ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

MODELING SPIKE-BAND EXTRACELLULAR BACKGROUND ACTIVITY USING JOHNSON'S S_U DISTRIBUTION

M.Sc. THESIS

Melih Yılmaz ÖGÜTCEN

Department of Applied Informatics

Information and Communication Engineering Programme

NOVEMBER 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

MODELING SPIKE-BAND EXTRACELLULAR BACKGROUND ACTIVITY USING JOHNSON'S S_U DISTRIBUTION

M.Sc. THESIS

Melih Yılmaz ÖGÜTCEN
(708191018)

Department of Applied Informatics

Information and Communication Engineering Programme

Thesis Advisor: Assoc. Prof. Dr. Murat OKATAN

NOVEMBER 2022

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

AKSİYON POTANSİYELİ BANDINDAKİ HÜCRE DIŞI ARKA PLAN ETKİNLİĞİNİN JOHNSON S_U DAĞILIMI KULLANILARAK MODELLENMESİ

YÜKSEK LİSANS TEZİ

Melih Yılmaz ÖGÜTCEN
(708191018)

Bilişim Uygulamaları Anabilim Dalı

Bilgi ve Haberleşme Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Murat OKATAN

KASIM 2022

Melih Yılmaz ÖGÜTCEN, a M.Sc. student of ITU Graduate School student ID 708191018, successfully defended the thesis entitled "MODELING SPIKE-BAND EXTRACELLULAR BACKGROUND ACTIVITY USING JOHNSON'S S_U DISTRIBUTION", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Dr. Murat OKATAN** Istanbul Technical University	.....
Jury Members :	**Prof. Dr. Neslihan Serap ŞENGÖR** Istanbul Technical University	.....
	Asst. Prof. Dr. Mehmet KOCATÜRK Istanbul Medipol University	….....

Date of Submission : 13 October 2022
Date of Defense : 10 November 2022

To my family and friends,

FOREWORD

I would like to express my deep appreciation and endless thanks to my advisor Assoc. Prof. Dr. Murat OKATAN for his unwavering support. And I want to thank my family and friends for being there for me. This work was supported by Research Fund of the Istanbul Technical University. Project Number: MAB-2020-42808.

November 2022

Melih Yılmaz ÖGÜTCEN
(Data Scientist)

TABLE OF CONTENTS

Page

ABBREVIATIONS

cdf	**:** Cumulative distribution function
iBMI	**:** invasive Brain-Machine Interface
JS_U	**:** Johnson's S_U Distribution
$JS_{U\ all}$	**:** non-fixed Johnson's S_U Distribution
$JS_U\ \gamma$	**:** γ fixed Johnson's S_U Distribution
$JS_U\ \delta$	**:** δ fixed Johnson's S_U Distribution
$JS_U\ \xi$	**:** ξ fixed Johnson's S_U Distribution
$JS_U\ \lambda$	**:** λ fixed Johnson's S_U Distribution
KS	**:** Kolmogorov-Smirnov
MSE	**:** Mean Squared Error
N	**:** Normal Distribution
pdf	**:** Probability density function
SCS	**:** Set of Candidate Solutions

SYMBOLS

$\theta_{l\|m}$, $\theta_{h\|m}$	: Lower and upper intermediate threshold, respectively
a, b	: Lower and upper truncation threshold, respectively
ϕ	: Scaling factor
μ	: Location parameter of the Normal distribution that is truncated by the truncation thresholds
σ	: Scale parameter of the Normal distribution that is truncated by the truncation thresholds
$\hat{\sigma}$	: Maximum likelihood estimate of σ
s	: Conventional standard deviation estimate
ξ	: Location parameter of JS_U distribution
λ	: Scale parameter of JS_U distibution
γ, δ	: Shape parameters of JS_U distribution

LIST OF TABLES

Page

LIST OF FIGURES

Page

MODELING SPIKE-BAND EXTRACELLULAR BACKGROUND ACTIVITY USING JOHNSON'S S_U DISTRIBUTION

SUMMARY

Extracellular neural recordings are collected in awake behaving subjects (e.g. rats) via microelectrode arrays that are chronically implanted into the brain. These recordings contain important information about the activity of individual neurons and the functioning of the brain.

In order to detect the action potentials, or spikes, in the recordings, the recordings are first bandpass filtered with cut-off frequencies that are appropriate for spike detection. Now the signal consists of spikes and background activity. Here, background activity is produced by neurons relatively far from the electrode. It is critical to distinguish the spikes from the background activity. In the literature, this problem is usually solved by amplitude thresholding. However, currently, the threshold value is usually determined by multiplying a coefficient determined subjectively by the researcher by the estimated standard deviation of the data under an assumption of Normality. By contrast, purely data-driven thresholding can be performed with truncation thresholds calculated with the truncated Normal distribution, truncation thresholds calculated with the truncated Johnson's S_U (JS_U) distribution, and an Otsu-based method.

In the present study, completely data-driven amplitude thresholding methods are examined and compared against each other. Realistic simulated extracellular neural recording data having 21 different firing rate levels were used in the analyses. The methods were compared in terms of the accuracy with which they allowed the standard deviation of the background activity to be estimated. The best method was found to be the Otsu-based method. In addition, the distribution of the data encompassed by the Otsu-based threshold values was examined. It was found that the truncated Normal distribution could pass the Kolmogorov-Smirnov (KS) test only at two firing rates, while the truncated JS_U distribution could pass the test at all firing rates tested.

JS_U is a distribution with 4 parameters (γ, δ, ξ, and λ) and therefore provides more flexibility in modeling than the Normal distribution. However, when computational costs are compared, the JS_U method has the highest computational cost. Reducing the cost is important in terms of being useful for real-time applications. Therefore, four different parameter-fixed JS_U distributions ($JS_U\,\gamma$, $JS_U\,\delta$, $JS_U\,\xi$, and $JS_U\,\lambda$) are proposed in this study, in which each parameter is fixed to its initial value (estimated before any truncation) and the other three different parameters are estimated after truncation. By fixing one parameter (except for λ), the calculation time can be reduced to approximately 1/7 of JS_U all. The thresholding performance of these methods has been studied and $JS_U\,\xi$ has been shown to give better results than JS_U all. However, the parameter-fixed models were not as successful as JS_U all in fitting the data encompassed by the Otsu-based threshold values.

All the analyses were also performed on a real data segment of 1 second duration. Due to the success of the Otsu-based method in simulated data, it was used as a reference.

It has been shown that standard deviation estimates computed using parameter-fixed methods are comparable to the estimate produced by the Otsu-based method. Moreover, the data encompassed by the Otsu-based threshold values could be successfully modeled using the JS_U γ, JS_U δ, and JS_U ξ distributions. These findings are important for the development of a fully data-driven and automated invasive brain-machine interface.

AKSİYON POTANSİYELİ BANDINDAKİ HÜCRE DIŞI ARKA PLAN ETKİNLİĞİNİN JOHNSON S_U DAĞILIMI KULLANILARAK MODELLENMESİ

ÖZET

Kompleks bir organ olan beyin büyük bir bilgi işleme cihazı olarak işlev görmektedir. Buradaki bilgi, beyindeki sinir hücreleri tarafından işlenmektedir. Beynin yapısını anlamlandırabilmek için sinir hücrelerinin aktivitelerinin incelenmesi ve anlamlandırılması kritiktir. Bu yüzden bilim adamları sinir hücrelerinin elektriksel özelliklerini inceleyebilmek için hücre dışı sinir kayıtlarını kullanmaktadır.

Hücre dışı sinir kayıtları deneklere (örn. sıçan) kronik olarak yerleştirilebilen mikro elektrot dizileri aracılığıyla alınabilmektedir. Bu kayıtlar içerisinde nöronlar tarafından üretilen kısa süreli ve büyük genlikli sinyaller aksiyon potansiyeli, bir diğer deyişle vuru, olarak adlandırılmaktadır. Laboratuvar ortamında anestezi altında olmayan ve davranış sergileyen bu deneklerden alınan kayıtlar tekil bir nöronun aktivitesini ve beynin işleyişini anlamlandırabilmek için önemli bilgiler içermektedir. Bu kayıtların analiz edilmesi invaziv beyin-makine arayüzü (iBMA) ve nörobilim deneyleri için de kritiktir.

Kayıtlar, ilk olarak aksiyon potansiyellerini tespit edebilmek için uygun kesim frekanslarında bant geçiren bir filtre ile filtrelenmektedir. Artık sinyal aksiyon potansiyeli ve arka plan aktivitesinden oluşmaktadır. Burada arka plan aktivitesi elektrottan göreceli olarak uzak olan nöronlar tarafından üretilmektedir. Elektrota göreceli olarak yakın sinyaller de asıl aksiyon potansiyelleridir ve bu sinyallerin arka plan aktivitesinden ayrımlandırılması kritiktir. Literatürde bu problem genellikle genlik eşikleme yöntemiyle çözülmektedir. Fakat mevcut eşik değeri genellikle araştırmacı tarafından belirlenen bir katsayı ile veriden hesaplanan standart sapmanın çarpımı kullanılarak belirlenmektedir ve bu yöntemler subjektif bir yaklaşım içermektedir. Eşik değerinin yüksek olarak belirlenmesi aksiyon potansiyeli verisinin, arka plan aktivitesi olarak sınıflandırılmasına ya da düşük olarak belirlenmesi arka plan aktivitesinin, aksiyon potansiyeli olarak sınıflandırılmasına sebep olabilir. Eşikleme veri işleme adımında ilk adım olduğundan ve sonraki adımları doğrudan etkileyebileceğinden dolayı kritiktir. Bu yüzden eşikleme için tamamen veriye dayalı, objektif yöntemlerin kullanılması bir ihtiyaçtır. Halihazırda, kırpık Normal dağılım ile hesaplanan kırpma eşikleri, kırpık Johnson's S_U (JS_U) dağılımı ile hesaplanan kırpma eşikleri ve Otsu-tabanlı eşik yöntemleri ile tamamen veriye dayalı eşikleme gerçekleştirilebilmektedir.

