

EM 알고리즘을 이용한 재귀적인 음소분리

강병옥, 정홍

한국전자통신연구원 네트워크 연구소 음성정보연구센터,
포항공과대학교 전자전기공학과

Recursive Segmentation of Speech Signals using Expectation-Minimization

Byung-Ok Kang, Hong Jung

SpeechTechnology Research Center, Network Laboratory, ETRI

EE DEPT, POSTECH

E-mail : bokang@etri.re.kr

요약

본 논문에서는 입력음성신호로부터 음소간의 경계를 찾는 문제를 풀기 위해 재귀적인 방식으로 EM 알고리즘을 적용한다. 즉, 예상되는 두 끝점 사이의 부분을 현재의 프레임 n 이라고 하면, 그 전 프레임 $n-1$ 에서 구해진 끝점이 주는 정보와 그 끝점으로부터 이어지는 음성샘플로부터 현재 프레임의 끝점을 구한다. 또한 현재의 프레임 n 에서 끝점을 추정해 내면, 그 추정된 끝점과 그 점 이후에 이어지는 음성샘플값으로부터 다음 프레임 $n+1$ 의 끝점을 구한다. 이러한 방식을 재귀적인 음소분리 방식이라고 한다. 그리고, 각 프레임에서 끝점을 구하기 위해서는 끝점의 좌표를 추정해야 할 파라미터로 하고, 그 주변의 음성샘플 값을 관찰 값으로 하여 EM(Expectation and Maximization) 알고리즘을 이용한다. 이 EM 알고리즘을 이용한 재귀적인 음소분리 방식을 실제 음성 DB 로부터 음소쌍을 추출하여 테스트 했을 때 약 5 회의 EM 반복 후에 경계값으로 수렴함을 볼 수 있었다.

1. 서론

본 알고리즘에서는 음소분리를 프레임 단위로 수행해 나간다. 즉, 예상되는 두 끝점 사이의 부분을 현재의 프레임 n 이라고 하면, 그 전 프레임 $n-1$ 에서 구해진 끝점과 그 끝점으로부터 이어지는 음성 샘플로부터 현재 프레임의 끝점을 구한다. 또한, 현재의 프레임 n 에서 끝점을 추정해 내면, 그 추정된 끝점과 그 점 이후에 이어지는 음성샘플 값으로부터 다음 프레임 $n+1$ 의 끝점을 구한다. 이러한 방식을 재귀적인 음소분리 방식이라고 한다.

그리고, 각 프레임에서 끝점을 구하기 위해서 끝점의 좌표를 추정해야 할 파라미터로 하고, 그 주변의 음성샘플 값을 관찰 값으로 하여 EM 알고리즘을 이용

한다.

본 논문의 구성은 다음과 같다. 제 1 장에서는 음소분리를 위한 통계적 모델을 세우고, 제 2 장에서는 EM 알고리즘을 이용하여 음소값 끝점을 추출하는 과정을 전개한다. 세부적으로 EM 알고리즘의 E-step 과 M-step 의 전개과정을 기술하였다. 그리고, 제 3 장에서 실험결과를 보여준다.

2. 음소 분리를 위한 통계적 모델

여기서는 EM 알고리즘을 이용하여 재귀적으로 음소 분리를 수행하기 위해서 통계적인 모델을 세우도록 한다. 본 알고리즘은 음성 신호로부터 특징 벡터를 추출해 특징 벡터열을 다루는 방식을 사용하지 않고, 각 음성 샘플 단위로 프로세싱한다.

