DBpia

Radial Basis Function Networks를 이용한 이중 임계값 방식의 음성구간 검출기

준회원 김 홍 익*, 정회원 박 승 권**

Voice Activity Detection Algorithm base on Radial Basis Function Networks with Dual Threshold

Hong-Ik Kim* *Associate Members*, Sung-Kwon Park** *Regular Members*

요 약

본 논문에서는 간단한 구조, 적은 계산량과 안정된 빠른 수렴속도를 가진 RBF (Radial Basis Function) 신경회로망을 이용한 이중 임계값 방식의 음성구간 검출기 알고리듬을 제안하고 시뮬레이션을 통해 유용성을 확인하였다 음성압축기에 사용되는 CELP (Code-Excited Linear Prediction) 파라미터들을 신경회로망 입력으로 하여 잡음에 강하게 반응하게 하였고, 음성구간 검출기의 성능향상을 위해 음성구간과 침묵구간에서 다른 임계값을 사용하는 이중 임계값 방식을 적용하였다.

실험 결과 이중 임계값을 이용한 RBF 신경망 음성구간 검출기는 G.729 Annex B 음성구간 검출기 보다 우수한 성능을 보였고, 기존의 MLP (Multi Layer Perceptron) 신경회로망을 이용한 음성구간 검출기와 비교하여 음성구간에서는 비슷한 성능을 보였으나 침묵구간에서 25% 정도의 성능향상을 보였다

Key Words voice activity detection, radial basis function network, dual threshold

ABSTRACT

This paper proposes a Voice Activity Detection (VAD) algorithm based on Radial Basis Function (RBF) network using dual threshold. The k-means clustering and Least Mean Square (LMS) algorithm are used to upade the RBF network to the underlying speech condition. The inputs for RBF are the three parameters in a Code Exited Linear Prediction (CELP) coder, which works stably under various background noise levels Dual hangover threshold applies in BRF-VAD for reducing error, because threshold value has trade off effect in VAD decision

The experimental result show that the proposed VAD algorithm achieves better performance than G.729 Annex B at any noise level

Ⅰ. 서 론

일상생활에서 많이 사용하는 양방향 음성 통신의 경우, 어느 한쪽의 입장에서 보면 실제 화자가 말하는 때가 있는 반면, 상대방의 말을 듣기만 할 때도 있다 이를 신호처리 관점에서 보면, 음성신호를 전송하는 구간과 수신하는 구간이 분리되어 있음을 뜻한다 이러한 특징을 이용하여 음성신호 구간만을 압축하여 전송하고 침묵구간에는 데이터를 보내지 않음으로써 데이터 전송 효율을 높이는 방법을 침묵압축 (Silence Compression) 이라고 하고, 침묵구간에 음성이 아닌 데이터를 송신하는 방법을 비연속통신 (DTX: Discontinuous Transmission) 기술이라고 한다. 이러한 침묵압축 및 비연속 전송기술은 디지털 음성

* 한양대학교 전자통신전파공학과대학원 응용통신연구실 (hongik@ihanyang ac kr)
** 한양대학교 전자전기컴퓨터공학부 (sp2996@hanyang ac kr)
논문번호 KICS2004-07-001, 접수일자 2004년 7월 15일

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

표현 및 음성압축 기술의 발전과, 기존의 회선기반의 음성통신에 패킷기반 음성통신(VoIP)이 더해짐으로써 가능하게 되었다.

음성구간 검출기(VAD: Voice Activity Detector)란 음성신호에서 음성(Active Voice)구간과 침묵(Silence) 구간을 구별해내는 것을 말한다. 이런 음성구간 검출이 선행되어야만 앞에서 말한 침묵압축 및 비연속 전송이 가능해진다. 음성구간검출 기술은 유무선 통신에서 많이 응용되어 이용하고 있고, 이를 통해 대역폭(Bandwidth)을 줄일 뿐 아니라, 대역을 재사용 가능하게 하고, 전송효율도 높인다. 무선통신에서는 음성구간에서만 전송을 하므로 대역폭을 줄이고, 통신 잡음을 줄이면서, 전송효율을 높인다. 음성구간 검출기는 거의 모든 양방향 통신에서 사용되고 있고 이를 이용한 많은 응용분야가 있다 [1, 2, 3].

국제표준으로는, 음성구간검출기만의 표준은 없으나, ITU-T의 G.729 Annex B와 G.723 Annex A, 그리고 GSM의 음성압축 기술인 RPE-LTP등에서 음성구간 검출기를 기반으로 침묵압축기법을 표준화하여 현재 널리 사용되고 있다.

