

# 능동형 기울기 가중치 제약에 기반한 환경소리 인식시스템용 DTW 알고리즘

정영진<sup>†</sup>, 이윤정<sup>\*\*</sup>, 김필운<sup>\*\*\*</sup>, 김명남<sup>\*\*\*\*</sup>

## 요 약

농자들은 청각장애를 가지고 있기 때문에 알람, 도어 벨, 사이렌, 경적, 전화 벨 등과 같은 유용한 소리정보를 인식할 수 없다. 이러한 문제를 해결하기 위해서는 주요한 환경소리 인식방법을 탑재한 휴대형 청각 보조 장구가 요구되어진다. 본 논문에서는 환경소리 인식시스템에 적용할 수 있는 능동형 기울기 가중치 제약 방식의 새로운 동적 시간정합 알고리즘을 제안하였다. 환경소리 인식방법은 크게 세 단계로 구성된다. 첫 번째 단계에서는 소리의 주파수 성분과 크기를 이용하여 시작점과 끝점을 추출한다. 두 번째 단계에서는 추출된 구간에 대하여 특징을 추출하며 세 번째 단계에서는 추출된 특징들을 분류한다. 실험 결과 제안한 방법의 인식률이 거의 90%가 되었다. 그리고 기존의 동적 시간정합 알고리즘과 비교하였을 때 인식율에 있어서 약 20%정도의 개선이 있었다. 따라서 제안된 방법을 사용하여 농자가 환경소리를 인식할 수 있는 휴대형 청각 보조 장구가 개발된다면 그들의 생활에 편리함을 줄 수 있을 것으로 판단된다.

## Active Slope Weighted-Constraints Based DTW Algorithm for Environmental Sound Recognition System

Young Jin Jung<sup>†</sup>, Yun Jung Lee<sup>\*\*</sup>, Pil Un Kim<sup>\*\*\*</sup>, Myoung Nam Kim<sup>\*\*\*\*</sup>

## ABSTRACT

The deaf can not recognize useful sound informations such as alarm, doorbell, siren, car horn, and phone ring etc., because they have the hearing impairment. To solve this problems, portable hearing assistive devices which have suitable environment sound recognition methods are needed. In this paper, the DTW algorithm for sound recognition system with new active slope weighting constraint method was proposed. The environment sound recognition methods consist of three processes. First process is extraction of start point and end point using frequency and amplitude of sound. Second process is extraction of features and third process is classification of features for given segments. As a result of the experiment, the recognition rate of the proposed method is over 90%. And, the recognition rate of the proposed method increased about 20% than the conventional algorithm. Therefore if there are developed portable assistive devices which use developed method to recognize environment sound for hearing-impaired persons, they could be more convenient in life.

**Key words:** Hearing impairment(청각장애), Environment sound recognition(환경소리 인식), Dynamic time warpping(DTW)(동적 시간정합), Active slope weighting constraint(능동형 기울기 가중치 제약)

※ 교신저자(Corresponding Author) : 김명남, 주소 : 대구광역시 중구 동인동 2가 101번지(700-422), 전화 : 053)420-5252, FAX : 053)420-5264, E-mail : kimmn@knu.ac.kr  
접수일 : 2008년 1월 15일, 완료일 : 2008년 3월 4일

<sup>†</sup> 연세대학교 대학원 의공학과 박사과정

(E-mail : microbme@gmail.com)

<sup>\*\*</sup> 준회원, 경북대학교 대학원 의용생체공학과 박사과정

(E-mail : zeros97@hanmail.net)

<sup>\*\*\*</sup> 준회원, 경북대학교 대학원 의용생체공학과 박사과정  
(E-mail : whitegleam@paran.com)

<sup>\*\*\*\*</sup> 종신회원, 경북대학교 의대 의공학교실 주임교수

※ 본 연구는 보건복지부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임. (과제고유번호 : 02-PJ3-PG6-EV10-0001) 또한 이 논문은 2008년도 두뇌한국21사업에 의하여 지원되었음.

## 1. 서 론

청각 장애인들은 일상생활에서 주요한 환경소리 정보를 인식하지 못하여 많은 불편을 겪는다. 이러한 장애인을 위한 많은 기기들이 있지만 휴대하면서 환경소리를 인식하여 장애인에게 알려줄 수 있는 장치 및 알고리즘에 대한 연구는 많이 이루어지지 않았다. 환경소리 인식 시스템 개발을 위해서는 청각 장애인들에게 중요한 환경소리 데이터베이스의 구축과 휴대형으로 개발 가능한 환경소리 정보인식 및 전달기술에 대한 연구가 진행되어야 한다. 최근까지 여러 연구자들이 환경소리 인식과 관련된 연구들을 진행해오고 있다[1,2]. Cowling은 기존의 음성인식방법들으로써 환경소리 인식에 적용한 결과에 대한 비교평가를 통하여 멜 주파수 cepstrum 계수(mel frequency cepstrum coefficients, MFCC)를 추출하여 동적 시간정합(dynamic time warping, DTW)을 사용하는 것이 높은 인식률을 가진다고 보고하였다[3]. Cowling은 8가지 환경소리를 대상으로 실험하여 70%의 인식율을 얻었으나 실제 시스템에 적용하기 위해서는 인식률의 개선이 필요하였다.

일반적인 환경소리 인식은 크게 세 단계로 구성되어진다. 첫 번째 단계는 환경소리의 유무를 판별하는 시작점 및 끝점 추출이다. 기존에는 시작점 및 끝점 추출은 소리의 크기만을 이용하거나 주파수 혹은 영교차율을 이용해 왔으나 크기를 이용할 경우 주파수 성분을 고려하지 못하는 단점이 있고 영교차율을 이용할 경우 크기 성분을 고려하지 못하는 단점을 가지고 있다[4]. 두 번째는 특징추출단계로서 멜 주파수 cepstrum 계수 연산[5], 웨이블릿 변환(wavelet transform, WT)[3] 및 위그너-빌 분포(Wigner-Ville distribution, WVD)[3] 등이 있다. 마지막으로 특징분류단계로서 동적 시간정합, 신경 회로망(artificial neural network, ANN)[6], 은닉 마코프 모델(hidden Markov model, HMM)[7] 등의 여러 가지 알고리즘이 존재한다.

본 논문에서는 환경소리 인식시스템에 적용할 수 있는 능동형 기울기 가중치 제약 방식의 새로운 동적 시간정합 알고리즘과 이를 적용한 환경소리 인식방법을 제안하였다. 제안한 방법은 먼저, 환경소리의 주파수와 크기를 동시에 고려하기 위하여 단 구간 퓨리에 변환(short-time Fourier transform, STFT)

을 사용하여 시작점 및 끝점 추출을 수행하였다. 그런 다음, 청각 특성을 잘 반영하고 있는 멜 주파수 cepstrum 계수를 사용하였다. 마지막으로 구조가 간단하며 연산량이 작아 휴대형 시스템에 적합한 동적 시간정합 알고리즘을 사용하였다. 그러나 기존의 동적 시간정합 알고리즘은 전체 경로를 제약함으로써 정상적인 탐색경로를 막아 버리는 문제점을 가지고 있다. 이러한 단점을 극복하기 위하여 능동형 기울기 가중치 제약을 통하여 상대적으로 탐색경로를 제한하는 새로운 동적 시간정합 알고리즘을 제안하였다. 또한 환경소리 인식방법이 휴대형 기기에 적용 가능하도록 시작점 및 끝점 추출 작업과 특징 추출 작업을 연관시켜 2회의 퓨리에 변환과정을 1회로 구현함으로써 연산량을 줄일 수 있었다. 제안한 방식과 기존 방식의 성능을 비교하기 위한 실험을 수행하였으며 이때 기존의 방식은 Cowling의 논문에서 언급된 멜 주파수 cepstrum 연산 방법을 그대로 이용하였다. 기존의 환경소리 인식 기술 비교논문[3]에서 사용한 Slaney의 알고리즘과 비교한 결과, 제안한 방법이 8종류 88개의 샘플들에 대한 인식률과 정확성면에서 기존의 방법에 비하여 우수함을 확인되었다.