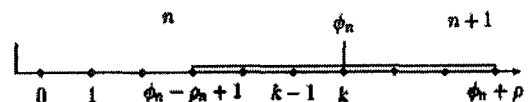


그림 3.1: 시간축에서 각 프로세스들의 정의

우선, 시간축에서의 각 프로세스들을 정의해 보자. 그림 2.1 에서 ϕ_n 은 프레임 n 에서의 끝점의 좌표이다. 따라서 현재의 프레임에서 끝점을 찾는 문제는 결국 ϕ_n 을 추정하는 문제인데, 이를 위해서는 $[0, K_n]$ 에서의 음성샘플을 필요로 한다. $S(\phi_n)$ 은 ϕ_n 주위의 점들의 집합으로서 $2\rho_n$ 의 일정한 크기를 갖는다. 즉, $S(\phi_n) = \{k | k \in [\phi_n - \rho_n + 1, \phi_n + \rho_n]\}$ 이고, $S(\phi_n) \subseteq [0, K_n]$ 이다. 결국 $S(\phi_n)$ 은 ϕ_n 의 함수임을 알 수 있다. 프레임 n 에서 음성샘플로서 ϕ_n 주위의 값을 $Y_n = \{Y_n(k) | k \in S(\phi_n)\}$ 로 정의한다. 또 하나의 프로세스로서 X_n 이 있다. X_n 는 노이즈가 제거된 원래의 음성신호로서 $X_n = \{X_n(k) | k \in S(\phi_n)\}$ 이고, $X_n - Y_n$ 는 가우시안 분포로 모델링이 가능하다.

여기서 $Z_n = \{X_n(k), Y_n(k)\}$ 이고, 역사 $Z_n = \{Z_n(k) | k \in S(\phi_n)\}$ 이다. Y_n 은 유일하게 관찰 가능한 프로세스로서 complete data인 Z_n 의 일부분이다.

각 프레임의 특성은 대개 그 직전 프레임에 의해서만 영향을 받는다고 할 수 있으므로 ϕ_n 은 Markov 프로세스로 모델링이 가능하고, 음성신호의 특성에 의해서 X_n, Y_n 은 확실히 Markov 프로세스이다.

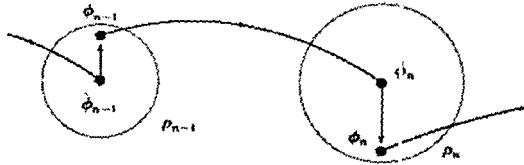


그림 3.2: 재귀적인 음소분리의 개념

그림 3.2는 재귀적인 음소분리의 개념을 보여주는 그림이다. 현재의 프레임에서 끝점을 찾기 위해서는 과거의 프레임 $n-1$ 에서의 결과로부터 두가지 데이터를 받아야 한다. 즉, ϕ_n 의 초기치가 되는 ϕ_{n-1} 와 그 주위의 필요한 데이터의 범위인 ρ_n 이다. 이들 정보와 해당범위 $S(\phi_n)$ 의 Y_n 으로부터, 몇 번의 순환을 거쳐서 ϕ_n 을 추정하게 된다. 마찬가지로 방식으로 현재의 ϕ_n 으로부터 다음 프레임 $n+1$ 의 끝점을 추정한다.

3. EM 알고리즘을 이용한 음소간 끝점의 추출

여기서는 앞 절에서의 모델을 기반으로 하여, 음소분리에 적용하기 위해 필요한 식을 유도하겠다. EM 알고리즘[1]을 적용하기 위해서 먼저 다음과 같은 posterior 확률식이 필요하다.

$$P(\phi_n | z^n) = \frac{P(z_n | \phi_n, z^{n-1})P(\phi_n | z^{n-1})}{P(z_n | z^{n-1})} \quad (3.1)$$

앞장에서 $\hat{\phi}_n$ 은 앞 프레임에서의 z^{n-1} 를 이용하여 얻어진 값이므로 $P(\phi_n | z^{n-1}) = P(\phi_n | \hat{\phi}_n)$ 이 성립한다. 이를 식 (3.1)에 대입하면,

$$P(\phi_n | z^n) = \frac{P(z_n | \phi_n, z^{n-1})P(\phi_n | \hat{\phi}_n)}{P(z_n | z^{n-1})} \quad (3.2)$$

