한국심리학회지 : 생물 및 생리
Korean Journal of Biological and Physiological Psychology
2002. Vol. 14, No. 1, pp. 75-91

주성분분석에 기초한 세포외 활동전위 신호의 분류*

이 준 열　　서 효 정　　이 춘 길†

서울대학교 심리학과

컴퓨터 기술의 발전과 세포활동을 계측하는 기법이 개선되면서, 복수의 전극을 통해 세포활동을 계측하고 계측된 세포활동의 아날로그 신호를 연속적으로 분석하는 것이 가능해 졌다. 이 신호를 분석하는데 있어서 신호를 그 기원에 따라 분류하는 것이 분석의 중요한 문제로 인식되고 있다. 본 논문에서는 다중채널의 전극을 이용해서 채집된 세포외 활동전위(extracellular action potential)를 분류하는 기법을 소개하고 이를 적용한 실례를 제시하였다. 여기서 적용한 전위분류 절차는 주성분분석(principal component analysis, PCA)에 기초한 것으로서, 우선 몇 개의 주성분으로 이루어진 특질차원 공간상에 각각의 활동전위들을 표상시킨 후 Maximum likelihood estimator(MLE)를 이용해 적절한 군집의 수를 결정하고 Learning vector quantization(LVQ) 알고리즘을 통해 각 군집의 중심점을 결정하는 방법을 사용하였다. 이렇게 결정된 각 군집의 중심점과 각 활동전위 사이의, 특질차원 공간상에서의 거리(Euclidean distance)를 계산함으로써 분류가 이루어지며, 중심점과 활동전위 간 거리의 분포를 이용해서 잡음을 제거하면 세포에 따른 활동전위의 분류가 완료된다. 이 모든 과정은 본 실험실에서 작성된 세포분류 프로그램(WAVESORTER)으로 이루어졌다. 따라서 본 논문에서는 WAVESORTER와 그 환경에 대한 소개와 이 프로그램을 통한 실제적인 세포분류과정을 기술하였다.

* 본 연구는 과학기술부 뇌신경정보학연구사업의 지원으로 이루어졌음.

† 교신저자 : 이춘길 / 151-742 서울시 관악구 신림동 산 56-1 서울대학교 심리학과 / cklee@plaza.snu.ac.kr

신경세포가 활동전위를 발생하는 빈도는 세포의 활동 수준을 나타내는 대표적인 지표로 사용되고 있다. 흔히 세포 바깥에 위치하는 전극을 통해서 계측되는 신호에는 잡음을 포함하여 여러 개의 신경세포들로부터 기원하는 전위들이 다양한 크기와 형태로서 섞여 있는데, 따라서 한 세포의 활동전위의 빈도를 세기 위해서는 우선 전극을 통해서 계측되는 신호를 분류해야 한다. 이 목적으로 사용하는 전통적인 방법은 전위변별기(windows discriminator)를 이용하는 것이다. 이 방법에서는 우선 전극을 통해 계측되는 세포 활동의 전기적 신호를 증폭한 후, 증폭된 신호의 최고점(peak)을 일정한 전위와 비교하여 일정 역치를 넘어서는 신호의 최고점(즉 활동전위의 최고점)이 있는지를 결정하는 것이다. 혹은 일정 역치를 넘어서면서 일정 역치를 넘어서지 않는 소위 'window' 내에 포함되는 신호의 최고점이 있는지를 결정하는 것이다(이춘길, 박정현, 1991). 즉, 활동전위의 크기를 기준으로 단일 신경세포에서 발생하는 활동전위를 분류해 내는 것이다. 활동전위의 최고점이 동일하더라도 다른 세포들에서 발생하는 신호들이 있을 수 있으므로, 전위변별기에 추가하여 'time window'를 사용하기도 한다. 후자의 방법은 일정 역치를 넘어서는 전위가 역치를 넘어서는 시점을 기준으로 하여 일정 시간의 경과(즉, time window) 후에 일정 수준의 전위에 도달하는지를 결정하는 것인데, 활동전위의 크기에만 전적으로 의존하는 전자의 방법에 비해서 단일 세포의 활동전위를 분리하는 데 있어서 상대적인 장점이 있다. 활동전위가 상승하는 패턴을 두 시점에서 파악하는 것이기 때문이다. 그러나, 여전히 완벽한 분류는 보장되지 않는다.

전위변별기를 통한 세포활동의 기록과정은 다음과 같다. 역치 기준을 만족시키는 활동전위가 있을 때, 변별기는 한 개의 TTL(Transistor-Transistor-Logic)신호를 발생시키고 이 신호를 컴퓨터에 접속하여 신호가 발생한 시점을 기록한다. 추후 이 기록들이 활동전위의 시기와 빈도에 대한 지표로 사용된다. 그러나 이 방식을 따르면 활동전위의 크기나 패턴이 어떠했는지에 대한 정보가 없어지게 되며 무엇보다도 전극을 통해서 한 개의 세포로 분류된 활동전위만이 기록되어서 전극 주위의 여러 세포들의 활동에 대한 정보를 낭비하게 된다. 또, 전위변별기(시간변별기를 포함하여)는 활동전위의 최고치에 의존하기 때문에 해석에 있어서 두 가지 점에서 오류의 여지가 남게 되는데, 첫째, 기록된 TTL이 과연 동일한 세포로부터 계측된 것인지에 대한 불확실성이다. 둘째 문제는 배경잡음(noise)에 대한 해결이 제대로 이루어 졌는가 하는 것이다. 크기가 작은 배경잡음의 경우 전위변별기의 역치값을 높게 잡아서 걸러낼 수 있지만, 동물의 움직임이나 울음으로 인해 유발되는 잡음은 그 크기가 매우 크므로 전위변별기의 역치값을 높게 잡더라도 제거할 수가 없다. 따라서 전위분류기의 TTL 출력에 의존하게 되면 동물의 움직임이나 울음이 세포활동인 것으로 오인될 여지가 있으며, 동물이 연속적으로 움직일 경우, 이때 채집된 세포활동을 버릴 수밖에 없는 상황이 유발되기도 한다. 최근 컴퓨터 등의 기술 발전으로 인해 세포의 활동을 채집, 해석하는 방법에 있어서 개선이 있었으며 위와 같은 문제점들이 해결될 수 있는 길이 열렸다.

세포활동을 채집하는 방법에 있어서 최근의 가장 큰 변화는 우선 세포활동의 채집에 사용되는 전극의 수에 있어서의 변화이다(Lewicki, 1998). 이전에는 주로 하나의 전극을 동물의 뇌 속에 내려 세포활동 측정에 이용했지만, 현재는 다중채널(multichannel)의 전극을 사용하여 여러 영역에서

세포활동을 동시에 측정하는 방법을 사용하기도 한다. 여러 전극을 사용할 경우 한번에 사용하는 전극의 수가 늘어나므로 한번의 실험을 통해 획득할 수 있는 정보의 양이 커질 뿐 아니라 각 전극에서 채집되는 세포들이 활동하는 시점 간의 관계를 분석하여 기능적인 연결에 대한 추론이 가능해지므로 실질적으로 얻어지는 정보의 양은 전극의 수에 따른 단순한 증가 이상이라 할 것이다. 세포활동을 채집하는 방법과 관련한 또 다른 변화는 활동전위를 컴퓨터로 받아들이는 방식의 변화이다. 종래에는 전위변별기를 통과한 TTL 신호만이 컴퓨터에 저장되었지만, 최근의 방식에서는 하나 혹은 그 이상의 전극에서 들어오는 세포활동에 대한 전위신호 자체가 높은 속도의 A/D(Analogue-to-Digital) 변환기를 통해 컴퓨터에 저장된다. 따라서, 세포활동의 파형에 대한 정보가 저장되므로 파형에 근거한 다양한 분석이 가능하게 된다.

세포활동의 아날로그 신호(analogue signal)를 획득할 수 있게 되면서 대두된 문제는 하나의 전극에서 채집된 전위신호의 기록으로부터 활동전위를 가려내고 각 활동전위가 어떤 세포들로부터 기원하는지를 분류해내는 문제이다. 본 논문에서, 먼저 세포분류의 몇 가지 방법에 대해 대략적으로 살펴본 후, 본 실험실에서 작성된 세포분류 프로그램(WAVESORTER)을 중심으로 세포분류의 실제적인 과정과 PCA를 기반으로 한 세포분류 방법을 설명하였다. 여기서 예시하는 활동전위의 기록은 각성상태의 고양이 상구(superior colliculus) 세포의 활동을 세포외 전극을 통해서 계측하고 25kHz의 속도로 A/D 변환하여 컴퓨터에 저장한 것들이다.

