

# 전자상거래를 위한 음성 숫자 인식 알고리즘

강홍석<sup>0</sup>      최두현<sup>\*</sup>

<sup>0</sup>경북대학교 정보통신학과

redstone@palgong.knu.ac.kr

<sup>\*</sup>서울대학교 전기공학부

New Voiced Number Recognition Algorithm for Electronic Commerce

Hong-Seok Kang<sup>0</sup>

Doo-Heyun Choi<sup>\*</sup>

Dept. of Information & Communication, Kyungpook National University

School of Electrical Engineering, Seoul National University

## 요 약

최근 인터넷의 보급이 대중화되면서 전자상거래 시장이 급속도로 커지고 있다. 또한, 무선 이동 장비의 보급이 증가하고 무선장비의 인터넷 접속 지원이 보편화되면서 기존의 유선망 전자상거래 개념이 무선망에 까지 확장되고 있다. 무선 단말기를 전자상거래에 이용하기 위해서는 음성인식에 관한 연구가 절대적이다. 본 논문에서는 전자 상거래 응용에 꼭 필요한 음성 숫자 인식 시스템에 대해 다루었다. 제안한 알고리즘은 입력 음성 신호를 주파수와 진폭을 이용하여 몇 개의 그룹으로 나눈 후, 그 그룹 내에서 패턴 비교를 통해 숫자 인식 확률을 높이는 다단계 인식 방법을 제안하고 한다. 개발된 시스템은 대부분의 개인 정보를 담고 있는 숫자 정보의 음성입력이 가능하게 하므로 많은 전자상거래 시장에 응용할 수 있을 것이다.

## 1. 서론

최근 새로운 교역의 장으로 급속히 확대되고 있는 전자상거래는 인터넷이 보편화되기 이전인 1980년대 후반부터 정형화된 거래문서를 구조화되고 표준화된 전자적 양식으로 대체하여 상거래 활동에 적용한 이른바 전자문서교환(Electronic Data Interchange: EDI)이나 PC 통신에 의한 홈쇼핑 및 홈뱅킹 등 다양한 형태로 존재해 왔다[1]. 그러나 이러한 거래 형태들도, 웹(Web) 기술을 탄생시킨 멀티미디어 기술의 발달과 인터넷의 보편화, 상용화에 따라 글로벌 시장을 대상으로 한 전자상거래로 전환되어 가고 있다. 전체 상거래에서 전자상거래가 차지하는 비중이 급속도로 증가하고 있는 상황이며, 현재는 주로 컴퓨터를 이용한 거래가 이루어지고 있으나 이는 장소에 많은 제약을 가지고 있다[2]. 이런 제약을 극복하기 위해 휴대용 개인 단말기를 통한 거래가 시도되고 있으며, 보다 편리한 서비스를 제공하기 위해, 단순한 키 입력이 아닌 음성을 이용한 거래에 대한 연구도 이루어지고 있다. 이러한 휴대 단말기의 전자상거래 응용에 대한 관심은 휴대 단말기가 인터넷 기술과 접목되면서 더욱 확산되어 가고 있다.

지금까지 실제 응용에 사용되는 음성 인식은, 음성 워드 프로세서와 같은 일부 대형 인식 시스템이

있기는 하지만, 주로 아주 좁은 범위에서 패턴의 비교를 통한 인식이 대부분이었다. 이러한 제약의 이유는 전체를 인식하는 시스템은 작은 컴퓨터로 처리하기에는 한계가 있고 실제 응용도 전체를 필요로 하지 않는 경우가 많기 때문이다. 본 논문에서는 대부분의 전자상거래에 필요한 음성 입력은 그 영역이 제한되어 있고 숫자를 이용하는 경우가 많다는 점에 착안하여, 현재 기술로 실제 실시간 응용이 가능한 음성 숫자 인식 시스템을 제안한다. 본 논문에서는 보다 높은 인식률을 얻기 위해 먼저 한국어 음성 숫자의 특성을 분석하여 음성 신호를 일차적으로 몇 개의 그룹으로 나누고, 그 그룹 내에서 패턴을 비교를 하는 다단계 인식 기법을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 일반적인 음성 인식 시스템의 구조와 방법에 대한 소개를 하였다. 3장에서는 제안한 시스템의 구성과 한국어 숫자 음성의 특징에 대해서 설명하였다. 4장에서는 실험 환경과 실험 결과에 대해 기술하였고, 5장에서 결론을 맺는다.

## 2. 관련연구

이 장에서는, 일반적인 음성 숫자 인식 방법에 대하여 소개하며, 사용되는 기법들 중에서 핵심이 되는 방법에 대해서 간단히 다룬다.

### 2.1 격리 단어 인식

격리 단어 인식 시스템이란, 발음되어진 인식 대상 단어로부터 테스트 패턴을 추출하고, 이를 기준 패턴과 비교하여 대응되는 단어로 인식하는 것을 말한다[3].

