

HTM 기반의 소리 인식을 이용한 부품의 양·불량 판별 시스템

배선갑[†], 한창영^{**}, 서대호^{***}, 김성진^{****}, 배종민^{*****}, 강현석^{*****}

요 약

하나의 프레스로 여러 종류의 부품을 소량으로 생산하는 중·소형 공장에서는 부품 생산 과정에 발생하는 소리가 다양하게 나타난다. 이에 우리는 제품의 생산 순간의 소리를 인식하여 제품의 양·불량을 판별하는 시스템을 계층형 시간적 메모리(HTM : Hierarchical Temporal Memory) 기술을 이용하여 개발하였다. HTM 이론은 인간 두뇌의 신피질(neocortex)의 동작 원리를 컴퓨터에 접목시킨 이론이다. 이는 실제계에 대한 시공간적인 패턴을 계층적으로 기억하는 것으로 기존의 인식 기술보다 여러 경우에 인식률이 뛰어난 것으로 알려져 있다. 우리는 이 HTM 기술을 소리 인식에 적용하여 부품에 대한 양·불량 판별 시스템을 개발하였다. 개발 결과를 검증하기 위해 실제 공장에서 부품 생산 순간의 다양한 소리들을 녹음하고, 소리 HTM 네트워크를 구성한 후, 학습과 훈련을 반복하여 해당 부품의 불량여부를 판정하도록 하였다. 그 결과 잡음이 많은 생산 현장에서도 판정의 정확도가 높은 것을 확인하였다.

A Quality Identification System for Molding Parts Using HTM-Based Sound Recognition

Sun-Gap Bae[†], Chang-Young Han^{**}, Dae-Ho Seo^{***}, Sung-Jin Kim^{****}, Jong-Min Bae^{*****}, Hyun-Syug Kang^{*****}

ABSTRACT

A variety of sounds take place in medium and small-sized manufactories producing many kinds of parts in a small quantity with one press. We developed the identification system for the quality of parts using HTM(Hierarchical Temporal Memory)-based sound recognition. HTM is the theory that the operation principle of human brain's neocortex is applied to computer, suggested by Jeff Hopkins. This theory memorizes temporal and spatial patterns hierarchically about the real world, which is known for its cognitive power superior to the previous recognition technologies in many cases. By applying the HTM model to the sound recognition, we developed the identification system for the quality of molding parts. In order to verify its performance we recorded the various sounds at the moment of producing parts in the real factory, constructed the HTM network of sound, and then identified the quality of parts by repeating learning and training. It reveals that this system gets an excellent and accurate results at the noisy factory.

Key words: Sound recognition(소리 인식), HTM, neocortex(신피질), Identification system(인식 시스템)

※ 교신저자(Corresponding Author): 강현석, 주소: 경상남도 진주시 가좌동 900번지 국립경상대학교 8동309호(660-701), 전화: 055)751-5998, FAX: 055)762-1944, E-mail: hskang@gnu.ac.kr

접수일: 2010년 3월 18일, 수정일: 2010년 6월 22일
완료일: 2010년 8월 31일

[†] 준회원, 경상대학교 컴퓨터 과학과
(E-mail: bsgap@hanmail.net)

^{**} 준회원, 경상대학교 컴퓨터 과학과
(E-mail: hancy@gnu.ac.kr)

^{***} 정회원, 경상대학교 컴퓨터 과학과
(E-mail: chorokh@naver.com)

^{****} 정회원, 연암공업대학 컴퓨터공학과
(E-mail: sjkim@yc.ac.kr)

^{*****} 정회원, 경상대학교 컴퓨터 과학과
(E-mail: jmbae@gnu.ac.kr)

^{*****} 종신회원, 경상대학교 컴퓨터 과학과

※ 이 연구는 2009년도 경상대학교 연구년제 연구교수 연구 지원비에 의하여 수행되었음.

1. 서 론

소규모로 다품종을 소량 생산하는 중소기업 공장에서는 제품의 양·불량 검사가 대부분 사람의 육안으로 이루어진다. 따라서 제품 검사의 신뢰성이 검사자의 능력에 따라 많은 영향을 받는다. 하지만 중소기업 회사의 경우, 부품마다 양·불량 검사를 하기 위해 자동화 시스템을 설치하는 것은 많은 비용과 복잡함으로 인해 쉽지 않은 것이 현실이다. 따라서 이를 극복하기 위하여 적은 비용으로 간편하게 주문자의 요구에 따라 쉽게 양·불량 판정이 가능한 시스템의 개발이 필요하다.

최근에 이미지 인식을 이용하여 제품의 양·불량을 판정하는 시스템[1]과 자동차 번호판 인식, 얼굴인식 등의 연구가 있어 왔다. 이 시스템들은 이미지를 촬영하는 카메라의 성능과 촬영 환경에 따라 원본 이미지의 품질이 많은 차이가 난다. 원본 이미지의 품질이 좋지 못하면 판정의 정확도와 인식이 크게 떨어지게 된다. 또한 이 시스템들은 특정 목적에 맞게 개발이 된 것이어서 하나의 프레스에서 소량으로 여러 종류의 부품을 생산하는 소규모 공장에서는 사용할 수 없다. 우리는 이들의 문제점들을 해결하기 위하여 부품의 생산 순간의 소리를 인식하여 제품의 불량 여부를 간편하게 판단할 수 있는 HTM 기반의 소리 인식을 이용한 부품의 양·불량 판별 시스템(QIPSR; Quality Identification system for molding Parts using HTM-based Sound Recognition)을 개발하였다.

기존의 소리 인식에 사용된 기술들은 인공신경망, HMM(Hidden Markov Model)[2], GMM(Gaussian Mixture Model)[2], 그리고 SIDE(Stabilized Inverse Diffusion Equation)[3]와 결합된 알고리즘이 주로 사용되어 왔다. 인공신경망을 이용한 소리 인식 연구는 일반적으로 상호 연결된 뉴런에 의해 연상 기억, 필터, 변환, 분류, 최적화 등의 기능을 수행하여 예측을 할 수 있게 한다. 하지만 소리 신호가 가지는 시계열 모형의 특징을 기억 구조로서 인공신경망의 학습 구조에 도입한 연구는 매우 드물고, 네트워크 학습을 위해서는 많은 시간이 필요하며, 이런 일련의 학습 과정은 새로운 자료가 추가되면 전체 과정을 모두 다시 수행해야 하는 단점을 가지고 있다[4]. HMM은 음성 및 소리 신호에 포함된 잡음에 민감하고, 현재

상태 이전의 상태에 영향을 받으므로 가변적인 환경에서의 인식이 민감하게 변화하는 단점이 있다. 따라서 HMM으로 인식한 후 또 다른 알고리즘을 추가 적용하여 인식을 높이는 방법을 추가로 사용하고 있다[12]. 그리고 GMM은 화자 인식, 음성 변조 등의 음성 처리 분야에 많이 응용되고 있다[11]. 하지만 이 역시 복잡한 확률적 계산과 인식을 높이기 위하여 또 다른 알고리즘을 추가하여 시스템을 구성해야 된다. 또한, SIDE 방법을 이용하는 시스템의 경우, SIDE와 결합된 알고리즘을 사용하는 경우는 소리를 음절 단위로 분절하는 것에 불과하고, 소리의 높이, 길이, 파형을 인식하기 위하여 유전자 알고리즘과 상대적인 음의 높낮이를 설정하는 등의 알고리즘을 따로 적용하고 있다[3].