세포분류의 단계

특질 분석(feature analysis)

세포를 분류하기 위해서는 먼저, 활동전위 파형의 어떤 특질을 분류에 사용할 것인지를 결정하여야 한다. 전통적으로 사용되어 온 특질들로는 주로 활동전위의 높이와 폭, 그리고 활동전위 최고점과 최저점 간 크기 등을 들 수 있다(Lewicki, 1998). 각각의 활동전위를 특질들로 이루어진 n차원의 공간상에 표시하게 되면 각 특질의 유사성에 의해 하나의 전극에서 도출된 활동전위 신호가 몇 개의 군집(cluster)으로 분류되는데, 이렇게 분류된 패턴에 근거하여 세포분류가 가능해진다(그림 1). 전위변별기를 이용한 활동전위의 탐지 역시 특질 분석의 일종이라 할 수 있는데, 활동전위의 최고점이라는 특질에 근거하여 활동전위와 배경잡음을 구분하기 때문이다. 하지만, 전위변별기를 이용하여 활동전위를 분류하는 것은 하나의 특질만을 사용한다고 볼 수 있다. 즉, 1차원 상에서 활동전위에 대한 변별이 이루어지는 것이다. 2차원 상에서 얻어질 수 있는 정보의 양은 1차원 상에서 얻어질 수 있는 정보량의 단순한 2배 이상의 효과를 가지므로 1차원 보다는 2차원, 2차원 보다는 3차원의 특질 공간 속에서 활동전위의 분류(즉, 세포의 분류)가 이루어질 때 활동전위의 보다 많은 측면이 고려되므로 보다 정확한 분류가 이루어 질 수 있다.

이와 같이 특질 차원을 이용해서 활동전위를 분류하고자 할 때에는 우선 신뢰롭고 변별력 있는 활동전위 파형의 특질차원을 결정하여야 하고, 다음으로 선택된 특질 차원에서 표현된 활동전위들을 일정한 기준에 근거하여 각각의 군집으로 분류하게 된다.

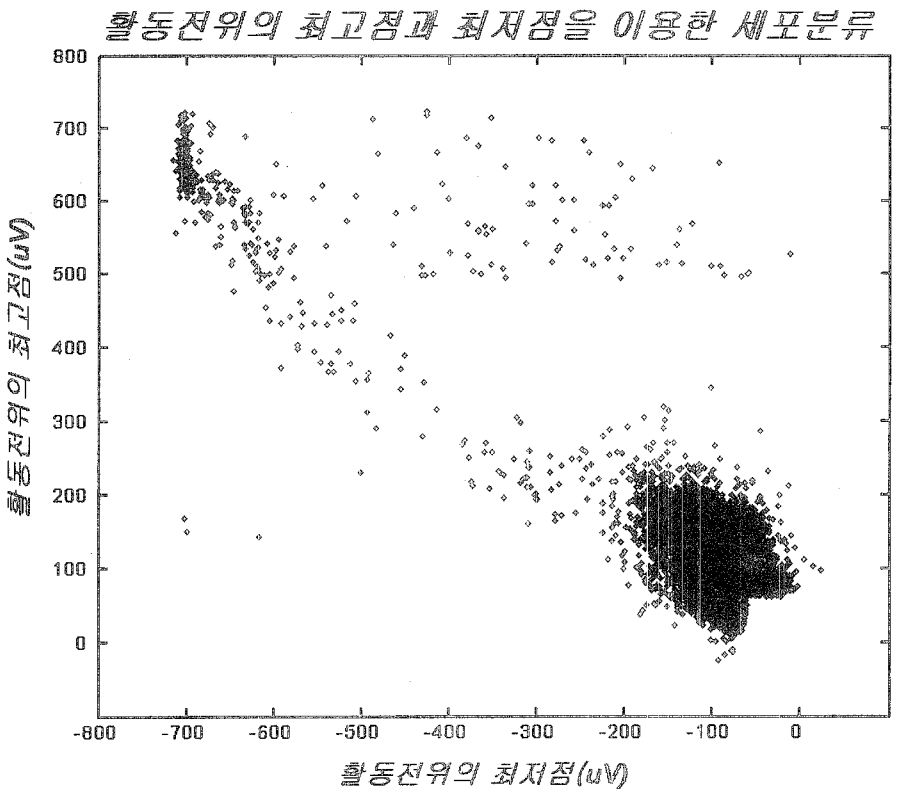

그림 1. 각 활동전위의 최고점과 최저점이라는 두 가지 특질 차원 공간상에 각각의 활동전위를 표상하였을 경우의 예. 각 점은 한 개의 활동전위를 나타낸다. 이 경우 두 개의 군집(왼쪽 위와 오른쪽 아래)이 형성되고 있음을 알 수 있다. 자료의 분포를 보았을 때 활동전위의 최고점과 최저점이 일정 수준 이상의 경우만 표시되어 있는 것을 알 수 있는데, 이는 활동전위를 추출할 때 일정 기준을 넘어서는 부분만 추출했기 때문에 나타나는 현상이다(본문참고). 본 자료는 고양이의 상구(superior colliculus)에서 채집된 실제 세포활동 자료이다(nr1101-channel 5).

주성분 분석(Principal component analysis, PCA)을 이용한 특질의 선택

활동전위를 분류함에 있어서 가장 적절한 특질은 각각의 파형에 있어서의 차이들을 가장 잘 설명해 주는 것이라 할 수 있다. 위에서 언급하였듯이, 전통적으로 세포분류에 사용되어 온 특질은 활동전위의 최고점과 최저점 간의 크기, 높이와 폭 등이었다. 그러나 이러한 특질들은, 얻어지는 활동전위의 특징, 실험상황(전반적인 배경잡음의 수준 등), 혹은 세포활동의 기록이 이루어지는 뇌의 영역 등에 따라 달라질 수 있기 때문에 실험에 따라 세포분류에 적절한 특질이 달라질 수 있다는 문제점이 존재한다.

최근, 주성분 분석(principal component analysis, PCA)을 이용하여 특질을 선택하는 방식이 세포분류에 효율적으로 이용될 수 있음이 보고 되고 있다. 이 방법은 특정 자료가 존재할 때 이 자료들 사이의 분산(variance)을 가장 잘 설명해주는, 서로 직교하는 축(axis)을 찾아내는 기법이라 할 수 있

다. 다시 말해서 자료의 분산을 극대화 시켜주는 방향을 잡아내는 직교인 벡터들의 순차적인 배열을 구하는 기법이다(Giri, 1995)[1]. 따라서 주성분 분석을 통해 활동전위들 사이의 분산을 가장 잘 설명해 줄 수 있는 특질을 잡아낼 수 있으며, 이 특질차원 상에서 각각의 활동전위를 분포시켰을 때 활동전위들은 서로 유사한 형태적 특성을 지닌 것끼리 군집을 이루게 된다. 이 때 특질의 개수는 자료에 있어서의 분산을 얼마만큼 설명해 주는가에 기준하여 선택한다. 그림 2a에서 알 수 있듯이 자료의 분산에 대한 설명력이 가장 큰 주성분을 포함하여 몇 개의 주성분을 선택하느냐에 따라 전체 자료의 분산에 대해 설명할 수 있는 정도(%)가 달라지므로 분산에 대한 설명력의 선택을 통해 주성분, 즉 특질의 개수를 결정할 수 있다. 주성분 분석에 의한 특질의 선택은 여러 가지 점에서 고전적 방법에 비해 강점을 가진다. 첫째는, 고전적 방식으로 미리 정해진 몇 개의 파라미터가 세포분류에 사용되는 경우, 매 실험의 조건이나 얻어진 자료의 조건에 따라 가장 적절한 파라미터를 임의로 사용해야 하지만, 주성분 분석을 사용하는 경우, 실험의 조건에 따라 활동전위의 전체적 특징이 매번 달라지더라도, 매 실험에서 얻어진 활동전위 파형들에서 가장 많은 분산을 설명해주는 차원이 자동적으로 찾아질 수 있다는 점이다. 즉, 자동성과 함께 임의성을 피할 수 있는 장점이 있다. 둘째는, 자료를 분류하는 것은 자료들 간에 존재하는 차이점에 근거하므로, 자료가 높은 변산을 보이는 특질을 사용하는 것은 분류의 변별력을 높인다. 주성분 분석은 자료에서 가장 많은 분산을 설명해주는 정도를 특질변수 선택의 기준으로 사용한다는 점에서 세포분류의 효율적인 방법으로 고려되어 질 수 있다. 셋째는, 고전적으로 사용되던 한정된 수(1~3개)의 특질을 벗어나, 더 많은 차원의 특질을 이용함으로써 또한 세포분류의 변별력을 높일 수 있다는 점이다.

