제19회 한국정보처리학회 춘계학술발표대회 논문집 제10권 제1호 (2003. 5)

데이터베이스 그룹화를 이용한 음성인식시스템의 성능향상에 관한연구

조태수, 권승호, 이동규, 한수영, 이두수
한양대학교 전자통신전파공학과
e-mail : rose@ihanyang.ac.kr

A Study on the Fast Speech Recognition System using DB Classification

Tae-Su Cho, Seung-Ho Kwon, Dong-Gyu Lee,
Soo-Young Han, Doo-Soo Lee
Dept. of Division of Electrical and Computer Engineering,
Han-Yang University

요 약

고립단어 인식에서 동적 패턴 정합법은 비교적 간단한 알고리즘과 최소의 하드웨어를 요구하므로 간단한 응용분야에 효율적으로 이용할 수 있다. 본 논문에서는 동적 패턴 정합법을 이용한 기존의 고립단어 인식시스템에 기준패턴 그룹화를 이용하여 연산량을 감소시켜 저가형 프로세서에서도 고속으로 동작할 수 있게 한다.

1. 서론

최근 생활수준의 발달과 함께 각종 디지털 미디어의 발달, 초고속 정보 통신망의 구축과 더불어 멀티미디어 통신을 통한 통신 판매, 음성인식 기술을 이용한 홈오토메이션 등의 생활 전반에 음성인식 및 합성기술에 관한 중요성도 점점 높아져가고 있다.

실제 생활 속에서도 개인용 컴퓨터를 이용하지 않더라도 가전제품 속에서도 음성처리 기술을 손쉽게 접해 볼 수 있게 되었다.

하지만 가전제품 등에 고가의 프로세서를 사용한다면 제품 가격 전반에 걸친 가격상승의 부작용을 나타낼 수 있다.

본 논문에서는 기존의 고립단어 인식시스템에 연산량을 감소시켜 저가형 프로세서에서도 고속으로 동작할 수 있게 하고자 한다. 저가형 프로세서를 사용하여 음성인식을 수행할 수 있다면, 이 음성인식 알고리즘을 사용한 제품에 전체 가격을 절감시키는 효과를 기대할 수 있을 것이다.

2. 음성인식

음성인식은 인식 방법에 따라 4가지로 구분할 수 있다. 첫째로 패턴정합법(Pattern Matching)에 의한 동적 정합(Dynamic Time Warping)은 입력패턴을 미리 정해진 기준 패턴과 비교하여 최적화된 유사성을 판단하는 방법이다. 둘째로 신경회로망을 이용한 방법은 각 음성별로 신경회로망을 구성하고 음성간의 변별력을 갖도록 학습을 수행하는 인식 방법이다. 그러나 이 방법은 새로운 패턴의 추가 시 인식 시스템을 다시 학습시켜야 하고 고도의 병렬계산 능력이 요구되기 때문에 실제 응용 시에는 적합하지 않다는 단점이 있다. 세 번째 방법인 벡터양자화 방법은 입력 패턴과 양자화 코드북 사이의 거리로 유사성을 판단하는 방법이지만 많은 학습자료가 필요하고 음성간의 동적인 변화 특성을 이용하지 못하기 때문에 인식률에 한계가 있다. 마지막으로 은닉마코프모델(Hidden Markov Model-HMM)은 학습기능을 이용하여 음성내의 변이를 흡수 할 수 있으며, 입력 패턴의 비선형 정합을 수행하는 특성이 있다[1].

2.1 화자인식

일반적으로 화자 인식은 크게 두 가지로 나누어 처리되고 있다.

첫째로 화자식별 (Speaker Identification)은 N명의 화자 중 가장 비슷한 사람을 찾아내는 과정이다. 둘째로 화자확인(Speaker Verification)은 지금 발성중인 화자가 인식시스템이 요청한 그 사람인지 아닌지(Yes-no task)를 결정하는 과정이다. 또 화자인식 시스템은 인식에 사용하는 문장의 종속여부에 따라 정해지지 않는 어휘로 인식을 수행하는 텍스트 독립형(Text Independant)과 정해진 어휘만을 발성해야 하는 텍스트 종속형(Text Dependant)으로 나눌 수 있다[2].

2.2 단어인식

음성인식은 인식하는 음성의 단위에 따라 고립단어 및 연속음성인식으로 나눌 수 있다.