## 2. 본 론

환경소리 인식방법은 전체적으로 그림 1과 같은 순서대로 수행된다. 먼저, 환경소리 신호에 대한 전처리와 창함수연산(windowing)을 수행한다. 그런 다음, 단 구간 퓨리에 변환으로써 소리의 시작점과 끝점을 검출하여 의미있는 구간을 분할한 다음, 이 구간에 대하여 멜 주파수 cepstrum 알고리즘을 사용하여 환경소리의 특징을 검출한다. 마지막으로 검출한 특징을 제안한 동적 시간정합 알고리즘을 이용하여 분류하여 환경소리를 인식한다.

### 2.1 전처리 및 창함수

입력된 환경소리 신호에는 직류성분을 포함한 저주파의 신호가 포함되어 있다. 이러한 저주파 신호는 감쇄 차이 또는 신호의 포화 같은 현상을 발생시키므로 때문에 신호 대 잡음비를 저하시킨다. 따라서 2차의 유한 임펄스 응답(finite impulse response, FIR) 필터를 사용해 이런 직류성분을 포함한 저주파의 신

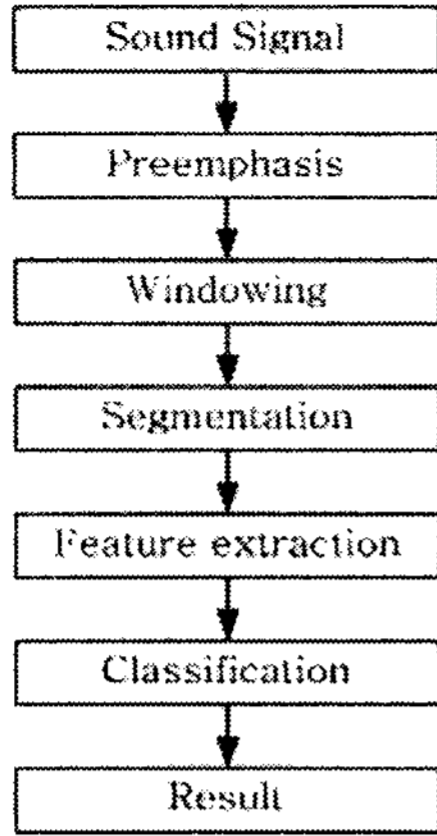


그림 1. 환경소리 인식방법의 순서도

호들을 제거할 필요가 있다. 또한 고주파 성분의 에너지는 저주파 에너지보다 상대적으로 크기가 작다. 전처리 과정은 이와 같은 단점을 보완하고자 고주파 에너지를 보상하기 위한 방법으로 사용한다. 고른 에너지 분포를 얻기 위해 전처리 과정을 수행하는  $z$  공간의 전달함수는 식(1)과 같은 유한 임펄스 응답 필터를 사용한다.

$$H(z) = 1 - az^{-1}, \quad 0 \leq a \leq 1 \quad (1)$$

여기서  $a$  는 전처리 과정의 변수로서 일반적인 값으로 0.95를 사용하는데, 고주파 스펙트럼의 크기에 대해 20dB 이상의 증폭효과가 있다[8].

환경소리 신호는 매우 짧은 동안 정상상태 특성을 반영하고 있으므로 창함수를 통한 단 구간 스펙트럼 분석을 수행할 수 있다. 정상상태라고 가정할 수 있는 매우 짧은 시간 동안의 신호 단위로 분석하게 되는데 이 구간을 프레임(frame)이라고 한다. 이때 환경소리 신호  $x(n)$ 는 연속된 프레임인  $x_t(n), t = 1, 2, 3, \dots, T$ 로 나누지게 되며 한 프레임의 길이는 대개 20ms 정도이다. 각 프레임을 추출할 때에는 환경소리 정보의 연속성을 보존하기 위해 일반적으로 10ms 정도를 중복(overlapping)시킨다. 환경소리 신호  $x(n)$ 에 대해 식(2)와 같이 프레임 처리하여  $x_t(n)$ 을 구하고 창함수 처리 후 식(3)과 같이  $x'_t(n)$ 을 구한다.

$$x_t(n) = \sum_{n=1}^N x(n - tQ), \quad (0 \leq n \leq N) \quad (2)$$

$$x'_t(n) = \sum_{n=1}^N w(n)x_t(n) \quad (3)$$

여기서  $w(n)$ 은 창함수이고,  $t$ 는 프레임의 위치를 의미한다. 이 때 각 프레임은 임의의  $Q$  만큼 이동(shift) 된다.

### 2.2 시작점 및 끝점 추출

동적 시간정합의 성능을 가늠하는 요소 중 하나가 시작점과 끝점 추출과정이다. 이는 소리의 시작과 끝을 검출하는 것으로 소리 인식방법에서 중요한 요소로서 인식률에 커다란 영향을 미친다.

기존의 영 교차율 혹은 에너지 값을 이용한 방법은 환경소리 구간을 그대로 찾아내지 못하는 경우가 있다[9]. 이러한 단점을 보완하기 위하여 단 구간 푸리에 변환을 이용하였다. 단 구간 푸리에 변환의 결과  $S(f,t)$ 에서  $f$ 는 주파수 단위를 말하며  $t$ 는 시간 단위를 나타낸다. 시작점 및 끝점 추출은 다음의 순서로 수행된다. 우선, 식(4)과 같이  $S(f,t)$ 구간 중에서 가장 큰 값을 가지는  $(f_n, t_n)$ 을 구한다.

$$A_n(0) = 20 \log_{10} |S(f_n, t_n)| \quad (4)$$

$A_n(0)$ 값을 기준으로  $\beta$ dB만큼 작아지는 지점  $(t_s, t_e)$ 을 구한다.  $t_s$ 는 이전구간에서  $t_e$ 는 이후구간에서 구하여 시작점과 끝점을 추출한다.

$$A_n(T) < A_n(0) - \beta \text{dB} \quad (5)$$

$$\text{START TIME} = t_n - t_s \quad (6)$$

$$\text{END TIME} = t_n + t_e \quad (7)$$

여기서  $\beta$ dB는 실험적으로 15dB로 설정하였다. 그림 2는 입력에 대한 시작점과 끝점의 추출과정이며 사각형 영역은 시작점과 끝점 추출 결과이다.