군집분석(cluster analysis)

위의 방식으로 결정된 특질 차원 상에서 각각의 활동전위를 표시했을 때 세포 분류를 통해 마지막으로 이루어져야 할 일은 각 군집의 경계를 구하는 일이다. 가장 흔히 사용되며 또한 간단한 방법 중 하나는 각각의 활동전위가 속해 있는 군집의 평균점까지의 거리를 이용해서 각 군집의 경계를 결정하는 것이다. 미리 결정된 특질차원 상에서 각 군집의 평균점을 구한 후 모든 활동전위와 이들 평균점 사이의 유클리드(Euclidean)거리를 계산하여 이 거리를 최소화 시키는 방향으로 각 활동전위를 분류하는 방법을 말한다. 그러나 이러한 방법은 특정 특질 차원 상에서 세포의 군집이 명확하게 결정되지 않을 경우에, 혹은 형성되는 군집의 모양이 구형(spherical)이 아닐 경우에 (유클리드 거리는 군집의 모양이 구형임을 가정

1) 주성분 분석은 다음과 같이 설명할 수 있다. n개의 변수들로 이루어져 있는 자료군 X는 n차원상의 무선 벡터 X로 표상 가능하다. 즉,

$$X=(X_1, X_2, X_3, \ldots, X_n$$

로 나타낼 수 있다. 본 논문에서 사용된 자료는 30개의 A/D point로 이루어진 활동전위의 파형이므로 총30개의 변수로 이루어져 있고 따라서 30차원 상의 무선 벡터로 표상 가능하다. 이 때 X의 첫 번째 주성분(principal component)을 Z이라 하면 Z_1은 요소들의 표준화된 선형 결합을 통해 만들어진 새로운 변수로 분산을 극대화시킨다는 특성을 가진다. 즉,

$$Z_1=L'X,\ L=(l_1, l_2, l_3, \ldots l_{30},\ L\in E^{30}$$

풀어서 쓰면,

$$Z_1=l_1X_1+l_2X_2+ \ldots +l_{30}X_{30}$$

과 같이 정의되는데, 여기서의 L은 $L'X$의 분산이 최대가 되도록 하는 값으로 결정된다. 마찬가지로 X의 두 번째 주성분을 Z_2라 한다면 이 때의 L은 $L'X$의 분산을 두 번째로 크게 만드는 값으로 결정된다.

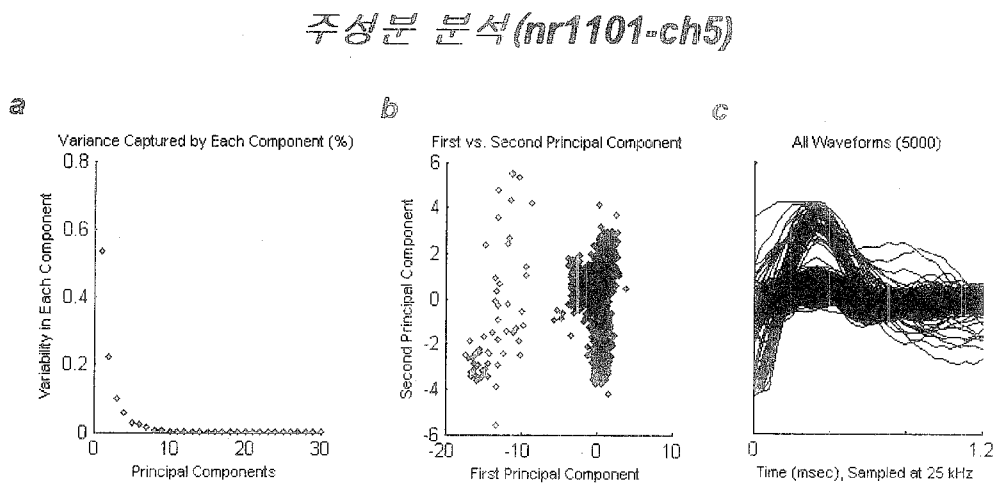

그림 2. 추출된 활동전위의 정렬 및 주성분 분석. a: 주성분 분석의 결과 각각의 주성분이 가지는 전체 자료의 분산에 대한 설명력을 보여준다. 각각의 활동전위는 30개의 A/D point로 이루어져 있기 때문에 추출될 수 있는 주성분의 전체 개수 역시 30개이다. 처음 3개의 주성분이 대부분의 분산을 설명하고 있음을 알 수 있다. b: 첫번째의 주성분과 두번째의 주성분으로 이루어진 2차원의 특질 차원 공간상에 5000개의 활동전위를 표시한 결과이다. 그림에서 2-3개의 군집이 형성되었음을 알 수 있다. 이 결과는 활동전위의 최고점과 최저점이라는 특질 차원 공간상에서 각 활동전위들을 표상한 그림 1의 결과와 비교할 수 있다. c: 추출된 5000개의 활동전위가 최고점을 기준으로 정렬됨을 보여준다. 25kHz의 A/D 변환률을 적용했을 때 30개의 A/D point는1.2ms의 길이에 해당하는 활동전위를 구성한다.

할 때 군집분류의 기준으로 사용이 가능하다) 문제의 여지가 있다. 따라서 오늘날에는 통계적 기법을 이용하여 베이지안 군집분석(Bayesian clustering analysis) 방법 등을 사용하기도 한다 (Lewicki, 1994).

WAVESORTER에 이용된 알고리즘

활동전위의 추출

WAVESORTER는 본 연구실에서 Matlab (The Mathworks Inc., U.S.A.)을 사용하여 작성한 프로그램이다. 이 프로그램은 SPIKER(Yu, 1999)라는 프로그램을 기반으로 하여 작성 되었으며, 따라서 세포분류의 기본적인 원리 또한 SPIKER에서 사용된 방식을 그대로 따랐다. 이 프로그램은 총 세 부분으로 나누어져 있으며 각각 활동전위 파형의 추출, 분류, 그리고 분류된 활동전위의 확인을 담당하였다. 우선 활동전위 파형의 추출을 담당하는 부분에서는 1에서 5개의 다중채널(Thomas Recording, Germany)을 통한 세포활동 기록으로부터 하나의 채널만을 선택하여 전극이 동일한 깊이에 위치하는 한 회기(session)의 실험결과를 추출하였다. 다중채널에서 채집된 세포활동의 기록은 25kHz의 A/D 변환과정을 거쳐 컴퓨터에 저장되었으므로 이렇게 저장된 파일에서부터 한 채널의 세포활동기록을 인출할 수가 있었다. 인출한 세포활동의 기록에서 우선 일정한 역치(threshold)를 넘어서는 활동전위들을 모두 추출하였다. 이 작

업을 통해 일차적으로 배경잡음이 걸러졌는데, 이때 사용되는 역치는 세포활동의 전반적인 수준을 통해 실험자가 직접 결정하거나 아니면 전체적인 배경잡음 수준을 고려하여 배경잡음 수준의 4 ~ 5 표준편차 이상의 것만을 선택하는 방식을 사용하였다. 역치는 활동전위의 최고점, 최저점 둘 다에 근거하여 고려되었다. 즉, 특정 역치를 넘어서는 세포활동이나 특정 역치보다 최저점이 작은 세포활동 두 경우를 모두 고려하였다. 추출되는 세포활동 하나하나는 총 30개의 A/D 자료로 이루어졌으며(25kHz의 AD변환에서 보았을 때 1.2ms에 해당) 최고점이나 최저점에 기준하여 추출되었으므로(최고점에서부터 좌우로 얼마, 혹은 최저점으로부터 좌우로 얼마 등), 추출된 각각의 활동전위는 최고점 혹은 최저점을 기준으로 정렬되었다(그림 2c).