여기서 Z_n 은 Markov 프로세스라 볼 수 있으므로 식 (3.2)는 다음과 같이 간단히 할 수 있다.

$$P(\phi_n | z^n) = \frac{P(z_n | \phi_n, z_{n-1})P(\phi_n | \hat{\phi}_n)}{P(z_n | z^{n-1})} \quad (3.3)$$

위 식 (3.2)에 EM을 적용하면, E-스텝과 M-스텝은 아래와 같다.

$$\begin{cases} E: Q(\phi_n | \hat{\phi}_n^{(p)}) = E[\log P(z_n | \phi_n, z_{n-1}) | y, \hat{\phi}_n^{(p)}, z_{n-1}] \\ M: \hat{\phi}_n^{(p+1)} = \operatorname{argmax}_{\phi_n} [Q(\phi_n | \hat{\phi}_n^{(p)}) + \log P(\phi_n | \hat{\phi}_n)] \end{cases} \quad (3.4)$$

3.1. E-스텝의 계산

먼저 E-step을 유도해 보자. 먼저 $P(z_n | \phi_n, z_{n-1})$ 는 다음과 같이 전개된다.

$$\begin{aligned} P(z_n | \phi_n, z_{n-1}) &= P(y_n, x_n | \phi_n, z_{n-1}) \\ &= P(y_n | x_n, \phi_n, z_{n-1})P(x_n | \phi_n, z_{n-1}) \end{aligned} \quad (3.5)$$

Y_n 은 X_n 에만 의존하므로, $P(y_n | x_n, \phi_n, z_{n-1}) = P(y_n | x_n)$ 가 성립하고, 두번째 텀은 $P(x_n | \phi_n, z_{n-1}) = P(x_n | \phi_n, x_{n-1})$ 와 같이 단순화가 가능하고, 또 다시 X_n 과 X_{n-1} 이 서로 독립이라고 가정하면, 아래와 같이 정리된다.

$$P(y_n | x_n, \phi_n, z_{n-1}) = P(y_n | x_n)P(x_n | \phi_n, x_{n-1}) \quad (3.6)$$

이제부터 $P(y_n | x_n)$ 와 $P(x_n | \phi_n, x_{n-1})$ 를 자세히 유도해 보겠다. $Y_n(k) - x_n(k)$ 가 $k \in S(\phi_n)$ 에 대해서 $N(0, \sigma)$ 인 i.i.d 가우시안 분포를 이룬다고 가정하면, 다음과 같이 전개가 가능하다.

$$L(P(y_n | x_n)) = -\frac{1}{2} \sigma^{-2} \sum_{k \in S(\phi_n)} (x_n(k) - y_n(k))^2 \quad (3.7)$$

가 현재 프레임에서의 끝점이라고 보면, $x_n = \{x(k) | k \in [\phi_n - \rho + 1, \phi_n + \rho]\}$ 은 두부분 $x_1 = \{x(k) | k \in [\phi_n - \rho + 1, \phi_n]\}$ 과 $x_2 = \{x(k) | k \in [\phi_n + 1, \phi_n + \rho]\}$ 으로 나뉘어 서로 다른 분포를 이루게 된다. 즉,

$$P(x_n | \phi_n) = P(x_1)P(x_2) \quad (3.8)$$

$X_n(k)$ 은 확실히 Markov 프로세스라고 볼 수 있다. 따라서, MRF와 Gibbs의 동일성에 대한 Hammersley-Clifford 정리[3]에 의하여 그 확률분포는 이스토펜셀 계열을 이룬다. X_1 과 X_2 를 평균이 0인 가우시안 분포를 이룬다고 모델링하면, 식(3.8)은 아래식과 같이 유도된다.

$$\begin{aligned} L(P(x_n | \phi_n)) &= -\frac{1}{2} [\sigma_1^{-2}(\phi_n) \sum_{k \in [\phi_n - \rho + 1, \phi_n]} (x_n(k) - x_n(k-1))^2 \\ &\quad + \sigma_2^{-2}(\phi_n) \sum_{k \in [\phi_n + 1, \phi_n + \rho]} (x_n(k) - x_n(k-1))^2] \end{aligned} \quad (3.9)$$