고립단어인식은 짧은 음성명령이나 간단한 음성제어 등에 주로 사용된다. 연속음성인식은 문장을 인식하기 때문에 사용자가 단어 단위로 끊어 발음하지 않아도 된다. 이 시스템은 고립단어인식에 비해 인식률이 낮고, 인식 어휘 수에도 제약이 많았다. 그러나 최근 알고리듬 개선 등을 통해서 인식률을 개선해 왔다. 수백어휘 이내의 단어를 인식하는 소 어휘 시스템은 인식률과 신뢰도가 높다. 그러나 어휘가 제한되어 특정 응용분야를 지원하는 시스템으로만 개발되고 있다. 대 어휘 시스템의 경우 수만 단어 어휘까지 인식 가능하지만 인식률이 낮고 말할 때 사용자가 발음에 주의를 기울여야 하는 불편이 있다[2].

3. 음성 특징 매개변수 검출 및 패턴정합

3.1 음성구간 검출

음성구간의 검출의 정확성에 따라 인식의 정확도에 큰 영향을 미치기 때문에 정확한 끝점검출이 필요하게 된다. 간단한 방법으로는 단구간 에너지(short-time energy)를 이용하여 에너지 값이 큰 부분은 음성구간으로, 작은 부분은 묵음구간으로 결정하는 방법이 있다[3].

그림 1은 에너지와 영교차율을 이용하여 음성끝점 검출하는 예를 나타내었다.

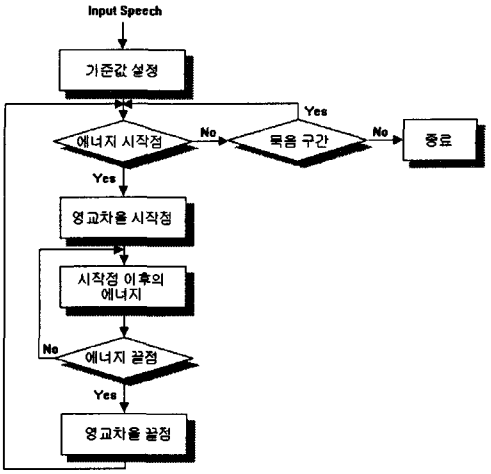

그림 1 음성끝점 검출의 흐름도

3.2 LPC(선형 예측 계수) 검출

LPC 방법은 음성 신호 처리에서 가장 널리 쓰이고 있는 알고리즘의 하나로 음성을 전극(all pole)모델로 가정하고 그에 따른 필터의 계수를 이용하여 음성 신호를 모델링하는 방법이다. 이 LPC 계수에 의해 구성되는 필터는 전극특성으로 가정하여 음성이 어떻게 생성되는가하는 것을 분석하는 것으로, 성도의 특성을 모델링하게 된다. 또한 실제 구현 시 쉽게 적용될 수 있기 때문에 많이 사용되고 있는 알고리즘이다[1].

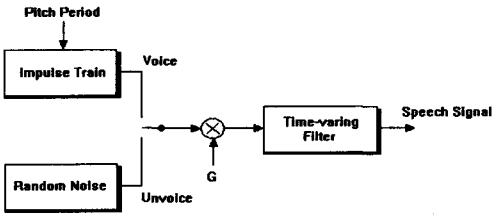

그림 2 음성 생성 모델

음성의 생성모델을 보면 그림 2와 같다. 우선 음원 성분인 유성음과 무성음을 생성한 뒤 여기신호의 크기를 조절한 후 성도성분을 나타내는 시변환 필터를 거치게 되어 음성신호가 생성되게 되는데 이때 성도를 나타내는 필터는 시변환 특성을 가지고 전극 구조이기 때문에 다음과 같은 구조를 갖게 된다.

$$H(z)=\frac{1}{1-\sum_{i=1}^{k}a_i z^{-i}}=\frac{1}{A(z)} \qquad \text{식 1}$$

위의 (식 1) 에서 성도모델에 관한 필터 계수 $a_i\,(i=1,\ldots,p)$가 LPC계수이다.

3.3 동적패턴정합(DTW)

패턴 정합 방법인 동적정합법(DTW: Dynamic Time-Warping)은 길이가 서로 다른 두 개의 자료에서 최적의 정합 경로를 서로 비교할 수 있는 방법으로, 비교적 간단한 알고리즘과 최소의 하드웨어를 요구하므로 간단한 응용분야에 효율적으로 이용할 수 있다. 이 기술은 고립단어 인식에서 기원되었으나 연속음성 인식에 역시 적용할 수 있다. 그러나 동적 프로그래밍(DP: Dynamic Programing)으로 인해 계산량이 많고, 수많은 화자내의 변위를 수용할 수 있는 기준패턴의 작성이 어려워 사용어휘가 제한되는 단점이 있다.