PCA를 사용한 특질의 추출

추출된 활동전위를 이용하여 전위분류(즉, 세포분류) 작업이 이루어졌는데, 우선 PCA를 통해 추출된 세포활동의 일정분산(일반적으로 95% 분산을 사용)을 설명해 줄 수 있는 만큼의 주성분을 선택하였다. 이 때, 선택된 주성분의 개수가 5개 미만일 경우, 주성분 분석의 장점을 활용하고 분류의 변별력을 높이기 위해서 최소한 5개의 특질차원을 사용하여 분류하도록 하였다. 그런 후 선택된 주성분의 개수에 해당하는 특질차원 상(n개의 주성분이 선택될 경우 n차원)에 모든 활동전위들을 표시하게 되면 각각의 활동전위들은 n차원상의 한 점으로 표상된다. 주성분분석을 사용하지 않고 추출된 세포활동의 파형 정보를 그대로 사용할 수도 있다. 본 프로그램(WAVESORTER)에서는 각 활동전위당 30개의 A/D 자료를 사용했기 때문에 이렇게 할 경우 각 활동전위는 30차원상에서 하나의 점으로 표상된다. 따라서 n차원상의 각 점(활동전위)들은 파형의 특성(결국 활동전위의 모양에 근거해서 분류가 이루어졌는데)에 따라 군집을 형성하게 된다.

군집 분석

일단 주성분분석에 의해서 각 활동전위들이 n차원 상에서 점들로 표상되고 나면, 이 점들의 분포를 가장 잘 기술하는 군집 구조(clustering structure)를 결정하게 된다. 보다 구체적으로는, 현재 자료가 몇 개의 군집에 의해 가장 잘 기술되는가 하는 것과, 각 군집의 n차원 상에서의 위치 즉 군집의 중심점이 어디인지를 결정하는 것이다. 먼저, 자료를 가장 잘 기술하는 군집의 개수를 결정하기 위한 기준으로 Maximum Likelihood Estimator(MLE)를 사용하였고, 이렇게 해서 결정된 군집들의 중심점을 결정하기 위해서는 Learning Vector Quantization(LVQ)의 방법을 사용하였다. 간단하게 기술하면,

(1) 미리 가능한 군집 개수의 범위를 결정한다. 각 군집 개수는 자료를 기술하는 하나의 모델이 되고, 4~5개로 범위를 결정하는 경우, M=4, 5로 표현할 수 있다. 이 경우, 4개의 군집에 의해 자료를 설명하는 모델과 5개의 군집에 의해 자료를 설명하는 모델, 이렇게 두 모델 중에서, 실제 자료의 구조를 산출했을 확률이 높은 모델을 선택하게 된다.

(2) 먼저 각 모델에 대해 LVQ 방법을 사용하여, 군집의 중심점을 결정하고, 자료를 분류한다. M=4의 모델을 예로 들면, 4개의 군집 중심점을 결정하고, 활동 전위들을 4개의 군집으로 분류한다(LVQ의 구체적인 알고리즘은 뒤에서 설명하겠다).

(3) 일단 각 모델에 의해 자료가 분류되고 나면, 이 모델이 주어진 자료의 구조를 얼마나 잘 기술하는지를 평가하는 지표값을 구하게 되는데, 이 때 MLE를 사용한다. MLE는 실제 자료 X를 얻었을 때, 이 X가 모델 M(k개의 군집구조)에 의해서 산출되었을 확률, P(M|X)를 구하는 것이다. 이를 Bayes 법칙으로 기술하면, 다음과 같이 기술될 수 있다.

$$P(M|X) = \frac{P(X|M) * P(M)}{P(X)}$$

모델마다 이 확률(실제 X라는 자료가 있을 때, 이것이 모델 M에 의해 산출되었을 확률)을 계산하고, 이 확률을 가장 높게 하는 모델을 선택하는 것인데, 실제 모든 모델마다 P(X)의 값은 동일할 것이므로, 모델을 평가하기 위한 MLE의 식은 다음과 같다.

$$MLE = V(X, M) = P(M) \times P(X|M)$$

각 모델에 대해서 MLE를 구하기 위해서는 P(M), 즉 주어진 모델이 일어나는 확률과 P(X|M), 즉 모델 M이 일어날 때, 이 모델이 실제자료 X를 산출할 확률을 계산하여야 하는데,

각 모델에 대해서 P(M)은 다음 공식에 의한다.

$$P(M) = \begin{cases} \frac{1}{C_1 * (k - k_0) + 1} & k < k_0 \\ \frac{1}{C_2 * (k - k_0) + 1} & k \geq k_0 \end{cases}$$

(k_0: 실험자에 의해 사전 설정된 군집의 개수, k: 모델 M의 군집 개수, C_1, C_2: 가중치. 이 수치가 커지면 커질수록 $P(M)$의 값은 작아진다. 본 프로그램에서는 이 가중치를 달리 적용했으며 현 군집의 개수가 사전 설정된 군집의 개수보다 클수록 $P(M)$값이 커지도록 하였다. 즉, C_2가 C_1보다 작다.)

P(M)은 주관적 확률인데, 군집 분석을 하기 전에 군집의 개수에 대한 정보가 거의 없으므로, 사용자의 사전 지식을 반영하는 주관적 확률을 사용하는 것이다. 이 경우, 사전 지식으로 미루어 가장 일어날 가능성이 높다고 생각되는 군집 개수를 k_0로 설정하고, 각 모델의 군집 개수가 이 값으로부터 멀어지는 정도에 따라 그 모델에 벌점을 주게 되는 것이다.

$P(X|M)$은 가정되는 모델 M이 실제 자료 X를 산출했을 확률로, 각각의 자료 x_i에 대해서 $P(x_i|M)$ 즉, 각각의 자료 x_i가 모델 M에 의해 산출되었을 확률을 구하고 이 값을 모든 자료들에 대해서 평균함으로써 얻어진다.

$$P(X|M) = \overline{P(x_i|M)} = \frac{\sum_{i=1}^{n} P(x_i|M)}{n}$$

(n: 분석에 사용된 자료의 총수)

$P(x_i|M)$은, 모델 M이 가정하고 있는 k개의 군집 각각에 대해서, 각 군집으로부터 자료 x_i가 산출될 확률을 구하고, 이 확률을 k개의 군집에서 더함으로써 구해진다.

$$P(x_i|M) = \sum_{j=1}^{k} P(x_i|c_j) P(c_j)$$

이러한 확률적 군집 분석에서는 각각의 군집을, 다변량 가우스(multivariate Gaussian)로 가정하며, 이러한 군집분포의 모양은 다차원 타원이 된

다. j번째 군집을 평균 c_j (군집의 중심점), 공분산 행렬(covariance matrix) C를 갖는 가우스로 모델화한다면, $P(x_i|c_j)$는 다음의 식에 의해 구해질 수 있다.

$$P(x_i|c_j) = \frac{e^{-\frac{1}{2}(x_i-c_j)^T * C^{-1} * (x_i-c_j)}}{\sqrt{(2\pi)^d * detC}}$$

(C: noise분포의 공분산행렬, c_j: j번째 군집의 중심점, d: 자료의 차원, $detC$: C 행렬의 행렬식)

$P(c_j)$는 j번째 군집 c_j 의 사전확률로, 다음의 공식에 의해 구해진다.

$$P(c_j) = \frac{\sum_{i=1}^{n} \frac{|x_i - c_j|}{\sum_{j=1}^{k} |x_i - c_j|}}{n}$$

즉, 각각의 자료에 대해서 이 자료가 k개의 군집의 중심점으로부터 떨어진 거리를 모두 합한 양에 대하여, 이 자료가 j번째 군집의 중심으로부터 떨어진 거리의 비율을 구하고, 이 비율을 모든 자료에 대해 평균한 값으로 얻어진다.

(4) 위 (3)의 공식들에 의해서, 각 모델마다 MLE를 계산하고, MLE의 값이 가장 높은, 즉, 실제 자료구조 X를 산출했을 확률이 가장 높은 모델(군집 개수)을 선택하게 된다.