Kırpık Normal yönteminde veri dağılımının ortanca değerinin üstündeki ve altındaki değerlerde yinelemeli olarak 'kırpma eşikleri' adı verilen bir çift eşik değeri aranmaktadır. Her yinelemede, veriler mevcut eşik adaylarıyla kırpılır ve kırpık Normal dağılım veriye maksimum olabilirlikle uyumlanır. Dağılımın uyumluluğunu test etmek için Kolmogorov-Smirnov testi (KS) kullanılmıştır. Burada KS testini geçen ve en geniş aralığı veren eşik değeri çifti asıl eşik değeri olarak belirlenmektedir.

Benzer yaklaşım kırpık JS_U yönteminde de kullanılmıştır, burada kırpık Normal dağılım yerine veri kırpık JS_U dağılımı ile uyumlanarak eşik değeri çifti aranmaktadır.

Bu çalışmada da tamamen veriye dayalı yöntemler incelenmiş, birbiriyle kıyaslanmıştır. Kıyaslama için arka plan aktivitesinin standart sapması bilinen ve farklı ateşleme hızlarına sahip 21 adet gerçekçi benzetim verisi kullanılmıştır.

Eşik değerleri arasında kalan örneklerin standart sapma değeri, arka plan dağılım karakteristiği için kritiktir. Bu yüzden gerçekçi benzetim verilerinde, arka plan standart sapmasının kestirilmesine göre eşikleme yöntemlerinin performansları kıyaslanmıştır ve en iyi kestiren yöntem Otsu-tabanlı yöntem olmuştur. Kruskal-Wallis testine göre her ateşleme hızında Otsu-tabanlı yöntemle kestirilen standart sapmanın, arka plan aktivitesinin gerçek standart sapmasından ayrışmadığı gözlemlenmiştir. Ayrıca Otsu-tabanlı eşik değerleri arasında kalan örneklerin olasılık dağılımlarına uygun olup olmadığı incelenmiştir ve kıyaslama için KS testi kullanılmıştır. Bu aralıkta sadece iki ateşleme hızında kırpık Normal dağılım KS testi geçerken, kırpık JS_U ile her ateşleme hızında arka plan aktivitesi başarılı bir şekilde modellenebilmiştir.

JS_U 4 parametreli (γ, δ, ξ ve λ) bir dağılımdır. ξ ve λ parametreleri dağılımın konum ve ölçek parametreleri iken γ, δ parametreleri dağılımın çarpıklığını ve basıklığını birlikte etkileyen şekil parametreleridir. JS_U ayrıca ağır kuyruklu ve çok basık bir dağılımdır ve bu yüzden verinin ortanca değerinden uzak değerleri modellemede Normal dağılıma nazaran daha fazla esneklik sağlamaktadır. Bu çalışmada ayrıca simüle veriler ve gerçek veri Pearson'un çarpıklık ve basıklık uzayında gösterilmiştir. Bu uzayda Johnson'un dağılımlarına karşılık gelen bölgeler belirtilmiştir. Eğer veri uzay içerisinde JS_U bölgesine karşılık gelmezse, eşik değerleri kırpık Normal dağılım ile hesaplanmıştır. Burada 21 adet simüle verinin 20'si ve gerçek veri JS_U bölgesine karşılık gelmektedir.

Fakat hesaplama maliyetleri karşılaştırıldığında JS_U yöntemi en yüksek hesaplama maliyetine sahip yöntemdir. Maliyetin azaltılması gerçek zamanlı uygulamalar için kullanışlı olabilmesi açısından önemlidir. Bu yüzden JS_U (JS_U all) dağılımında her bir parametre başlangıçtaki değerine (herhangi bir kırpma olmadan tüm veriden kestirilen) sabitlenerek, diğer üç farklı parametrenin kestirildiği dört farklı parametre-sabitli JS_U dağılımları (JS_U γ, JS_U δ, JS_U ξ ve JS_U λ) bu çalışmada önerilmiştir. Başlangıçtaki değerler için ξ ve λ veriden hesaplanan konum ve ölçek değerlerine, γ ve δ ise sırasıyla çarpıklık ve basıklık değerlerine sabitlenmiştir. Parametre sabitlenerek (JS_U λ yöntemi hariç) hesaplama süresi JS_U all 'a göre yaklaşık 1/7'sine kadar düşürülebilmektedir. Parametre sabitli yöntemler arasında en hızlı yöntem JS_U δ'dır. Ancak parametre sabitlemeye rağmen en hızlı sonuçlar Otsu-tabanlı yöntem ile elde edilmiştir. Ayrıca parametre sabitli yöntemlerin eşikleme performansları incelenmiştir ve JS_U ξ ile JS_U all 'dan daha iyi sonuçlar elde edilebilmiştir. Ancak simüle verilerde Otsu-tabanlı eşik değerleri arasında kalan örneklerin dağılımını modellemede JS_U all kadar başarılı olamamışlardır.

Bu çalışmada analizlerde ek olarak simüle veriler yerine 1 sn uzunluğunda gerçek veri kullanılmıştır ve eşiklemeler gerçek veri üzerinde de gerçekleştirilmiştir. Simüle veriler her ne kadar gerçek verinin bazı karakteristik özelliklerini kapsasalar da, gerçek veriler farklı genliklere sahip aksiyon potansiyeli verilerini içerdiklerinden dolayı fizyolojik olarak simüle verilerden ayrışmaktadır. Otsu-tabanlı yöntemin simüle verilerdeki başarısından dolayı bu yöntem burada referans olarak kullanılmıştır. Otsu-tabanlı yönteme göre parametre sabitli yöntemlerin de gerçek veri üzerinde iyi

sonuçlar verdiği bu çalışmada gösterilmiştir. En yüksek olabilirlikle kestirilen standart sapmalarda en başarılı sonuç kırpık Normal dağılım ile hesaplanan kırpma eşikleri ile elde edilmiştir. Gerçek veride Otsu-tabanlı eşik değerleri arasında kalan örnekler JS_U γ, JS_U δ ve JS_U ξ ile modellenebilmiştir. Fakat simüle verilerde olduğu gibi gerçek veride de kırpık Normal dağılım ile modellenememiştir. Bu bulgular tamamen veriye dayalı, tam otomatik bir invaziv beyin-makine arayüzü geliştirilebilmesi için önemlidir.

1. INTRODUCTION

The brain, which has a complex structure, is the center of the nervous system. In order to understand the functions of the brain electrically, studies on electrophysiology have been carried out for years. The brain consists of two types of cells: the nerve cells called neurons and the glia, which mostly support neurons. Neurons communicate through fibers called axons through electrical and chemical signals [1]. Glia cannot generate impulses and they are thought to mostly provide metabolic and functional support for the neurons [2].

Electrophysiological recordings from neurons contain essential information to make sense of the brain's complex structure [1, 3]. These recordings can be taken with the help of both extracellular and intracellular electrodes. Extracellular neural recordings can be recorded with invasive microelectrode arrays chronically implanted in the brain of mobile and unanesthetized subjects (e.g. rats) in the laboratory environment [3-7]. Analysis of extracellular neural recordings is important in developing invasive brain-machine interfaces (iBMI) and neuroscience experiments [4-8].

These recordings are bandpass filtered to detect the action potentials, or spikes, fired by neurons located in the vicinity of the recording electrodes. However, these spikes are embedded in background noise, which consists of the activity of thousands of neurons in the background far from the electrode, the thermal noise of the tissue, and the noise generated by the recording system (microelectrode and electronic circuit) [9]. As a result, spikes need to be separated from the noise. The spikes of neurons that are near the recording electrode appear with a higher amplitude than the background activity, and the problem of detecting the spikes of individual neurons is usually solved by amplitude thresholding [10-14].

Amplitude thresholding is an important problem as it is one of the first steps in data analysis and will separate the signal from the background noise. The inconsistency of the method used in this step may adversely affect subsequent steps [15]. Therefore, it is critical to use effective and robust methods for amplitude thresholding.

A few different approaches exist for computing the amplitude threshold. In one of the most commonly used approaches, the standard deviation obtained of the spike band extracellular neural recordings is multiplied by a coefficient selected by the researcher to determine the threshold values [10, 14]. However, the use of a data-independent coefficient causes the neuron activity to be classified as background activity if the calculated threshold value is high, and the background activity is classified as neuron activity if the calculated threshold value is low. In addition, as the neural firing rate increases in extracellular neural recordings, the estimated standard deviation increases [10]. Therefore, it is necessary to determine the threshold values in an optimum way and to present an objective quantitative approach based on data.

A more principled alternative method is proposed in [16]. This method computes the amplitude threshold by maximizing the signal-to-noise ratio SNR of a behavioural variable in suprathreshold data. This method is compared with maximum-likelihood estimation of the threshold in [17]. In the maximum-likelihood method, the amplitude threshold can be chosen as the threshold value that brings the joint probability of the behavioural data to the highest value, under a probability model that models the behavioural response probability as a function of the suprathreshold firing rate. These two methods were compared on 63 neural recordings obtained from the motor cortex of two rats and the results showed that obtained thresholds were not different from each other ($P \geq 0.05$; Kruskal-Wallis test).

In another method that is entirely data driven and that does not require behavioral data a pair of threshold values called 'truncation thresholds' are iteratively searched for in the values above and below the median value of the data [18]. At each iteration, the data are truncated with the current threshold candidates and the truncated Normal distribution is fit to the truncated subthreshold data by maximum likelihood. Goodness-of-fit is tested using the Kolmogorov-Smirnov (KS) test. The threshold pair that gives the widest range and passes the KS test is used as the final threshold values.

Finally, another entirely data driven but non-parametric thresholding method is proposed in [19]. In this method, the amplitüde thresholds are searched for separately in the positive and the negative parts of the data, with a logic similar to the Otsu thresholding method. For the upper threshold value, a variable threshold is scanned from 1 μV to the maximum amplitue in the data with steps of 1 μV and the absolute difference between the variance of the background (subthreshold) activity and the

variance of the suprathreshold data is calculated as a function of the thresholds tested. The threshold value corresponding to the local maximum closest to zero is determined as the final threshold value. The same algorithm is also applied to calculate the lower threshold.

Accurate estimation of the standard deviation of the background activity is critical in revealing the characteristics of the noise. The study by [19] showed that the Otsu-based thresholding method was more accurate in estimating the standard deviation compared to the truncation thresholds method in simulated data. The comparison was performed using simulated extracellular neural recording data, where the firing rate ranged between 0-100 Hz [20].