따라서 이러한 인식 알고리즘들을 하나의 프레스에서 여러 종류의 제품을 소량으로 생산하는 중소기업 공장에서 양·불량 판정시스템을 개발하는데 사용하기엔 적합하지 않다. 이러한 공장들에서는 생산하는 제품의 품종이 바뀌는 순간 그 품종에 맞게 적용이 가능해야 한다. 따라서 추가 프로그램이나 비용이 들지 않으면서도 가변적인 환경에서 사용 가능한 시스템이 되기 위해서는 인간 두뇌의 소리 인식 기능에 비견할 수 있는 적합한 기술이 필요하다. 인간의 소리 인식 능력은 주로 두뇌의 신피질의 기능에서 비롯된다고 보고 있다[5]. 그러므로 이를 흉내낸 계층형 시간적 메모리(HTM : Hierarchical Temporal Memory) 기법[6]이 이러한 환경에서 소리를 인식하는데 적합할 것으로 판단된다. 무엇보다, 제품이 생산될 때의 소리는 시간적으로 일정한 간격으로 발생하는 시계열 데이터의 일종이므로 시간적 변화에 따른 패턴의 인식에 강건한 HTM 이론이 소리를 이용한 본 연구에 적합하다고 판단된다. 따라서 우리는 인간이 소리를 인식할 때와 비슷한 방법으로 제품의 생산 순간의 소리를 HTM 기술을 이용하여 인식하게 하여 제품의 양·불량을 판별하는 시스템을 개발하였다.

우리는 개발된 QIPSR을 검증하기 위해 공장의 프레스에 소리를 채집하는 녹음 장치를 설치하고 제품의 생산 순간의 소리를 채집하였다. 그리고 이들 소리들의 특징을 그룹화 하고, 적절한 HTM 네트워크 구조를 결정된 후, 각 소리에 대하여 훈련시켜 적합한 파라미터를 설정하였다. 그리고 QIPSR을 현장의 소리에 대하여 테스트해 보았다. 즉, 여러 가지 제품

의 생산 시점의 소리와 잡음이 섞인 소리를 채집하여 추론하여 인식하는 실험을 수행하였다. 그 결과 제품의 생산 순간의 소리가 다른 잡음과 섞여도 인식률이 높은 것을 확인하였다.

논문의 구성은 다음과 같다. 2장에서는 HTM 이론의 전반적인 내용을 기술하고 3장에서는 우리가 개발한 QIPSR의 개념을 설명한다. 4장에서는 QIPSR의 구현을 보이고 그 결과를 실제 환경에 적용해 분석한다. 5장에서는 기존의 부품 판별 시스템들과 QIPSR과 비교한다. 그리고 마지막 6장에서는 결론과 향후 과제를 논한다.

2. HTM을 이용한 소리 인식

이 장에서는 개발에 사용된 HTM 기술의 전반적인 내용과 소리 인식 방법을 기술한다.

2.1 HTM 이론

HTM 이론은 미국의 Numenta사에서 2007년 중반에 개발된 기술로서, 기존의 경험적 탐색을 기본으로 하는 인공지능(AI)이나 단순 뉴런들의 연결로 보는 인공신경망(ANN)과는 달리, 시스템을 일정한 공통 연산을 수행하는 메모리 노드들의 계층 네트워크(hierarchy network)로 구성한다[5,6]. 이때, 네트워크의 기본 단위인 노드(node)는 신피질(neocortex)의 기본 단위인 신피질 칼럼(neocortex column)에 해당하는 것으로 본다. 이렇게 구성된 HTM 네트워크는 실제계의 시간적 패턴 정보와 공간적 패턴 정보를 이용하여 효율적으로 지능적 판단을 할 수 있게 한다. 또한, HTM은 일종의 메모리 시스템으로 문제

마다 각자 다른 알고리즘을 수행하는 것이 아니라 문제를 해결하는 방법을 학습한다. 그리고 학습 과정에서 감각 신호의 특성들 사이의 시공간적 관계성이 계층적 메모리 구조에 형성된다. 이러한 학습이 끝나면 새로운 입력 패턴을 인식할 수 있게 된다.

2.2 HTM 네트워크

HTM에서 네트워크는 노드들의 계층으로 구성된다. 데이터는 계층의 바닥으로부터 감각(Data Sensor)을 통해 입력되고, 입력된 데이터는 노드들의 계층 구조를 통해 위로 올라가면서 처리 된다. 이때 계층의 각 노드는 학습과 추론의 두 단계로 이루어진 알고리즘을 사용한다[7,8].

학습 단계에서는 네트워크가 훈련 패턴에 노출되어 범주(Category)들에 패턴들을 사상시키는 모델을 구축하고, 학습이 완료되면 네트워크는 추론 단계로 넘어간다. 추론 단계에서는 네트워크가 이전에 본 일이 없는 새로운 패턴에 대해 범주들에 대한 간단한 분포를 생성한다. 전형적인 노드들의 계층적 구조를 그림 1의 (가)에 개략적으로 나타내었다. 그림에서 보는 것과 같이 HTM 네트워크는 Zetal, ZetalTop, Data Sensor, Effector 등 몇 가지 유형의 노드들로 구성된다[6]. 이들 중 가장 핵심적인 것이 Zetal 노드와 ZetalTop 노드이다. 그 중에서 Zetal 노드가 계층의 대부분을 구성하게 된다. Zetal 노드는 그림 1의 (나)와 같이 정보를 공간적 풀러(Spatial pooler)와 시간적 풀러(Temporal pooler)의 두 가지 구성요소로 처리하게 된다. 공간적 풀러는 데이터의 공간적 패턴을 분류하는 역할을 수행한다. 즉, 입력 벡터들의 분산된 표현들을 분류하는 방법으로 학습한다. 시간적