입력 단어 음성 패턴 T 에 대해서, 모든 단어 음성의 기준 패턴 R 과 DP 매칭을 행한 다음, 거리 $D(T,R)$ 을 최소로 하는 R 의 패턴을 T 의 패턴으로 선택한다. 즉 단어 음성 패턴 R 에 의해서 표시되는 단어에 대응하는 것으로 판단한다.

$$D(T,R)=\min_{F(k)} D(T,R) = \frac{1}{N}\min_{F(k)}\sum_{k=1}^{K} w(k)d(a^{i(k)},b^{j(k)}) \quad \text{식 2}$$

(식 2)에서 특징 벡터 a^i 와 b^j 의 거리 $d(a^i,b^j)$ 는 (식 3) 와 같은 유클리드 거리로서 주어질 수 있다.

$$d(a^i,b^j)=\|a^i-b^j\|_2=\left\{\sum_{m=1}^{M}(a_m^i-b_m^j)^2\right\}^{\frac{1}{2}} \quad \text{식 3}$$

거리를 생각하는 2개의 벡터의 시계열 a^i 와 b^j 를 대응시키는 것은 i 축과 j 축 상에서 평면상의 쌍을 식 으로 표시하고, 여기서 $F(k)$ 를 시간 변환 함수라고 부른다[2].

4. 제안한 음성인식 알고리즘

본 논문에서 제안한 음성인식 알고리즘은 기준패턴의 음성을 그 에너지에 따라 유무성음을 분리하고 그 구분된 유성음 구간을 라벨링 함으로서 전체 인식기의 연산량 감소를 구현하였다. 하지만 음성 라벨링 단계는 음성의 구간 검출 및 에너지 파라미터의 추출의 단계에서 정확한 파라미터의 검출을 전제로 하기 때문에 이를 보안하기 위해 피치의 주기에 따른 가변윈도우를 사용하여 보다 정확한 파라미터 검출을 구현하였다.

그림 20은 본 논문에서 제안된 음성인식 시스템의 구조를 보여주고 있다.

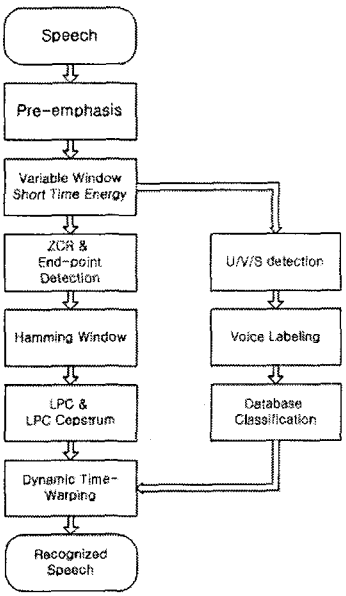

그림 3 제안된 음성인식 시스템

5. 실험 및 결과

본 논문의 알고리즘을 모의실험하기 위해 IBM PC에 마이크가 장치된 16비트 A/D변환기를 인터페이스 시켰다. 알고리즘을 구현하기위한 도구로는 Matlab6.0을 사용하여 파형분석을 하였고, 전체 인식률과 속도를 측정하기위하여 Win32api를 사용하여 인식기를 구현하였다.

실험은 일반 실험실 환경에서 20명의 화자가 각각 10개의 지정된 단어를 발성한 음성 시료를 8khz로 샘플링하고 16비트로 양자화하여 사용하였다. 인식을 위한 특징벡터로는 14차 LPC Mel-Cepstrum을 사용하였다.

5.1 윈도우 영향이 제거된 에너지의 추출

기존의 에너지 파라미터 추출 방법은 윈도우 사이즈를 고정 시켜 에너지 파라미터를 구함으로써 정확한 프레임 에너지를 구할 수 없었다. 따라서 본 논문에서는 피치주기를 먼저 구하고 그 주기에 따라 에너지를 구하는 방법을 제안하였다.

제안한 알고리즘의 블럭도는 그림 4에서 보여 주고 있다.