여기서 σ_1 와 σ_2 는 표준편차를 의미한다. 결국, 식(3.7)과 식(3.9)에 의하여, 식(3.5)는 식(3.10)과 같이 정리된다.

$$\begin{aligned} L(P(z_n | \phi_n)) &= -\frac{1}{2} [\sigma^{-2}(\phi_n) \sum_{k \in [\phi_n - \rho + 1, \phi_n + \rho]} (x(k) - y(k))^2 \\ &\quad + \sigma_1^{-2}(\phi_n) \sum_{k \in [\phi_n - \rho + 1, \phi_n]} (x_n(k) - x_n(k-1))^2 \\ &\quad + \sigma_2^{-2}(\phi_n) \sum_{k \in [\phi_n + 1, \phi_n + \rho]} (x_n(k) - x_n(k-1))^2] \end{aligned} \quad (3.10)$$

위 식 (3.10)의 기대값을 구하기 위해서는 $X(k)$ 의 모멘트를 반드시 구해야만 한다. 이는 앞절에서의 가정을 다시 사용하여,

$$y_n(k) = x_n(k) + w_n(k), \quad \forall k \quad (3.11)$$

이고 $w_n(k)$ 는 $N(0, \sigma)$ 이고, $E[w_n(i) w_n(j)] = 0, \forall i \neq j$ 인
가우시안 노이즈로 가정하면 문제는 간단해진다. 즉,
모든 $X_n(k)$ 의 모멘트는 $Y_n(k)$ 의 모멘트로 바꿀 수 있다.
또한 $Y_n(k)$ 는 ergodic 이므로, 모든 $Y_n(k)$ 의 모멘트는
그것의 샘플 평균으로 바꿀 수 있다. 이러한 가정하에
서 식 (3.7)의 기대값은 아래와 같이 구해진다.

$$E[L(P(y_n | x_n)) | y_n, \hat{\phi}_n^{(p)}] = -\frac{1}{2} \sigma^{-2} [2 \sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k)^2 - \frac{1}{\rho} (\sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k))^2], \quad (3.12)$$

식 (3.9)의 기대값은 조금 복잡하여, 다음의 두가지
경우에 의해서 식이 나뉘게 된다. 즉,

$\hat{\phi}_n - \rho + 1 < \hat{\phi}_n^{(p)} \leq \hat{\phi}_n$ 혹은 $\hat{\phi}_n + 1 < \hat{\phi}_n^{(p)} \leq \hat{\phi}_n + \rho$ 의
두 경우에 의해 식 (3.13)과 식 (3.14)과 같이 다른

형태로 계산된다. $\hat{\phi}_n - \rho + 1 < \hat{\phi}_n^{(p)} \leq \hat{\phi}_n$ 일 경우에는,

$$E[L(P(x_1)) | y_n, \hat{\phi}_n^{(p)}] = -\frac{1}{2} \sigma_1^{-2} (\hat{\phi}_n) (\rho - 1) \left[\frac{1}{\hat{\phi}_n^{(p)} - \hat{\phi}_n + \rho - 1} \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n^{(p)}\}} y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2) \right. \\ \left. + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} - 1} \sum_{k \in \{\hat{\phi}_n^{(p)} + 2, \hat{\phi}_n + 1\}} (y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2)) \right. \\ \left. - \frac{1}{2} \sigma_2^{-2} (\hat{\phi}_n) \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n^{(p)}\}} y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2) \right. \\ \left. + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} - 1} \sum_{k \in \{\hat{\phi}_n + 2, \hat{\phi}_n + \rho\}} (y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2)) \right]$$