### 2.3 특징 추출

환경소리의 특징을 추출하기 위해서 그림 3과 같은 멜 주파수 캡스트럼 계수 연산 알고리즘을 사용하였다. 멜 주파수 캡스트럼 계수를 구하기 위해서는 이산 푸리에 변환을 거쳐 이를 멜 필터 बैं크에 통과시킨다. 이후 대수 에너지(logarithm energy)를 구한다. 이러한 동적 대수(dynamic logarithm) 압축 과정을 통하여 크기 성분은 살리고 중요성이 떨어지는 위상 성분은 버리게 된다. 따라서 이로부터 추출된 음성 특징은 동적(dynamic)측면에서의 변화에 덜 민

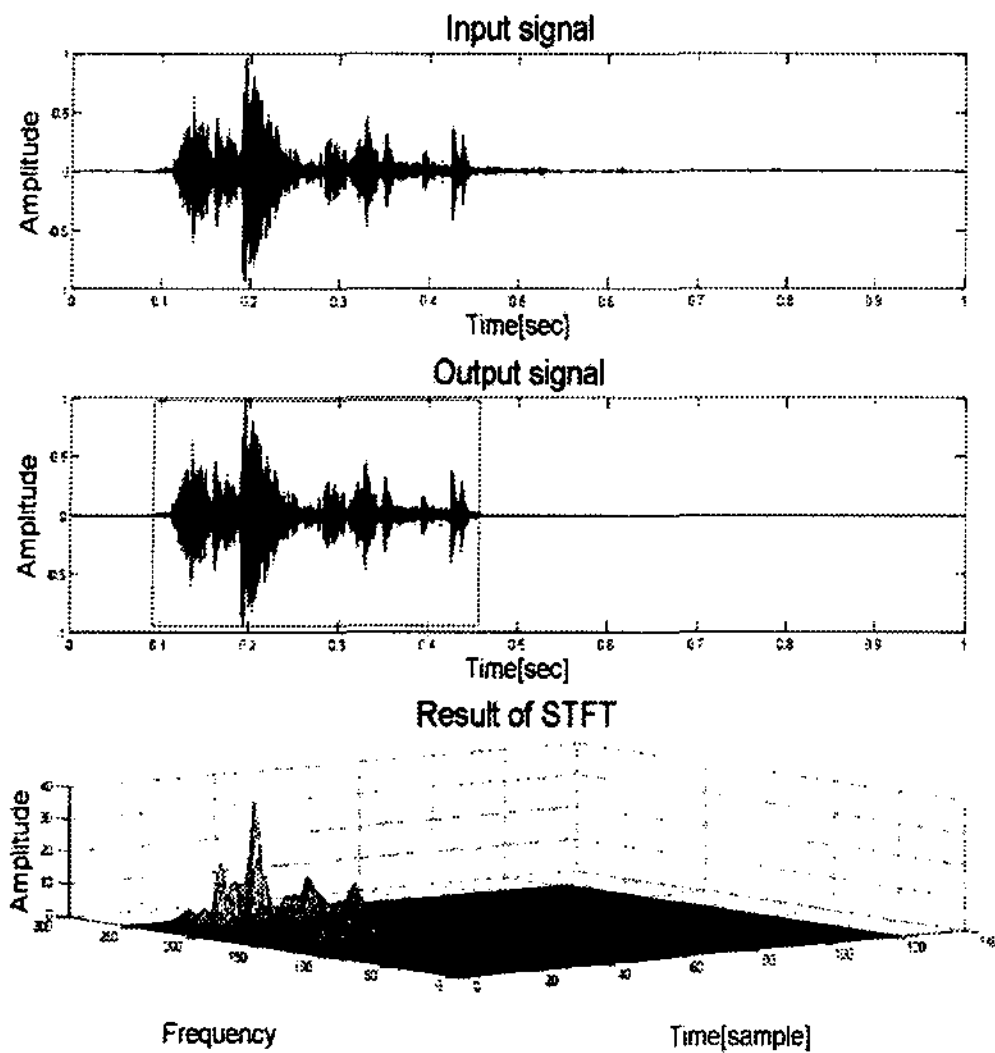


그림 2. 시작점과 끝점 추출

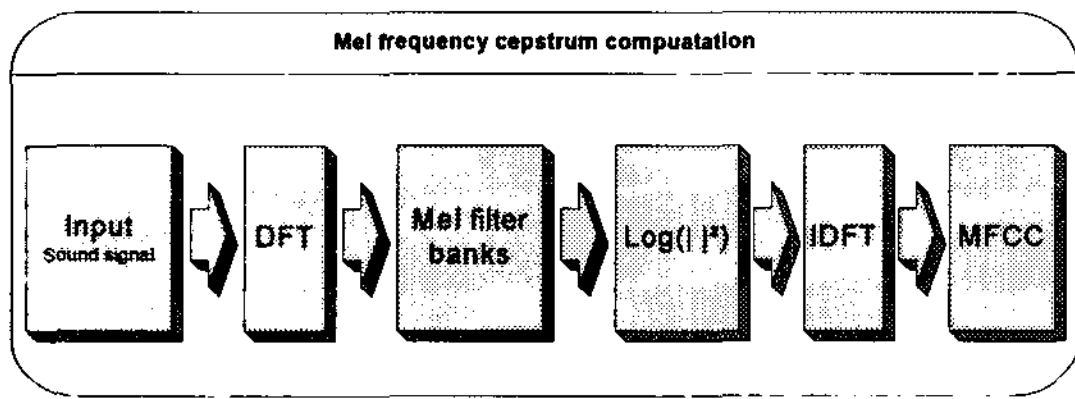


그림 3. 멜 주파수 켈스트럼 계수의 연산 순서

감하게 된다. 대수 에너지를 역 이산 푸리에 변환을 통하여 멜 주파수 계수를 추출하게 된다.

2.4 동적 시간정합

동적 시간정합은 시간 축 상에서의 비선형 신축을 허용하는 패턴 매칭이다. 동적 시간정합 알고리즘은 두 열의 각 성분에 대한 거리척도를 비용으로 설정한 다음 두 열이 이루는 격자 상에서 각 열의 시작 성분에서 시작하여 끝 성분에 이르기까지 비용 테이블에 최소 비용을 순환적으로 선택하여 저장하는 점화식을 이용하는 동적 계획법으로 매핑(mapping) 함수를 찾아가면서 두 열을 비교하는 알고리즘이다. 한편, 매핑 함수의 궤적은 앞의 동적 계획법에서 최적 탐색경로를 찾는 것과 같이 탐색 과정에서 탐색 비용을 택하는 경로를 별도의 경로 테이블에 매 단계마다 저장하고 끝 성분에서 최종 최소 비용을 구한 후에 역 추적(back tracking)하여 찾게 된다[10]. 그러므로 동적 시간정합 알고리즘은 열의 길이가 일치하지 않

는 두 열의 유사도를 측정하는 매칭 알고리즘(matching algorithm)으로 적당하다.

2.4.1 제약 조건

환경소리 인식에 적용된 동적 시간정합알고리즘의 성능을 결정하는 것은 다음과 같은 5가지 제약조건(constraints)을 가지고 있다[10].