Although the estimation of the standard deviation of the background activity becomes less accurate at high firing rates in the truncation thresholds method [18], the Otsu-based method produced results close to the true standard deviation at all frequencies examined [19]. However, while the truncation thresholds are defined on the basis that the data they encompass obey a truncated Normal distribution, whether the data that lie between the Otsu-based thresholds obey a known noise probability distribution has not been examined until the present thesis. Here, the data between the Otsu-based threshold values are fit with both the truncated Normal distribution and the truncated Johnson's S_U (JS_U) distribution (which will be discussed in detail Section 2.3.1) [21]. The KS test for goodness-of-fit is used at $P \geq 0.05$. While the P values obtained with the truncated Normal distribution were above 0.05 at only 0 and 5 Hz, the P values obtained with the truncated JS_U distribution passed the KS test at all firing rates. In addition, threshold values are also calculated by using the truncation thresholds approach with the truncated JS_U distribution instead of the truncated Normal distribution [21]. This demonstrates that the truncated JS_U distribution can be used to compute the truncation thresholds. However, the JS_U distribution is a 4-parameter distribution [22] and maximum likelihood estimation of the parameters is computationally costlier.

The aim of this thesis is to test whether the JS_U distribution can be used in place of the Normal distribution in computing the truncation thresholds and whether the subthreshold data encompassed by the Otsu-based thresholds obey the JS_U distribution in both simulated and real extracellular neural recordings.

Existing amplitude thresholding algorithms are reviewed in detail in Section 1.1 and the structure of the thesis is discussed in Section 1.2.

1.1 Related Works

In the study conducted in [9], Yang et al. used in vivo array recordings from the hippocampus, cortex surface, and spinal cord of rats. They showed in a lumped circuit model of the microelectrodes implanted into the brain that the noise has four different components: background neural noise, electrode-electrolyte interface noise, tissue thermal noise, and electronic noise. To confirm that the noise originates from different sources, two different electrodes, one coated with carbon nanotube (CNT) and the other uncoated, were applied to the hippocampus region of rats. Afterwards, a euthanizing drug was injected and the records until death were recorded. This experiment was repeated several times. After death, the noise was reduced by more than 80%. This experiment shows that the majority of the reduced noise is made up of neural noise. In their proposed approach to detect spikes, they first filter the data with a 300 Hz high pass filter. Then they apply the Hilbert transform to the filtered signal. Here, while the neural noise has an exponential distribution ($e^{-\frac{Z}{2\sigma_n^2}}$ where σ_n^2 is the noise variance and Z is the instantaneous energy in Hilbert space), the spike data posesses a power-law distribution ($Z^{-\lambda_2}$). With the EC/PC (exponential component/power component) linear decomposition method, the recordings are divided into two different components as signal and noise. These components are individually represented by density functions where the sum of the sample probabilities due to the noise and the spikes is 1. Finally, they calculated a spiking probability map by taking the ratio of the power-law component to all components, that shows probability of spiking as a function of time. On the spiking probability map, samples with a probability higher than 0.5 were labeled as spikes and those below them as noise. They used the data they synthesized between 5-300 Hz to test the performance of the proposed approach. For each data, a linear regression was calculated between the threshold values corresponding to 0.5 on the spiking probability map and the EC/PC crossing point they predicted at linear decomposition, and obtained R^2=0.935 performance. They also added that the threshold probability (0.5) in spiking probability map may vary from channel to channel. However, it is a

subjective approach that the 0.5 threshold probability value remains static for each data and this shows that the study is not a completely data-driven method.

In another study [10], simulated data obtained from 594 different average spike shapes from the neocortex and basal ganglia were used. The method proposed in [23] was used for thresholding. This method estimates the standard deviation of the background noise and computes the threshold as k times that estimate, as shown in Equation 1.1., where k=4. Here x is the spike band neural recording and Φ^{-1} is the inverse of the standard cumulative Normal distribution.

In addition, their standard deviation estimate has been compared against the conventional standard deviation formula (*s*, Eq. 2.10). They generated a 10-second long background noise which has unit standard deviation and added spike data at different firing rates (0-50 Hz) to these data. When the standard deviation estimates calculated from these data were compared, $\hat{\sigma}_o$ gave results closer to the unit standard deviation. On the other hand, *s* moved away from the unit standard deviation faster as the firing rate increased.

$$\hat{\sigma}_o = \frac{median(|x|)}{\Phi^{-1}(0.75)} \qquad Thr = 4 \times \hat{\sigma}_o \qquad (1.1)$$

An alternative method for computing the amplitude thresholds was developed by Okatan and Kocatürk [18]. This method, which is referred to as the truncation thresholds method here, is examined in detail in Section 3. The truncation thresholds method has been compared with the method used by [10] for thresholding spike data [18, 24]. Thresholds were computed for simulated extracellular neural recordings with firing rates in the range 0 Hz to 50 Hz (explained in detail in Section 2.2). Truncation thresholds were found to be superior in a number of aspects [18]. In particular, the standard deviation of the background activity was estimated more accurately by the truncation thresholds method [18, 24].

In a further study, a method based on the Otsu threshold [25] was used for thresholding extracellular neural recordings [19]. The Otsu-based method was compared with the truncation thresholds method in simulated data [19]. Otsu-based approach was shown to be more accurate than truncation thresholds in estimating the standard deviation of the background noise. This method is discussed in detail in Section 2.3.2.

Finally, in the present thesis, truncated JS_U distribution is used in computing the truncation thresholds instead of the truncated Normal distribution [21]. The JS_U distribution, which has four parameters, is more flexible than the Normal distribution. When the ratio of estimated to true standard deviations is examined, it does not have good performance compared to either the truncated Normal distribution or the Otsu-based method. In addition, the distribution of the data encompassed by the Otsu-based threshold values are modeled with truncated Normal and truncated JS_U distributions at 21 different frequencies (0-100 Hz with steps of 5 Hz) in [21]. While the truncated JS_U distribution passes the KS test ($P>0.05$) at all frequencies, the truncated Normal distribution passes the KS test only at 0 and 5 Hz. Although the truncated JS_U distribution is a more flexible model of the background activity, its estimate of the noise standard deviation is not sufficiently accurate in the simulated data examined. In addition, fitting the JS_U distribution is computationally costly. These disadvantages are discussed in detail in Section 2.3.4.

1.2 Structure of Thesis

In Section 2, the structure of extracellular neural recordings and simulated data are discussed. Then, truncation thresholds with Normal distribution [18], truncation thresholds with JS_U distribution [21] and Otsu-based thresholding method [19] are examined. The data in between the thresholds are modeled with truncated Normal and truncated JS_U distributions, and the results are compared in Section 3. In addition, in Section 3, because using truncated JS_U has higher computation cost, the effect of fixing one of the four parameters of the JS_U distribution on the computation time, on calculated truncation thresholds and on estimated standard deviations is determined. In addition to simulated data, the same analyzes are performed on real data to test whether the truncation thresholds can be computed for real data under the JS_U distribution. Finally, the obtained findings are evaluated and conclusions are drawn in Section 4.

2. EXTRACELLULAR NEURAL RECORDINGS AND AMPLITUDE THRESHOLDING

2.1 Real Data

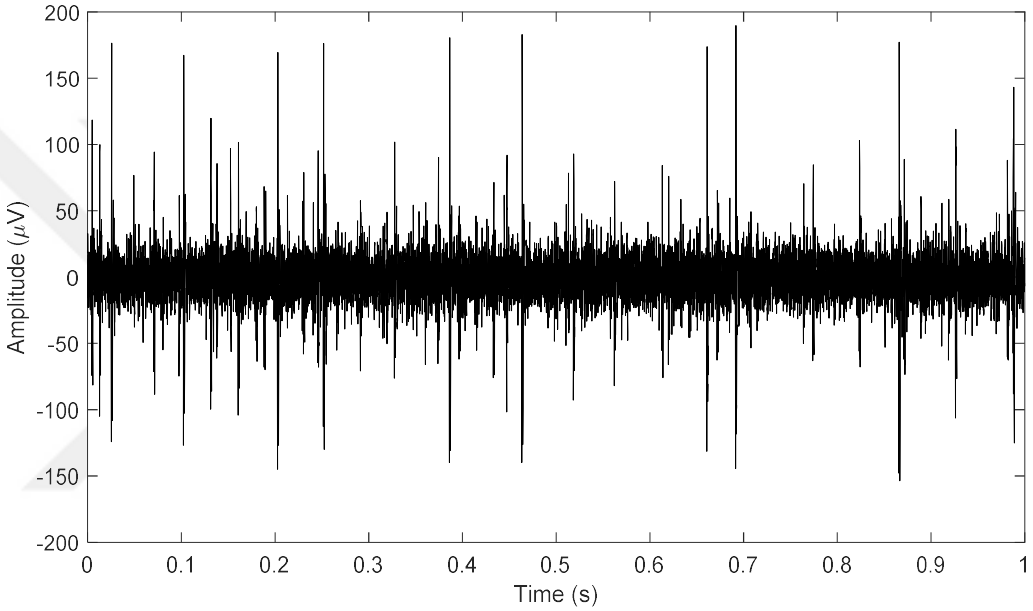

Figure 2.1 : Spike band extracellular recording from the primary motor cortex of a rat.

Extracellular neural recordings can be taken with micro electrode arrays from the primary motor cortex (M1) of the unanesthetized rat during behavior. These raw recordings are then passed through a bandpass filter to detect spikes [26]. Figure 2.1 shows a piece of extracellular neural recording from the primary motor cortex of a rat that is 1s in length, sampled at F_s=40 kHz and passed through a 4th-order Butterworth band-pass filter with cut-off frequencies of 400 Hz – 8 kHz. This piece of recording is freely available in the data of [20]. In this recording, brief but high-amplitude signals represent spikes from nerve cells. Inspection of Figure 2.1 reveals that the recording contains spikes with various amplitudes. It is not trivial to tell the boundary between the background noise and the spikes. To quantify the performance of different thresholding methods in separating spikes from the background noise, ground truth

data are needed. For this reason, realistic simulation data were generated using previously recorded real recordings as explained below (Fig. 2.2), without performing any new laboratory experiments [24].