신경회로망은 다양한 신호처리 분야에서 비선형 처리를 가능하게 해 준다. 신경회로망의 대표적인 기능 중에 하나가 분류(Classification)기능이다. 분류기는 혼재된 분포를 갖는 입력을 명확히 구분된 출력으로 보여 준다 [4]. 음성구간 검출기도 일종의 분류기라고 할 수 있으며, 신경회로망을 적용할 수 있다. 이러한 신경회로망의 특징을 이용한 음성구간 검출기에 대한 연구는 계속되고 있다 [1]. Jotado Ikedo가 제안한 음성구간에서 침묵구간으로 변할 때 음성잔존구간 임계값을 적용하는 MLP (Multi Layer Perceptron) 신경회로망 음성구간 검출기는 ITU-T G.729 Annex B 표준에서 구현한 음성구간 검출기와 비교하여 오차율면에서 대략 50% 정도 줄어드는 성능을 보였다 [1]. 하지만 MLP는 학습속도가 느리고 국부 최소점을 가지기 때문에 수렴을 항상 보장하지도 못하고, 한번학습이 되면 갱신을 위하여 전체 학습패턴이 필요하기 때문에 변화가 있는 곳에서는 사용하기 어려워서 음성구간 검출기로서는 적합하지 않은 신경회로망이다 [5, 6]. 또한 임계값을 음성에서 침묵구간으로 변화시에만 적용하여 상대적으로 침묵구간에서 음성구간으로 변화 시 성능은 고려하지 않았다.

신경회로망 중에서 RBF(Radial Basis Function) 신경망은 하나의 은닉층을 가지고 있고, 은닉층 각각의 노드(Node)는 중심점(Center)을 중심으로 방사형(Radial)으로 대칭인 응답을 출력한다. 노드 각각의 중심점은 입력 공간상에 존재하는 벡터이며 각 노드의 입력과 출력간의 함수는 비선형적이다. 최종 출력은 은닉층에 있는 각각의 노드의 출력들을 가중치를 곱해서 더한 값이 된다. RBF 신경망의 장점은 단순한 구조와 선형의 학습 알고리듬을 사용함으로써 학습속도가 빠르다는 것이다 [5, 6]. 이러한 신경망 특징은 음성구간 분류기에 필요한 안정된 고속분류가 가능하게 한다.

본 논문에서는 신경회로망 음성구간 검출기 구조로 가우시안 RBF를 사용하고 입력은 잡음의 영향을 적게 받는 음성압축기에 사용되는 CELP (Code-Excited Linear Prediction) 파라미터를 사용하였으며 음성잔존구간이 적용된 이중 임계값 방식의 분류기 알고리듬을 이용하여 그 성능을 분석하였다. 또한 RBF 신경망의 학습방법으로 매번 입력이 들어올 때마다 중심점과 가중치를 계속해서 갱신하는 하이브리드 방법을 사용하여 계산량이 입력에 들어오는 시간 동안 분산되는 효과와 학습이 이루어진 후 입력채널의 특성이 변화더라도 들어오는 입력신호로부터 입력의 특성을 추적하여 신경망이 입력에 따라 이중적으로 변할 수 있게 하였다.

Ⅱ. 음성구간 검출기 및 신경회로망

1. 음성구간 검출기의 기본 구조

신경회로망을 사용하여 음성구간 검출기를 설계할 경우 그림 1과 같이 음성신호 입력과 구간나누기, 신호분석, 신경망 테스트를 거쳐 입력 신호가 음성구간 신호인지 침묵구간 신호인지 판별하는 과정을 거치게 된다.

그림 1. 신경회로망을 이용한 음성구간 검출기 구조

2. 신경 회로망 (Neural Network)

1) RBF (Radial Basis Function) 신경망

RBF 신경망은 그림 2와 같이 입력층과 비선형 처리를 하는 하나의 은닉층, 그리고 가중치를 갖는 출력층으로 구성되어 있다. RBF 신경망에서 입력층과 은닉층을 연결하는 전달함수는 비선형이고, 은닉층과 출력층 사이를 연결하는 전달함수는 선형이다. 은닉층

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

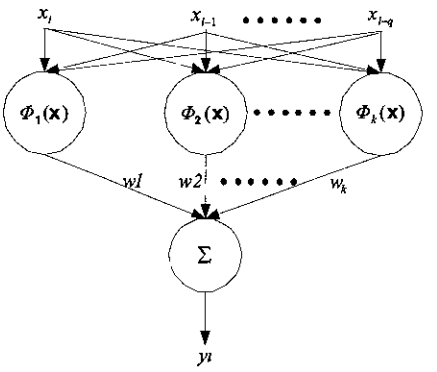

그림 2 RBF 신경망의 구조

의 각 노드의 중심점은 입력 공간상에 존재하는 벡터이며 각각의 노드는 중심점을 중심으로 방사상으로 대칭인 응답을 출력한다 최종 출력은 은닉층에 있는 각각의 노드의 출력들을 가중치를 곱해서 더한 값으로

$$y_i = \sum_{n=1}^{k} w_{i,n} \Phi_n(x) \quad (1)$$

과 같이 구해진다

$$\Phi_n(x) = exp(-\frac{1}{2S_n}\|X - C_n\|^2),\ n = 1, 2, \ldots, k \quad (2)$$