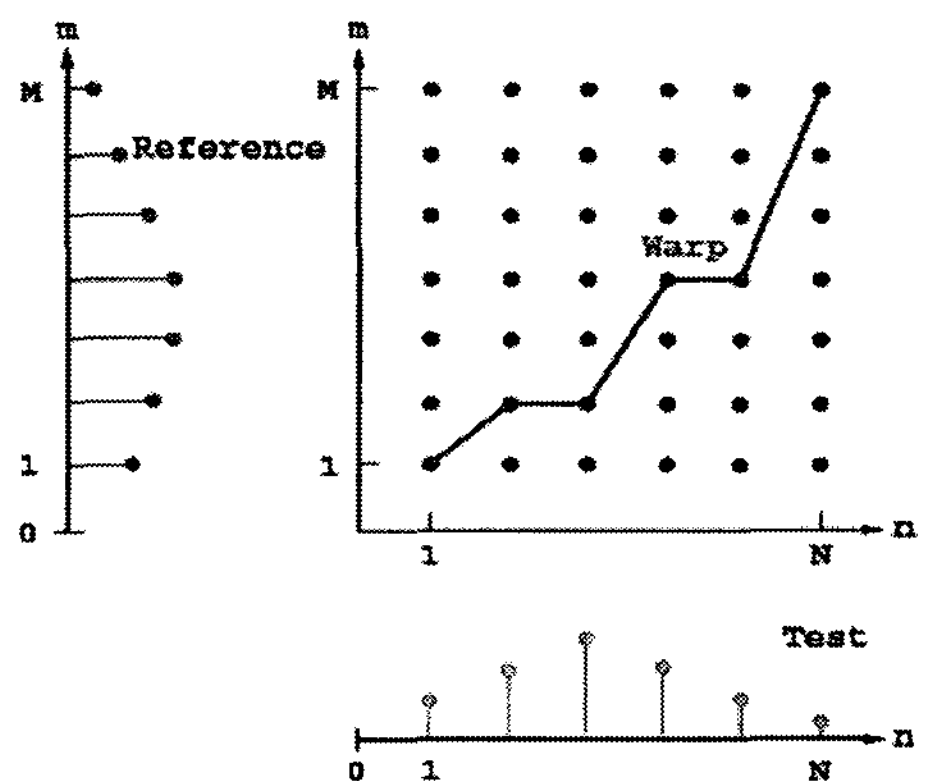
- 첫째 : 끝점 제약(endpoint constraints) 조건
- 둘째 : 단조 증가 제약(monotocity constraints) 조건
- 셋째 : 국부 제약(local continuity constraints) 조건
- 넷째 : 전역 경로 제약(global path constraints) 조건
- 다섯째 : 기울기 가중치(slope weighting) 조건

동적 시간정합 알고리즘은 참조(reference, R)특징과 시험(test, T)특징의 각각의 위치에 대응하는 좌표에 거리 값을 구하여 위의 5가지 조건에 부합하면서 가장 작은 거리 값을 찾아서 최적의 검색경로를 찾고 그 총거리를 구하는 것이다.

2.4.2 거리 계산

두 패턴간의 유사도를 측정하는 방식으로는 유클리디안 거리(Eucliden distance)계산 방식과 특성에 맞춘 거리를 사용하는 방식 두 가지로 분류된다. 그림 4와 같이 두 패턴의 각각의 위치에 대한 거리를 구한다.

$$d(m,n) = \sqrt{(R(m) - T(n))^2} \tag{8}$$



$R(1) = 1, T(1) = 1$  : 시작점  
 $R(m) = M, T(n) = N$  : 끝점

그림 4. 동적 시간정합 알고리즘

이렇게 계산된 거리는  $m=n=0$ 부터 시작해서 식 (9)의 국부탐색경로 범위에서 최소 거리 값을 가지는 방향으로 진행하면서 누적거리를 구해나간다.  $d$ 는 각 위치에 대한 거리 값이며  $D$ 는 각 위치를 거치면서 누적된 거리 값이다.

$$D(m,n) = D(m,n) + \min \begin{cases} d(m+1,n+1) \\ d(m+1,n) \\ d(m,n+1) \end{cases} \quad (9)$$

그러나 국부 연속 제약조건에서 각각의 노드의 진행 거리가 다르기 때문에 이에 상응하는 보정을 하기 위해 가중치(weight)  $W$ 를 곱해준다.

$$D(m,n) = D(m,n) + \min \begin{cases} d(m+1,n+1)W_1 \\ d(m+1,n)W_2 \\ d(m,n+1)W_2 \end{cases} \quad (10)$$

이렇게 누적된 최종거리 값인  $D$ 를 그대로 사용하게 될 경우 짧은 길이의 패턴의 누적거리는 상대적으로 작게 나오므로 이것을 정규화 시켜야 한다.

$$W_{dist\_sum} = \sum_{k=1}^K W_{(k)n} \quad (11)$$

$W_{(k)n}$ 은 각 위치에 적용된 국부 경로 가중치이며  $n$ 은 각각의 국부 경로에 따른 가중치를 의미한다. 거리 계산을 정규화 시키기 위한 기준으로 진행해온 거리에 비례 관계를 가지는 가중치 값의 합인  $W_{dist\_sum}$ 으로 하였다.

$$D = \sum_{k=1}^K d(m(k),n(k))W_k \quad (12)$$

$d(m(K),n(K))$  기준벡터의 프레임벡터  $m(K)$ 와 테스트 벡터의 프레임 벡터  $n(K)$ 간의 국부거리를 나타내며  $W_k$ 는 경로 따른 경사 가중치를 말한다.

$$Dist = \frac{D}{W_{dist\_sum}} \quad (13)$$

패턴의 길이를 보상하기 위해서 전체 경로 가중치의 합으로 나누어 보상해 준다.  $Dist$ 는 두 패턴간의 거리 값을 나타낸다.