$\hat{\phi}_n + 1 < \hat{\phi}_n^{(p)} \leq \hat{\phi}_n + \rho$ 일 경우에는,

$$E[L(P(x_1)) | y_n, \hat{\phi}_n^{(p)}] = -\frac{1}{2} \sigma_2^{-2} (\hat{\phi}_n) (\rho - 1) \left[\frac{1}{\hat{\phi}_n^{(p)} - \hat{\phi}_n - 1} \sum_{k \in \{\hat{\phi}_n + 2, \hat{\phi}_n^{(p)}\}} y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2) \right. \\ \left. + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} + \rho - 1} \sum_{k \in \{\hat{\phi}_n^{(p)} + 2, \hat{\phi}_n + \rho\}} (y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2)) \right. \\ \left. - \frac{1}{2} \sigma_1^{-2} (\hat{\phi}_n) \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n + 1\}} y_n(k)^2 (y(k)^2 - 2y(k)y(k-1) + y(k-1)^2) \right]$$

와 같이 얻어진다.

위 두 경우를 한 식으로 묶기 위해서, 유닛, 스텝 함수
 $u(x)$ 를 사용하여 다시 정리하면,

$$E[L(P(z, \hat{\phi}_n)) | y, \hat{\phi}_n^{(p)}] = -\frac{1}{2} \sigma^{-2} [2 \sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k)^2 - \frac{1}{\rho} (\sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k))^2] + u(\hat{\phi}_n - \hat{\phi}_n^{(p)}) \left\{ -\frac{1}{2} \sigma_1^{-2} (\hat{\phi}_n) (\rho - 1) \left[\frac{1}{\hat{\phi}_n^{(p)} - \hat{\phi}_n + \rho - 1} \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n^{(p)}\}} (y(k) - y(k-1))^2 + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} - 1} \sum_{k \in \{\hat{\phi}_n^{(p)} + 2, \hat{\phi}_n + \rho\}} (y(k) - y(k-1))^2 \right] - \frac{1}{2} \sigma_2^{-2} (\hat{\phi}_n) \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n + 1\}} (y_n(k) - y_n(k-1))^2 \right\} + (1 - u(\hat{\phi}_n - \hat{\phi}_n^{(p)}))$$

$$\left\{ -\frac{1}{2} \sigma_2^{-2} (\hat{\phi}_n) (\rho - 1) \left[\frac{1}{\hat{\phi}_n^{(p)} - \hat{\phi}_n - 1} \sum_{k \in \{\hat{\phi}_n + 2, \hat{\phi}_n^{(p)}\}} (y(k) - y(k-1))^2 + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} + \rho - 1} \sum_{k \in \{\hat{\phi}_n^{(p)} + 2, \hat{\phi}_n + \rho\}} (y(k) - y(k-1))^2 \right] - \frac{1}{2} \sigma_1^{-2} (\hat{\phi}_n) \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n + 1\}} (y(k) - y(k-1))^2 \right\} \quad (3.15)$$

와 같다. 식 (3.15)를 식 (3.4)에 대입하면,

$$Q(\hat{\phi}_n | \hat{\phi}_n^{(p)}) = -\frac{1}{2} \sigma^{-2} [2 \sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k)^2 - \frac{1}{\rho} (\sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k))^2] + u(\hat{\phi}_n - \hat{\phi}_n^{(p)}) \left\{ -\frac{1}{2} \sigma_1^{-2} (\hat{\phi}_n) (\rho - 1) \left[\frac{1}{\hat{\phi}_n^{(p)} - \hat{\phi}_n + \rho - 1} \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n^{(p)}\}} (y(k) - y(k-1))^2 + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} - 1} \sum_{k \in \{\hat{\phi}_n^{(p)} + 2, \hat{\phi}_n + \rho\}} (y(k) - y(k-1))^2 \right] - \frac{1}{2} \sigma_2^{-2} (\hat{\phi}_n) \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n + 1\}} (y_n(k) - y_n(k-1))^2 \right\} + (1 - u(\hat{\phi}_n - \hat{\phi}_n^{(p)})) \left\{ -\frac{1}{2} \sigma_2^{-2} (\hat{\phi}_n) (\rho - 1) \left[\frac{1}{\hat{\phi}_n^{(p)} - \hat{\phi}_n - 1} \sum_{k \in \{\hat{\phi}_n + 2, \hat{\phi}_n^{(p)}\}} (y(k) - y(k-1))^2 + \frac{1}{\hat{\phi}_n - \hat{\phi}_n^{(p)} + \rho - 1} \sum_{k \in \{\hat{\phi}_n^{(p)} + 2, \hat{\phi}_n + \rho\}} (y(k) - y(k-1))^2 \right] - \frac{1}{2} \sigma_1^{-2} (\hat{\phi}_n) \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n + 1\}} (y(k) - y(k-1))^2 \right\} \quad (3.16)$$