식(2)에서 사용된 C_n 은 베이시스 함수의 중심벡터, S_n은 분산, $\|X-C_n\|$은 입력벡터 X와 중심점 C_n간의 유클리드 거리를 나타낸다. 따라서 식(2)를 식(1)에 대입하면 식(3)과 같이 가우시안 RBF 신경망의 출력을 얻을 수 있다

$$y_i = \sum_{n=1}^{k} w_{i,n} exp(-\frac{1}{2S_n}\|X - C_n\|^2) \quad (3)$$

RBF 신경망의 설계에 있어 중요한 두 가지는 k개의 가우시안 RBF 각각에 대해 중심점 C_n을 계산하는 것과 출력층의 가중치를 결정하는 것이다

3 음성/침묵 구간 판정

신경회로망의 출력을 적절한 임계값과 비교하여, 분석한 신호구간이 음성구간인지 침묵구간인지를 판정해야 한다 이 과정은 그림 3에서 보듯이 임계값(Threshold)에 의한 판정과 음성 잔존구간(Hangover) 합이 있다

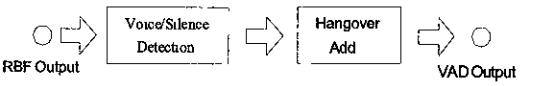

그림 3 신경회로망 출력을 이용한 음성/침묵구간 판정과정

임계값은 신경회로망이 출력이 나타내는 분포중에서 음성 및 침묵구간으로 구분하는 기준값을 말한다 음성 잔존구간은 음성구간에서 침묵구간으로 바뀌는 순간의 침묵구간을 음성구간으로 판별하는 것을 말한다. 왜냐하면 음성구간의 음절과 음절 사이의 매우 짧은 구간이 침묵구간과 마찬가지로 전력 레벨이 낮고 피치가 불안정할 수 있는데, 이를 침묵구간으로 잘못 판정할 수 있기 때문이다 이런 오류를 줄이기 위해서 음성구간에서 침묵구간으로의 변환을 늦추는 방법이다

Ⅲ. RBF 신경회로망을 이용한 음성구간 검출기

1. 음성구간 검출기 구조

음성신호 특징 파라미터를 사용하여 분류기(Classifier)로서의 역할을 할 신경회로망으로, 그림 4에서 보듯이 하나의 은닉계층을 사용한 가우시안 RBF 구조를 사용하였다 신경회로망의 입력계층에는 앞에서 설명한 3개의 음성 특징 파라미터를 사용하였고, 은닉계층에는 30개의 뉴런을 두었고, 출력계층에 하나의 뉴런으로 결과를 출력하게 된다. 앞의 2장에서 설명한 그림1의 구조와 같은 형태이다 여기서는 음성신호 분석을 통한 신경회로망의 입력 파라미터와 RBF 신경회로망의 구조에 대해서 알아본다

RBF 신경회로망을 이용한 음성구간 검출기는 RBF

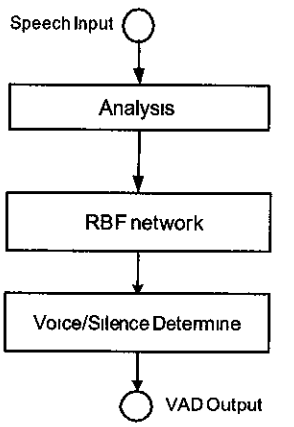

그림 4 음성구간 검출기 구조

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

신경회로망의 출력값을 임계값으로 판단하여 결정하기 때문에 RBF 신경회로망의 출력을 일반화(normalize) 해야 임계값을 찾기도 쉽고 판단하기도 쉬워진다.

$$OUT = \frac{1}{1 + e^{-IN}} \tag{4}$$

논문에서는 식 (4)처럼 시그모이드함수 (Sigmoid Function)를 정규화 하는 활성화 함수로 사용하였다 활성화 함수의 사용으로 RBF 신경회로망의 출력값은 항상 0에서 1사이의 값을 가지게 된다

2. 음성신호 특징 파라미터

음성구간 검출기에서는 신경회로망의 입력변수로 사용하는 음성신호의 특징 파리미터로 세 가지를 사용하였는데, 이들은 단구간 평균전력(Short-time Average Power), 영차 ML 파라미터(Zero-order Maximum Likelihood Parameter), 그리고 피치 길이의 차(Pitch Period Difference) 이다. 이들 값은 음성구간과 침묵구간을 검출하는 음성신호 특성 파라미터이다 [1]

단구간 평균전력은

$$E = 10log\left[\frac{1}{N}\sum_{n=m-N+1}^{m} s_w^2(n)\right] \tag{5}$$

와 같이 구해진다 즉, 식 (5)에서처럼 N개의 표본크기를 갖는 구간내의 각 표본을 자승하여 모두 더하고 대수값을 취하면 단구간 평균전력이 된다.