### 2.5 제안한 능동형 기울기 가중치 제약을 갖는 동적 시간정합 알고리즘

제약조건들 중에서 4번 제약조건은 전역 경로 제

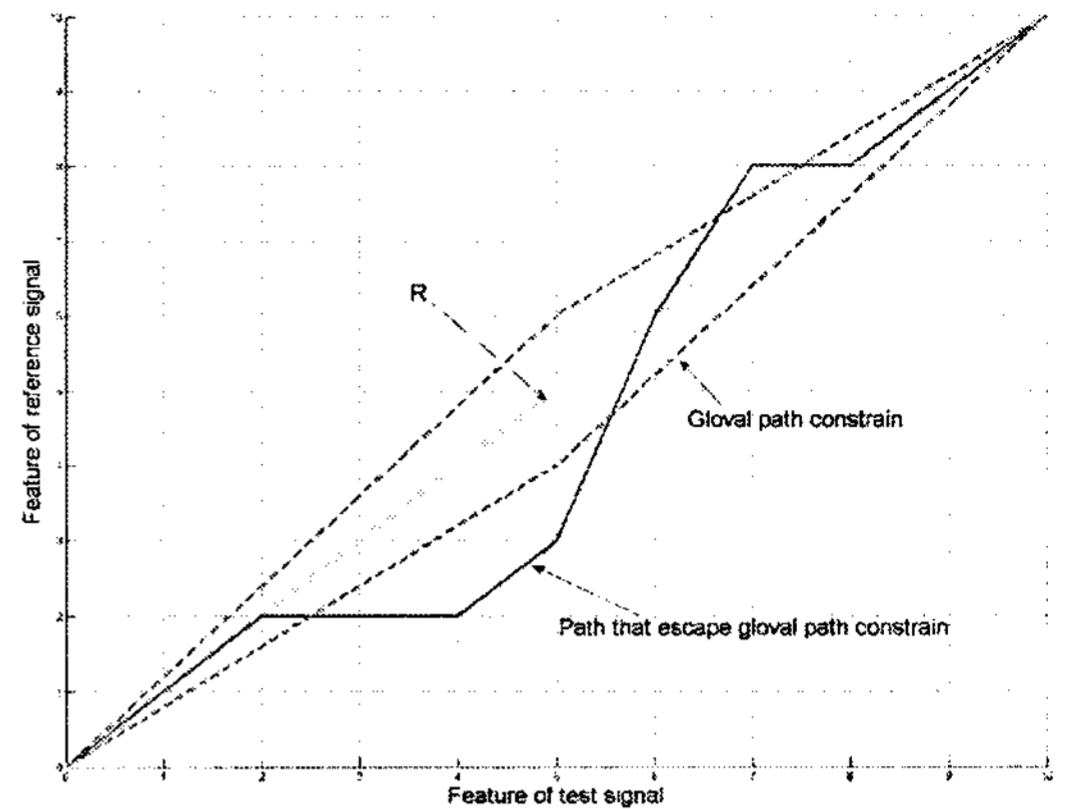


그림 5. 동적 시간정합의 단점

약을 통하여 동적 시간정합의 검색 경로를 축소함으로써 연산량을 감소시킨다. 그러나 그림 5와 같이 특징음의 길이가 상대적으로 길어지게 된다면 그 영역에 대한 정보 손실이 발생 될 수 있다.

이러한 정보손실을 줄이고 정확한 검색 경로를 찾기 위해 전역경로 제한조건을 제거하였으며 전역경로 제한을 통해 얻을 수 있었던 동적 시간정합의 지나친 변형과 속도 향상을 5번 제약조건인 경로 가중치를 이상적인 탐색경로와의 거리를 기준으로 능동적으로 변화시킴으로서 검색 경로를 찾게 하였다. 즉, 특징이 입력되면 지역 제한 조건에 맞추어 최적 국부 탐색경로를 찾는다. 이때 국부 탐색경로는 기존의 방법과는 다르게 이상적인 탐색경로  $R$ 과의 거리 값을 기준으로 특정 연산을 거친  $K_n$  값을 탐색 가중치 값에 곱해주게 된다. 탐색경로 가중치 값을 따로 저장해두고 총거리에 이동하는 거리를 계산하여 더한다. 이와 같은 작업을 끝점에 도달 할 때까지 반복하여 최종 거리 값을 구하고 또 다른 테스트 환경소리 특징과도 같은 방법으로 최종 거리 값을 구한다. 모두 구해진 최종 거리 값들 중 가장 작은 것을 선택하여 그것이 임계 거리 값보다 크지 혹은 작은지를 판단하여 환경소리를 인식한다.

이상적인 검색 경로는 시작점의 위치와 끝점의 위치를 직선으로 이어주는 것이다. 다시 말하면 두 가지 같은 특징 신호는 동적 시간정합 연산을 거치게 되면 일직선의 검색 경로  $R$ 이 생긴다.

이를 기준으로 하여 현재의 위치( $X_c, Y_c$ )의 값과  $R$ 와의 거리 값을 기준으로 하여 cosh의 해당하는 기울기 가중치 값을 각각  $K_1, K_2, K_3$ 에 반영된다. cosh은

일정 범위까지는 매우 작은 값을 가지고 있으나 그 범위를 벗어나면 점진적으로 매우 큰 값을 가지게 된다. 실험적인 방법을 통하여  $\alpha$ 를 구하였다.