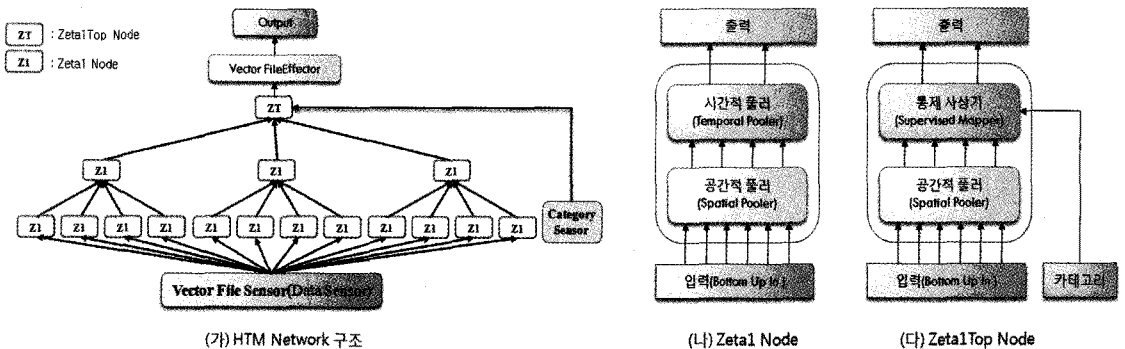


그림 1. HTM 네트워크의 구조

풀러는 데이터의 동시발생(Coincidence)을 기초로 시간적 패턴을 그룹화 하는 알고리즘을 수행한다. 이는 추론 결과가 이전 입력 값뿐만 아니라 최근의 입력 값을 기초로 하여 계산하는 시간 기반의 알고리즘이다. 그리고 그림 1의 (다)에 나타난 ZetaTopNode는 공간적 풀러와 통제 사상기로 이루어져 있다. ZetaTopNode의 공간적 풀러는 ZetaNode와 같은 역할을 하며, 통제 사상기는 범주 정보를 이용하여 최종적으로 판단을 하게 된다. 보다 상세한 것은 [8]을 참조하기 바란다.

2.3 HTM 네트워크에서의 학습과 추론

HTM 네트워크의 동작은 크게 학습과 추론 단계로 이루어진다. 학습(Learning)은 HTM 네트워크에서 아래로부터 계층별로 이루어진다. 즉, 학습 단계에서 공간적 풀러는 입력 패턴을 정량화하여 이들의 동시발생들을 찾는다. 공간적 풀러의 출력은 동시발생의 목록들이고, 이들을 시간적 풀러에 보내게 된다. 시간적 풀러는 목록들을 받아서 시간적으로 동시에 발생하는 패턴들을 그룹으로 묶음으로써 학습을 완료한다. 계층별로 학습이 완료되면 위 계층으로 보낼 데이터를 결정하기 위해 입력 데이터들에 대해서 추론하는 추론(Inference) 단계로 이행된다. 추론은 아래로부터의 입력 패턴에 대해 학습 단계에서 학습한 동시발생에 대한 그룹들에 대한 범주의 분포를 발생시킨다.

이렇게 하여 네트워크의 모든 노드들이 전체적으로 학습이 완료되면 훈련이 끝난 HTM 네트워크가 완성된다. 이제 이를 이용하여 새로운 입력 데이터를 인식할 수 있게 되는데, 이것은 최상위 계층에서 추론을 하는 것을 의미한다. 최상위 계층의 추론 결과는 입력 데이터의 패턴 범주들에 대한 확률 분포를 나타낸다.

2.3 HTM 기술을 이용한 소리 인식

HTM 기술은 주식가격 예측, 이미지 인식, 음성 인식 등의 여러 응용에 이용될 수 있다. 현재까지 이 기술을 이용하여 개발된 시스템들은 주로 이미지 인식에 집중되어 있다. 그러나 이 기술을 음성 인식에 적용한 예는 네 사람의 음성을 녹음하여 학습을 하게 한 후, 이들의 음성이 누구의 음성인지를 구분하는

정도의 응용이 전부이다[9]. 즉, 이 기술을 적용하여 기계 소리를 인식하여 제품의 양·불량을 판단을 하는 시스템은 아직 보고된 바가 없다. 그래서 우리는 음성 인식에 HTM 기술을 응용한 예를 바탕으로 하여 프레스에서 제품이 생산될 때의 소리의 패턴을 인식하는 시스템인 QIPSR을 개발하였다. 프레스에서 제품이 생산되는 순간의 소리를 분류하여 소리의 파형을 분석한 결과 양품이 생산되는 순간의 소리 파형과 불량품이 생산되는 순간의 소리 파형이 약간 다르다는 것을 확인하였다. 그래서 우리는 이들 소리를 녹음하여 전처리 과정을 거쳐서 64개의 실수형 데이터로 만들어 HTM 네트워크의 입력 데이터로 사용하였다. 그리고 이를 위한 적절한 네트워크를 구성하고 파라미터를 조정하여 판정의 성공률이 높은 시스템을 완성하였다. 다음 장들에서 QIPSR에 대한 상세한 설계와 구현을 보인다.

3. 부품의 양·불량 판별 시스템의 구조

이 장에서는 우리가 개발한 HTM 기반의 소리 인식을 이용한 부품의 양·불량 판별 시스템(QIPSR; Quality Identification system for molding Parts using HTM-based Sound Recognition)의 구조를 설명하고 그 처리 과정을 알아본다.

3.1 시스템 구조

그림 2는 전체 시스템의 구조를 나타낸 것이다. 생산 라인에서 소리 데이터를 녹음하게 되고, 녹음된 소리 데이터는 전처리에 의하여 전처리 과정을 거쳐 네트워크의 입력 데이터를 생성하게 된다. 생성된 데이터는 HTM 엔진 조정기(Handler)와 데이터베이스 조정기에 동시에 보내진다. HTM 엔진 조정기는

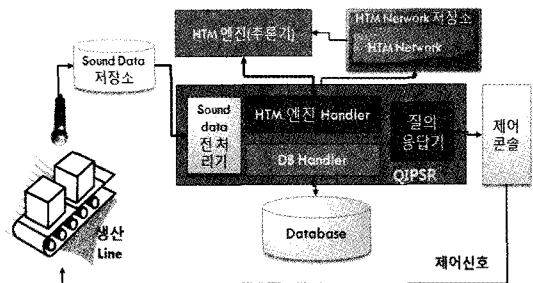


그림 2. 시스템 구조도

네트워크를 구성하여 HTM 엔진으로 넘겨 추론을 하게 된다. 저장소는 추론된 결과를 받아 HTM 네트워크의 전반적인 결과물을 XML 파일로 HTM 네트워크 저장소에 저장하게 된다. 그리고 데이터베이스 조정기는 소리 데이터 전처리기에서 생성된 소리 데이터를 데이터베이스에 저장하고, HTM 엔진 조정기의 요청에 따라 훈련 데이터와 실험 데이터를 데이터베이스로부터 가져오는 역할을 한다. 또한, 질의 응답기는 HTM 엔진 조정기가 보내온 결과를 제어 콘솔로 보내고, 제어 콘솔에서 외부로부터의 명령을 받는 역할을 하게 된다.