선형예측계수는

$$F = log\sum_{i=0}^{p} \alpha_i\alpha_i \tag{6}$$

과 같은 식을 사용하여 구한다. [1]

여기서 p는 선형 예측 차수 이고, α_i는 선형예측계수이다. F는 영차(zero order) ML 파라미터라고도 불리며 이는 음성파형과 평면(flat)파형과의 거리와 관계된 파라미터이다

현재구간과 이전구간 사이의 피치 길이 차이는 피치의 안정도(Pitch Stability)와 관련이 있다. 이 파라미터는 GSM에서 사용하는 RPE-LTP 또는 G 729의 음성구간 검출기에서도 사용되는 파라미터이다. 그러나 정확한 피치 길이를 측정하기가 힘들기 때문에

$$P = max\left[\frac{\log\sum_{t=0}^{N-1}\{r(t)r(t-\tau)\}}{\log\sum_{t=0}^{N-1}\{r(t)r(t)\}}\right] \tag{7}$$

과 같은 근사식을 이용한다 [1].

여기서 $20 \le \tau \le 160$ 이고, $r(t)$는 시간 t 에서의 선형 예측 오차 신호이다. 또한 시간 $t=0$는 분석구간의 시작 부분을 의미한다. 피치가 안정적인 경우, P는 큰 값을 가질 것이다.

3 신경회로망 학습

가우시안 RBF 신경망에서 학습해야 할 파라미터는 표준편차, 각 레디얼 베이시스 함수의 중심점, 그리고 가중치 이렇게 세 가지 값들이다. 먼저 표준편차는 각각의 레디얼 베이시스 함수마다 다르게 지정할 수 있지만 동일한 값으로 설정해도 신경망의 성능에 영향이 없다는 것이 밝혀져 있다. 표준편차값은 학습시 수렴속도등에 어느 정도 영향을 끼치지만 신경망의 일반적 기능에는 영향을 끼치지 않는다. 따라서 RBF 신경망의 설계시 표준편차 σ_k값은 고성상수를 사용하며 나머지 두개의 파라미터인 중심점과 가중치를 결정하는 것이 학습알고리듬의 핵심을 이룬다.

RBF 신경망의 학습 알고리듬에는 첫째는, 최소자승오차(Least Square: LS) 방법에 의해 중심점과 가중치를 일괄적으로 계산해내는 방법이고, 둘째는 레디얼 베이시스 함수들을 직교하게 설계하여 첫째 방법의 과도한 계산량을 감소시킬 수 있는 직교 최소 자승 오차(Orthogonal Least Square OLS) 방법이다 셋째는 입력벡터들의 통계적인 측성을 이용하는 하이브리드 알고리듬으로 중심점의 계산시에는 k-평균 클러스트링(k-means clustering) 방법을, 가중치의 계산시는 기존의 LMS(Least Mean Square)나 RLS(Recursive Least Square) 방법을 사용하는 것이다 [7, 8, 9].

본 논문에서는 가우시안 RBF 신경망의 학습 시 세 가지의 알고리듬 중 하이브리드 알고리듬을 사용하였다. LS 알고리듬과 OLS 알고리듬은 훈련열이 모두 수신된 후에 한번에 학습이 이루어지므로 중심점과 가중치의 계산에 많은 시간이 소요된다 또한 채널의 특성이 시간적으로 서서히 변하는 경우에는 처음부터 다시 학습을 해야만 한다 반면에 하이브리드 알고리듬은 매번 입력이 들어올 때마다 중심점과 가중치를 갱신해가는 방법을 사용하므로 계산량이 전체 입력이

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

들어오는 시간 동안 분산되는 효과를 얻을 수 있다

하이브리드 알고리듬에서 사용하는 두 개의 부 알고리듬인 k-평균 클러스터링 알고리듬과 LMS 알고리듬은 모두 선형 학습 알고리듬으로 빠른 수렴속도를 가지고 있다. RBF 신경망과 같은 기능을 하는 다층 퍼셉트론의 학습에 사용되는 오류역전파 알고리듬이 매우 느린 수렴속도를 가지고 있음에 비해 하이브리드 알고리듬으로 학습된 RBF 신경망은 빠른 수렴 속도를 보인다