(5) Learning Vector Quantization(LVQ)을 이용한 중심점 결정

MLE를 이용해 군집의 개수가 결정되면 다음으로 각 군집의 중심점을 구해야 하는데, 본 프로그램에서는 역시 SPIKER의 방법을 따랐으며 자율학습알고리즘(unsupervised learning algorithm)의 하나인 LVQ가 사용되었다(Yu, 1999). 이 알고리즘에서는 미리 결정된 군집의 개수만큼 우선 무선적으로 중심점의 위치를 결정한 후 세포활동 자료를 넣어 각 자료와 무선적으로 선택된 중심점들 사이의 거리를 비교한다. 각 세포활동 자료에 대해 그 거리가 최소가 되는 군집의 중심점을 그 자료와 중심점 사이의 거리 보정값[2]만큼 자료쪽 방향으로 이동시킨다. 즉, 세포활동자료 자체를 입력하여 각 군집의 중심점과 거리를 비교한 후 주어진 증가함수에 따라 중심점을 조금씩 움직이는 것이다. 이러한 작업이 총 10회 반복되는데, 각 작업마다 이전 작업에서 얻어진 중심점과 현 중심점 사이의 거리를 얻어낸 후 거리의 최대값과 현재의 중심점들 사이의 거리에 있어서의 중앙치(median)를 비교하여 그 비율이 0.05보다 작으면 그 시점에서 중심점의 값이 안정되었다고 보고 10회가 되기 전에 작업의 반복을 중단한다. 그렇지 않을 경우 10회의 반복작업을 통해 중심점의 값을 안정시킨다.

WAVESORTER를 이용한 세포분류의 실제

세포활동을 몇 개의 군집으로 분류하고 각 군집의 중심점을 결정하는 작업은 추출된 전체 활동전위를 대상으로 이루어지는 것이 아니라 전체 가운데 무선적으로 선택된 일정 수의 활동전위들을 대상으로 해서 이루어진다. 이런 방법을 사용하는 이유는 첫째, 자료의 중복성을 어느 정도 제거함으로써 분류작업의 계산적, 시간적 효율성을 높이기 위한 것이고, 둘째 전체 자료에 무선적으로 일정한 비율 이상으로 일관되게 나타나는

2) $\frac{\text{현 세포활동자료와 중심점사이의 거리}}{\sqrt{\text{iteration 횟수}}}$

활동전위의 파형을 주되게 고려하기 위한 것이다. 따라서 우선 추출된 활동전위 전체가 아닌 전체의 20~50%의 활동전위를 대상으로 세포분류작업을 한 후 이를 기준으로 하여 전체 활동전위들에 대한 분류작업을 실시하였다. 일정 샘플을 대상으로 세포분류작업을 실시하게 되면 실험자가 정해놓은 범위(2에서 7)내에서 군집의 개수가 결정되며, 각 군집의 중심점을 기준으로 거리가 계산된 뒤 분류작업이 완료된다(그림 3). 이와 같이 일차적으로 이루어지는 분류작업을 통해서 앞에서 언급한 바 있는 군집의 개수를 결정하였으며 각 군집의 중심점을 구하였다. 일단 군집의 개수와 중심점의 결정이 이루어지면, 아래에 설명한 잡음 제거(prune)과정을 거쳐서 각 군집에 있어서의 잡음을 제거하였으며, 동시에 중심점의 위치를 이동시켜 각 군집에 최적화된 중심점을 찾아내었다.

무선적 샘플을 이용하여 군집의 개수와 중심점의 결정이 이루어진 후에는 전체 세포활동에 대한 분류작업이 이루어지는데, 이 때에는 이미 결

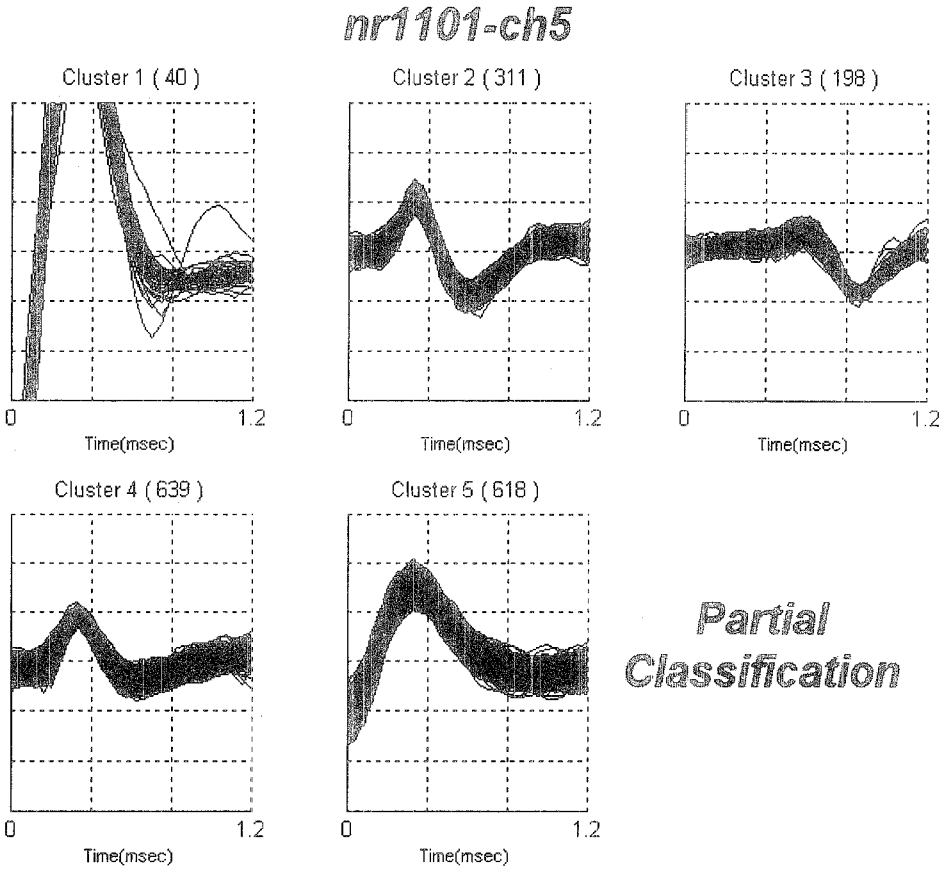

그림 3. 일정 수의 샘플을 대상으로 세포분류작업이 이루어진 예. MLE 알고리즘과 LVQ알고리즘을 통해 자료의 특성에 가장 적절한 군집의 수를 자동적으로 결정하며, 각 군집의 중심점을 구하게 된다. 그림 2c에서 사용된 고양이 상구 세포들의 활동전위를 이용하여 실제 분류 작업이 진행된 후의 결과를 보여주는데, 이 경우 5개로 군집이 형성되었음을 알 수 있다.

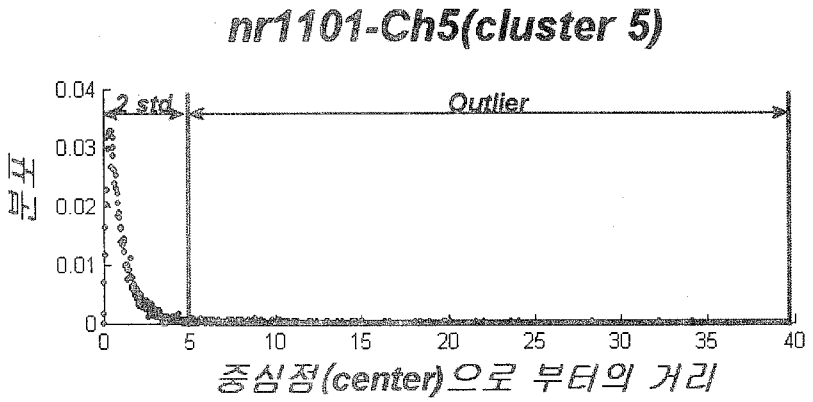

그림 4. 중심점과 각 활동전위 간 거리의 분포. 중심점과 각 활동전위간 거리 분포에서 2 표준편차 이상의 영역에 존재하는 활동전위가 잡음으로 제거됨을 보여준다. 여기서의 기준은 절대적인 것이 아니며, 실험자가 자료의 성질에 따라서 결정해야 하는 부분이다. 위 그림은 그림 3의 5번째 군집을 대상으로 한 잡음제거 과정을 보여준다.