위 식은 다음의 식 (3.17)과 같은 함수를 정의하여 사
용하면 간단한 형태로 정리된다.

$$\Gamma(x, y; a, b; \sigma; n) = -\frac{1}{2\sigma^2(b-a+1)} \sum_{k \in [a, b]} (x(k) - y(k-1))^2$$

이를 이용하여 다시 정리하면,

$$Q(\hat{\phi}_n | \hat{\phi}_n^{(p)}) = -\frac{1}{2} \sigma^{-2} [2 \sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k)^2 - \frac{1}{\rho} (\sum_{k \in \{\hat{\phi}_n - \rho + 1, \hat{\phi}_n + \rho\}} y_n(k))^2] + u(\hat{\phi}_n - \hat{\phi}_n^{(p)}) \left\{ (\rho - 1) [\Gamma(y, y; \hat{\phi}_n - \rho + 2, \hat{\phi}_n^{(p)}; \sigma_1(\hat{\phi}_n); 2) + \Gamma(y, y; \hat{\phi}_n^{(p)} + 2, \hat{\phi}_n; \sigma_1(\hat{\phi}_n); 2) + \Gamma(\hat{\phi}_n + 2, \hat{\phi}_n + \rho; \sigma_2(\hat{\phi}_n); 2)] + (1 - u(\hat{\phi}_n - \hat{\phi}_n^{(p)})) \left\{ (\rho - 1) [\Gamma(y, y; \hat{\phi}_n - \rho + 2, \hat{\phi}_n; \sigma_1(\hat{\phi}_n); 2) + \Gamma(\hat{\phi}_n + 2, \hat{\phi}_n^{(p)}; \sigma_2(\hat{\phi}_n); 2) + \Gamma(\hat{\phi}_n^{(p)} + 2; \hat{\phi}_n + \rho; \sigma_2(\hat{\phi}_n); 2)] \right\} \right\} \quad (3.18)$$

위 식에서 표준편차 $\sigma_1(\hat{\phi}_n)$ 와 $\sigma_2(\hat{\phi}_n)$ 를 구하기 위해,
다시 가우시안 가정과 $Y_n(k)$ 가 ergodic 이라는 것을 이
용하면,

$$\sigma_1^2(\hat{\phi}_n) = E(x(k) - x(k-1))^2 = E(y(k) - y(k-1))^2 \\ = \frac{1}{\rho - 1} \sum_{k \in \{\hat{\phi}_n - \rho + 2, \hat{\phi}_n + 1\}} (y(k) - y(k-1))^2$$

$$= -\Gamma(y, y; \hat{\phi}_n - \rho + 2, \hat{\phi}_n; \frac{1}{2}; 2) \quad (3.19)$$

$$\sigma_2^2(\hat{\phi}_n) = E(x(k) - x(k-1))^2 = E(y(k) - y(k-1))^2 \\ = \frac{1}{\rho - 1} \sum_{k \in \{\hat{\phi}_n + 2, \hat{\phi}_n + \rho\}} (y(k) - y(k-1))^2$$

$$= -\Gamma(y, y; \hat{\phi}_n + 2, \hat{\phi}_n + \rho; \frac{1}{2}; 2) \quad (3.20)$$