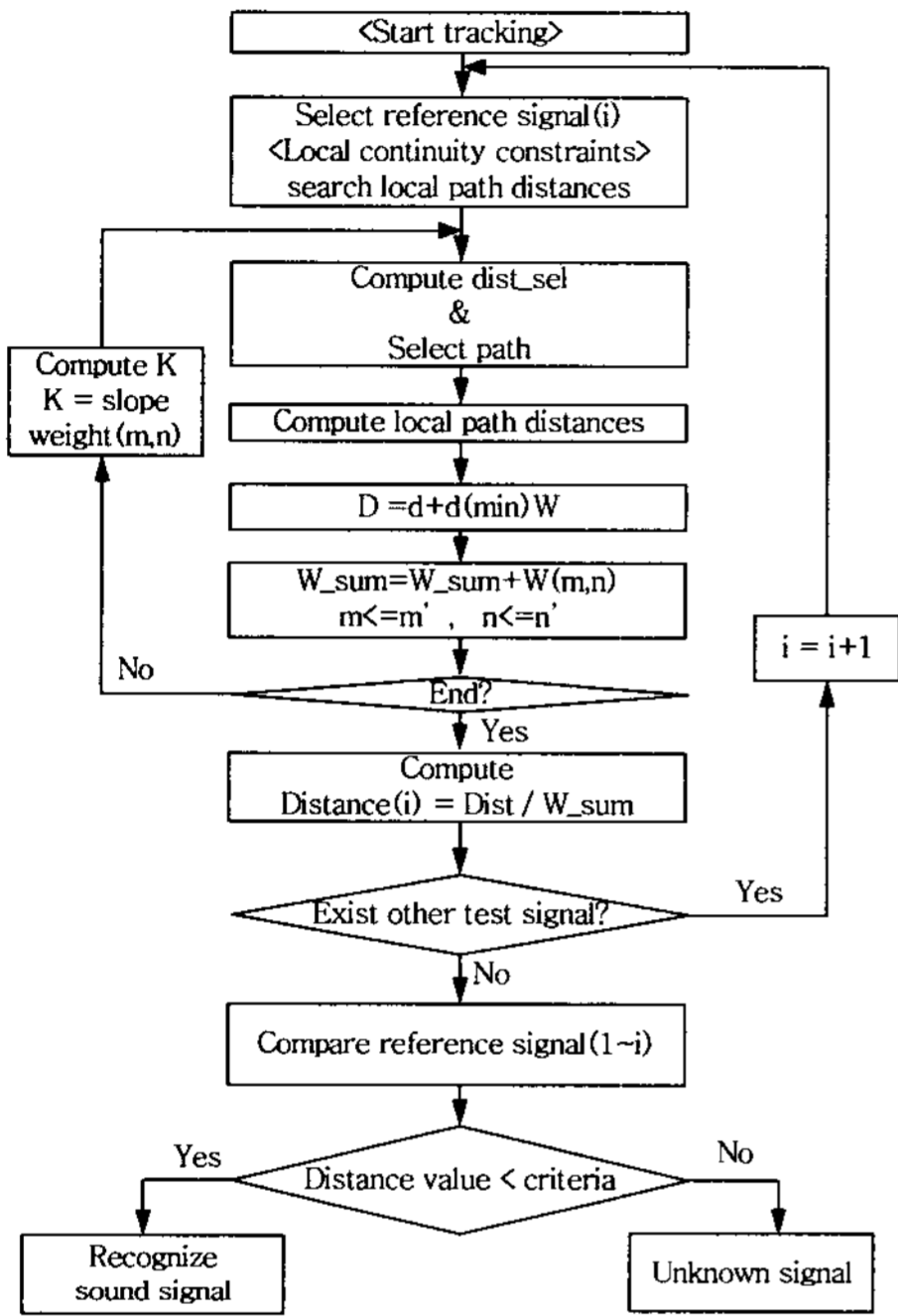


그림 6. 제안한 동적 시간정합 알고리즘

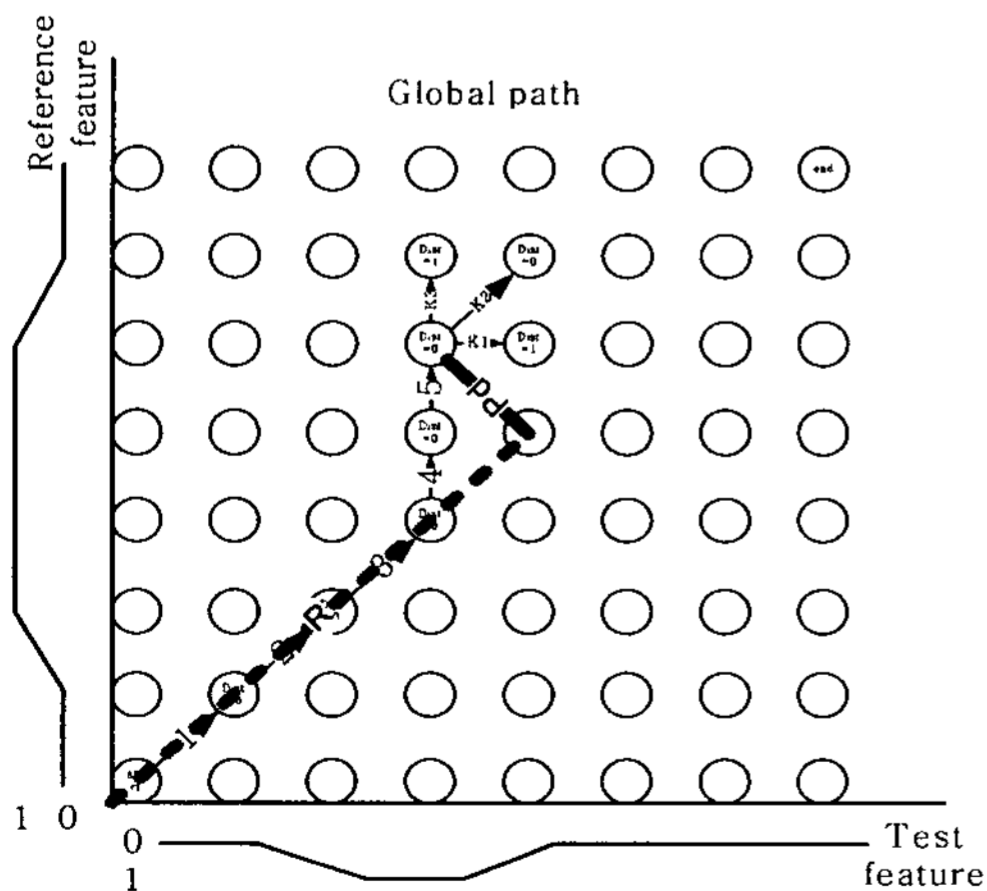


그림 7. 가중치 값 계산

$$K_1 = 1 + \cosh(\alpha \tan^{-1}(Pd/D_0)) \quad (14)$$

$$K_2 = 1 \quad (15)$$

$$K_3 = 1 - \cosh(\alpha \tan^{-1}(Pd/D_0)) \quad (16)$$

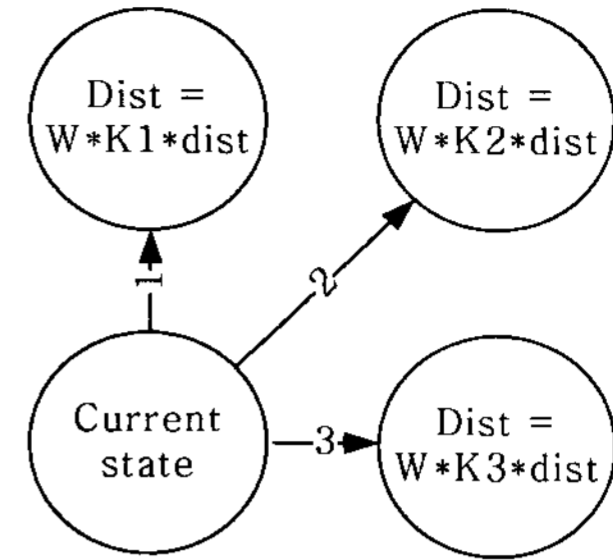


그림 8. 동적 시간정합 거리 계산

이상의 제안한 동적 시간정합 알고리즘과 가중치 계산 방법은 각각 그림 6과 7에서 보였다. 그림 8과 같이 적용된 지역 경로 제한은 동시에 전역 경로 제한의 역할을 함으로서 전역 경로 제한에 의해 제한되어진 절대적 이동 경로에 유연성을 확보할 수 있으며 동시에 전역 경로 제한에서 오는 지나친 변형이나 속도 향상도 기대할 수 있다

### 3. 실험 결과 및 고찰

일상적인 생활에서 주요한 정보를 제공하는 화재 감지기, 경찰차, 소방차, 앰블런스, 자동차경적, 전화기, 초인종, 벨의 환경소리에 대한 데이터베이스를 구축하였다. 이들 환경소리의 종류는 표 1에서 보였다. 환경소리는 매우 다양하며 비록 같은 소리라도 음원이 발생하는 거리의 변화 혹은 속도의 변화에 따라 소리의 특성이 조금씩 달라지기 때문에 각각의 종마다 7~9 개의 변형된 실험용 데이터베이스를 생성하여 총 88개의 실험용 환경소리 데이터베이스를 구축하였다.

제안한 방법에 대한 실험에서는 멜 주파수 캡스텀 계수의 특징을 추출하고 제안된 동적 시간정합 알고리즘을 적용하였으며 이때 시뮬레이션 도구로는 Mathwork사의 MATLAB ver. 7.1 프로그램을 이용하였다. 제안한 방법의 성능을 평가하기 위하여

표 1. 환경소리의 종류

Type of sound	
Doorbell	Ambulances
Car horn	Electronic phone
Smoke alarm	Fire engines sirens
Police sirens	Bells



Cowling의 논문[3]에 사용된 Slaney의 동적 시간정합 알고리즘을 비교대상으로 하였다.

기본 8개의 데이터와 88개의 변형 데이터는 전처리 과정을 거친 후 시작점과 끝점을 찾아 구간을 검출한다. 검출된 구간에 대해 멜 주파수 캡스트럼 계수 특징을 추출한 후 동적 시간정합 알고리즘을 통하여 비교한다. 비교된 일정 신호에서 기준 거리 이상의 값은 알 수 없는 신호로 가정하고 그 이하의 값을 가진 신호를 인식하게 된다. 또한 위의 과정에서 마이크에서 백색잡음이 입력과정에서 생길 수 있으므로 이것을 고려해서 일반적인 마이크의 신호 대 잡음비가 일반적으로 40dB 이상임을 감안하여 40dB, 30dB, 20dB, 10dB, 5dB 까지 잡음이 첨가되었을 경우 대해서도 실험하였다. 이때 백색잡음은 각각의 환경소리에 따라 새롭게 생성하여 적용하였다.