3.2 소리 HTM 네트워크를 위한 전처리 과정

소리 데이터는 스트림 데이터의 일종으로서 시간적으로 연속적인 값을 가지고 있다. 이 소리 데이터를 HTM 네트워크에 적용하기 위하여 전처리 과정을 거쳐야 한다. HTM 네트워크의 입력은 벡터 센서로부터 입력이 되기 때문이다. 이 벡터 센서의 입력은 64개의 실수 값으로 이루어지게 된다. 그렇게 하기 위하여 먼저 소리의 녹음은 8 kHz 모노 형식으로 녹음된다. 이 녹음된 소리를 512 바이트씩 분할하여 하나의 해밍 윈도우(Hamming window) 크기로 하

여 FFT(Fast Fourier Transformation) 변환을 한다. 다음으로 이 윈도우를 80바이트씩 이동하면서 연속적으로 소리 데이터를 추출한다. 그리고 이들을 차례로 FFT 변환하여 변환된 값들을 누적시켜 겹치게 된다. 이렇게 하면 소리에 대한 스펙트로그램(Spectrogram)이 형성된다. 다음으로 이 스펙트로그램 값의 로그 멜(Log Mel) 스펙트럼을 생성하여 64 삼각 필터를 적용하여 64개의 주파수 빈(Bin)을 생성한다 [9]. 이 64개의 주파수 빈은 소리의 진폭과 위상을 나타내는 실수 값으로 소리 HTM 네트워크의 입력으로 쓰이게 된다. 이 전처리 과정으로 HTM의 특징인 시간적 패턴 정보를 잘 표현 할 수 있게 된다. 그림 3은 소리 데이터의 전처리 과정을 나타낸 것이다.

3.3 소리 HTM 네트워크의 구성

우리는 소리 HTM 네트워크를 총 3계층으로 구성하였으며, 그림 4에 보인 것과 같은 구조를 갖게 했다. 바닥 계층으로 입력되는 입력 데이터를 64개로 하면 계층 1의 각 노드에는 16개씩 입력이 되어 4개의 Zeta1Node 노드들로, 계층 2에는 2개의 노드들로 구성되고, 최상위 계층은 한 개의 Zeta1TopNode로 구성되어 적합한 네트워크가 된다. 그리고 소리

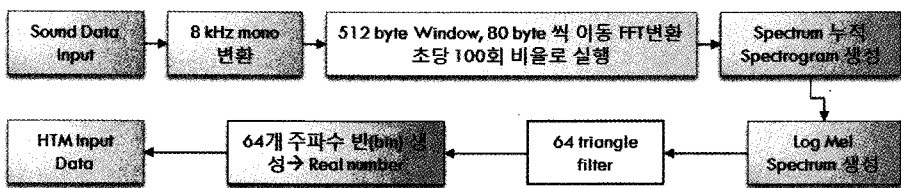


그림 3. 소리 데이터의 전처리 과정

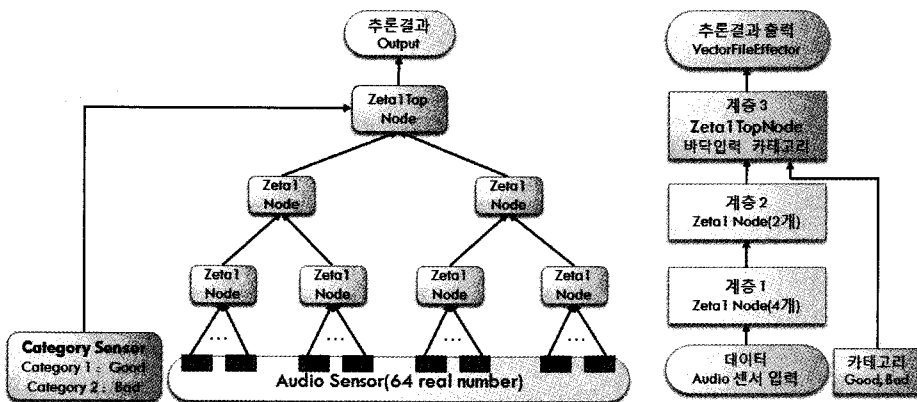


그림 4. 소리 HTM 네트워크의 구조

HTM 네트워크의 입력 데이터는 64개의 실수 값으로 이루어진 주파수 빈이다. 이 주파수 빈의 값을 오디오 센스가 네트워크에 입력으로 받아들인다. 받아들인 64개의 실수 값은 계층 1의 네 개의 ZetaNode에 각각 16개씩 입력된다. 입력을 받은 각 ZetaNode는 공간적 패턴과 시간적 패턴을 학습하게 된다[6]. 학습은 각 계층별로 이루어진다. 계층 1에서 학습이 마무리 되면 계층 2로 출력을 보내게 되는데, 계층 2에서는 2개의 ZetaNode가 존재한다. 계층 2에서의 학습이 끝나면 학습된 결과를 최상위 계층에 있는 ZetaTopNode에 보낸다. ZetaTopNode는 자식 노드의 입력과 범주 센서로부터 받은 범주들에 대하여 상대적 확률 분포를 발생시켜 추론하게 된다. 일반적인 HTM 네트워크의 학습과 추론에 관한 더 자세한 내용은 참고문헌 [7,8]을 참조하기 바란다.

3.4 파라미터 결정

구성된 소리 HTM 네트워크가 최대의 정확도를 얻기 위해 우리는 소리 데이터 전처리에 적용되는 파라미터들과 네트워크에 관여하는 파라미터들로 분리하여 실험을 통해 조정하였다. 먼저 전처리에 적용되는 파라미터는 윈도우(Window)의 크기와 윈도우의 이동에 관한 파라미터(Parameter)이다. 이때 윈도우의 크기는 해밍 윈도우(Hamming window)로서 512 바이트를 기본 단위로 하여 FFT 변환을 하기 위한 단위이다. 그리고 윈도우 이동에 관한 파라미터 중 중요한 것은 윈도우 진행 크기(Advanced size)이다. 실험 결과 이 두 파라미터가 시스템의 정확도에 많은 영향을 미치는 것으로 나타났다.

다음으로 네트워크에 관여하는 파라미터이다. 이 파라미터는 시스템의 정확도를 높여 주어져서 중요하게 다루어져야 한다. 그러므로 여러 번 반복 실험을 하여 최적의 값을 찾아야 한다.