4 이중 임계값 음성구간 검출기 알고리듬 및 음성 잔존구간

신경회로망을 이용한 음성구간 검출기에서 임계값을 뉴런의 출력 범위 사이에서 여러 가지를 임계값을 적용하여 보았다 그 결과 임계값이 작을수록 음성구간에서의 오차가 작게 나타나며 임계값이 커질수록 침묵구간에서의 오차가 작아짐을 볼 수 있었다 [1]

이중 임계값 방식은, 음성구간 검출기에서 보듯이 음성구간에서는 임계값이 작을수록, 침묵구간에서는 임계값이 클수록 오차율이 작게 나타나는 점을 감안하여 음성구간과 침묵구간에서 각기 다른 임계값을 사용하는 방법이다. 임계값을 조정하는 알고리듬으로는 그림 5와 같다.

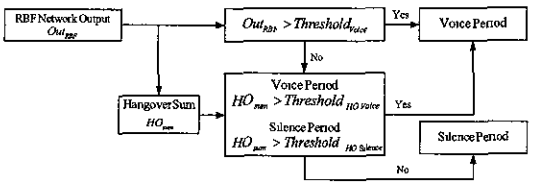

그림 5 이중 임계값 음성구간 검출기 알고리듬

본 논문에서는 음성구간 검출 방법으로 음성 잔존구간을 사용하였는데, 음성 잔존구간의 경우는 그림 5의 흐름도와 같은 방법으로 적용하여 음성구간과 침묵구간을 판별한다

이중 임계값 음성구간 검출기 알고리듬을 살펴보면, 음성구간이 이어지다가 출력(OUT_{RBF})이 임계값 보다 작게 나타나면, 즉시 침묵구간으로 판정하지 않고 이전 출력 값들과의 합(HO_{SUM})을 임계값과 비교하기 전에, 이 임계값을 이전의 출력상태에 따라 음성구간이면 $Threshold_{HO\,Voice}$, 침묵구간이면 $Threshold_{HO\,Silence}$ 로 선택하는 알고리듬 이다. 이런 경우 음성구간과 침묵구간에서 각각에 유리한 임계값이 사용되어 성능개선을 얻을 수 있다.

이때 출력(OUT_{RBF})은

$$HO_{SUM}(n) = \sum_{i=n-k}^{n} Out_{RBF}(i) \qquad (8)$$

과 같이 구해진다. 이때 k는 음성잔존구간의 길이이며 n은 현재 입력이다.

이렇게 할 경우, 잔존구간 길이보다 작은 길이를 갖는 오차는 없어지게 된다. 반대로 음성구간에서 침묵구간으로의 전이는 잔존구간 길이만큼 늦게 일어난다. 이로 인해 전체 데이터 중에서 침묵구간의 비율이 낮아진다 또한 침묵구간에서의 오차율도 다소 증가하게 된다 그러나, 음성구간에서 발생하는 오차를 줄이는 효과를 가져온다

Ⅳ. 실험을 통한 성능 분석

1 실험 환경 및 방법

그림 6에서와 같은 순서로 모의실험을 하였다.

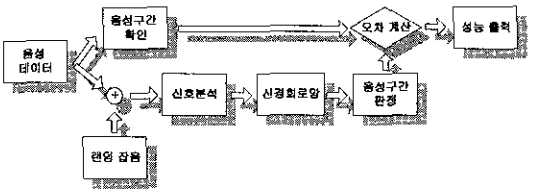

그림 6 모의실험 과정

실험에 사용하는 음성신호 데이터는 남성과 여성의 전화 대화내용으로 총 길이는 40,000구간 정도인 wave형식으로 저장된 파일이다. 즉, 8kHz 표본화에, 16bit 선형 양자화한 파일을 사용한다. 음성신호의 크기는 -24dBov 에서 -40dBov 사이였으며, 평균 -30dBov 였으며, 여기에 실제 응용환경에서와 같은 환경설정을 위해 음성 파일에 가우시안 잡음을 삽입하였다. 잡음의 크기는 -63, -58, -53, -48dBov로 하였다 음성신호 분석에 있어서는 블록 크기는 10ms, 그리고 20ms 크기의 해밍창함수를 사용하였다. 선형예측에 사용되는 과거 표본의 수(예측차수: prediction order)는 10개를 사용하였다. 신경회로망의 경우는, 하나의 은닉층을 사용한 가우시안 RBF 구조이며, 입력층, 은닉층, 그리고 출력층에 각각 3개, 30개, 그리고 1개의 뉴런을 사용하였다. 입력층의 경우 3개의 음성신호의 특징파라미터 단구간 평균전력, 영차 ML 파라미터, 그리고 피치 길이 차가 뉴런의 입력으로 사용된다. 학습 알고리듬으로는 중심점의 계산시에는 k-평균 클러스터링 방법을, 가중치의 계산시는 기존의

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

LMS를 사용하였다

논문에서 제시한 알고리듬을 적용한 실험 결과를 기존의 MLP 음성구간 검출기, ITU-T G.729 Annex B VAD와 성능 비교를 하였다.