정된 중심점까지의 거리를 계산하여 각 활동전위들을 군집에 귀속시키는 작업만이 필요하므로 비교적 짧은 시간 내에 분류작업의 완료가 가능하다. 이렇게 전체 세포활동에 대한 분류작업이 완료되면, 동일하게 잡음 제거(prune)과정을 거쳐서 각 군집의 극단치(outlier)를 제거해 준다.

잡음 제거의 원리는 각 군집에 속하는 세포활동들과 중심점 사이의 거리에 대한 분포를 구한 후 미리 결정된 기준(보통 2 표준편차)에서 벗어나는 거리에 해당하는 세포활동들을 제거하는 방식이다(그림 4). 주성분으로 이루어진 특질차원 혹은 A/D 자료 개수에 해당하는 차원 상에서의 거리를 이용해 세포분류가 이루어지기 때문에 잡음제거 과정 역시 이와 같은 거리에 있어서의 분포를 이용해 이루어졌다. 각 군집의 중심점과 각 군집에 속하는 활동전위 사이의 n차원(n개의 주성분이 선택되었을 경우) 상에서의 거리에 대한 분포를 계산한 후 일정 기준 내에 들지 못하는 거리를 가진 활동전위를 배경잡음으로 선택하였다. 이때 잡음제거는 1회로 끝나는 것이 아니라 실험자가 일정 기준을 정해 두고 이에 근거하여 여러 차례 시행한다. 잡음제거 기준을 엄격하게 적용할 경우 실제 세포활동을 잡음으로 판단할 가능성이 높아지고 기준을 느슨하게 적용할 경우 잡음을 세포활동으로 판단할 가능성이 높아지므로 기준의 설정은 여러 차례에 걸친 경험과 이론적 바탕에 근거하여 이루어져야 한다.

분류된 세포활동의 확인

앞에서 언급한 바와 같은 절차에 따라 하나의 전극에서 얻어진 전위의 기록을 통해 2개 이상의 세포에 대한 활동을 분류해 내었다면, 마지막으로 분류된 각 세포활동에 대한 확인작업이 이루어진다. 이 작업을 위해 사용되는 부분이 WAVESORTER의 마지막 부분인 'VIEWER'부분이다. 비록 각 활동전위가 가지는 파형의 특성에 의해 세포활동을 분류해 내었다고 하더라도 여전히 오류의 가능성은 존재하는데, 우선 활동전위

와 유사한 패턴을 가지며, 파형의 최고점이 높게 나타나는 배경잡음이다. 예로써 동물의 울음으로 인해 유발되는 잡음이나 동물의 움직임에 의해 유발되는 잡음, 가령, 보상으로 제공되는 물을 핥을 때 발생하는 잡음 등을 들 수 있다. 이런 잡음들은 비교적 그 파형의 최고점이 크기 때문에 파형의 최고점의 크기에 근거해 이루어지는 파형의 추출 작업에서 배제되지 못하는 부분들이다. 더군다나 이런 배경잡음은 대부분의 실험상황에서 유발되는 것이며 피험동물의 특정 행동과 상관되어 일어나는 경우가 많기 때문에 이러한 잡음을 세포활동으로 보게 될 경우 세포활동기록의 해석에 있어서 중대한 오류를 범할 수 있다. 따라서 이러한 잡음에 대한 제거는 필수적인 것이라 할 것이다. 또 한가지 들 수 있는 경우는 전극과의 물리적 접촉으로 세포가 죽을 때 발생하는 (Ca^{++}의 유입에 의해 발생하는 것으로 생각되는) 활동 전위이다. 일반적으로 활동이 기록되고 있는 세포가 죽을 경우 활동전위의 빈도가 높아지고 크기가 증가한다. 이러한 세포활동은 동물의 특정 행동과 상관되어 일어나는 것이 아니기 때문에(물론 실험에 따라서는 특정 행동과 상관되어 일어나는 현상일 경우도 있겠으나) 마찬가지로 해석에 있어서 주의가 요구된다.

이와 같은 몇 가지의 문제점을 해결하기 위해 사용할 수 있는 방법 중 하나는 세포활동의 파형 하나하나를 분류된 세포활동에 대한 전위열(spike train)과 비교하는 것이다. WAVESORTER에서 제공하는 VIEWER를 통해 이와 같은 작업이 가능하며, 세포분류알고리즘을 통해 해결해 낼 수 없는 부분들에 대한 해결이 가능하다. 세포활동의 아날로그 기록에 분류된 세포활동의 전위열을 겹쳐 놓은 후 비교과정을 통해 어떤 세포군집이 특정 배경 잡음(sucking noise 등)을 반영하고 있는지, 혹은 세포의 죽음으로 인한 고주파 활동을 반영하고 있는지 확인할 수 있다. 특정 배경 잡음이나 죽음으로 인한 고주파 활동은 그 패턴이 일반적인 세포활동과 구분되므로 세포분류작업을 거

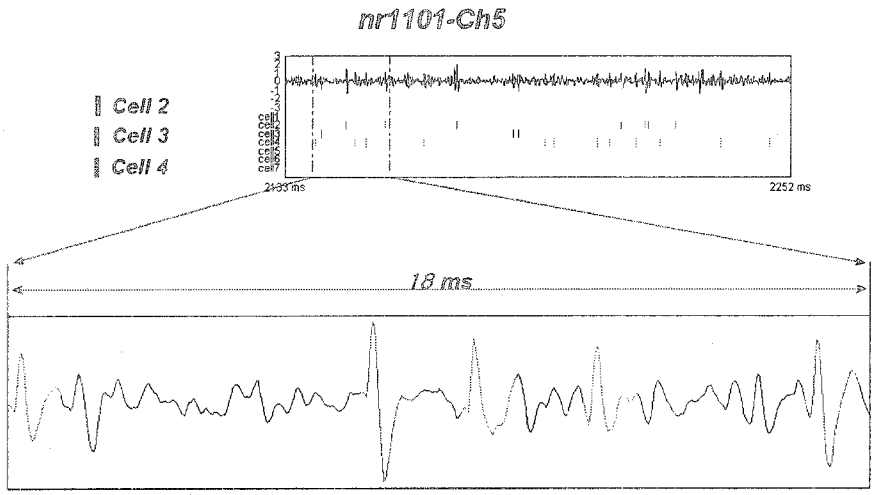

그림 5. WAVESORTER의 VIEWER를 통한 세포활동의 아날로그 기록과 활동전위열(spike train) 사이의 비교. 그림 3의 군집 2, 3, 4에 해당하는 3개의 세포만이 나타나 있으며, 일부분을 확대한 것이 아래의 그림이다. 세포활동의 파형이 유사한 것들끼리 하나의 세포군으로 묶여져 있음을 볼 수 있다.