2.2 Realistic Simulation Data

The realistic simulation data that are used in this thesis are available at [20]. These data were produced as the sum of noise and action potential processes. White noise with a mean of 0 and standard deviation estimated by the truncation thresholds method from real data was used as noise. Convolutions of Bernoulli processes with a 7 ms piece of real data segment were used as the action potential processes. The 7 ms-long data segment contains an action potential between -2ms and 5 ms with a peak at 0ms, shown in the bottom panel of Figure 2.2. These data are defined in Eq. 2.1.

$$y_f[n] = \eta[n] + a_f[n] \tag{2.1}$$

where $0 \leq n \leq N-1$ is discrete time with a sampling frequency of F_s=40kHz. For these data, which have a length of 10 seconds, $N=4\times10^5$. η[n] is white Gaussian noise (N(0, σ)) with 0 mean and standard deviation σ=12.25 μV. $a_f[n]$ represents the action potential process with firing rate *f*.

In order to generate the noise component, first, the truncation thresholds were calculated for the data in Figure 2.1 using the Truncation Thresholds Software [18], which is freely available at SciCrunch.org under RRID: SCR_014637. The truncated Normal probability distribution was fit by maximum likelihood to the activity between these thresholds. The mean and the standard deviation of the distribution were found to be $\hat{\mu}_\eta = -0.30$ μV ve $\hat{\sigma}_\eta$=12.55 μV. Here $\hat{\sigma}_\eta$ is used as the standard deviation of the simulated Gaussian noise.

To generate the action potential process, a Bernoulli process was initialized as an array of 0's with a sampling frequency of F_s=40 kHz. Thus, each sample occupies Δ=250 μs on the time axis. By using 21 different firing rates, distinct Bernoulli processes were produced by setting each sample to 1 with probability f/F_s where f ranged between 0 Hz and 100 Hz with steps of 5 Hz. The convolutions of the resulting Bernoulli processes at each firing rate with the action potential waveform shown in Figure 2.2 was calculated (Eq. 2.3). This procedure generated a simulated action potential process

at each frequency *f*. The action potential waveform is defined as follows for a sampling frequency of F_s=40kHz.

$$A_k = \begin{cases} 0, & k < -80 \; or \; k > 200 \\ A(k\Delta), & -80 \leq k \leq 200 \end{cases} \tag{2.2}$$

Here Δ represents the sampling period (F_s^{-1}), and A(.) represents the action potential in Figure 2.2. $a_f[n]$ is defined as follows.

$$a_f[n] = \sum_{j=-\infty}^{+\infty} B_j(f) A_{k-j} \tag{2.3}$$

where, $B_j(f)$ is the Bernoulli random variable with probability f/F_s. The action potential and the convolution of the Bernoulli process at 21 different frequencies the simulated data. In this thesis, all operations involving simulated data were performed using these data.

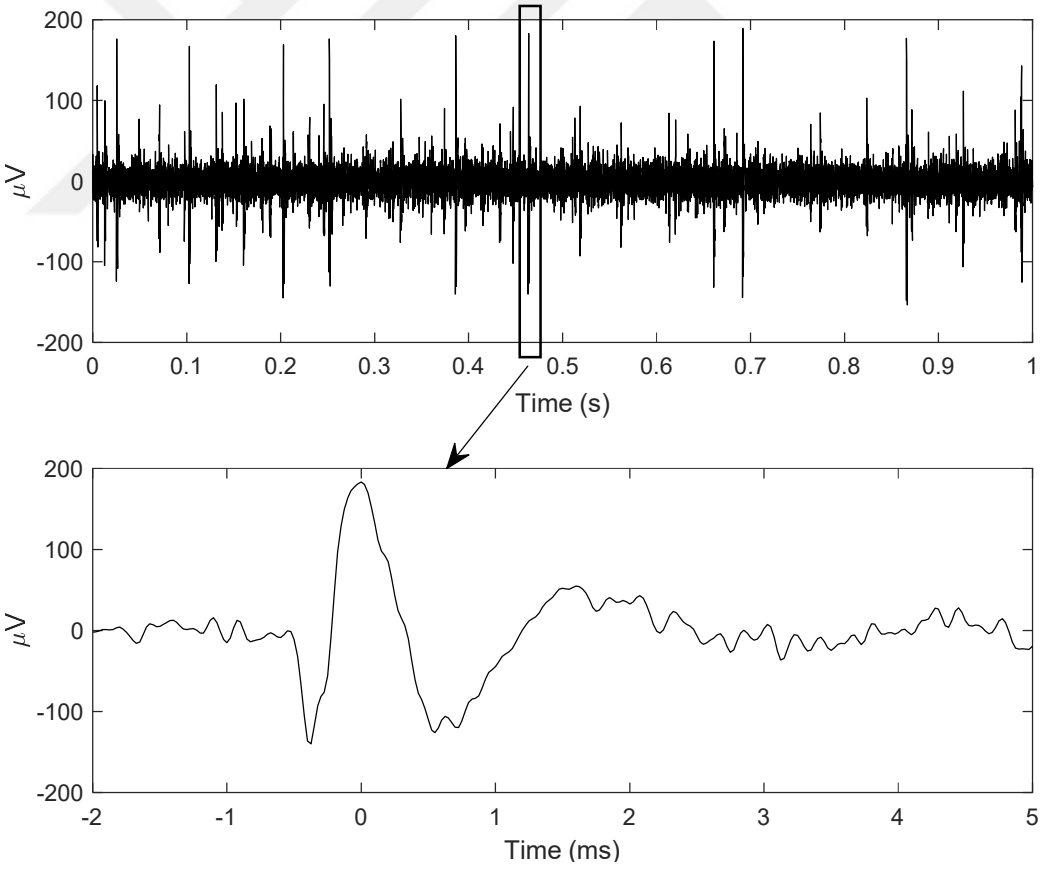

Figure 2.2 : The real action potential waveform that was used in generating the simulated data.

2.3 Thresholding Algorithms

In order to determine the threshold values, the algorithms in the literature generally use the standard deviation of the data multiplied by a coefficient k (kσ), which has a value between 3 and 5 [10-14]. Although this method provides fast results, its disadvantages are as follows [18].

1) The use of a coefficient (k) determined independently of the data in the threshold calculation is a subjective approach.
2) Standard deviation estimates increase with firing rate, which increases the computed threshold.
3) It is assumed that the same threshold can be used for the negative part of the data after negation, in view of the underlying assumption that the noise has a Normal distribution with zero mean. However the noise distribution may not be perfectly symmetrical, necessitating the use of non-symmetrical thresholds for the positive and negative parts of the data.
4) The distributional assumptions concerning the subthreshold data are not verified.

To avoid these disadvantages, alternative thresholding methods have been developed, such as the truncation thresholds with Normal distribution [18] or JS_U distribution [21], and Otsu-based thresholding [19]. These methods are presented in Sections 2.3.1 and 2.3.2.

2.3.1 Truncation thresholds with Normal distribution and JS_U distribution

The truncation thresholds are calculated using the Normal distribution in [18]. In the present thesis, the same method is extended to use a more flexible distribution, the JS_U distribution, in place of the Normal distribution [21]. The two methods differ only in virtue of the distribution used.

In the method that uses the Normal distribution, the data are truncated by two threshold values and the distribution of the subthreshold data is modeled as a truncated Normal distribution. The latter is defined in Eq. 2.4.

$$f(x|a,b,\mu,\sigma) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \tag{2.4}$$

where, $\phi(\cdot)$ is the standard Normal probability density function and $\Phi(\cdot)$ is its cumulative distribution function. a and b are the lower and upper truncation thresholds, respectively, which are discovered from the data.

The JS_U distribution is one of the four distributions specified in the Johnson system and is defined as the unbounded distribution [22]. This distribution is a four-parameter distribution that is obtained from the Normal distribution. The parameters of JS_U distribution ξ, and λ correspond to the location and scale respectively while γ and δ are two shape parameters that jointly determine the skewness and kurtosis. JS_U distribution can fit leptokurtic (kurtosis>3) and skewed data [27]. The probability density function (pdf) of the JS_U distribution is as follows [28].

$$f(x|\gamma,\delta,\xi,\lambda) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+\left(\frac{x-\xi}{\lambda}\right)^2}} e^{-\frac{1}{2}\left(\gamma+\delta sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right)^2} \tag{2.5}$$

Here, γ and δ together affect the skewness and kurtosis of the distribution. In this way, it provides a more flexible distribution than the Normal distribution.

Let $F(x|\gamma,\delta,\xi,\lambda)$ denote the cumulative distribution function (cdf) of the JS_U distribution. Then, the pdf of the truncated JS_U distribution with truncation points $a < b$ is given by Eq. 2.6, where $I(x) = 1$ if $a \le x \le b$ and $I(x) = 0$, otherwise:

$$f_{tr}(x|\gamma,\delta,\xi,\lambda,a,b) = I(x)\frac{f(x|\gamma,\delta,\xi,\lambda)}{F(b|\gamma,\delta,\xi,\lambda) - F(a|\gamma,\delta,\xi,\lambda)} \tag{2.6}$$

The pdf and the cdf of the JS_U distribution are calculated using the Johnson Curve Toolbox in this thesis [29].

Truncation thresholds are calculated in two steps. In the first step, the intermediate threshold values ($\theta_{l|m}$ and $\theta_{h|m}$) are calculated, and in the second step, the truncation thresholds (a and b) are calculated using the intermediate threshold values.

2.3.1.1 Calculation of intermediate thresholds ($\theta_{l|m}$ and $\theta_{h|m}$)

The calculation of the intermediate threshold values consists of the following steps.