2 가우시안 RBF 신경회로망을 이용한 음성구간 검출기 알고리듬의 성능분석

1) 임계값에 따른 음성구간 검출기 오차율 분석

RBF 신경회로망을 이용한 음성구간 검출기의 임계값을 정하기 위해서 RBF 신경회로망의 임계값에 따른 음성구간과 침묵구간의 오차율을 알아 보아야 한다. 그림 7은 RBF 신경회로망의 임계값에 따른 오차율을 나타낸다.

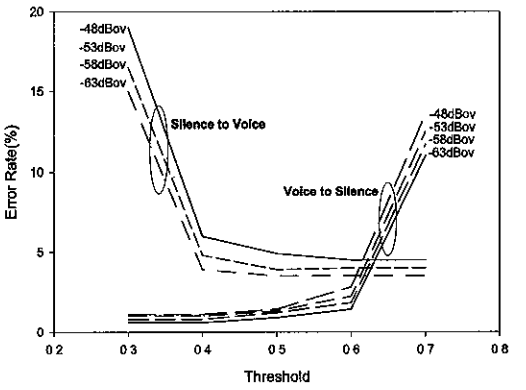

그림 7 RBF 신경회로망을 이용한 음성구간 검출기의 임계값에 따른 오차율

실험결과 음성구간과 침묵구간의 오차율은 임계값에 대해서 절충효과가(trade off) 나타났다. 임계값이 커 질수록 침묵구간에서 음성구간으로 변환 오차율은 줄어드는 반면 음성구간에서 침묵구간으로 오차율은 증가하였다. 효과적인 임계값은 0.4~0.6 사이의 값이었고 논문에서는 임계값을 0 54로 정하고 실험을 하였다.

2) 음성잔존구간에 따른 음성구간 검출기 오차율 분석

음성 잔존구간을 사용할 경우, 잔존구간의 길이를 정해야 한다 그림 8은 음성잔존구가에 따른 음성구간과 침묵구간에서 발생하는 오차길이에 따른 분포이다.

그림 8에서 보듯이, 음성구간의 오차는 음성 잔존구간의 길이가 길어질수록 오차발생률이 작아짐을 알 수 있다. 하지만 음성 잔존구간 길이가 길어지면 침묵구간에서 오차발생률이 커지게 된다. 따라서 이러한

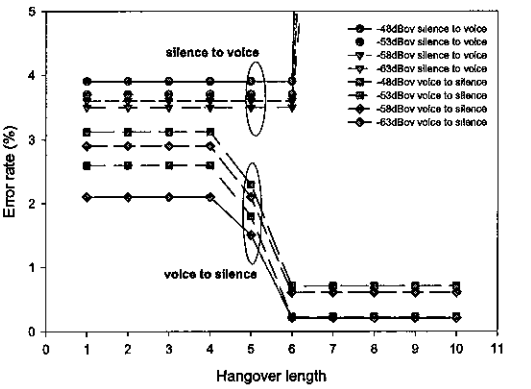

그림 8 음성 잔존 구간에서 길이에 따른 오차율 분포

특성을 모두 고려하면 효율적인 음성 잔존구간의 길이는 5에서 6구간이 적당한 것으로 보이며 앞으로의 모든 실험에서는 6구간으로 고정하여 사용하기로 한다.

$$HO_{SUM}(n) = \sum_{i=n-5}^{0} Out_{RBF}(i) \qquad (9)$$

따라서 논문에서는 식(9)와 같이 RBF 신경회로망의 출력(Out_{RBF})이 정해진다

3) RBF 노드 수에 따른 음성구간 검출기 오차율분석

RBF 신경회로망을 이용한 음성구간 검출기를 설계하기 위해서 RBF 신경회로망의 은닉층 노드 수를 정해야 한다 그림 9는 RBF 신경회로망의 은닉층 노드수와 음성구간 검출기 오차율과의 관계를 나타낸다

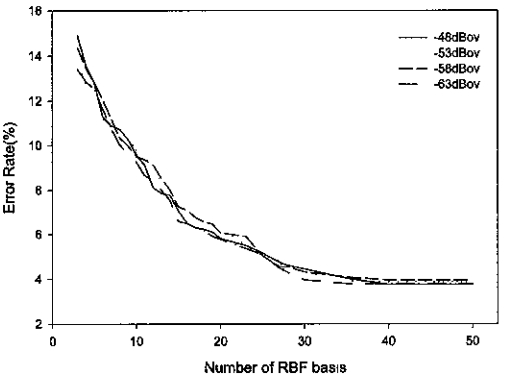

그림 9 RBF 노드수에 따른 음성구간 검출기 오차율

실험결과 RBF 신경회로망의 은닉층 노드수가 증가할수록 오차율은 감소하였다 하지만 노드수가 25~30개 이상부터 오차율은 거의 감소하지 않음을 확인했