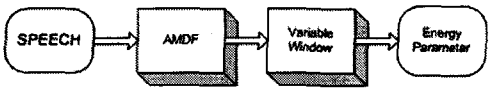

그림 4 제안된 음성파라미터 검출

그림 5는 가변윈도우를 사용한 모의실험 결과를 보여주고 있다. 그림 (a)에서는 입력된 음성을 나타내고, 그림 (b)에서는 256프레임의 고정된 윈도우를 사용한 결과를 보여주고, 그림 (c)에서는 피치 주기에 따른 가변 윈도우를 사용한 결과를 보여주고 있다. 제안된 가변윈도우를 사용하는 방법은 입력 음성의 전이구간에서 정확한 에너지 파라미터를 구하는 결과를 보여주고 있다.

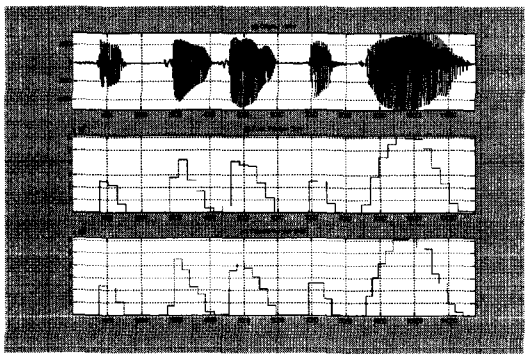

그림 5 제안된 에너지 추출 방법의 결과 파형

5.2 검출된 에너지를 이용한 기준패턴 그룹화

제안된 가변윈도우를 사용한 에너지 파라미터를 이용하여 음성의 U/V/S를 구별하고 구별된 유성음 구간에서 에너지를 라벨링 함으로써 기준패턴을 구룹화 하였다.

그림 6은 유성음 구간에서의 에너지 라벨링 결과를 보여주고 있다.

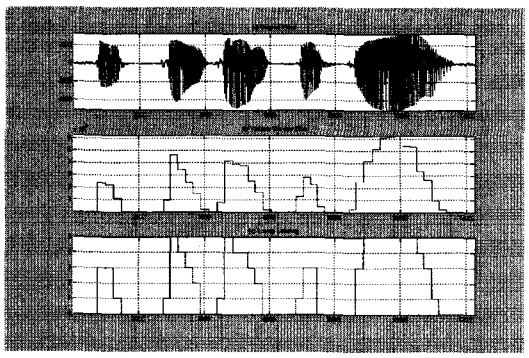

그림 6 기준패턴 그룹화

5.3 검색된 그룹 내에서의 음성인식 결과

실험결과 표1과 표2에서는 제안한 방법이 기존의 방법에 비해 인식률은 거의 비슷하였고 처리시간은 제안한 방법에서 76%정도로 약 25%정도의 연산량을 감소시켰다.

표 1 전체 인식률

	기존의 방법	제안한 방법
인식률	93%	92%

표 2 전체처리시간

	기존의 방법	제안한 방법
처리시간	3.84	2.92

6. 결론

최근 생활수준의 향상과 음성인식기술의 발달로 개인용 컴퓨터를 이용하지 않더라도 가전제품 속에서도 음성인식 기술을 손쉽게 접해 볼 수 있게 되었다.

본 논문에서는 기존의 고립단어 인식시스템에 연산량을 감소시켜 저가형 프로세서에서도 고속으로 동작할 수 있게 하고자 하였다. 저가형 프로세서를 사용하여 음성인식을 수행할 수 있다면, 이 음성인식 알고리즘을 사용한 제품에 전체 가격을 절감시키는 효과를 기대할 수 있을 것이다.

참고문헌

[1] S. Funui, "Digital Speech Processing, Synthesis and Recognition" , Marcel Dedder, Inc., 1992

[2] L. R. Rabiner & R.W.Schafer, "Digital Processing of Speech Signal" , Prentice-Hall, Englewood Cliffs, N.J., U.S.A., 1978

[3] L. R. Rabiner & Biing-Hwang Juang, "Fundamentals Of Speech Recognition" , Prentice-Hall AT&T, U.S.A, 1993

[4] Vlandal, "Dynamic Time-Warping Method for Isolated Speech Sequence Recognition" , IEEE, 2001

[5] Guoqing Chen , "Discovering similar time series patterns with fuzzy clustering and DTW methods", IEEE, 2001

[6] 조태수, "윈도우 영향이 제거된 에너지 파라미터에 관한 연구",전자공학회 하계종합학술대회,2001