1) In order to calculate the lower intermediate threshold $\theta_{l|m}$, the upper intermediate threshold value of the data is temporarily equated to the median (med) of the data. All in-between samples form the set of candidate solutions (SCS) for the lower intermediate threshold value of the data. Since the data are bandpass filtered and roughly symmetrical about the median, the cardinality of this set is about half the number of samples in the data. However, all these candidate solutions are expensive to test exhaustively, so only the median of the SCS is tested and the SCS is updated depending on the test result, as explained below. This update reduces the size of the SCS by one half at each iteration. This method is referred to as the bisection method [30], which is a search algorithm. As a result, $\theta_{l|m}$ is found in about $\log_2(N/2)$ iterations, instead of the N/2 iterations that exhaustive search would have required, where N represents the number of distinct voltage values in the data.
2) $\theta_{l|m}$ is searched for in the SCS iteratively, thus $\theta_{l|m}[i]$ denotes the solution candidate at the ith iteration of the bisection method. The truncated pdf in Eq. 2.4 or 2.6 is fit to the data contained between a=$\theta_{l|m}[i]$ and b=med by maximum likelihood (ML) to estimate the model parameters. When the truncated Normal distribution is used as the noise distribution model, the parameters are the location (μ) and the scale (σ) parameters in Eq. 2.4. When the truncated JS_U distribution is used instead, the parameters are γ, δ, ξ and λ (Eq. 2.6).
3) At each iteration, the goodness-of-fit is assessed using the Kolmogorov-Smirnov (KS) test at level 0.05. Depending on whether the fit passes the KS test, the SCS for the lower intermediate threshold is updated as in Eq. 2.7.

$$new\ SCS\ for\ \theta_{l|m} = \begin{cases} KS\ test\ is\ passed, & (\theta_{l|m}[-], \theta_{l|m}[i]) \\ KS\ test\ is\ not\ passed, & (\theta_{l|m}[i], \theta_{l|m}[+]) \end{cases} \quad (2.7)$$

Here, $\theta_{l|m}[-] = \max_{j<i} \theta_{l|m}[j]$ for which the KS test was not passed, while $\theta_{l|m}[+] = \min_{j<i} \theta_{l|m}[j]$ for which the KS test was passed. At the first iteration, $\theta_{l|m}[-]$ is the minimum sample in the data and $\theta_{l|m}[+]$ is med.

4) The algorithm continues until only one value remains in the SCS. When the search is finished, $\underset{i\in G}{min}\,\theta_{l|m}[i]$ is chosen as $\theta_{l|m}$, where G is the set of iteration numbers where the KS-test was passed with $P \geq 0.05$.

5) To obtain the upper intermediate threshold value $\theta_{h|m}$, the data are multiplied by -1, $\theta_{l|m}$ is computed for the resulting data using the procedure explained above, and $\theta_{h|m}$ is obtained as $-\theta_{l|m}$. This computation too takes approximately $\log_2(N/2)$ iterations.

2.3.1.2 Calculation of truncation thresholds (a and b) using $\theta_{l|m}$ and $\theta_{h|m}$

1) The data between $\theta_{l|m}$ and $\theta_{h|m}$ are fit by the truncated pdf.
2) If the fit passes the KS test at level 0.05:
   a. The distances between the median of the data and the intermediate threshold values are expanded by multiplying them by a factor ϕ that is calculated from the data by the bisection method as shown in Eq. 2.8. The set of candidate solutions of ϕ is defined further below.

$$tt[i] = \left[\theta_{l|m}\ \theta_{h|m}\right] \cdot \phi[i] + med \cdot (1 - \phi[i]) \tag{2.8}$$

   where, $\phi[i]$ is the median value of the SCS of ϕ at iteration i.

   b. Let $x_{l|m} < \theta_{l|m}$ and $x_{h|m} > \theta_{h|m}$ denote the samples beyond the indicated limits, then the SCS of ϕ is $\{(med - x_{l|m})/(med - \theta_{l|m})\} \cup \{(x_{h|m} - med)/(\theta_{h|m} - med)\}$. In order to obtain the optimum ϕ, the bisection method is used.

   c. At each iteration, SCS is updated according to Eq. 2.9.

$$new\ SCS\ for\ \phi = \begin{cases} KS\ test\ is\ passed, & (\phi[\mathrm{i}], \phi[-]) \\ KS\ test\ is\ not\ passed, & (\phi[+], \phi[i]) \end{cases} \tag{2.9}$$

   Here $\phi[-] = \underset{j<i}{\min}\,\phi[j]$ for which the KS test was not passed, while $\phi[+] = \underset{j<i}{\max}\,\phi[j]$ for which the KS test was passed. At the first iteration $\phi[-]$ is the maximum value in the SCS and $\phi[+]$ is 1.

3) If, at step 1, the fit fails the KS test at level 0.05:

a. This time, the distances between the median of the data and the threshold value are narrowed by multiplying $med - \theta_{l|m}$ and $\theta_{h|m} - med$ by a ϕ value with a SCS between 0 and 1.
b. Let $\theta_{l|m} < x_{l|m} < med$ and $med < x_{h|m} < \theta_{h|m}$ denote the samples within the indicated limits, then the SCS of ϕ becomes $\{(med - x_{l|m})/(med - \theta_{l|m})\} \cup \{(x_{h|m} - med)/(\theta_{h|m} - med)\}$.
c. The new solution set for ϕ is determined using Eq. 2.9, except that at the first iteration $\phi[-]$ is 1 and $\phi[+]$ is the minimum value in the SCS.

4) The SCS gets exhausted iteratively and the widest interval for which the KS-test was passed is taken as the final solution giving the truncation thresholds.

Usually, the truncated Normal distribution or the truncated JS_U distribution fail to pass the KS test when they are fit to the data between $\theta_{l|m}$ and $\theta_{h|m}$ (at step 1). Thus, the algorithm proceeds to shrinking the interval [$\theta_{l|m}$, $\theta_{h|m}$]. During the shrinking, the cardinality of the initial SCS of ϕ is at most N. Thus, it takes at most $\log_2(N)$ iterations to complete the shrinking. Together with the computation of $\theta_{l|m}$ and $\theta_{h|m}$, it takes a maximum of $3\log_2(N)$-2 iterations to compute the truncation thresholds.

2.3.2 Otsu-based thresholding

Another completely data-driven thresholding method was proposed in [19]. In image processing applications, Otsu thresholding method enables finding the optimum threshold value by minimizing the variance within the classes and maximizing the variance between the classes on the histogram of the picture in order to separate the background and the object from each other [25] In the Otsu-based thresholding method, firstly, the positive and negative values of the data are separated using the 0 threshold value. Then, to calculate the upper threshold value, positive values are truncated starting from the maximum value of the data, down to a threshold value of 1 μV with steps of 1 μV and the variances of the subthreshold (background) activity and the suprathreshold activity are calculated at each threshold value. The absolute value of the difference between the two variances is assigned to an array. The threshold value that gives the local maximum closest to 0 in this array is used as the final upper threshold value. The same approach is applied on the negative part of the data to calculate the lower threshold value. This method is referred to as the Otsu-based method in this thesis.

Using the simulated data that were explained in Section 2.2, the ratio of the estimated standard deviation of the background activity to the actual standard deviation value was computed and compared for the Otsu-based method and the methods that use the truncated Normal and the truncated JS_U distributions [21]. The results have shown that the ratios obtained using the Otsu-based method are indistinguishable from 1 (P=0.78; Kruskal-Wallis test) at all firing rates, whereas this was not the case for the other methods.

2.3.3 Choice of probability distribution

There are three different distributions in Johnson's distribution system: S_L (log-normal), S_U (unbounded) and S_B (bounded). To choose the correct distribution for a given data set, Johnson [22] has located these distributions in the Pearson system [31, 32]. Here, the square of the skewness (β_1) and traditional kurtosis (β_2) can determine which distributions are suitable for the data. Figure 2.3 shows the locations of the real and the simulated data, in the Pearson system.

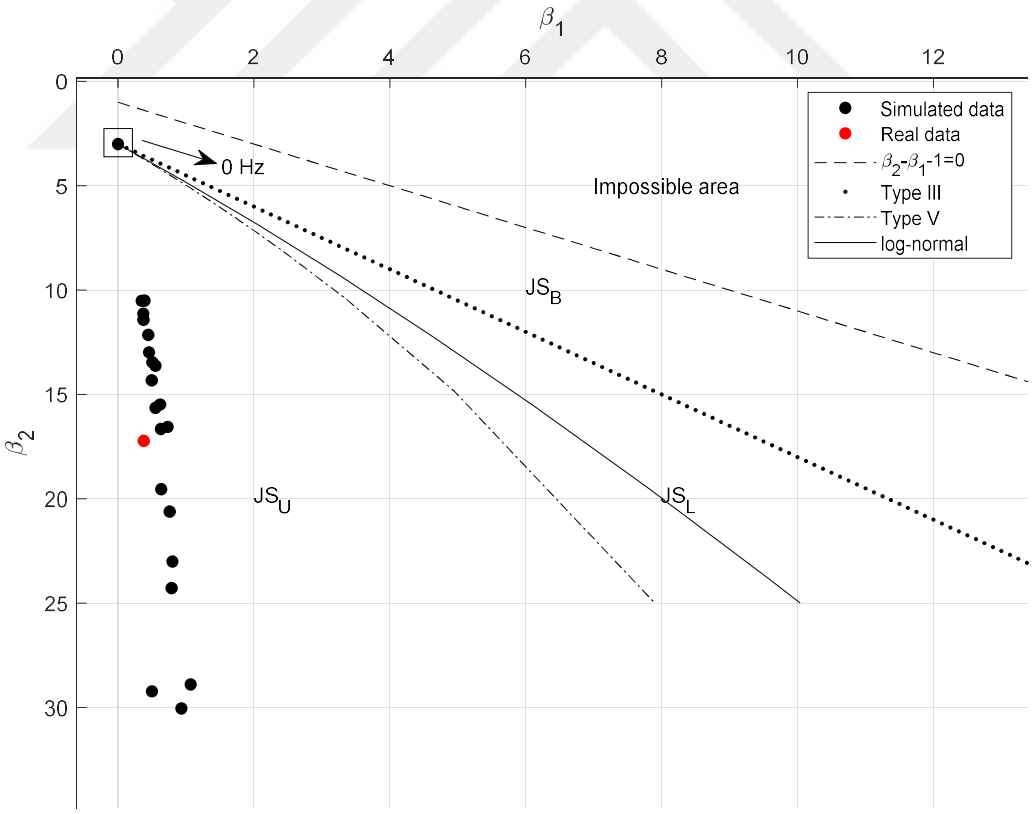

Figure 2.3 : Johnson's distributions, along with the real and the simulated data in the Pearson system.

As mentioned in Section 2.2, the data generated at 0 Hz consists of Gaussian noise only. In Figure 2.3, the position of 0 Hz is $\beta_1=4.6\times10^{-6}$ and $\beta_2=3$. Here β_1 is almost 0.