시작점과 끝점이 추출과정에서 단 구간 퓨리에 연산을 거치게 되는데 이 결과를 멜 주파수 캡스트럼 계수를 구하는 과정에서 그대로 이용함으로써 중복적인 연산과정을 생략할 수 있었다. 즉, 시작점과 끝점 과정에서 계산된 단 구간 퓨리에 변환 값을 이용하여 멜 주파수 캡스트럼 계수를 구한다. 그림 9와 그림 10은 신호 대 잡음비가 30dB와 5dB인 8개의 환경소리에 대한 특징을 나타낸다. 일반적으로 신호 대 잡음의 비가 감소함에 따라 동적 시간정합의 거리 값도 증가한다. 그러나 그림 9의 초인종소리의 경우를 보면 그 특징이 매우 뚜렷한 것을 확인할 수 있으나 백색잡음이 증가하면서 그림 10에서 보면 특징이

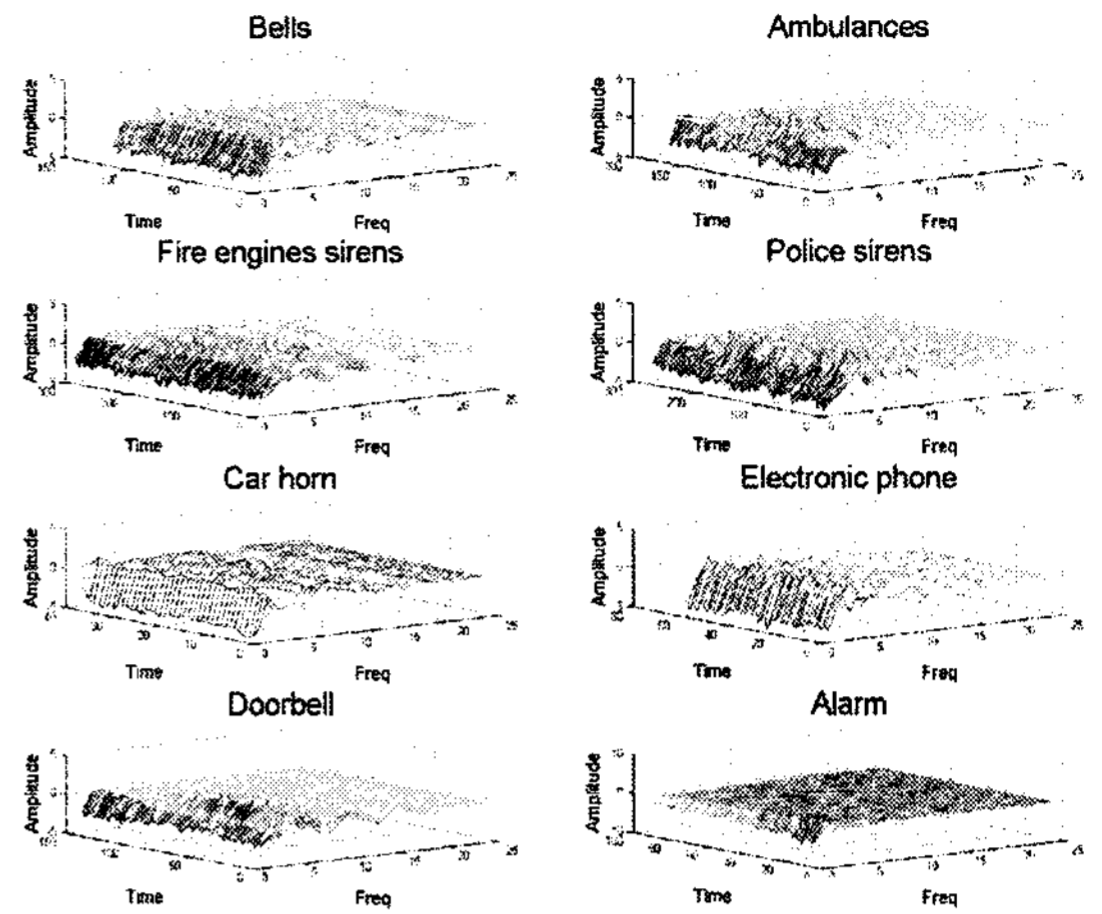


그림 10. 5dB의 신호 대 잡음비를 갖는 환경소리 신호의 멜 주파수 캡스트럼 특징

상당부분 평평해짐을 확인할 수 있다. 이것은 백색잡음의 주파수 특성이 전 주파수 영역에 걸쳐있기 때문에 백색잡음의 비율이 커짐에 따라 전체 주파수 특성이 평탄해 지는 것이라 추측된다. 이렇게 특징이 평탄해 지는 경우 신호 대 잡음비가 감소하여 패턴의 거리 값이 줄어들게 되며 이것으로 인하여 인식의 오류를 초래 할 수 있다.

환경소리 인식실험에 있어서 음원이 멀어지거나 가까워짐에 의한 속도변화 및 장애물로 인한 소리의 변형 등 실제적 환경에서 소리가 변형되는 경우를 고려하였다. 8가지의 기준 환경소리(화재감지기, 경찰차, 소방차, 앰블런스, 자동차경적, 전화기, 초인종, 벨)와 8종류 70가지의 인식성능 테스트용 환경소리를 실험에 사용하였다. 실제 가상 실험에서는 2종류의 미지신호를 함께 적용하여 총 88개의 데이터를 사용하였으며 이 미지신호의 최소 거리 값은 동적 시간정합의 거리에 대한 임계값을 결정하는데 사용된다. 그 결과를 확인해보면 제안한 알고리즘이 Slaney 동적 시간정합 알고리즘에 비해 마이크 등으로 인한 시스템 자체의 잡음에 상대적으로 강하며 또한 같은 잡음환경에서 더욱 정확한 인식률을 나타냄을 표 2를 통해 확인할 수 있다. 인식률의 형태를 확인해보면 30dB 백색잡음에 의한 인식률이 상승하는 경우도 찾을 수 있었다. 이는 백색잡음에 의한 거리 값의 감소 현상이 발생하였기 때문이다.

