응용통계연구 (2008)
**21**(6), 959–967

# 숨은마코프모형을 이용하는 음성 끝점 검출을 위한 이산 특징벡터

이재기[1] · 오창혁[2]

[1]영남대학교 통계학과, [2]영남대학교 통계학과

(2008년 7월 접수, 2008년 9월 채택)

**요 약**

본 연구의 목적은 숨은마코프모형을 사용하여 음성구간의 끝점을 검출하는 문제에서 소음의 환경에서도 강건하며 계산의 부하가 적은 이산형 특징벡터를 제안하고 이의 성질을 실증적으로 밝히는 것이다. 제시된 특징벡터는 일차원의 소리 신호의 에너지의 변화율을 나타내는 경사도이며 숨은마코프모형과 관련된 계산에서의 부하를 감소하기 위하여 세 개의 값으로 이산화하였다. 여러 소음 수준의 끝점 검출의 실험에서, 제시된 특징벡터가 잡음 환경에서도 강건함을 보였다.

주요용어: 끝점 검출, 이산화, 숨은마코프모형, 백색잡음.

## 1. 서 론

컴퓨터 등의 기계를 이용한 음성인식은 인공지능에서의 중요한 문제로 인식되어 많은 연구가 이루어져 왔다. 음성인식을 하기 위해서는 먼저 음성신호가 포함된 소리 신호를 마이크와 같은 기계로 입력받게 된다. 발음된 음성을 마이크 등으로 채집하면 음성신호 구간과 비음성신호 구간이 혼재되어 있는 것이 일반적이다. 흔히 음성신호 구간은 음성구간, 비음성신호 구간은 소음구간이라고 불리어진다. 음성인식의 첫 단계는 채집된 신호에서 음성구간을 추출하는 것이다. 음성구간의 추출은 음성이 시작하는 시작점과 음성이 끝나는 종점을 검출하는 문제로 귀결될 수 있다. 음성구간의 시작점과 종점을 각각 시작 끝점과 종료 끝점이라고 부르며, 통틀어서 끝점이라고 부른다. 끝점 검출의 정확성은 후행되는 인식과정에서 인식률의 향상과 효율적인 계산을 위해 필수적인 것으로 알려져 있다.

이러한 음성 끝점 검출에 대한 다양한 연구가 이루어져 왔으며 Rabiner와 Sambur (1975)는 에너지와 영교차율을 이용하여 음성구간을 검출하는 방법을 제안하였고, Lamel 등 (1981)은 이를 향상시키는 방법을 제안하였다. Kaiser (1990)는 Teager 에너지를 제안하였으며 Ying 등 (1993)은 Teager 에너지를 사용하여서 음성구간을 추출하는 방법을 제안하였다. 또한, Teager 에너지는 음성의 피치 검출에 사용되기도 하였는데 자세한 것은 Chen과 Hu (2007)를 참조하기 바란다.

석종원과 배건성 (1996)은 웨이블렛 변환을 이용하여 새로운 특징벡터를 제안하였으며 이를 이용하여 잡음 환경 하에서의 음성검출 방법을 제안하였다. 이용형과 오창혁 (2001)은 계산의 효율성을 위하여 pre-emphasis나 밴드패스 필터링이 요구되지 않는 이산형 1차원 특징벡터를 제안하였으며 제시한 특징

[1](712–749) 경상북도 경산시 대동, 영남대학교 통계학과, 석사. E-mail: jaikylee290@hotmail.com
[2]교신저자: (712–749) 경상북도 경산시 대동, 영남대학교 통계학과, 교수. E-mail: choh@yu.ac.kr

벡터는 1차원 이산형이므로 연산속도가 빠르고 알고리즘의 구성을 간단하게 만들 수 있는 장점을 가진다. 이와 같은 방법은 잡음이 없는 조용한 환경 즉, 신호대잡음비율인 SNR이 큰 환경에서는 끝점 검출이 원활하게 이루어지지만, SNR이 작은 환경에서는 배경 잡음구간과 음성구간의 에너지 차이가 크지 않아서 끝점 검출률이 낮아지게 된다. 박기상 등 (2002)은 SNR에 따라서 대역에너지를 세분화하여 음성의 끝점검출 방법을 제안하였다. 정대성 등 (2003)은 스펙트럼 차이를 이용해서 SNR에 따른 가변적인 끝점검출 방법을 제안하였다. 홍정우와 오창혁 (2008)은 멜주파수의 크기를 평활하는 일차원 특징벡터 평균파워가 기존의 고차원 멜주파수 보다 다양한 수준의 소음환경에서 끝점검출이 우수함을 보였다.