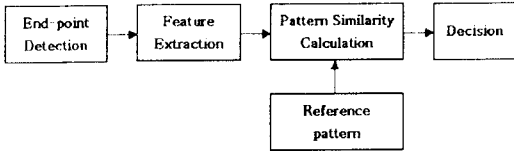


그림 1. 패턴인식 모델에 의한 숫자 인식 시스템

### 2.2 음성 신호 처리

음성인식 시스템에서는, 음성에 포함되어 있는 정보를 적절히 나타내기 위해서 저역통과(Low Pass) 필터를 거치고, 이 신호를 샘플링 한 후 양자화하여 디지털 신호를 생성한다. 이런 과정 후 끝점 검출을 하여 단어 부분을 찾고, pre-emphasis를 하여 잡음 제거 및 dynamic range를 감소시키는 효과를 얻는다. 이 신호를 여러 블록으로 나눈 후, FFT(Fast Fourier Transform)를 수행하여 신호를 처리한다[4].

#### (1) 끝점 검출

격리 단어 인식에서 끝점 검출의 정확성 여부는 단어인식의 정확도에 큰 영향을 미치므로 정확한 끝점 검출이 필요하다. 간단한 방법으로 short-time energy를 이용하여 이것이 일정 값보다 큰 부분은 음성 구간이며, 작은 부분은 묵음 구간으로 인식한다[5].

#### (2) 특징 추출

끝점 검출에 의하여 여러 개의 구간으로 나누어진 음성신호는 그 구간의 특징을 잘 나타낼 수 있는 이산 벡터(discrete vector)의 형태로 변환되는데, 이 과정을 특징 추출이라 한다.

이러한 특징에는, energy와 영 교차율과 같은 간단한 것부터 short-time spectrum과 LPC, 그리고 FBO(Filter Bank Output)나 homomorphic model 등과 같은 다양한 종류가 있다.

### 2.3 음성신호의 유사도 측정

#### (1) Distortion Measure

음성인식 시스템에서 입력음성과 기준패턴의 유사도를 측정하기 위해서는 두 신호의 차이를 측정할 필요가 있는데, 이를 distortion measure라고 한다.

#### (2) DTW (Dynamic Time Warping) Algorithm

같은 단어라 하더라도 사람마다 발음하는 데는 속도의 차이가 있으며, 한 사람이 같은 단어를 발음하더라도 다르기 마련이다. 그러므로, 테스트하고자 하는 음성을 기준이 되는 음성과 동일한 시간 축 상에서 비교한다면 추출한 음성 특징이나 distortion

measure가 아무리 우수하더라도 좋은 결과를 얻기 어려울 것이다. DTW는 정확한 음성인식을 위해서 발음 속도의 변화에 대해 time alignment와 같은 적절한 보상을 해주는 것을 의미한다[6].

### 3. 제안한 숫자 음성 인식 시스템

#### 3.1 숫자 음성 인식 시스템의 구성

이 장에서는, 제안한 숫자 음성 인식 시스템의 구성도와 방법에 대해 알아본다.

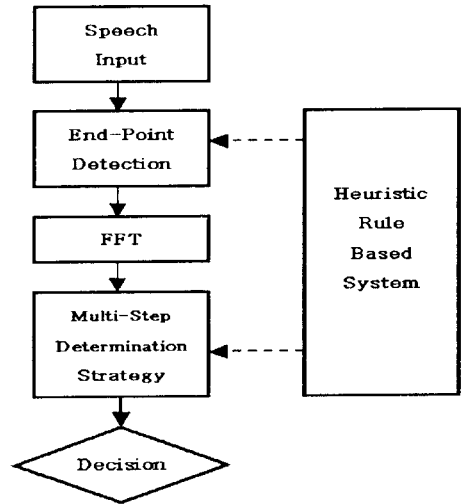


그림 2. 숫자 음성 인식 시스템 블럭도

- 음성 입력 : 마이크로 0에서 9까지의 숫자를 입력 받는다.
- 끝점 검출 : Short-time energy를 이용한 끝점 검출 방법은 잡음이 섞일 경우 구분하기 힘들며 짧은 구간 뒤는 잡음이 발생시 구분이 힘들다. 다음 그림처럼 음성의 peak 값을 중심으로 양쪽의  $\Delta X$  만큼을 음성으로 간주하였다.

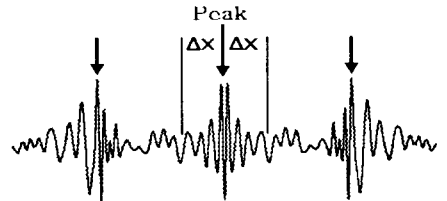


그림 3. 음성 인식 끝점 검출 방법

- Heuristic Rule Base System : 입력 음성으로부터 peak 값에서  $\Delta X$  만큼을 계산하여 음성인 부분과 묵음 부분을 구분한다.
- Multi-Step Determination strategy: 끝점 검출을 하여 얻은 음성 부분을 FFT를 수행한 후 특정 주파수에서 Amplitude를 갖는 숫자들을 그룹으로 구분한다. 0에서 9까지 10개의 숫자들은 4개의 그룹으로 구분하였다.