네트워크에 관여하는 파라미터의 첫째는 최대 거리(Max distance)이다. 최대 거리는 하나의 기준 센서로부터 원본 벡터 w 와 비교 대상인 입력 벡터 x 사이의 유클리디안(Euclidean) 거리 $d^2(x, w) = \sum_{i=1}^n (x_i - w_i)^2$ 로 계산된 것의 최대값이다. 이는 학습하는 동안에 입력된 데이터 벡터들이 한 기준점에서 최대 거리값을 설정하면 입력되는 데이터 센서의 벡터들의 거리를 각각 계산하여 이것과 비교하게 된다. 최대 거리의

값보다 작을 경우는 같은 동시발생으로 취급하고, 최대 거리 보다 클 경우에는 새로운 동시발생으로 취급한다. 따라서 최대 거리를 크게 설정해 줄 경우, 속도는 빠르지만 정확한 비교는 어렵다. 대신 최대 거리를 작게 설정할수록 학습 속도는 느리지만 좀 더 정확한 비교를 할 수 있다. 하지만 적용하는 분야에 따라 적절한 최대값을 찾아야 추론의 정확도를 높일 수 있어서 최대값의 변화에 따른 실험으로 정할 수밖에 없다.

둘째는 시그마(Sigma)이다. 시그마는 추론 과정에서 필요한 파라미터로서 동시발생 사건의 정규 분포의 표준편차이다. 이는 동시에 발생하는 경우의 수의 범위를 지정한다. 이 값이 작은 것은 동시발생 사건이 밀집되어 있음을 의미하며, 큰 값은 넓게 분포함을 뜻한다. 대개 추론시 잡음이 많이 섞인 소리일 경우는 이 값을 크게 설정하고, 깨끗한 소리일 경우는 이 값을 작게 설정하면 정확도가 높게 나타난다. 우리가 추론 과정에 사용한 소리 데이터는 훈련 데이터와 미세한 차이가 있는 것으로 판정의 정확도를 높이기 위해 적당한 시그마를 설정하는 것이 중요하다. 그래서 최대거리와 시그마 파라미터를 변화시키면서 실험을 반복하여 최적의 값을 찾았다. 그 결과들은 다음 장에서 보다 상세하게 알아본다.

3.5 훈련 데이터와 테스트 데이터

훈련에 사용된 데이터는 부품 하나가 생산되는 순간의 소리를 녹음하는 장치(초지향성 마이크)를 프레스에 직접 부착하여 연속적으로 녹음을 하였다. 부품은 자동차의 문의 잠금장치에 사용되는 부품(TMLock)을 채택하여 실험하였다. 정상적인 부품의 모양은 그림 5와 같다. 정상적인 부품이 생산되는 순간에 발생하는 소리는 거의 동일하지만 불량품 생산 순간에 발생하는 소리는 불량유형에 따라 여러 가지로 다양하다. 그래서 불량품 유형 중 세 유형을 분류하고 각 불량유형이 생산된 순간에 발생한 소리

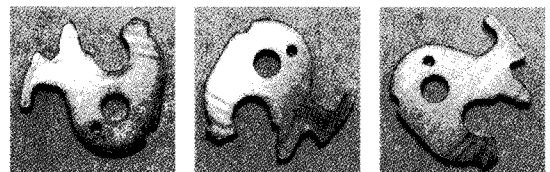


그림 5. 부품 이미지

를 녹음된 데이터들 중에서 찾아서 분류하였다. 그리고 이들 세 유형의 소리를 훈련 데이터와 실험 데이터로 하여 실험하였다. 각 유형별로 녹음된 소리의 5초간의 데이터를 훈련 데이터로 사용하였다. 실험 데이터는 녹음된 파일에서 훈련 데이터로 사용한 부분을 제외한 데이터를 무작위로 5초 간격으로 4번 반복하여 총 20초간의 데이터를 추출하여 사용하였다. 그림 6은 불량품의 발생 유형 1, 2, 3의 훈련 데이터와 실험 데이터의 개략적인 파형, 상세 파형 그리고 각 유형에 해당하는 불량 부품의 이미지를 나타낸 것이다.

4. 시스템 구현

이 장에서는 우리가 개발한 QIPSR의 구현 환경 및 결과 인터페이스 화면을 보이고, 최적 시스템이

될 수 있도록 하는 파라미터 조정 작업에 대하여 설명한다.

4.1 구현 환경 및 결과

개발된 QIPSR의 구현 환경은 다음과 같다. Windows XP를 기본 운영 체제로 구성하였고, HTM 엔진인 NuPIC 1.6.1과 Vitamin D 툴킷[10]을 이용하였으며, Python 2.5 언어를 개발 언어로 사용하였다. 그림 7은 개발한 시스템의 사용자 인터페이스의 하나인 인식 결과 화면의 예이다.

4.2 파라미터 조정 실험

파라미터의 조정 실험은 두 단계로 나누어 진행하였다. 첫째는 소리의 전처리 과정에 관여하는 파라미터 변화 실험, 둘째는 HTM 네트워크에 관여하는 파