이용형과 오창혁 (2001)은 연속되는 프레임에서의 에너지 변화의 양에 관련된 특징벡터를 제시하고 이에 기초한 이산형 숨은마코브모형을 사용하여 끝점 검출에 관한 실험에서, Acero 등 (1993)의 방법에 비해 끝점 검출률이 우수하며, 또한 끝점 검출을 위한 연산회수가 대폭 감소됨을 보였다. 예컨대 프레임의 개수가 100, 각 프레임에 대한 신호값이 80개인 상황에서 필요한 연산의 회수는 이용형과 오창혁 (2001)의 방법이 Acero 등 (1993)의 방법에 비해 5% 정도에 불과함을 보였다. 이러한 연산회수의 개선은 관측치의 이산화에 있으며, Acero 등 (1993)은 연속 관측치인 잡음 소거 로그 에너지와 델타 로그에너지를 사용하는 연속형 숨은마코프모형을 가정하였다. 이들의 연구에서는 둘 다 저소음 환경에서 각각 제시한 방법에 대하여 실험을 하였다.

본 연구에서는 이용형과 오창혁 (2001)이 제시한 특징벡터에 근거한 소음환경에서 끝점 검출을 향상시킬 수 있도록 새로운 특징벡터를 제시하고 이를 숨은마코프모형을 이용하여서 음성구간을 검출하는 방법을 제시한다. 본 연구의 구성은 다음과 같다. 2절에서는 소음 환경에서 강건한 끝점 검출을 위한 새로운 특징벡터를 제시하고, 3절에서는 숨은마코프모형에 관한 소개를 한다. 4절에서는 제시한 특징벡터와 숨은마코프모형을 이용한 음성구간 추출 방법의 성능에 대한 비교실험을 한다.

## 2. 이산특징벡터와 숨은마코프모형

음성구간의 추출을 위한 첫 단계는 음성신호로부터 특징벡터를 얻는 것이다. 이 논문에서는 음성신호의 단구간에너지에 기초하여 단순선형회귀모형으로 적합한 직선의 기울기를 특징벡터로 정의한다. 음성구간추출은 숨은마코프모형을 이용하여 이루어지며 이러한 구간추출의 과정을 서술한다. 먼저 특징벡터를 제안한다.

### 2.1. 이산특징벡터

입력된 음성 신호집합을

$$S = \{s(1), s(2), \dots\} \tag{2.1}$$

라고 하자. 그리고 프레임 $F(n)$은 음성신호집합 $S$의 부분집합으로 주어진다. 즉, 프레임 번호 $n = 1, 2, \dots$에 대하여

$$F(n) = \{s(t) : (n-1) \times (u-v) + 1 \le t \le (n-1) \times (u-v) + u\}, \tag{2.2}$$

단, $u$는 프레임 크기, $\nu$는 중첩된 구간의 크기이며 $v < u$이다. 각 프레임 $F(n)$에 대해서 절대값 단구간에너지 $E(n)$은 다음으로 정의된다.

$$E(n) = \sum_{t \in F(n)} |s(t)| . \tag{2.3}$$

이 논문에서는 절대값 단구간 에너지를 그냥 단구간 에너지라고 부르기로 한다. 여기서 모두 $N$개의 프레임이 있다고 하자. 이들 $N$개의 프레임에서 연속된 $(2l+1)$개의 프레임에 대하여 특징벡터를 정의한다고 하자. 프레임의 처음부분과 마지막 부분에 대한 프레임의 크기를 조절하기 위하여 $n=1,2,\ldots,N$에 대하여 $l_n=\max\{a : n-a>0,\ n+a<n,\ a\le l\}$이라고 하자. 이때 $(2l_n+1)$개의 $(n-l_n,\ E(n-l_n)),\ldots,(n,\ E(n)),\ldots,(n+l_n,\ E(n+l_n))$에 대하여 기울기 $v(n)$은 선형회귀모형의 최소제곱추정값으로 정의한다.