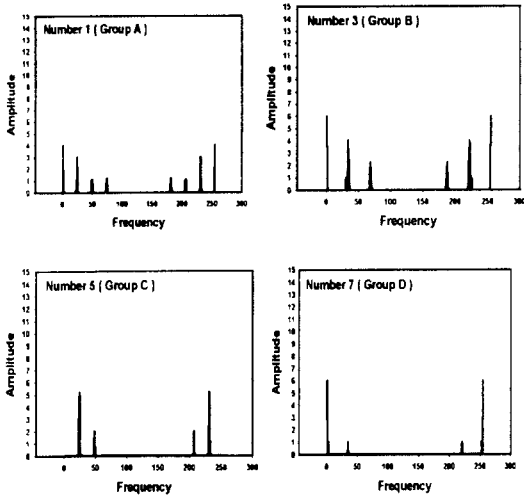


그림 4. 10개의 숫자를 4개의 그룹으로 구분

다음의 표는 특정 영역의 주파수에 속한 숫자들을 4개의 그룹으로 표현하였다.

GROUP	Frequency (Hz)	Number
Group A	0, 20-30, 45-55, 70-80	0, 1
Group B	0, 20-30, 60-70	2, 3, 4
Group C	20-30, 45-55	5
Group D	0, 20-40	6, 7, 8, 9

표 1. 특정 주파수로 구분한 4개의 그룹

동일한 그룹에 속한 숫자들은 특정 주파수의 진폭의 임계값과 비교하여 숫자를 인식한다. 임계값과 비슷한 값의 진폭을 가지면 잘못된 숫자 인식의 가능성이 있으므로 임계값과 진폭 사이에 일정한 차이가 있어야 인식하도록 하였다.

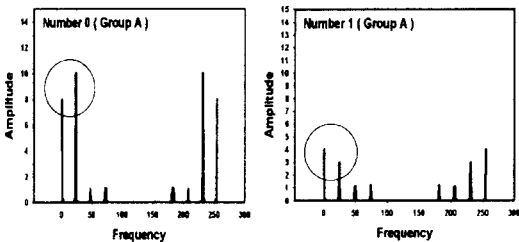


그림 5. 그룹(Group A)의 숫자 0과 1의 진폭 비교

#### 4. 실험환경 및 결과

본 실험은 Visual C++을 이용하여 구현하였고, 실험 테스트는 잡음이 적은 조용한 상태에서 실험을 하였으며 그림 6은 테스트 실험의 예이다.

실험의 결과 그룹간의 인식은 높은 편이었으나, 동일 그룹간의 인식율은 저조했다.



그림 6. 숫자 음성 인식 구현 실험 예

#### 5. 결론 및 향후 연구 방향

본 논문은 전자상거래에 이용할 수 있는 음성 숫자 인식 알고리즘에 대해 살펴보았다. 제안한 알고리즘은 음성 입력을 음소단위로 분리하여 각 단어의 주파수 성분을 이용하여 4개의 그룹으로 분리하고 그 그룹 내에서 다시 인식을 시도하는 다단계 인식 기법을 채택하고 있다. 제안된 시스템을 이용한 실험에서 그룹간의 분리는 비교적 잘 되었으나 그룹내의 인식에서는 다소 오류를 보였다. 제안된 알고리즘을 보완하여 숫자 인식에 대해서 만이라도 신뢰할 수 있는 결과를 얻는다면 전자상거래의 차원을 한 차원 올리는 계기가 될 것이다. 멀티미디어 기술의 발달과 인터넷의 상용화에 따라 급속도로 증가하고 있는 전자상거래를 이동성과 휴대성이 뛰어난 개인 단말기를 이용하여 단순한 키 입력이 아닌 음성을 이용한 거래는 매우 유용할 것이다.

본 논문에서 제안한 음성 숫자 인식 알고리즘을 보완하기 위해서는 다음과 같은 연구가 필요하다.

- 동일 그룹의 숫자 비교를 위한 임계값 설정에 대한 연구.
- 개인 단말기 사용 시 주위 환경 혹은 시스템 잡음에 대처 방안에 대한 연구.
- 연속 음성 입력에서의 음소 분리에 대한 연구
- 전자상거래에 뿐만 아니라 여러 응용에 이용 방안 모색.

#### 6. 참고 문헌

- [1]. 김철환, 김규수, "전자 상거래", 문원출판, 1999.6.
- [2]. 홍민수, "전자상거래 전략", 내하출판사, 1999.10.
- [3]. 이성환, "패턴인식의 원리", 흥광과학 출판사, 1994.1.
- [4]. Emmanuel C. Ifeachor and Barrie W. Jervis, "Digital Signal Processing", Addison-Wesley, 1996.
- [5]. 이형세, "음성 인식 기법", 청문사, 1999.2.
- [6]. C. Myers and L. R. Rabiner, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognitions", IEEE Trans. on Speech, Signal Processing, vol. ASSP -28, pp. 623 -635, Dec. 1980.