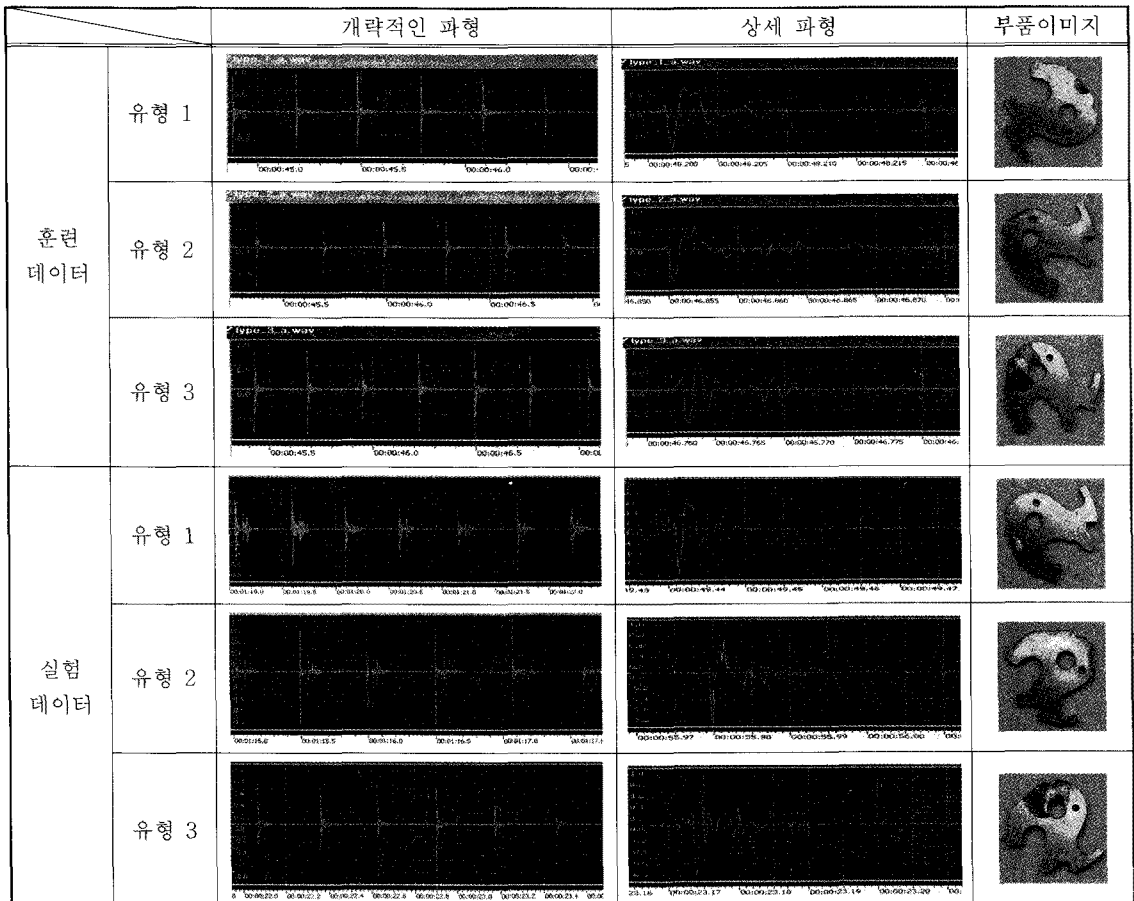


그림 6. 훈련 데이터 및 실험 데이터 파형의 예

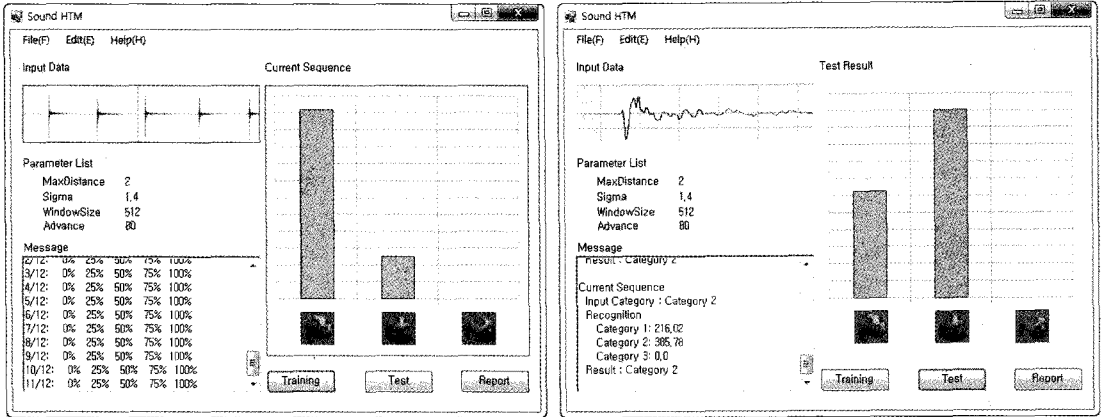


그림 7. QIPSR의 인식 결과 화면

라미터 변화 실험으로 하였다. 먼저 소리의 전처리 과정에 필요한 파라미터인 FFT 변환을 위한 헤밍 윈도우의 크기와 진행 크기를 변화시키면서 실험을 수행하였다. 그 다음으로 HTM 네트워크에 관여하는 파라미터인 최대 거리와 시그마 값을 변화시켜 추론의 정확도를 조정하는 실험을 수행하였다.

4.2.1 윈도우 크기

소리 데이터를 HTM 네트워크의 입력 데이터로 변환하기 위하여 FFT 변환을 하게 된다. 이 때 필요한 것이 헤밍 윈도우이다. 이는 소리 패턴을 추출하기 위한 최소 단위의 크기이기 때문에 그 크기가 소리의 패턴을 결정하는데 영향을 준다. 따라서 추론의 정확도에 많은 영향을 미칠 것이라는 가정으로 헤밍 윈도우 크기를 변화시켜 가면서 실험을 수행하였다. 그 결과 이 윈도우의 크기에 따라 정확도가 많이 차이가 나는 것을 확인하였다. 윈도우의 크기는 표 1에서 보는 것과 같이 512 바이트에서 가장 높은 정확도를 보였다.

표 1. 윈도우 크기에 따른 정확도 (단위: %)

Window Size	유형1	유형2	유형3	평균
64	68.3	66.2	68.5	67.7
128	85.5	85.6	83.2	84.7
256	90.7	89.6	87.4	89.2
512	94.2	95.7	95.9	97.3
768	93.3	92.9	94.5	93.6
1024	90.2	89.2	88.5	89.3

4.2.2 윈도우 진행 크기(advanced size)

소리 데이터는 시간적으로 연속적인 값이므로 특정한 패턴을 찾기 위하여 FFT 변환에 사용한 윈도우를 이동하게 된다. 이 윈도우의 이동 간격을 잘 조정하면 뚜렷한 패턴의 특징을 찾을 수 있을 것이다. 따라서 추론의 정확도에도 영향을 미칠 것이라는 가정으로 실험을 수행하였다. 윈도우의 크기는 가장 정확도가 높았던 512 바이트로 고정하고, 윈도우 진행 크기를 변화시키면서 실험을 하였다. 그 결과는 그림 8과 같았다. 보는 바와 같이 윈도우 진행 크기가 80일 때 정확도가 가장 높았다.

4.2.3 최대 거리

앞의 실험으로 가장 정확도가 높은 윈도우의 크기는 512 바이트, 윈도우 진행 거리는 80 바이트로 고정한 상태에서 최대 거리 파라미터를 변경시키면서 실험하였다. 그 결과는 표 2에 나타나 있다.

최대 거리 값을 변화시키면서 실험해 본 결과, 위

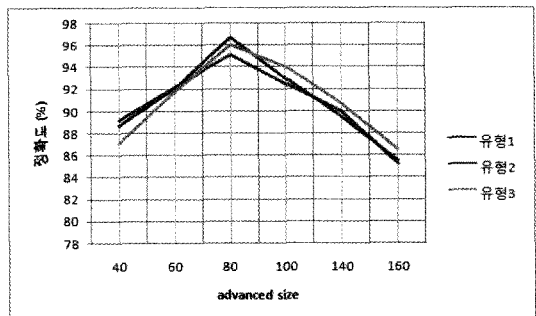


그림 8. 윈도우 진행 크기에 따른 정확도

표 2. 최대 거리 변화에 따른 정확도 (단위 : %)

최대거리	유형1	유형2	유형3	평균
1	91.4	92.1	90.2	91.2
2	95.2	98.2	97.5	96.9
4	92.5	94.2	94.7	93.8
8	87.9	84.3	89.0	87.1
16	81.2	81.7	83.0	81.9
32	72.6	77.0	72.0	73.8
64	58.3	63.5	55.3	59.0

에서 보는 바와 같이 최대 거리가 2일 때 평균적으로 정확도가 가장 높은 것으로 나타났다.