$$v(n)=\frac{1}{M_n}\sum_{i=-l_n}^{l_n} iE(n+i), \tag{2.4}$$

단, $M_n=l_n(l_n+1)(2l_n+1)/3$이다. 기울기 $v(n)$은 발성되는 소리의 크기에 따라 값의 차이가 많이 날 수 있으므로 즉, 데이터 집합 간에 변동성이 클 수 있으므로, 녹음의 상태 등에 의한 음량의 크기에 따른 기울기의 변동을 보정하기 위하여 표준화를 적용한다. 여기서는 이를 위하여 표준화절대기울기 $\eta(n)$을 다음으로 정의한다.

$$\eta(n)=\frac{1}{\hat{\sigma}}\,|v(n)-\hat{\mu}|, \tag{2.5}$$

단, $\hat{\mu}$와 $\hat{\sigma}$는 전체 $N$개의 기울기 $v(1),\ldots,v(N)$의 평균과 표준편차이다. 표준화절대기울기 $\eta(n)$에 대하여 각 프레임의 이산에너지 $\delta(n)$을 다음으로 정의한다.

$$\delta(n)=\begin{cases}3, & \eta(n)\ge 10,\\ 2, & 5\le \eta(n)\le 10,\\ 1, & \text{기타.}\end{cases} \tag{2.6}$$

이산에너지 $\delta(n)$의 값을 정할 때 경계값으로 사용되는 5와 10은 여러 음성 데이터에 대한 구간 추출에서 좋은 결과를 가져온 경험적 값이다.

한편, 이용형과 오창혁 (2001)은 식 (2.3)의 절대값 단구간 에너지 $E(n)$에 기초하여, 기울기 특성을 나타내는 이산 특징값 slope$(n)$, 두 연속된 단구간 에너지의 차이가 급격히 변하는지를 나타내는 shock$(n)$ 그리고 각 프레임에서의 단구간 에너지의 크기를 이산화하여 나타내는 sigma$(n)$을 정의하였다. 특징벡터 slope$(n)$은 세 개의 연속된 프레임에서의 단구간 에너지 값이 계속하여 증가하거나 혹은 감소하는 경우에 따라 정의되었다. 특징벡터 shock$(n)$은 연속된 두 프레임의 단구간 에너지의 차이가 일정크기 이상이면 1 또는 $-1$로, 그 이외의 경우에는 0으로 정의되었으며 이는 큰 에너지 변화를 특징짓는 값이다. 특징벡터 sigma$(n)$은 각 프레임에서 에너지의 값의 크기가 어떤 범위에 들어가는지를 나타내는 값이다. 이들 세 개의 특징벡터 slope$(n)$, shock$(n)$ 그리고 sigma$(n)$을 바탕으로 obsFeat$(n)$가 정의되었다.

$$\text{obsFeat}(n)=\begin{cases}0, & \text{if sigma}=0;\ \text{shock}=0; & \text{slope}=-1,\ 1,\\ 1, & \text{if sigma}=1;\ \text{shock}=-1,\ 0,\ 1; & \text{slope}=1,\\ 2, & \text{if sigma}=0;\ \text{shock}=-1,\ 1; & \text{slope}=-1,\ 1,\\ 3, & \text{if sigma}=2;\ \text{shock}=-1,\ 0,\ 1; & \text{slope}=-1,\ 1.\end{cases} \tag{2.7}$$

특히 특징벡터 slope$(n)$은 다음과 같이 정의된다.

$$\text{slope}(n)=\begin{cases}1, & \text{for } \xi(n)\ge 0,\\ -1, & \text{for } \xi(n)<0,\end{cases}$$