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

다 이를 통해서 RBF 은닉층의 노드수를 늘일수록 오차율은 줄어들지만 노드를 많이 증가시켜도 오차율이 어느정도 이상은 줄어들지 않고 RBF 신경회로망의 계산량만 증가하기 때문에 이들 사항을 고려하여 가장효과적인 은닉층의 노드수를 결정해야 한다는 것을 확인했다 논문에서는 RBF 신경회로망의 은닉층 노드 수를 30개로 하여 실험을 하였다.

4) 음성구간 검출기 성능비교

그림 10은 논문에서 제시한 RBF 신경회로망을 이용한 음성구간 검출기와 기존 MLP 신경회로망 음성구간 검출기, ITU-T G 729 Annex B 음성구간 검출기의 오차 발생률 비교를 하였다

실험결과 논문에서 제시한 RBF 신경회로망 음성구간 검출기가 기존의 MLP 신경회로망 음성구간 검출기의 오차율은 음성구간에서 거의 비슷하게 나타났으나 침묵구간에서는 25% 정도의 성능 개선을 보였고, G 729 Annex B 음성구간 검출기 보다는 50%이상의 성능향상을 보였다 이러한 특징은 잡음크기가 클수록 더 확연하게 드러났다

이러한 결과는 RBF 신경회로망에 이중 음성 잔존구간 임계값이 적용되어 기존의 신경회로망에서 사용한 단일 임계값 구조보다는 우수한 성능을 나타내었고, CELP 파라미터, RBF 신경회로망, 이중 음성 잔존구간 임계값을 이용한 음성구간 검출기의 기능이 G 729 Annex B 음성구간 검출기 보다 우수하다는 것을 나타낸다.

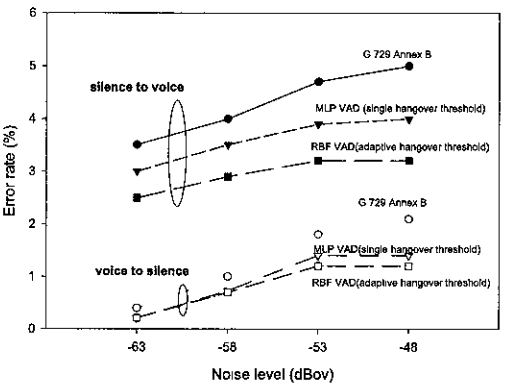

그림 10 수정된 음성 잔존구간 적용 시 음성구간 오차율

논문에서 제안된 RBF 신경회로망은 음성구간 검출기로서 이용할 경우 기존의 MLP 신경회로망 음성구간 검출기 보다 빠른 학습속도, 입력에 따라 이중적으로 변하는 신경회로망이라는 특징뿐만 아니라 이중 임계값을 적용하여 우수한 성능을 보이기 때문에 기존 제안되었던 신경회로망 음성구간 검출기 보다는 우수한 성능을 보인다는 것을 확인했다. 하지만 RBF 은닉층의 개수가 증가할수록 오차율이 줄어드는 성질 때문에 은닉층의 개수를 늘여야 하지만 이를 통해서 계산량이 늘어나는 단점을 확인할 수 있었다.

V. 결 론

본 논문에서는 이중 임계값 방식의 RBF 신경회로망을 이용한 음성구간 검출기에 대해 제안하고 이를 적용하여 그 유용성을 확인하였다.

RBF 신경회로망은 분류기로서의 우수한 성능을 보이면서 빠른 학습속도와 단순한 구조를 가지고 있어 적용하기 쉽고, 입력에 따라 이중적으로 학습되어가는 하이브리드 방법을 사용한 RBF 신경회로망 음성구간 검출기라는 장점도 있다. 또한 이중 임계값 알고리듬을 사용하여 음성구간과 침묵구간에서 오차율에 상호 유리한 임계값을 가지게 하여 성능을 향상시켰다.

실험결과, 이중 임계값이 적용된 RBF 신경회로망 음성구간 검출기 시스템은 복잡도를 크게 증가시키지 않으면서 성능 향상을 확인할 수 있었다 논문에서 제시한 RBF 신경회로망 음성구간 검출기는 기존의 MLP 신경회로망 음성구간 검출기의 오차율은 음성구간에서 거의 비슷하게 나타났으나 침묵구간에서는 25% 정도의 성능 개선을 보였고, G 729 Annex B 음성구간 검출기 보다는 50% 이상의 성능 향상을 보였다 이러한 특징은 잡음크기가 클수록 더 확연하게 드러났다. 또 학습속도가 빠르고 시간에 따라 입력채널의 특성이 변화더라도 이중적으로 신경 회로망이 변할 수 있는 특징을 이용한다면 다양한 변화가 예상되는 실시간 고속 음성구간 검출기에도 효과적으로 사용될 수 있다

참 고 문 헌

[1] Jotaro Ikedo, "Voice Activity Detection Using Neural Network," *IEICE Trans. Commun*, Vol E81-B, No 12, Dec 1998.