Therefore, the data at 0 Hz can be modeled with a Normal distribution (the β_2 of a Normal distribution is 3). The real data and the simulated data in the other 20 frequencies can be modeled with JS_U according to Figure 2.3. The flowchart we designed based on the location of the simulated data in the Pearson system is as in Figure 2.4.

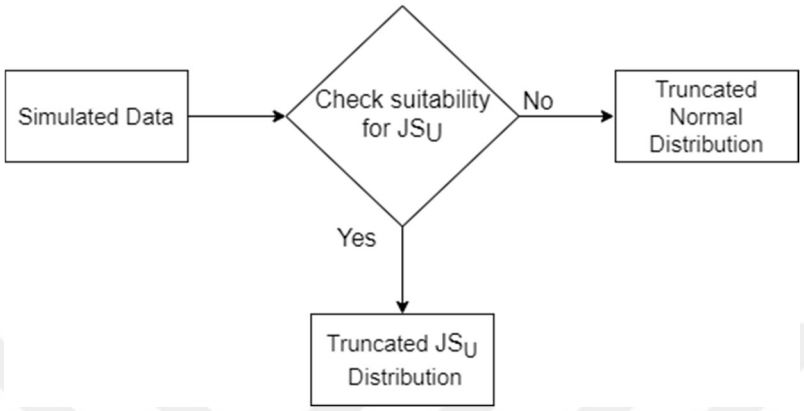

Figure 2.4 : The flowchart of the proposed approach.

According to this flowchart, if the data are not suitable for JS_U, the method suggested in [18] is used to determine the truncation thresholds, otherwise the method explained in [21] is used. That is, while truncated Normal distribution was used for 0 Hz, truncated JS_U distribution was used for the remaining 20 simulated recordings and for the parameter-fixed JS_U models, which are explained in Section 2.3.4.

2.3.4 Truncation thresholds with parameter-fixed JS_U distribution

Estimation of all parameters of the JS_U distribution is computationally more expensive than the other methods. The total time it takes to compute the truncation thresholds for the 21 simulated recordings is measured in 10 independent runs. Table 2.1 shows the average of these 10 measurements. (Details of computations are given on Appendix A).

Table 2.1: Computation time of the methods considered.

Method	Duration (mean±std)
Truncation Thresholds with Truncated Normal Distribution	1373.957 ± 2.697 s
Truncation Thresholds with Truncated JS_U Distribution*	21078.932 ± 106.515 s
Otsu-Based Thresholding	28.969 ± 0.65 s

*The truncated Normal distribution is used for the data at 0 Hz in computing the truncation thresholds with the JS_U distribution, as explained in Section 2.3.3.

It can be seen that using the truncated JS_U distribution takes much longer than the alternatives to compute the thresholds. To check whether fixing one of the four

parameters of the JS_U distribution and estimating the remaining three reduces the computation time, the initial values (γ_0, δ_0, ξ_0 and λ_0) calculated from the data are kept constant for one parameter at a time and calculations are performed by estimating the remaining parameters. Here, the initial parameters correspond to the skewness, kurtosis, location, and scale of data, respectively. These developments are carried out in the Johnson Curve Toolbox [29]. Average execution times measured under these conditions are shown in Table 2.2. As in Table 2.1, the results show the average of 10 independent runs in each case.

Table 2.2: Computation time of the parameter-fixed methods.

Method	Duration (mean±std)
Fixed γ in Truncated JS_U Distribution	2904.562 ± 4.245 s
Fixed δ in Truncated JS_U Distribution	2829.403 ± 9.536 s
Fixed ξ in Truncated JS_U Distribution	3198.691 ± 9.720 s
Fixed λ in Truncated JS_U Distribution	59436.750 s (only one trial)

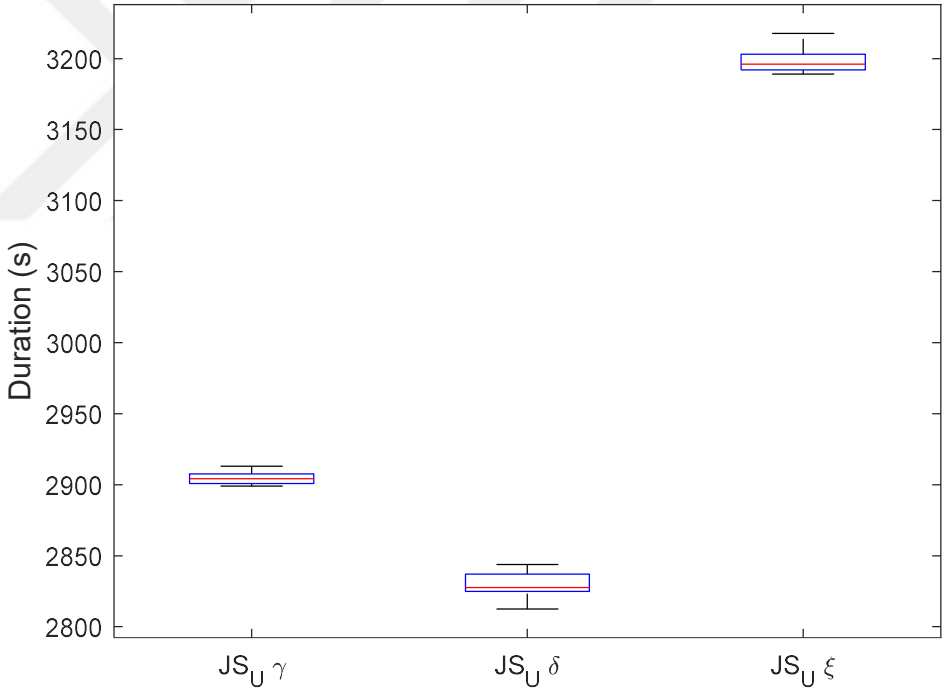

Figure 2.5 : The boxplot of total durations of the parameter fixed JS_U methods obtained as a result of 10 trials.

According to the results in Table 2.2, fixing the parameters γ, δ, or ξ reduces the computation time of the algorithm for JS_U to approximately 1/7 of the duration it takes when all parameters are estimated. Since the time for λ was too long, a single measurement was deemed sufficient to make the point that fixing λ was not reducing the computation time. In Figure 2.5, box diagrams of the total times obtained as a result

of 10 trials are shown. The shortest time was obtained for the case where δ was fixed ($P = 2.49\times10^{-6}$; Kruskal Wallis with Tukey-Kramer post-hoc multiple comparison test). Although parameter fixing reduced the computation time substantially, computation with truncated Normal distribution still takes shorter than any of the fixed parameter cases. Furthermore, Otsu-based thresholding method continues to take the shortest duration among all the methods considered. The frequency dependence and performance of the final threshold values obtained are examined in detail in Section 3.

2.4 Calculation of Standard Deviations

Accurate estimation of the standard deviation of the background activity is important for explaining the noise characteristic. The conventional standard deviation is calculated as follows.

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2.10}$$

where N is the number of samples, x_i is the i^{th} sample and $\bar{x}$ is the mean of data.

The estimated standard deviation and the estimated mean value for Normal distribution are obtained from the results of maximum likelihood estimation. The log-likelihood of the truncated Normal distribution parameters is calculated as follows.

$$l(\mu, \sigma | a, b, x_{1:N}) = \sum_{i=1}^{N} log\big(f(x_i | a, b, \mu, \sigma)\big) \tag{2.11}$$

where, $a \leq x_i \leq b$ is the i^{th} sample in the data truncated at a and b, and N is the number of samples in the truncated data.

The estimated standard deviation for JS_U distribution is calculated as in Eq. 2.12. For the cases where the parameter is fixed ($JS_{U\,\gamma}$, $JS_{U\,\delta}$, $JS_{U\,\xi}$, and $JS_{U\,\lambda}$), the inital value (γ_0, δ_0, ξ_0 and λ_0) obtained from all the data is used in the calculation of the standard deviation [28].

$$\hat{\sigma}_{JS_U} = \sqrt{\frac{\hat{\lambda}^2}{2}\left(e^{\hat{\delta}^{-2}} - 1\right)\left(e^{\hat{\delta}^{-2}} cosh\left(\frac{2\hat{\gamma}}{\hat{\delta}}\right) + 1\right)} \tag{2.12}$$

3. RESULTS

This section first presents the amplitude thresholds obtained using the truncation thresholds and Otsu-based thresholds on simulated data. The standard deviation of the background activity is estimated using each method and the estimates are compared against the true value. Frequency-dependent abrupt changes in the thresholds and standard deviation estimates are noted and explained. Finally, it is shown that the methods also work on real data.

3.1 Obtained Truncation Thresholds

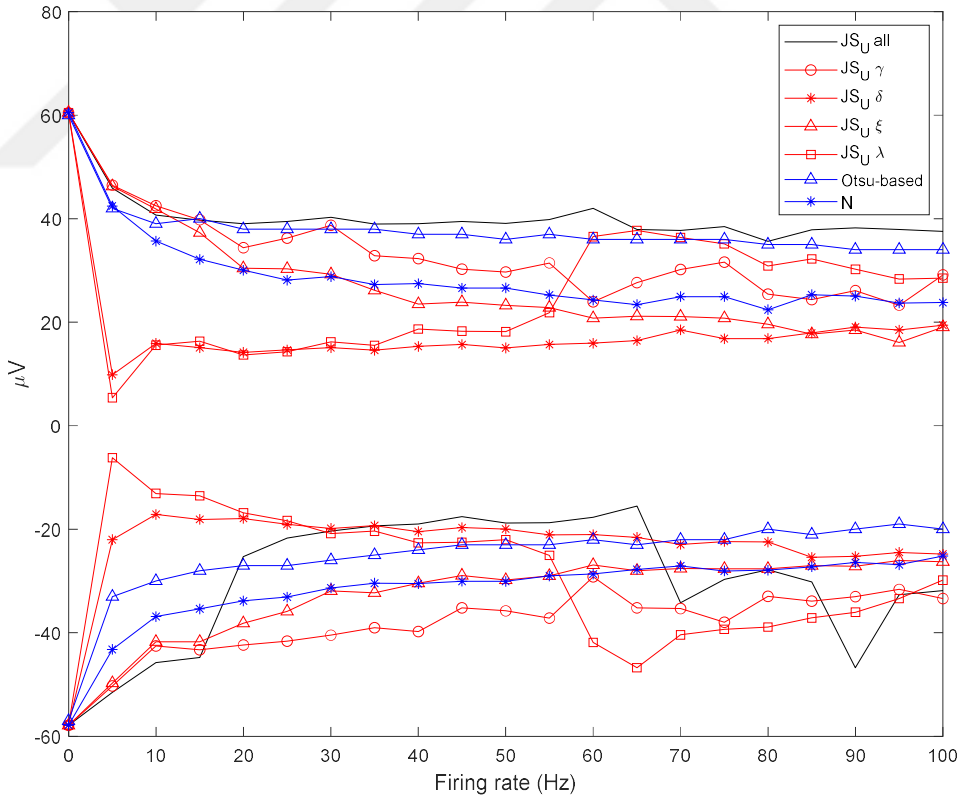

Figure 3.1 : Truncation thresholds.