단, $\xi(n) = \{E(n-1) - E(n)\} \times \{E(n) - E(n+1)\}$. 한편, 프레임 $n$에 대한 Teager 에너지 $\psi(n)$은 다음과 같이 정의된다.

$$\psi(n) = E(n)^2 - E(n-1)E(n+1). \tag{2.8}$$

한편, slope$(n)$이나 $\psi(n)$은 연속된 세 개의 프레임의 에너지에 기초하고 있으나, 식 (2.4)에서 정의된 기울기 $v(n)$은 $(2l_n+1)$개의 단구간 에너지 $E(n)$과 관련된 점을 적합하는 선형회귀직선의 기울기이다. 또한, obsFeat$(n)$은 연속된 세 에너지에 관한 기울기, 연속된 두 에너지의 변화, 에너지의 크기에 대한 함수인 반면, 식 (2.6)의 $\delta(n)$은 기본적으로 기울기 $v(n)$에만 의존하는 이산 특징벡터이다.

특징벡터 $\delta(n)$은 연속되는 프레임의 에너지의 크기의 미세한 변화를 평활하는 특징을 가지고 있으므로 음성신호를 구분해 낼 때 소음으로부터 강건성을 가질 것으로 예측된다. 백색잡음과 같은 종류의 소음은 음성에 비해 고주파로 구성되고 이들은 $\delta(n)$을 위한 기울기를 구하는 과정에서 필터링되는 효과가 있기 때문이다. 이러한 잡음 필터링의 성질은 홍정우와 오창혁 (2008)에서 연속적인 값의 특징벡터를 사용하는 경우에 나타남이 보여졌다.

### 2.2. 숨은마코프모형

숨은마코프모형은 이중확률과정 $\{(X_n, Y_n),\ n = 1, 2, \dots\}$이다. 관측되지 않는 확률과정 $\{X_n\}$은 상태공간 $\{1, \dots, q\}$, 전이확률행렬 $A=(a_{ij})_{q\times q}$ 그리고 초기확률분포 $\Pi = (\pi_1, \dots, \pi_q)$를 가지는 마코프연쇄이다. 한편, 관측 가능한 확률과정 $\{Y_n\}$은 각 $n$에 대하여 $Y_n$의 분포는 오직 $X_n$에만 의존하며, $Y_n,\ n = 1, 2, \dots,$은 서로 독립이라고 가정한다. 즉, 각 시점 $n$에서의 관측변수 $Y_n$의 확률분포는 마코프연쇄 $X_n$의 상태에만 의존한다고 가정한다. 관측기호공간을 $K = \{1, \dots, M\}$이라고 하자. 시점 $n$에서 마코프연쇄의 상태가 $X_n = j$로 주어진 경우 관측치 $Y_n = k$에 관한 조건부 확률은 $b_j(k) = P(Y_n = k \mid X_n = j)$로 나타내고, 조건부 확률을 행렬 $B=(b_j(k))_{q\times m}$로 나타내기로 한다. 이산형 숨은마코프모형은 전이확률행렬 $A$, 관측확률행렬 $B$, 초기확률분포 $\Pi$를 사용하여 $\lambda = (A, B, \Pi)$로 나타내기로 한다.

특징벡터 $\delta(n)$ 혹은 obsFeat를 사용하는 끝점검출 모형에서는 $\delta(n)$ 혹은 obsFeat이 관측가능한 값 $Y_n$이 된다. 한편, 관측변수 $Y_n$이 연속형인 경우에는 $X_n = j$가 주어진 조건 하에서의 $Y_n$의 분포는 흔히 혼합정규분포가 사용된다. 본 논문에서는 연속형 관측치 Teager 에너지 $\psi(n)$에 대하여 관측변수 $Y_n$의 분포로써 단일 정규분포를 가정한다. 즉, $X_n = j$로 주어진 경우 관측치 $Y_n$의 확률분포는 평균과 분산이 각각 $\mu_j$, $\sigma_j^2$인 정규확률분포라고 가정한다. 관측되지 않는 확률과정 $X_n$은 '끝점상태', '소음상태' 그리고 '신호상태' 등의 상태를 가지는 마코프연쇄로 가정한다. 끝점은 시작 끝점과 종료 끝점의 두 가지로 구별되나 음성에너지와 관련된 특성이 비슷하므로 끝점 상태 하나로 묶었다.