[2] Jaw Won Kim, Min Sik Seo, "A Voice Activity Detection Algorithm for Wireless Communication Systems with Dynamically Varying Background Noise," *IEICE Trans. Commun*, Feb 2000

[3] Dong Enqing, Liu Guizhong, Zhou Yatong, Cai

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

Yu, "Voice Activity Detection Based on Short-Time energy and Noise Spectrum Adaption," *Signal Processing, 2002 6th International Conference*, vol.1, pp 464-467, Aug. 2002.

[4] Jacek M. Zurada, "Introduction to Artificial Neural Systems," West publishing Company, 1992.

[5] D.R. Hush, B.G. Horne, "Progress in Supervised Neural Networks: What's New Since Lippmann?," *IEEE Signal Processing Magazine*, pp.8-39, January 1993.

[6] William J.Phillips, Caner Tosuner, William Robertson, "SPEECH RECOGNITION TECHNIQUE USING RBF NETWORKS," *WESCANEX 95. Communications, Power, and Computing. Conference Proceedings. IEEE.*, Vol.1, pp.185-190, May 1995.

[7] Chen S., Grant P.M., Cowan C.F.N, "Orthogonal least squares algorithm for training multi-output radial basis function networks," *Artificial Neural Networks, Second International Conference*, pp336-339, Nov. 1991.

[8] Zhang, Y.-M. Li, X.R. "Hybrid training of RBF networks with application to nonlinear systems identification," *Decision and Control, Proceedings of the 35th IEEE*, vol.1, pp 937-942, Dec. 1996.

[9] SIMON HAYKIN, "Neural Networks A Comprehensive foundation," Prentice Hall, pp298-305, 1999.

[10] Chedsada Chinrungrueng and Carlo H. Sequin, "Optimal Adaptive K-menas Algorithm with Dynamic Adjustment of Learning rate," Neural Networks, IEEE Transactions on, Volume:6, Issue:1, pp. 157-169, January 1995.

[11] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," Mar. 1996.

[12] ITU-T Recommendation G.729 Annex B, "A silence compression scheme for G.729 Optimized for terminals conforming to recommendation V.70," Nov. 1996.

[13] ITU-T Recommendation G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," Mar. 1996.

[14] ITU-T Recommendation G.723.1 Annex A, "Silence compression scheme," Nov. 1996.

[15] John Scourias, "Overview of the Global System for Mobile Communications," University of Waterloo, May 1995.

[16] A. M. Kondos, "Digital Speech," John Wiley & Sons, 1994.

[17] ITU-T Recommendations P.56, "Objective measurement of active speech level," March 1993.

[18] Richard V. Cox, " Three New Speech Coders from the ITU Cover a Range of Applications," *IEEE Comm. Magazine*, Sep. 1997.

김 홍 익 (Hong-Ik Kim) 준회원

1996년 2월 : 한국한공대학교 전자공학과 공학사

1996년 3월~2000년 8월: 삼성항공 전산팀 근무

2003년 2월 : 한양대학교 전자통신전파공학과 공학석사

2002년 6월~2002년 12월: ETRI 무선인터넷 보안팀 위촉연구원

2003년 3월~현재 : 한양대학교 전자통신전파공학과 박사과정

<관심분야> Digital Signal Processing, CATV Multimedia Systems, Broadband Communication Systems

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr

박 승 권 (Sung-Kwon Park) 정회원

1982년 2월: 한양대학교 전자통신전파공학과 공학사

1983년 8월: Stevens Institute of Technology, 전자공학과 공학석사

1987년 12월: Rensselaer Polytechnic Institute, 전자공학과 공학박사

1984년 1월~1987년 8월: Rensselaer Polytechnic Institute, Electrical, Computer and Systems Engineering Dept., Research Assistant

1987년 9월~1992년 8월: Tennessee Technological University, Electrical Engineering Dept., 조교수

1992년 9월~1993년 1월: Tennessee Technological University, Electrical Engineering Dept., 부교수

1993년 3월~현재: 한양대학교 공과대학 전자전기컴퓨터공학부, 교수

<관심분야> 지능형 데이터 방송, CATV Multimedia Systems, Digital Signal Processing

Copyright (C) 2005 NuriMedia Co., Ltd.
www.dbpia.co.kr