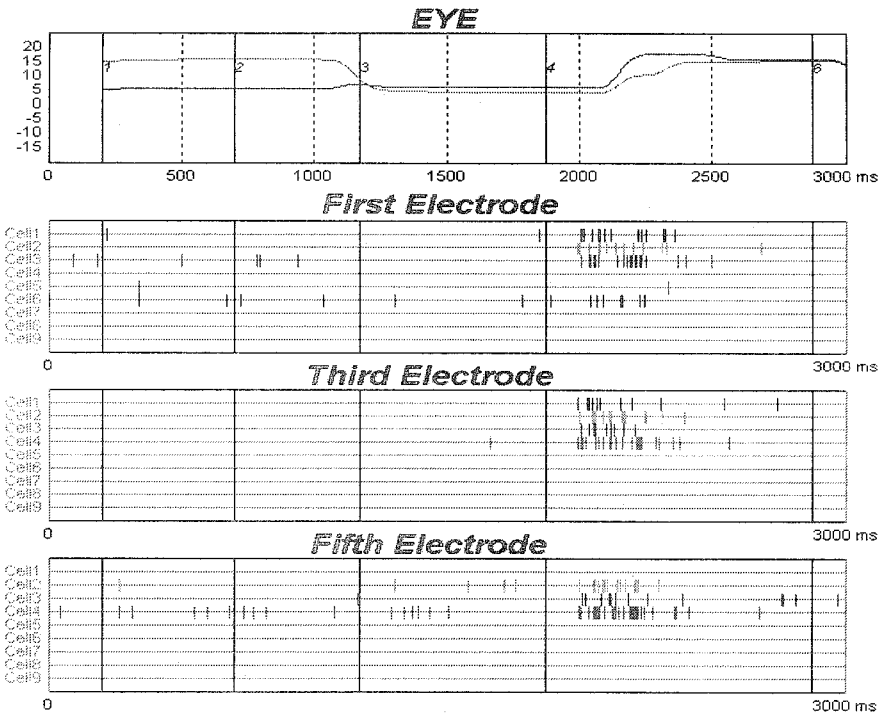

그림 6. 세포분류 작업을 통해 얻어지는 활동전위열. 최상단은 공막자기법(scleral search coil technique)으로 계측된 피험동물의 눈 움직임을 보여주며 아래 세 개의 패널은 각각 세 개의 전극에서 채집된 세포활동의 분류 결과를 보여준다(사용할 수 있는 5 개의 전극 가운데 1, 3, 5번의 전극이 사용되었다). 이 자료는 한 시행의 자료를 보이는데, 시행의 실험 조건을 첫 패널에서 숫자로 표시하고 있다. 시행은 '1'에서 시작하여 500ms 경과한 '2'의 시점에 실험동물(고양이)이 응시하고 있는 전면의 중앙에 LED 불빛이 제시된다. 이 중앙의 자극이 켜져 있는 곳을 향해서 동물이 시선을 이동을 하여 시선의 위치가 이 자극과 일정 거리만큼 가까워지면 실험 상태는 '3'의 상태로 바뀌게 되고 이후 일정 기간 동안(이 경우 700ms) 이 자극에 눈이 머물고 있으면 외곽에 두 번째 LED에 불빛이 들어온다. 이 시간이 '4'로 표시된 시점이다. 이 외곽의 시각 자극을 향해 동물이 도약안구운동을 하고 있다. 시간 '4' 직후에 세포활동이 강하게 일어나는 것을 볼 수 있는데, 이 반응들은 시각 자극에 의해 유발된 시각 반응이다. 그림 1, 그림 2, 그림 3, 그림 4, 그림 5에 보인 세포활동 자료는 위 그림의 아래 패널에 있는 5번 전극('Fifth Electrode')에서 채집된 자료와 동일하다. 이 시행에서, 5번 전극에서는 2, 3, 4번 세포(군집 2, 3, 4)만이 활동하고 있음을 알 수 있다.

치게 되면 대부분의 경우 독립적인 군집을 형성하게 된다. 따라서 위와 같은 비교과정을 거친 후 배경잡음으로 인정되는 군집을 제거하면 되는 것이다(그림 5, 6).

세포활동의 역치에 근거한 활동전위의 추출, 주성분 분석에 근거한 특질차원의 선택, MLE와 LVQ를 통해 군집의 개수 및 중심점 결정, 그리고 아날로그 세포활동기록과 활동전위열의 비교를 통한 배경잡음의 제거에 이르는 모든 과정을 거쳤을 때 하나의 전극에서 기록된 세포활동의 기

록은 1에서 7개 가량의 세포들로 분류되었다. 다중채널의 전극을 이용할 경우 다른 전극들에서 기록된 세포활동 신호에 대해서도 동일한 작업을 수행하게 되면 그림 7과 같이 각 채널별로 세포활동의 기록을 활동전위열의 형태로 나타낼 수 있었다. 그림에서도 알 수 있듯이 피험 동물의 특정 행동(시각자극을 향한 도약안구운동, 그림참고)이나 반응(시각자극에 대한 시각 반응, 그림참고)에 즈음하여 각 채널에서의 세포활동들이 강하게 나타나고 있다. 이와 같은 자극이나 행동에 대한 반응들을 중심으로 세포활동간 시간적 상관을 분석해 볼 수 있었으며, 이와 같은 분석을 통해 동물의 행동이나 반응을 일으키는데 있어서 세포들이 어떠한 패턴으로 신호를 주고받는지 추론해 볼 수 있었다. 세포활동이 분류되고 나면 세포활동간의 시간적 관계에 대한 분석이 이루어지는데 이를 위해 현재 사용하고 있는 방법으로는 상호교차상관 히스토그램(cross correlation histogram), Joint peri-stimulus time histogram(JPSTH), Gravity representation 등이 있다(Aertsen, Gerstein, & Palm, 1989; Gerstein, & Aertsen, 1985; Gerstein, & Dayhoff, 1985; Gochin, & Gerstein, 1989).

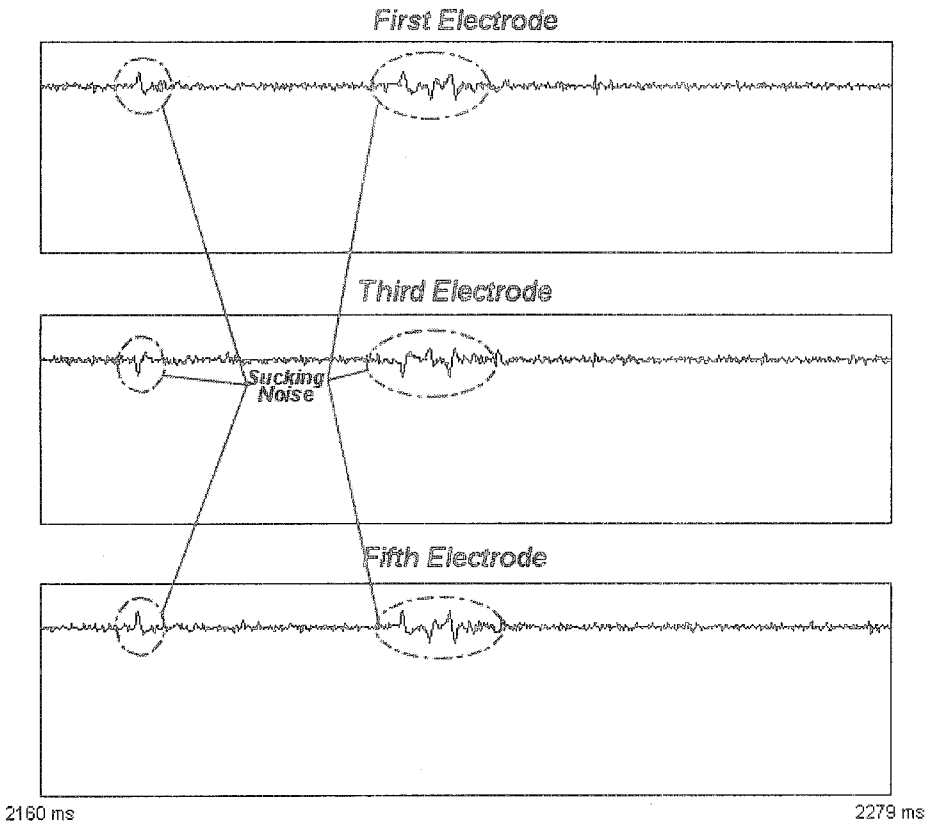

그림 7. 특정 잡음의 확인. 위 그림은 동물이 실험과제의 보상으로 주어지는 식수를 핥을 때 나타나는 잡음을 보여준다. 세 개의 전극에서 동일한 패턴의 잡음이 나타나고 있음을 확인할 수 있다. 이와 같은 아날로그 신호의 확인은 분류되어 나온 군집 중 특정 군집이 잡음으로 이루어진 것일 경우 제거가 가능하도록 해준다.