In Figure 3.1, the threshold values obtained by JS_U, JS_U with fixed parameters (fixed γ, δ, ξ and λ respectively), Normal distribution (N) and Otsu-based method are shown at 21 different firing rates. Since truncated Normal distribution is used rather than the

truncated JS_U distribution at 0 Hz, the values of all curves are equal at that point. The results given by the Otsu-based method were also essentially identical to those of the other methods at 0 Hz (a_K = -57 µV, a_N = -57.867 µV, b_K = 60 µV, b_N = 60.459 µV). Note that the Otsu-based method tests the thresholds only at integer levels.

In Figure 3.2, the noise widths obtained from the difference between the threshold values in Figure 3.1 are compared. JS_U γ gives the widest mean noise width (71.656 µV), while JS_U δ gives the narrowest mean noise width (41.125 µV).

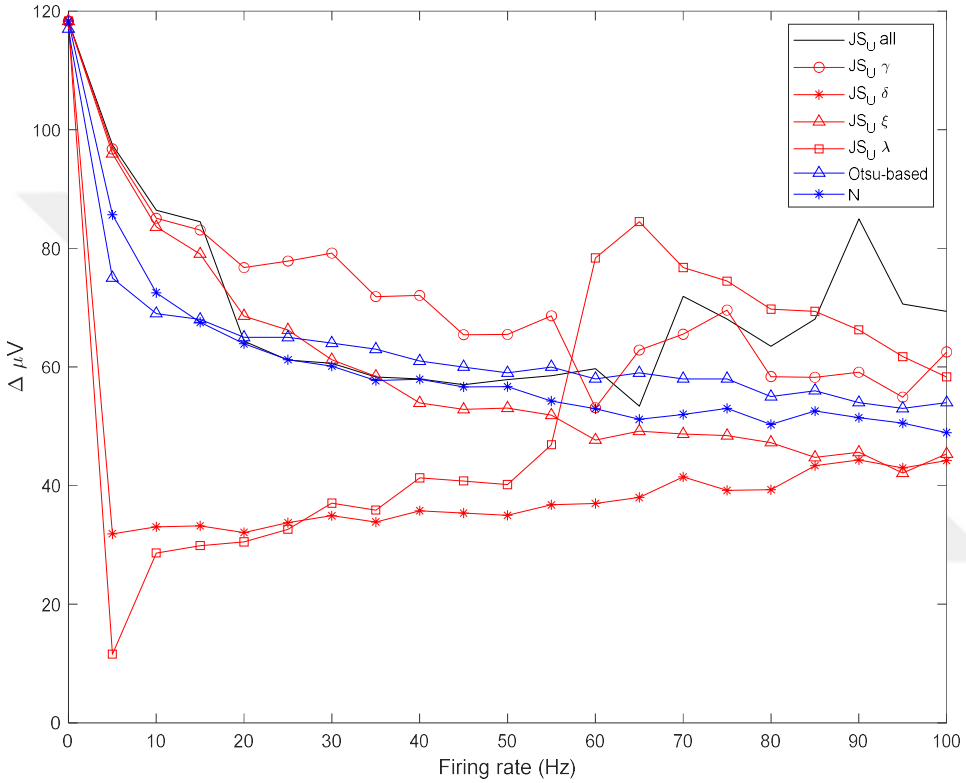

Figure 3.2 : The difference between upper and lower thresholds.

3.1.1 Frequency-dependent abrupt changes in $JS_{U\ all}$ and $JS_{U\ \lambda}$ cases

In the lower threshold values, the $JS_{U\ all}$ has a high variation in the 15-20 Hz and 65-70 Hz ranges compared to the other lower threshold values. In order to explain this variation, tt[i] values defined in Eq. 2.8 are shown for 19 iterations in Figure 3.3, $1 \leq i \leq 19$.

When the change of threshold values in 19 iterations is examined, there is a separation between the threshold values from the 5th iteration for 15-20Hz and from the 3rd iteration for 65-70Hz. The possible reasons for these variations are discussed in detail in Section 4.

In order to examine the dependence of the deviations in the threshold values on the $JS_{U\ all}$ parameters, the frequency-dependent variation of $JS_{U\ all}$ initial parameters (γ_0, δ_0, ξ_0 and λ_0) and estimated parameters ($\hat{\gamma}$, $\hat{\delta}$, $\hat{\xi}$ and $\hat{\lambda}$) are shown in Figure 3.4. The Pearson correlation coefficient between the lower threshold values and the differences of estimated and initial parameters ($\hat{\gamma}-\gamma_0$, $\hat{\delta}-\delta_0$, $\hat{\xi}-\xi_0$ and $\hat{\lambda}-\lambda_0$) are calculated as 0.7563, -0.3861, 0.6726, and -0.5604 respectively. Here, it is clearly seen that the frequency-dependent variation of $\hat{\gamma}$ and $\hat{\xi}$ highly correlates with the lower threshold values. In particular, γ seems to have settled to a different local optimum between 25-65 Hz (Fig. 3.4). ξ and λ also seem to have done the same to a lesser extent (Fig. 3.4). This may explain the abrupt change in the lower threshold in the range of 25-65 Hz.

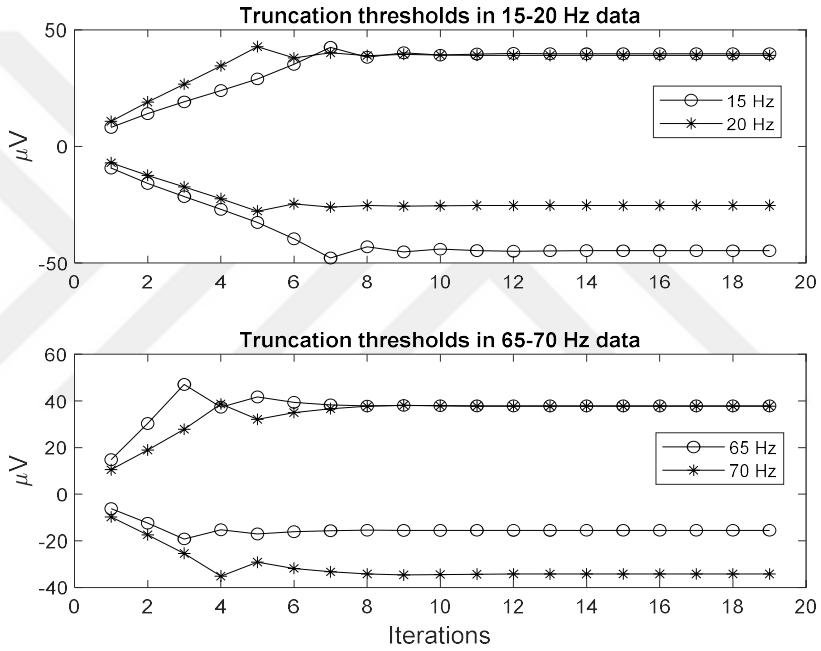

Figure 3.3 : Truncation thresholds in intermediate iterations (Eq. 2.8) for $JS_{U\ all}$.

Abrupt frequency-dependent change is also observed for the truncation thresholds computed in the $JS_{U\ \lambda}$ case (Fig 3.1). For f>55 Hz, upper threshold values increased, lower threshold values decreased and a wider range was obtained (Fig 3.2). In order to explain this change, the threshold values found at intermediate iterations of tt[i], $\theta_{l|m}$ and $\theta_{h|m}$ are shown for 55-60 Hz (Fig. 3.5). Here, the threshold values in iterations for 55 Hz are in a narrower range, while for 60 Hz this range is widened. The reason for this situation is that in first iteration of the 55 Hz case, the fitted $JS_{U\ \lambda}$ distribution between $\theta_{l|m}$ and $\theta_{h|m}$ passed the KS test and therefore the ϕ parameter of Eq. 2.8 is searched between 0 and 1. On the other hand, the fitted $JS_{U\ \lambda}$ distribution between $\theta_{l|m}$ and $\theta_{h|m}$ in the 60 Hz case couldn't pass the KS test and the ϕ parameter is searched

between 1 and the max value of the SCS (see Section 2.3.1.2). Therefore, the threshold values yield a wider range at 60 Hz compared to 55 Hz for $JS_{U\,\lambda}$.