숨은마코프모형에는 '평가', '해독' 그리고 '추정'의 세 가지 규범적 문제가 있다. 크기가 $N$인 관측벡터열 $y = \{y_1, \dots, y_n\}$이 주어졌다고 하자. 이때 평가문제는 모형 $\lambda$가 주어진 경우 관측값 $y$의 확률 $p(Y = y \mid \lambda)$을 구하기 위한 연산횟수 축소 문제이다. 평가문제를 위하여는 전진절차 또는 후진절차 알고리즘이 있다. 해독문제는 주어진 모형과 관측열에 대하여 관측되지 않은 마코프모형의 상태를 추정하는 문제이며 우도함수를 최대화하는 경로를 추정하는 반복적인 방법으로 Viterbi 알고리즘이 있다. 추정문제는 주어진 모형 $\lambda$와 관측열 $y$에 대하여 모수를 추정하여 개정된 모형 $\lambda'$을 얻는 것에 관한 것이다. 모수 추정을 위한 해법 중 하나는 Baum-Welch 알고리즘이며 반복 절차에 의해 국소 최우추정치로 수렴하는 추정치를 얻을 수 있음이 알려져 있다. 이산형 숨은마코프모형에 관한 보다 자세한 내용은 Rabiner (1989)를 참조하기 바란다. 음성의 끝점검출을 위하여 숨은마코프모형을 가정하는 경우 이들 세 가지 알고리즘을 사용하게 된다. 음성인식에서 Baum-Welch 알고리즘을 사용하여 여러 개의 자료로

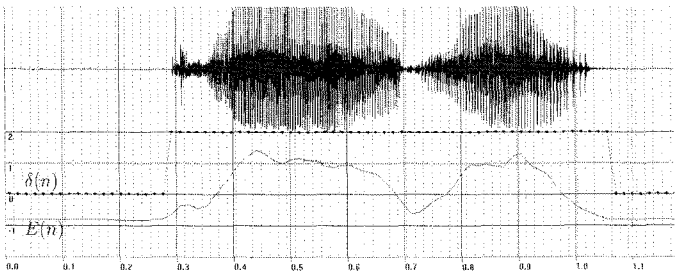


**그림 3.1.** 파형 '팔월'과 $E(n)$과 $\delta(n)$의 예시(조용한 상태)

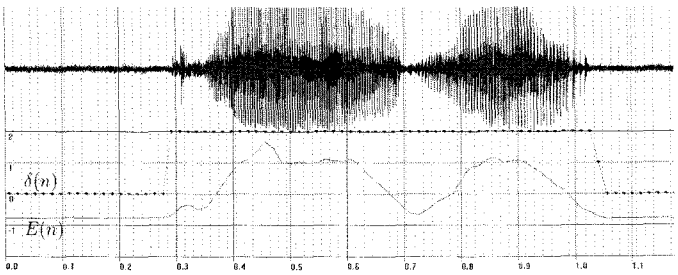


**그림 3.2.** 파형 '팔월'과 $E(n)$과 $\delta(n)$의 예시(SNR: $20dB$)

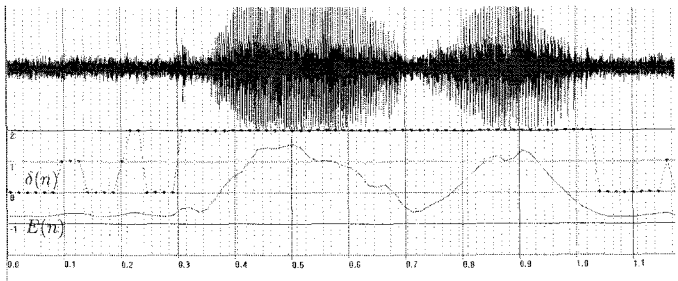


**그림 3.3.** 파형 '팔월'과 $E(n)$과 $\delta(n)$의 예시(SNR: $10dB$)

부터 모형의 모수를 개정하는 것을 '훈련'이라고 부르며, Viterbi 알고리즘을 사용하여 주어진 모형에 대하여 하나의 자료로부터 숨은 확률과정의 상태를 추정하는 것을 '해독'이라고 부른다.