4.2.4 시그마

우리는 최대 거리를 가장 정확도가 높은 2를 적용하여, 시그마 값을 변화시켜 가면서 실험을 수행하였다. 그 결과를 표 3에 나타내었다. 표에서 보는 바와 같이 시그마 값이 1.4일 때 평균적으로 가장 높은 정확도를 보였다. 이는 최대 거리의 제품근의 값일 때 제일 정확도가 높았다.

표 3. 시그마 값에 따른 정확도 (단위 : %)

Sigma Category \	1.0	1.4	2.0	2.8	4.0	5.6	8.0
유형 1	94.5	97.4	95.3	95.6	94.4	90.2	87.7
유형 2	95.3	98.9	95.2	95.2	94.0	93.2	85.4
유형 3	95.5	98.2	96.0	95.0	93.3	92.2	89.0
Average	95.1	98.2	95.5	95.3	93.9	91.9	87.4

4.2.5 최적의 파라미터

최종적으로 정리하면 판정의 정확도가 가장 높은 파라미터는 윈도우의 크기는 512 바이트, 윈도우의 이동거리는 80바이트, 최대거리는 2, 시그마 값은 1.4인 경우가 전체적으로 정확도가 높았다.

전체적으로 표 3의 결과에 의하면 각 카테고리별로 98%이상의 높은 정확도를 보였다. 실제 현장에서는 두세 번 이상 계속해서 불량이 일어났을 때 프레스를 점검하는 등의 대처를 할 수 있기 때문에, 이러한 시스템의 성능 결과만으로도 큰 효과를 볼 수 있어 매우 유용한 시스템임이 확인되었다.

한편, 소리는 하나의 프레스에서 발생하지만 여러 가지 소리가 섞여서 발생된다. 프레스의 제품 생산

과정을 살펴보면 유압 장치를 이용하여 압력을 높이고, 순간적으로 철판을 눌러 부품을 생산한다. 이러한 과정에서 유압을 높이는 소리와 컨베이어벨트의 움직이는 소리 등이 섞여서 제품을 생산하는 순간의 소리를 추출하는 것이 상당히 어려운 작업이었다. 이를 해결하기 위하여 우리는 유압을 측정하는 장치를 설치하고, 프레스의 유압이 누르는 순간의 압력이 되는 순간부터 부품이 생산되는 순간까지 발생하는 소리를 녹음할 수 있도록 하였다. 이렇게 하여 부품이 생산되는 순간만의 소리를 분리하여 녹음할 수 있었다. 그리고 이를 이용하여 앞에 언급한 전처리 과정을 거치면서 일정한 패턴으로 인식할 수 있게 하였다. 결과적으로 우리는 잡음이 많은 현장 환경에서도 높은 인식 성능을 확인할 수 있었다.

5. 기존 연구와의 비교

지금까지 소리 인식 기술을 기반으로 한 각종 응용들이 다양하게 개발되어 왔다. 예를 들어 실시간으로 음악을 판별하는 시스템[11], 소아의 울음소리의 스펙트럼 특징을 분석하여 질병을 진단하는 시스템[4], 주성분 분석 기법을 이용한 심음 인식 시스템[12], 환경 소리 인식 시스템[13,14] 등이 있다.

실시간으로 음악을 판별하는 시스템은 FM 라디오에 수신된 신호 중에서 순수하게 음악 부분만 추출하는 시스템으로 통계적 모델 기반의 클러스터링 방법인 GMM을 사용하여 음악 부분을 추출하고 있는데, 이는 학습을 통한 방법으로 추론을 하고 있지만 역시 복잡한 계산식으로 부품 판별 시스템에의 적용에는 어려움이 많기 때문에 우리는 클러스터링 방법으로 HTM 기술을 이용하였다. 소아의 울음소리의 스펙트럼 특징을 분석하여 질병을 진단하는 시스템의 예는 소아의 울음소리의 스펙트럼 분석으로 정상 아와 비교하여 질병 여부를 판정하는 시스템으로 단순한 특정 주파수 대역으로 한정하여 비교하고 있다. 또한, 주성분 분석 기법을 이용한 심음 인식 시스템은 심장 박동 소리를 주성분 분석(Principal Components analysis) 방법으로 심장의 소리를 인식하고 분류하는 것으로, 기준이 되는 심장 박동 소리의 데이터베이스를 만든다. 인식하는 과정은 데이터베이스에 저장된 소리와 비교하여 추론하는 방식을 취하고 있다. 그리고 능동형 기술기 가중치 제약에 기반

한 환경 소리 인식 시스템[14]은 청각 장애자들을 위하여 개발된 것으로 주변의 소리를 인식하여 알려주는 시스템이다. 이 역시 알람, 도어 벨, 경적, 전화 벨 소리 등의 주변 소리를 녹음하여 데이터베이스화하고, 새로운 소리를 인식하게 된다. 그런데, 이 시스템들의 공통적인 특징은 특정한 하나의 문제 해결에 초점이 맞추어져 있어서 대상이 바뀌면 또 다르게 프로그래밍을 해야 하는 번거로움과 추가 비용이 소요된다는 것이다. 따라서 이러한 소리 인식 기술들을 다품종 소량 생산하는 소규모 기업에서 양·불량 판정을 위한 시스템에 적용하기엔 적합하지 않다.

그에 비하여 본 시스템은 생산되는 부품의 종류가 바뀌는 순간에 제품 생산 순간의 소리를 녹음하여 학습, 추론하는 과정을 통해 바로 적용이 가능하기 때문에 다품종 소량 생산의 소규모 공장의 중소 제조업의 가변 환경에서 보다 적응력이 좋다. 그리고 부품 교체시도 거의 추가 비용이 들지 않으며, 설치된 시스템을 변화 없이 그대로 사용 가능하다. 또한, 일반 PC에 녹음 센서를 부착하여 사용하기 때문에 설치비용이 저렴하며, 사용자가 상황에 맞게 파라미터를 임의로 조정하여 정확도를 높일 수 있도록 하여 효율적으로 시스템을 운용할 수 있다. 뿐만 아니라 학습 후에는 부품에 대한 양·불량 판단이 네트워크를 통해 아래에서 위로 불과 몇 단계(우리의 시스템에서는 3단계)의 계층만으로 전파되면 판단이 가능하기 때문에 매우 빠르게 추론이 가능하여 실시간으로 현장에서 활용할 수 있다.