3.2. M-스텝의 계산

M-step 은 $P(\phi_n | \hat{\phi}_n)$ 텀을 포함한다. 이 값은 ϕ_n 과 $\hat{\phi}_n$ 사이의 추정 오차를 의미하고, 다음과 같은 예러 가우시안으로 가정할 수 있다. 즉,

$$L(p(\phi_n | \hat{\phi}_n)) = -\frac{1}{\sigma_3(\phi_n)} (\phi_n - \hat{\phi}_n)^2 \quad (3.21)$$

식 (3.21)과 식 (3.18)을 식 (3.4)에 대입해 정리하면, $L(\phi_n) =$

$$-\frac{1}{2} \sigma^{-2} [2 \sum_{k \in (\phi_n - \rho + 1, \phi_n + \rho)} y_n(k)^2 - \frac{1}{\rho} (\sum_{k \in (\phi_n - \rho + 1, \phi_n + \rho)} y_n(k))] + u(\phi_n - \phi_n^{(p)}) + \Gamma(y, y; \phi_n - \rho + 2, \phi_n^{(p)}; \sigma_1(\phi_n); 2) + \Gamma(y, y; \phi_n^{(p)} + 2, \phi_n; \sigma_1(\phi_n); 2) + \Gamma(\phi_n + 2, \phi_n + \rho; \sigma_2(\phi_n); 2) + (1 - u(\phi_n - \phi_n^{(p)}))(\rho - 1) + \Gamma(y, y; \phi_n - \rho + 2, \phi_n; \sigma_1(\phi_n); 2) + \Gamma(\phi_n + 2, \phi_n^{(p)}; \sigma_2(\phi_n); 2) + \Gamma(\phi_n^{(p)} + 2, \phi_n + \rho; \sigma_2(\phi_n); 2) + \frac{1}{\sigma_3(\phi_n)} (\phi_n - \hat{\phi}_n)^2 \quad (3.22)$$

이고, 결국 구하고자 하는 해는

$$\phi_n^{(p+1)} = \arg \min_{\phi_n} L(\phi_n) \quad (3.23)$$

로부터 얻어진다.

4. 실험결과

앞에서 전개한 EM 을 이용한 재귀적인 음소분리를 실제로 구현하여 몇 개의 음소쌍에 대해서 음소분리의 성능을 테스트 해 보았다. 여기서 음소쌍은 TIMIT 데이터베이스로부터 몇 개의 쌍을 선택하였다. TIMIT 은 16kHz, 16bits 인 문장 데이터베이스이다. 실험을 단순화 하기 위해서 과거의 프레임으로부터 얻는 정보인 ϕ_n 과 ρ 는 주어진 값을 사용하였고, EM 순환 횟수는 5 회로 하였다. 아래의 그림들은 각 음소쌍에 대해서 EM 순환 을 거치면서 음소간의 경계를 찾아가는 과정들을 보여 준다. 각 경우에 대해서 마지막 그림은 최종결과와 정 답과의 비교를 보여준다. 그림 3.3 과 그림 3.4 의 경 우는 '목음-sh' 음소쌍에 대한 음소분리 결과를 보여 준다. 정답인 hand segmentation 결과는 2180 번째 샘플이 경계가 되는데, 그림 3.3 은 ϕ_n 이 2800 으로 정답 의 오른쪽, 그림 3.4 는 1300 으로 정답의 왼쪽에서 출 발하여 결국 2122 인 같은 곳에서 경계값으로 수렴함 을 볼 수 있다. 정답과는 3.6ms 차이로서 거의 일치한 다고 볼 수 있다. 정답은 hand segmentation 결과는 6300 번째 샘플이고, 6700 을 ϕ_n 으로 하여 결국 6173 번 째 샘플을 경계값으로 찾아내었다. 이는 7.9ms 차이로 서 역시 좋은 결과를 보여준다.