## 3. 실험 및 결과

실험을 위한 음성 데이터를 얻기 위하여 일상적인 상태의 조용한 사무실 환경에서, 10명의 20대 성인 남녀가 각각 10개의 지정 단어를 녹음하여 모두 100개의 음성데이터를 얻었다. 실험을 위해 선택된 단어는 약한 파열음(/p/, /t/, /k/)이 음성의 시작 부분에 존재하는 것이다. 그리고 종료 부분에는 이들 파열음을 포함한 모음 등의 다양한 형태를 가지는 음들이다. 이러한 약한 파열음은 에너지가 유성음에

표 **3.1.** 여러 가지 소음 수준에서 시작 끝점 검출률(단위: %)

| 시간길이 | $\psi(n)$ | | | obsFeat | | | $\delta(n)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| ms | Clean | 20dB | 10dB | Clean | 20dB | 10dB | Clean | 20dB | 10dB |
| 30 | 79 | 79 | 66 | 77 | 75 | 59 | 70 | 74 | 82 |
| 45 | 95 | 94 | 79 | 96 | 86 | 73 | 96 | 96 | 97 |
| 60 | 97 | 98 | 83 | 98 | 96 | 79 | 97 | 98 | 99 |
| 75 | 98 | 99 | 88 | 100 | 100 | 90 | 98 | 98 | 100 |
| 90 | 99 | 99 | 94 | 100 | 100 | 92 | 99 | 99 | 100 |

표 **3.2.** 여러 가지 소음 수준에서의 종료 끝점 검출률(단위: %)

| 시간길이 | $\psi(n)$ | | | obsFeat | | | $\delta(n)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| ms | Clean | 20dB | 10dB | Clean | 20dB | 10dB | Clean | 20dB | 10dB |
| 30 | 60 | 26 | 4 | 79 | 25 | 2 | 71 | 56 | 30 |
| 45 | 83 | 43 | 12 | 95 | 37 | 6 | 84 | 72 | 46 |
| 60 | 91 | 60 | 18 | 97 | 54 | 12 | 92 | 77 | 59 |
| 75 | 95 | 67 | 25 | 98 | 65 | 15 | 97 | 86 | 70 |
| 90 | 97 | 69 | 32 | 99 | 68 | 22 | 98 | 96 | 73 |

비해 작아 안정적인 특성을 나타내지 못하여 끝점 검출이 어렵다고 알려져 있다. 보다 자세한 내용은 Rabiner와 Sambur (1975)를 참조하기 바란다. 또한, 종료부분에 모음이 존재하는 경우에는 에너지의 감소가 완만하여 끝점 검출이 어렵다고 알려져 있다. 실험에 사용된 단어는 김경태 등 (1990)에서 선택한 "팔월, 팽이, 페인트, 펜치, 평야, 폐품, 포크, 풀숲, 풍습, 필요"의 10개 단어이다. 녹음을 위한 설정은 1채널, 11025Hz, 16bit, PCM방식이며 RIFF WAVE 형식으로 파일을 저장하였다.

저장된 음성신호는 식 (2.1)의 $s(1)$, $s(2)$,...로 표시된다. 한편, 식 (2.2)의 프레임 $F(n)$을 얻기 위하여 프레임의 폭 $u$는 시간 길이 25ms에, 프레임 중첩 길이 $v$는 시간 길이 10ms에 대응되도록 정하였다. 또한, 녹음된 원래의 음성신호 데이터에 백색 소음을 추가시켜 각각 20dB, 10dB의 SNR를 가지는 음성신호 데이터를 만들었다. 그림 3.1, 3.2 그리고 3.3은 조용한 상태의 음성신호와 각각 10dB와 20dB의 백색잡음이 추가된 음성신호의 그림이다. 이들 그림에는 식 (2.3)의 단구간에너지 $E(n)$과 식 (2.6)의 이산특징벡터 $\delta(n)$도 함께 표시하였다.