표 3과 표 4는 기존의 동적 시간정합 방법과 제안한 동적 시간정합 방법을 각각의 소리에 대한 인식률

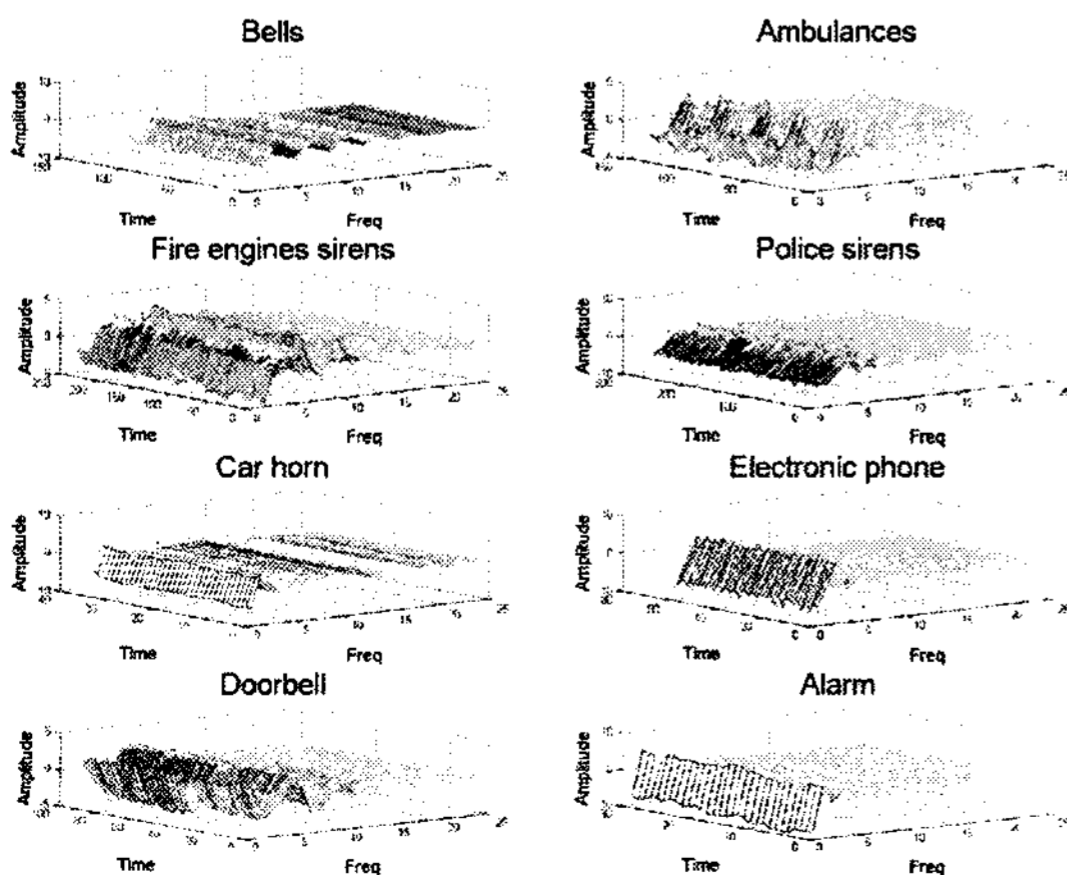


그림 9. 30dB의 신호 대 잡음비를 갖는 환경소리 신호의 멜 주파수 캡스트럼 특징





#### 4. 결 론

본 논문에서는 환경소리 인식시스템에 적용할 수 있는 능동형 기울기 가중치 제약 방식의 새로운 동적 시간정합 알고리즘과 이를 적용한 환경소리 인식방법을 제안하였다. 제안한 시간정합 알고리즘은 마이크 등으로 인한 시스템 자체의 잡음에 의한 인식을 저하를 상대적으로 개선시키며 이를 이용한 제안한 환경소리 인식방법은 기존의 동적 시간정합 알고리즘을 이용한 인식방법에 비하여 인식률과 정확성에서 더욱 우수하다는 것을 확인하였다.

제안한 동적 시간정합 알고리즘은 5가지 제약조건 중 전역 경로 제약조건을 절대적인 경로 제한이 아닌 상대적인 경로 제한을 함으로써 경로 제한의 유연성을 부여하였다. 기존의 동적 시간정합 알고리즘과 비교하였을 때 인식율에 있어서 약 20%정도의 개선이 있었으며 또한 표준편차가 상대적으로 낮아 정확성도 비교적 높음을 확인하였다. 따라서 청각 장애인을 위한 휴대용 환경소리 인식시스템에 제안한 인식방법을 적용할 경우 향상된 인식률을 바탕으로 청각 장애인의 삶의 질을 한층 더 끌어 올릴 수 있을 것이라고 생각된다. 또한 이러한 환경소리 인식방법은 산업현장에서의 기계의 오작동 혹은 자동차 내의 소음을 통한 고장 진단 등 여러 방면에 활용 가능성이 있다고 생각된다.

#### 참 고 문 헌

[1] Y. Yoda, "Environmental Sound Recognition by Multilayered Neural Networks," *The Fourth International Conference on Computer and Information Technology*, pp. 123-127, Sept. 2004.

[2] Y. Hattori, "Repeat Recognition for Environmental Sounds," *13th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 83-88, Sept. 2004.

[3] M. Cowling, "Comparison of Techniques for Environmental Sound Recognition," *Pattern recognition letters*, Vol.24, No.15, pp. 2895-

2907, Nov. 2003.

[4] T. Zhang and C. J. Kuo, "Audio Content Analysis for Online Audiovisual Data Segmentation and Classification," *IEEE Trans. on Speech and Audio Processing*, Vol.9 No.4, pp. 441-457, May. 2001.

[5] J. C. Wang, J. F. Wang, and Y. S. Weng, "Chip Design of MFCC Extraction for Speech Recognition," *The VLSI journal*, Vol.32, No.1, pp. 111-131, Nov. 2002.

[6] T. Zeppenfeld, R. Houghton, and A. Waibel, "Improving the MS-TDNN for Word Spotting," *Proc. of IEEE Conf. on Acoustics Speech and Signal Processing*, Vol.2, pp. 475-478, April 1993.

[7] H. Gish and K. Kenney, "A Segmental Speech Model with Applications to Word Spotting," *Proc. of IEEE Conf. on Acoustics Speech and signal Processing*. pp. 447-450, April 1993.

[8] 이건상, 양성일, 권영현, *Speech Recognition*, 한양대학교 출판부, pp. 25-38, 12월 2001.

[9] C. H. Lee, C. H. Chou, H. C. Han, and R. Z. Han, "Automatic Recognition of Animal Vocalizations using Averaged MFCC and Linear Discriminant Analysis," *Pattern Recognition Letters*, Vol.27, Issue 2, pp. 93-101, Jan. 2006.

[10] 한학용, 패턴이식 개론: Matlab 실습을 통한 입체적 학습, 한빛 미디어, 2005.



정 영 진

2005년 2월 금오공과대학교 전자공학부(공학사)  
 2007년 2월 경북대학교 대학원 의용생체공학과(공학석사)  
 2007년 3월~현재 연세대학교 대학원 의공학과 박사과정

관심분야 : Time-series analysis, Neuroimaging



이 윤 정

- 2003년 2월 경북대학교 전자전기공학부(공학사)
- 2005년 2월 경북대학교 대학원 의용생체공학과(공학석사)
- 2005년 3월~현재 경북대학교 대학원 의용생체공학과 박사과정

관심분야 : 생체신호처리, 의용전자기기

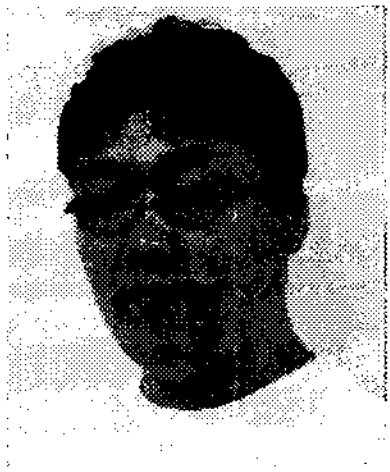


김 명 남

- 1988년 2월 경북대학교 전자공학과(공학사)
- 1990년 2월 경북대학교 대학원 전자공학과(공학석사)
- 1995년 2월 경북대학교 대학원 전자공학과(공학박사)
- 1996년~현재 경북대학교 의대

의공학교실 주임교수

관심분야 : 생체신호처리시스템, 의학영상처리



김 필 운

- 2001년 2월 경운대학교 컴퓨터공학과(공학사)
- 2003년 2월 경북대학교 대학원 의용생체공학과(공학석사)
- 2004년 3월~현재 경북대학교 대학원 의용생체공학과

박사과정

관심분야 : 의학영상처리, 의용전자기기, 생체신호처리