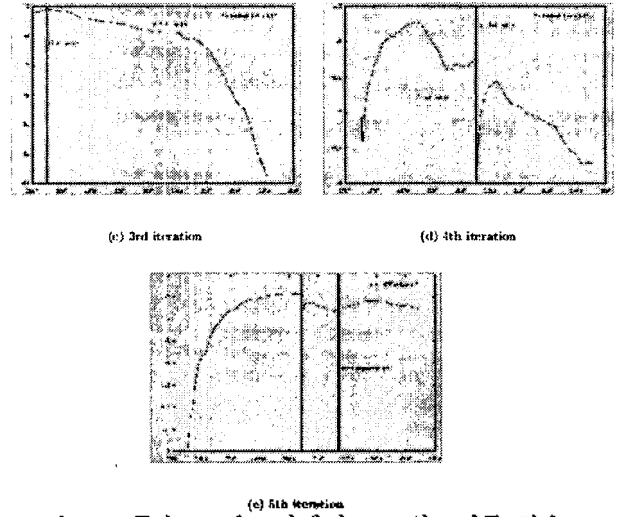
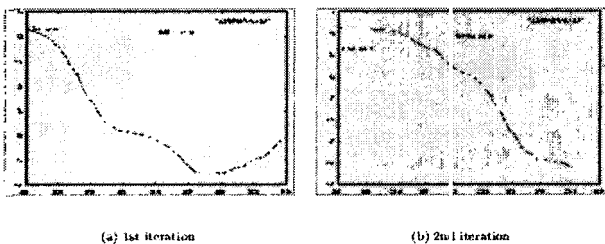


그림 3.3: '목음-sh' 음소쌍에의 EM 알고리즘 적용

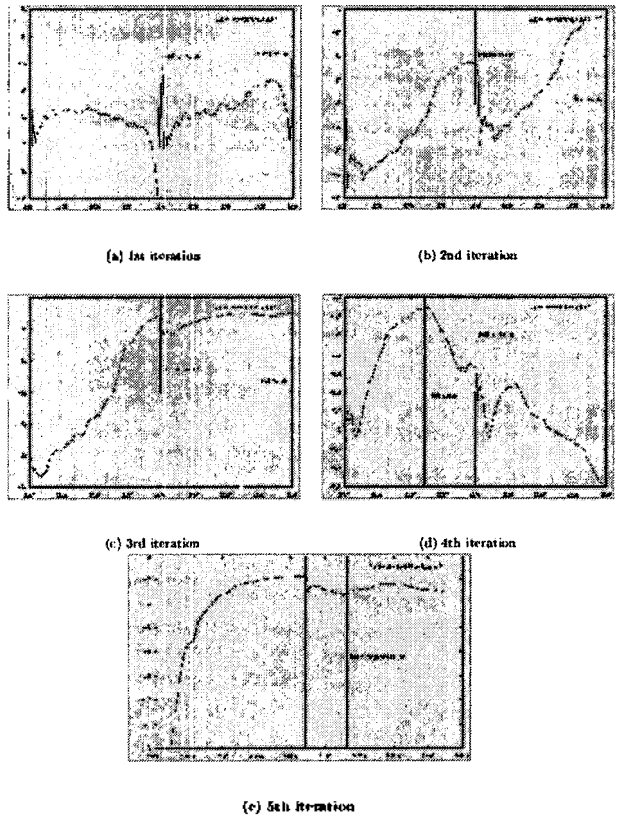


그림 3.4: '목음-sh' 음소쌍에의 EM 알고리즘 적용(2)

참고문헌

- [1] P. Green, "On use of the EM algorithm for penalized likelihood estimation", J. Roy. Statist. Soc. Ser. B, 52 (3): 443-452, 1990.
- [2] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", J. Roy. Statist. Soc. Ser. B, vol 39, no. 1, 1976.
- [3] J.E.Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems", J. Royal Statistical Society. Series B, vol. 36, pp. 192-236, 1974.