세 가지 특징벡터, Teager 에너지 $\psi(n)$, 이용형과 오창혁(2001)의 obsFeat 그리고 본 연구에서 제시한 $\delta(n)$을 사용하여 끝점 검출율을 비교하기 위한 실험을 하였다. 마코프연쇄의 상태는 $Q = 3$개로 하였으며, 이산 관측변수 $\delta(n)$과 obsFeat에 대하여는 관측기호의 개수는 $M = 3$ 또는 4로 두었으며, 연속 관측변수 $\psi(n)$에 대하여는 혼합성분의 개수가 1인 혼합정규분포를 가정하였다. 음성데이터 100개 중에서 30개는 모형의 모수를 추정하기 위한 데이터로 즉, 훈련용으로 사용하였으며, 끝점 검출율 조사에는 나머지 70개의 데이터를 사용하였다. 음성구간의 기준 끝점을 정하는 것은 시각과 청각을 사용하여 이루어졌다. 즉, 각 음성파형에 대하여, 사람의 귀에 들리는 음성이 시작하는 점과 눈으로 보아서 파형이 시작되는 점을 시작 끝점, 귀로 들어서 음성이 끝나는 점과 눈으로 보이는 파형의 진동이 끝나는 점을 종료 끝점으로 정하였다.

실험에서는 음성구간의 기준 끝점과 세 가지 특징벡터를 사용하여 검출한 시작 끝점과 종료 끝점 간의 시간 차이가 지정된 범위 내에 들어가는지를 조사하였다. 표 3.1은 시작 끝점 검출율이며, 표 3.2는 종료 끝점 검출율이다. 사용한 데이터는 원 녹음 자료와 20데시벨의 백색잡음을 추가한 자료, 10데시벨의 백색잡음을 추가한 자료이며, 특징벡터 $\psi(n)$, obsFeat 그리고 $\delta(n)$에 의한 검출률을 조사하였다. 표

3.1에서 'Clean'은 녹음된 원 자료에 관한 것이며, '20dB'과 '10dB'은 각각 20dB과 10dB의 백색잡음을 추가한 자료에 관한 것이다.

표 3.1의 시작 끝점 검출률을 살펴보면, 'Clean' 상태에서의 검출율은 세 가지 특징벡터의 경우 모두 비슷하며 높은 검출율을 나타내고 있다. 소음이 추가된 '10dB' 혹은 '20dB'의 상태에서는 $\psi(n)$과 obsFeat가 비슷하게 검출률이 감소하는 반면, $\delta(n)$은 100% 근처의 검출률을 나타내고 있다. 이는 백색잡음의 추가에도 불구하고 $\delta(n)$은 시작 끝점의 검출에서 검출율이 감소하지 않음을 나타내는 것이다.

표 3.2는 음성 구간의 종료 끝점의 검출에 관한 결과이다. 세 가지 특징벡터의 'Clean' 상태에서의 검출률에서는 obsFeat가 약간 높음을 나타내고 있다. 그러나, 실제 음성의 인식에 적용되는 90ms 범위 내에 드는 검출률은 세 가지 특징벡터 모두 우수함을 나타내고 있다. 그러나, 백색잡음이 추가된 경우에는 $\psi(n)$과 obsFeat는 검출률이 크게 떨어지고 있으며, 특히 obsFeat의 감소의 정도가 더 크다. 예를 들어 이들 두 특징벡터에 대하여 90ms 범위 내에서의 검출율은 '20dB' 경우에서 70%에 못 미치며, 이는 'Clean' 상태에서의 검출율 99% 혹은 97%에 비해 크게 감소한 것이다. 이에 반하여 $\delta(n)$은 그 감소의 정도가 훨씬 적으며, 90ms 범위 내에서의 검출률을 보면 '20dB' 경우의 검출률 96%는 'Clean' 상태의 검출률 98%와 거의 같음을 볼 수 있다. 더욱이 '90ms'와 '10dB'에서의 검출률은 $\psi(n)$, obsFeat 그리고 $\delta(n)$의 각각에 대하여 32%, 22% 그리고 73%로 나타나고 있어 $\delta(n)$이 주어진 백색잡음 환경에서 매우 강건함을 나타내고 있다.

