(zero crossing rate) LPC(linear predictive coding)[3], 혹은 HMM(hidden markov model)[4]등 비교적 간단한 파라미터와 방법을 이용하여도 만족할 만한 결과를 얻을 수 있다. 그러나, 잡음이 음성신호에 첨가되면 원래의 음성 신호를 왜곡시킨다. 특히, 무성음은 잡음신호와 유사하기 때문에 이들 사이의 구별이 아주 어렵게 되어 음성구간 추출에 있어 많은 문제점이 있다. 만일, 잘못된 음성구간 추출정보를 이용하여 잡음을 추정하여 제거한다면 음성신호 성분의 특성을 감쇠시키거나 잡음을 잔존시켜 음성인식 시스템의 성능저하를 가져온다.

또한, 음성인식을 위한 잡음처리를 위해 다양한 방법들은 시도하고 있으며, 대표적인 방법으로 전처리 과정을 통한 음질개선 방식으로 스펙트럼 차감법[5], comb 필터링[6], winner 필터링[7], 스펙트럼 사상법[8], bayesian 추정법[9]을 들 수 있다. 이들 잡음을 제거하기 위한 여러 방법 중에서 스펙트럼 차감법이 가장 많이 쓰이는 방법이다. 이 방법을 이용한 잡음제거의 성패는 음성신호에 잡음이 첨가된 정도에 따라 결정되는 차감가중치와 잡음을 얼마나 정확하게 추정하느냐에 달려있으며, 음성구간 추출과 마찬가지로 이들의 설정 또한 음성인식 시스템의 성능에 커다란 영향을 미친다.

따라서 본 논문에서는 잡음을 효율적으로 제거하기 위해 가능한 한 정확한 음성구간을 추출한 다음, 이 결과를 이용하여 잡음을 제거하는 방안에 대해 기술한다. 먼저, 제안하는 음성구간 추출은 차분 파워 스펙트럼(differential power spectrum)[10], 음향심리학적(psychoacoustic) 절대 가청 주파수 임계치(*ATH* : absolute threshold of hearing)[11] 효과와 미소 차분치 추정 및 제거를 거친 음성의 주요 신호성분을 대상으로 하였다. 이렇게 추출된 음성구간 정보를 이용하여 잡음신호의 변화에 대해 유연하게 적응할 수 있는 잡음추정 방법을 제안한 다음, 음성신호에 잡음이 첨가된 정도에 따라 적용되는 차감 가중치를 이용하여 잡음을 제거하는 방법을 제안한다.