6. 결론 및 향후 과제

지금까지 우리는 소리 HTM 네트워크를 기반으로 제품의 양·불량 판별 시스템 QPSR을 개발한 결과를 보이고 실제 데이터를 이용해 실험 결과를 분석하였다. 우리가 개발한 QPSR을 현장에 적용하여 판정의 정확도를 확인한 결과 98%에 이르는 것으로 확인되었다. 이는 100%에 미치지 못하는 것지만 현장에서 두세 번 이상 연속적으로 불량이 발생하면 프레스를 정지하는 등의 조치를 취함으로써 충분히 보완이 될 것으로 생각한다. 그리고 본 시스템은 간단하게 소리 데이터를 녹음할 수 있는 센서와 일반 PC만으로 설치가 가능하므로 적은 비용으로 쉽게 사용이 가능하다. 또한, 각 개별 공장에서 생산하는 부품이

다르더라도 각각의 생산 시점에 발생하는 소리를 녹음하여 학습, 훈련, 추론을 거쳐서 인식하므로 동일한 시스템을 사용할 수 있다. 따라서 새로운 다른 부품을 생산하더라도 추가 비용이 들지 않는다. 뿐만 아니라 사용자가 상황에 맞게 파라미터를 임의로 조정하여 정확도를 높일 수 있도록 하여 효율적으로 시스템을 운용할 수 있다.

향후 더 많은 데이터 수집 및 실험으로 다양한 부품들의 인식에 대한 정확도를 높이는 작업이 필요하며, 일반 사용자가 보다 쉽게 활용할 수 있도록 하는 사용자 학습 지원 시스템과 부품의 이미지를 이용한 인식 시스템을 개발하여 이와 연동하여 판정하게 하면 보다 더 정확도가 높은 시스템이 될 것으로 기대된다.

참고 문헌

- [1] 이상학, 서명호, 정태충, “이미지 프로세싱을 이용한 자동 너트 검사 장비 개발,” 한국정보처리학회논문지, 제11-A권 제4호, pp. 235-242, 2004. 4.
- [2] Alain Dufaux, “Detection and Recognition of Impulsive Sound Signals,” Institute of Microtechnology University of Neuchatel Switzerland, Phd. Thesis, 2001.
- [3] 형아영, 이준환, “SIDE를 이용한 자동 음악 채보 시스템,” 정보처리학회논문지 B 제16-B권 제2호, pp.141-150, 2009. 4.
- [4] 김봉현, 이세환, 조동욱, “소아 감기와 소아 폐렴 간의 울음소리 스펙트럼 특징 분석,” 정보처리학회논문지 B 제15-B권 제4호, pp.301-306, 2008. 8.
- [5] Jeff Hawkins, ‘On Intelligence’, New York, Henry Holt, 2005.
- [6] Jeff Hawkins and Dileep George, “Hierarchical Temporal Memory,” Numenta, Inc., March. 2007. http://www.numenta.com/Numenta_HTML_Concepts.pdf
- [7] Numenta Inc., “Numenta Platform for Intelligent Computing,” June. 2008. http://www.numenta.com/for-developers/software/pdf/nupic_plugin_guide.pdf

- [8] Dileep George and Bobby Jaros, "The HTM Learning Algorithms," Numenta Inc., March. 2007. http://www.numenta.com/for-developers/education/Numenta_HTM_Learning_Algos.pdf
- [9] Numenta Inc., "Speech Processing with Hierarchical Temporal Memory," June. 2008.
- [10] Vitamin D, Inc., "Vitamin D Toolkit Reference Guide," June. 2008. <http://www.vitamindinc.com/downloads/Vitamin%20D%20Toolkit%20Reference%20Guide.pdf>
- [11] 강현우, "FM 라디오 환경에서의 실시간 음악 판별 시스템 구현," 정보처리학회논문지 B 제 16-B권 제2호, pp. 151-156, 2009. 4.
- [12] 이상민, 홍성홍, "주성분 분석 기법을 이용한 심음 인식," 전자공학회논문지 제38권 SC편 제5호, pp. 59-69, 2001. 9.
- [13] Selina Chu, Shrikanth Narayanan, and C.-C. Jay Kuo, "Environmental Sound Recognition With Time - Frequency Audio Features," IEEE Transactions On Audio, Speech, And Language Processing, Vol. 17, No. 6, pp. 1142-1158, August. 2009.
- [14] 정영진, 이운정, 김필운, 김명남, "능동형 기울기 가중치 제약에 기반한 환경소리 인식시스템용 DTW 알고리즘," 한국멀티미디어학회 논문지 제11권 제4호, pp. 471-480, 2008. 4.



배 선 갑

1987년 경북대학교 물리학과 졸업(학사)
2004년 경상대학교 대학원 컴퓨터과학과 졸업(석사)
2010년 8월 경상대학교 대학원 컴퓨터과학과 졸업(공학박사)

관심 분야: 패턴인식, 지능시스템, 내장형 데이터베이스



배 중 민

1980년 서울대학교 수학교육과 졸업(학사)
1983년 서울대학교 대학원 계산통계학과 졸업(석사)
1985년 서울대학교 대학원 계산통계학과 졸업(박사)

1982년~1984년 한국전자통신연구소 연구원
1997년~1998년 Virginia Tech. 객원연구원
1984년~현재 경상대학교 컴퓨터과학부 교수
관심분야: XML, 데이터베이스 통합, 기계학습



한 창 영

2004년 경상대학교 자연과학대학 컴퓨터과학과 졸업(학사)
2007년 경상대학교 대학원 컴퓨터과학과 수료(석사)
관심 분야: 패턴인식, WEB, UNIX, 데이터베이스



강 현 석

1981년 동국대학교 전자계산학과 졸업(학사)
1983년 서울대학교 대학원 전산학과 졸업(석사)
1989년 서울대학교 대학원 전산학과 졸업(박사)

1981년~1984년 2월 한국전자통신연구원 연구원
1984년 3월~1993년 2월 전북대학교 전임강사, 부교수, 교수
1993년 3월~현재 경상대학교 컴퓨터과학과 교수, 컴퓨터정보통신연구소 연구원
관심분야: 멀티미디어, 내장형 데이터베이스, 지능시스템



서 대 호

2009년 경상대학교 컴퓨터과학과 졸업(학사)
2010년 경상대학교 대학원 컴퓨터과학과 수료(석사)
관심 분야: XML, 기계학습, 데이터베이스 통합



김 성 진

1979년 경북대학교 공과대학 전자공학과 졸업(공학사)
1981년 경북대학교 대학원 전자공학과 졸업(공학석사)
1995년 영남대학교 대학원 컴퓨터공학전공(공학박사)

1983년~1985년 삼성전자(주) 시스템개발부
1996년~1997년 University of California, Santa Cruz 연구교수
1985년 3월~현재 연암공업대학 컴퓨터공학과 교수
관심분야: 멀티미디어, 병렬처리, 운영체제