특징벡터 $\delta(n)$의 이와 같은 잡음 강건 성질은 인접한 여러 프레임의 에너지를 상대적으로 비교하기 때문에 나타나는 것이라고 할 수 있다. 즉, 동질적 잡음이 존재하는 경우에, 음성 신호 등으로 생성되는 추가적 에너지 증가를 검출해 내는 특징 때문이라고 할 수 있다.

## 참고문헌


김경태, 강성훈, 이용주, 정유현, 박찬경, 이정철, 류준형, 한희일 (1990). <대어휘 연속음성 인식을 위한 음소인식 기술 개발>, 최종연구보고서, 과학기술처.

박기상, 석수영, 정호열, 정현열 (2002). 대역에너지를 이용한 잡음음성의 끝점검출 알고리즘, <한국음향학회: 춘계 학술대회지>, 91–94.

석종원, 배건성 (1996). Wavelet 변환을 이용한 잡음 음성의 끝점 검출, <대한전자공학회, 학술대회 논문집(신호처리합동)>, **9**, 69–72.

이용형, 오창혁 (2001). HMM based endpoint detection for speech signals, <한국통계학회: 추계학술발표회 논문집>, 75–76.

정대성, 김정곤, 김형순 (2003). 화자인식을 위한 강인한 끝점 검출 알고리즘, <대한음성학회: 학술대회지>, 137–140.

홍정우, 오창혁 (2008). 숨은마코프모형을 이용하는 음성구간 추출을 위한 특징벡터, <한국통계학회 논문집>, **15**, 293–302.

Acero, A., Crespo, C., Torre, C. de la and Torrecilla, J. C. (1993). Robust HMM-based endpoint detector, In *EUROSPEECH '93*, 1551–1554.

Chen, N. and Hu, Y. (2007). Pitch detection algorithm based on Teager energy operator and spatial correlation function, In *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, **5**, 2456–2460.

Kaiser, J. F. (1990). On a simple algorithm to calculate the 'energyapos' of a signal, In *ICASSP '90*, **1**, 381–384.

Lamel, L. F., Rabiner, L. R., Rosenberg, A. E. and Wilpon, J. G. (1981). An improved endpoint detector for isolated word recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **29**, 777–785.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition, In *Proceedings of the IEEE*, **77**, 257–285.

Rabiner, L. R. and Sambur, M. R. (1975). An algorithm for determining the endpoints of isolated utterances, *The Bell System Technical Journal*, **54**, 297–315.

Ying, G. S., Mitchell, C. D. and Jamieson, L. H. (1993). Endpoint detection if isolated utterances based on a modified Teager energy measurement, In *ICASSP '93*, **2**, 732–735.

# A Discrete Feature Vector for Endpoint Detection of Speech with Hidden Markov Model

Jeiky Lee[1] · Chang Hyuck Oh[2]

[1]Dept. of Statistics, Yeungnam University; [2]Dept. of Statistics, Yeungnam University

(Received July 2008; accepted September 2008)

**Abstract**

The purpose of this paper is to suggest a discrete feature vector, robust in various levels of noisy environment and inexpensive in computation, for detection of speech segments and is to show such properties of the feature with real speech data. The suggested feature is one dimensional vector which represents slope of short term energies and is discretized into three values to reduce computational burden of computations in HMM. In experiments with speech data, the method with the suggested feature vector showed good performance even in noisy environments.

Keywords: Endpoint detection, discretization, hidden Markov model, white noise.

[1]Master of Science, Dept. of Statistics, Yeungnam University, Kyungsan, Kyungbuk 712-749, Korea. E-mail: jaikylee290@hotmail.com

[2]Corresponding author: Professor, Dept. of Statistics, Yeungnam University, Kyungsan, Kyungbuk 712-749, Korea. E-mail: choh@yu.ac.kr