WAVESORTER를 사용한 세포분류과정에서 고려해야 할 사항들

중첩된 활동전위의 문제

둘 이상의 세포가 거의 시간차를 두지 않고 활동할 경우 전극에 기록되는 세포활동의 파형은 두 활동전위 파형의 중첩으로 인해 크기가 커지는 현상을 보이거나 아니면 오히려 그 크기가 매우 작아지는 현상을 보이기도 한다. 이러한 경우 WAVESORTER의 알고리즘으로는 해결이 불가능한데, 중첩에 의해 나타나는 파형의 경우 만일 두 세포의 활동 시기가 매우 규칙적으로 상관되어있다면 독립된 하나의 군집으로 분류가 되겠지만, 그렇지 않을 경우 outlier로 배척되기가 십상이다. 따라서 이와 같은 문제를 해결하기 위해서는 우선 기존의 알고리즘을 통해 분류되어 나온 한 세포활동의 전형적인 패턴(template)을 세포활동의 아날로그 신호에서 빼 주거나, 또는 신경망을 이용해 각 군집의 활동전위 형태에 대한 학습이 이루어지도록 하여 분리해 내는 방법을 사용할 수 있다. 다른 방법으로는 이미 분류되어 나온 각 세포활동 사이의 조합을 통해 중첩 파형을 찾아내는 기법을 들 수 있다. 이 문제와 관련한 논의는 Lewicki의 1994, 1998년 논문을 참고하기 바란다.

군집의 경계 결정에 있어서의 문제

본 프로그램에서 군집의 경계 결정을 할 때 기본적으로 가정하는 것은 특질 차원 상에서 활동전위들의 분포가 구형이라는 것이다. 이와 같은 가정은 군집의 경계가 뚜렷할 경우 매우 적절한 것이지만, 그렇지 않을 경우(군집이 겹쳐져 있거나 군집의 분포 모양이 구형이 아니거나) 문제의 여지가 있다. 물론 이러한 문제는 세포분류과정에서 outlier 제거 과정을 통해 어느 정도 해결이 가능하다. 즉, 일정 수준의 outlier 제거를 통해 각 군집에 속해있는 활동전위가 하나의 세포 이상의 것일 가능성을 배제할 수 있다는 것이다. 하지만, 이렇게 했을 경우 다른 세포의 활동으로 분류될 수 있는 활동전위를 잡음으로 처리하게 되므로 정보의 손실을 가져온다.

이러한 문제를 해결할 수 있는 한 방법은 각 군집의 분포가 다변량 가우스(multivariate Gaussian) 분포를 따른다고 가정한 후 이를 각 군집의 경계 결정에 활용하는 것이다. 이와 같이 각 군집의 분포를 설정한 후 각각의 활동전위가 각 군집에 속할 확률을 Bayes 규칙을 이용해 계산한다. 본 프로그램의 방식이, 거리에 근거하여(구형 분포를 가정할 경우) 이분법적으로 활동전위들이 각 군집에 속하는지의 여부를 계산하는 것이라면, 베이지안 군집분석(Baysian clustering)은 활동전위들이 각 군집에 속할 확률을 계산하기 때문에 군집의 경계를 결정할 때 신뢰구간을 계산해 낼 수 있다(Cheeseman & Stutz, 1996; Lewicki, 1998). 따라서 더 정확한 군집의 분류가 가능하다.

결 론

컴퓨터 기술과 세포활동기록의 기술이 발전함으로 인해 여러 단일세포활동의 아날로그 신호에 대한 직접적인 기록이 가능해지면서 세포분류에 대한 관심이 증가하고 있으며, 그 중요성에 대한 인식 또한 확산되고 있다. 본 논문에서는 세포분류 기법을 개관하고, 일반적으로 사용되는 기법인 주성분분석에 기반한 세포분류의 절차를 기술하였다.

세포분류를 위해서 먼저 세포활동의 아날로그 기록으로부터 활동전위의 크기를 역치 기준으로

하여 활동전위를 추출하였다. 이렇게 추출된 활동전위들은 주성분분석을 통해 자료의 분산을 가장 잘 설명해주는 몇 개의 주성분으로 표현하여, 주성분의 개수에 해당되는 차원 상에서 활동전위의 형태적 특성에 근거하여 군집을 이루도록 하였다. 여기에서 필요한 작업은 군집의 개수가 몇 개인가 하는 것과 각 군집의 중심점을 어떻게 결정할 것인가 하는 것인데, MLE와 LVQ를 통해 군집의 개수와 각 군집의 중심점이 결정되었다. 군집의 개수와 중심점이 결정된 이후 모든 활동전위는 주성분들로 이루어진 차원 상에서 각 군집의 중심점들과 거리가 계산되고 그 거리가 최소인 군집의 구성원으로 분류되었다. 이와 같은 일차적 분류 후 각 군집의 극단치를 제거하는 과정을 거쳐 각 군집의 구성원이 되는 활동전위를 확정한 후 세포활동의 아날로그 신호와 분류된 활동전위열 사이의 직접 비교를 통해 주성분분석에 기초한 세포분류 알고리즘으로는 찾아내기 힘든 잡음을 제거하였다.

위에서 열거한 세포분류 과정은 본 실험실에서 작성된 프로그램인 WAVESORTER를 통해 이루어졌는데, WAVESORTER는 SPIKER의 세포분류 알고리즘을 기술적으로 향상시켜 작성된 프로그램이다. 또한, SPIKER와는 달리 세포활동의 아날로그 기록에서의 활동전위 추출 기능을 포함시켰으며, VIEWER를 통해 세포활동의 분류 결과로 나온 활동전위열과 세포활동의 아날로그 기록을 직접적으로 비교하여, SPIKER의 알고리즘으로는 해결할 수 없었던, 동물의 움직임이나 울음 등으로 인해 유발되는 잡음을 제거할 수 있었다.

참고문헌

이춘길, 박정현 (1991). 신경활동의 컴퓨터 분석을 위한 'window 변별기'의 설계와 제작. 한국심리학회지: 생물 및 생리, 3, 150-155.

Aertsen, A. M. H. J., Gerstein, M. K., & Palm, G. (1989). Dynamics of neuronal firing correlation: Modulation of "effective connectivity". *Journal of Neurophysiology, 61,* 900-917.

Cheeseman P., & Stutz J. (1996). Bayesian classification (autoclass): theory and results. *Advances in Knowledge Discovery and Data Mining.* 153-80. CA: AAAI Press.

Duda, R. O., & Hart, P. E. (1973). *Pattern Classification and Scene Analysis.* John Wiley & Sons; New York.

Gerstein, G. L., & Aertsen A. M. H. J. (1985). Representation of cooperative firing activity among simultaneously recorded neurons. *Journal of Neurophysiology, 54,* 1513-1528.

Gerstein, G. L., Perkel, D. H., & Dayhoff, J. E. (1985). Cooperative firing activity in simultaneously recorded populations of neurons: detection and measurement. *Journal of Neuroscience, 5,* 881-889.

Giri, N. C. (1995). *Multivariate Statistical Analysis.* Marcel Dekker, Inc.; New York.

Gochin, P. M., Kaltenbach, J. A., & Gerstein, G. L. (1989). Coordinated activity of neuron pairs in anesthetized rat dorsal cochlear nucleus. *Brain Research, 497,* 1-11.

Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network, 9,* R53-78.

Lewicki, M. S. (1994). Bayesian Modeling and Classification of Neural Signals. *Neural Computation, 6,* 1005-1030.

Yu, A. J. (1999). *Classification of extracellular microelectrode recordings from the human brain.* Caltech Summer Undergraduate Research Fellowship.

한국심리학회지 : 생물 및 생리
Korean Journal of Biological and Physiological Psychology
2002. Vol. 14, No. 1, pp. 75-91

Classification of extracellular action potentials based on Principal Component Analysis

Joonyeol Lee Hyojung Seo Choongkil Lee

Department of Psychology, Seoul National University

The development of the computer technology and cell recording techniques have made possible continuous recording and sorting of extracellulary-recorded action potentials. At the same time, a reliable sorting of action potentials became a significant focus of interest. In this study, we describe and summarize a method of sorting of extracellulary-recorded spikes based on the Principal Component Analysis(PCA). In this method, a strategic number of principal components are chosen and each cell is represented in the feature space formed by the chosen components. The number of data cluster(i.e., cells) in the space is determined by the Maximum Likelihood Estimator(MLE) and the center of each cluster is determined by the Learning Vector Quantization(MLQ) method, an unsupervised learning algorithm. The distances between every cell pair and the center of each cluster are calculated. According to the Euclidean distance from the center, the data are sorted into each cluster. Removing the outliers of each cluster based on the distribution of the distance completes the sorting process. A computer program, 'WAVESORTER' was written using the Matlab(The Mathworks Inc.) to realize all the phases of the sorting processes. In this paper, the logic and routines of the 'WAVESORTER' is described.