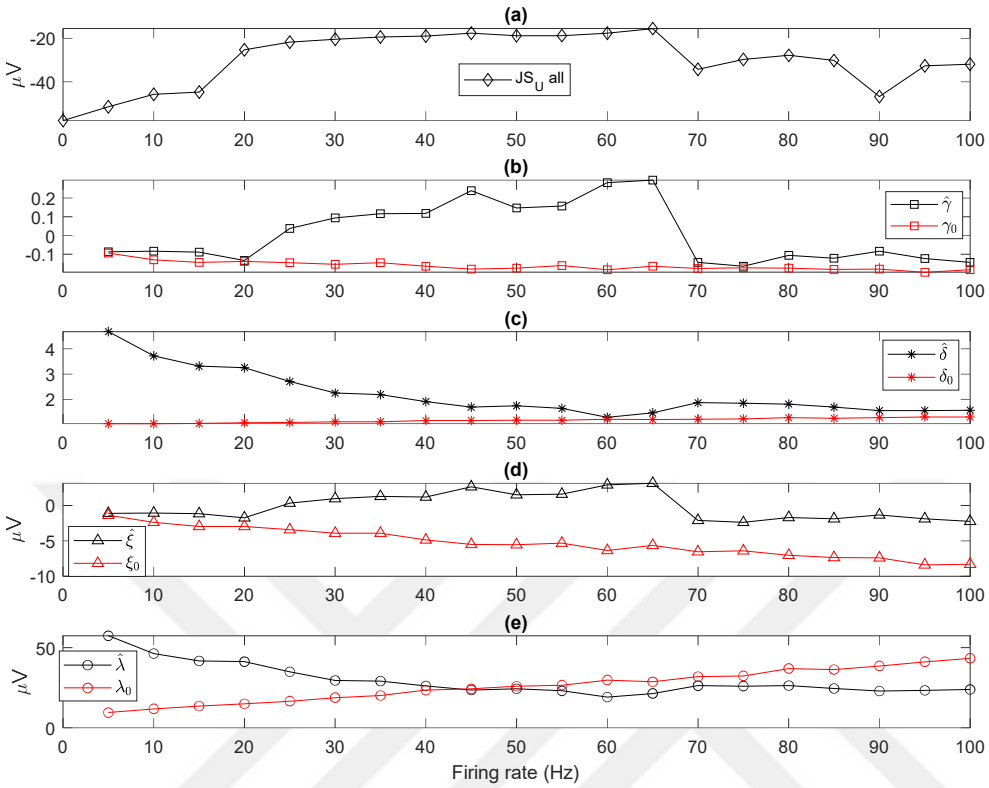

Figure 3.4 : (a) Lower threshold values for $JS_{U\,all}$ (b, c, d, e) Estimated and initial γ, δ, ξ and λ respectively at each firing rate.

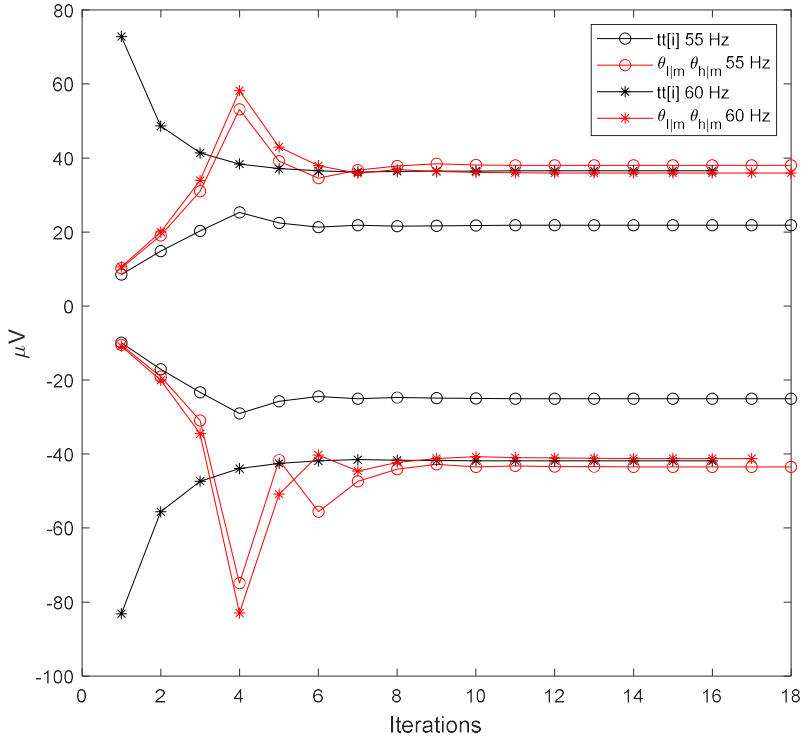

Figure 3.5 : Truncation thresholds in intermediate iterations (Eq. 2.8) for $JS_{U\,\lambda}$.

3.2 Estimated Standard Deviation of the Background Activity

The data are truncated with the threshold values specified in Figure 3.1, the ratio of the estimated standard deviations (Eq. 2.11 and Eq. 2.12) to the actual standard deviation is shown in Figure 3.6. Here, the conventional standard deviation formula in Eq. 2.10 is used for the Otsu-based method.

Table 3.1: The MSE values for estimated standard deviation.

Ratio	MSE with 1
$\hat{s}_K/\sigma$	5.78×10^{-5}
$\hat{\sigma}_N/\sigma$	0.0109
$\hat{\sigma}_{JS_U\,\xi}/\sigma$	0.0145
$\hat{\sigma}_{JS_U\,\gamma}/\sigma$	0.0537
$\hat{\sigma}_{JS_U\,all}/\sigma$	0.1361
$\hat{\sigma}_{JS_U\,\delta}/\sigma$	0.9794
$\hat{\sigma}_{JS_U\,\lambda}/\sigma$	24.8310 (without Inf value)

Regression analysis [19] and Kruskal-Wallis test [21] show that $\hat{s}_K/\sigma$ is indistinguishable from 1 (P=0.78; Kruskal-Wallis test). Except for the Otsu-based method all ratios moves away from 1. The Mean Squared Error (MSE) values calculated to measure the distance of each standard deviation ratio from 1 are shown in Table 3.1.

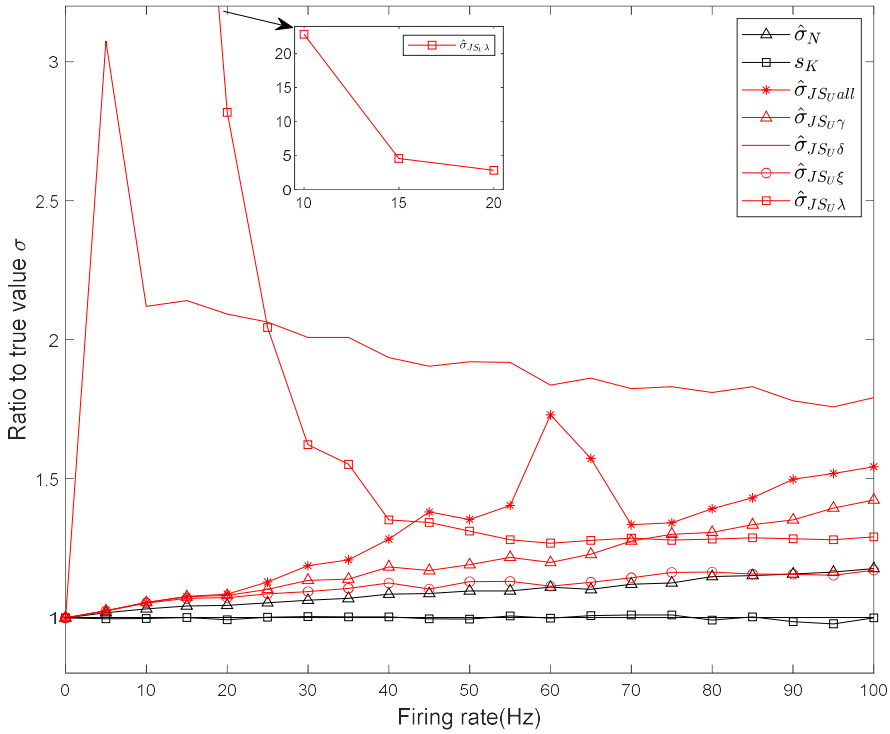

Figure 3.6 : Ratio of standard deviation estimates to σ.

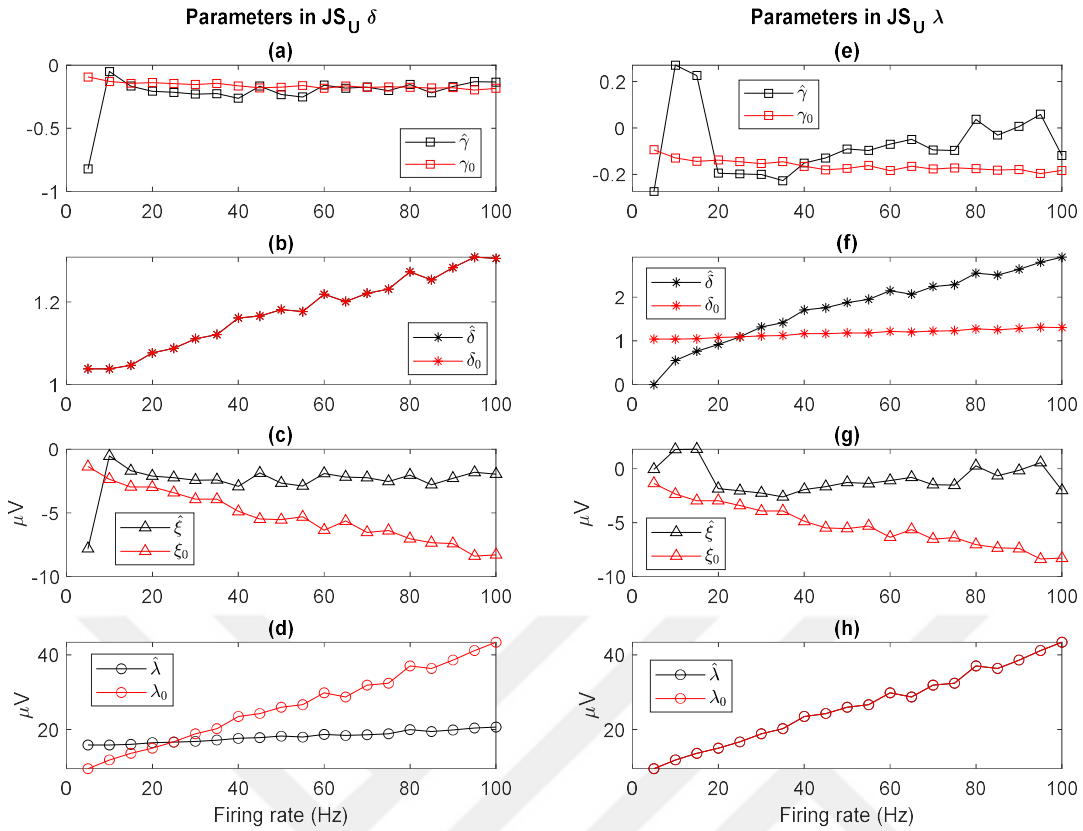

Figure 3.7 : Estimated parameters and initial parameters (γ, δ, ξ, and λ) at each frequency for $JS_{U\,\delta}$ (a,b,c,d) and $JS_{U\,\lambda}$ (e,f,g,h) respectively.

For the cases where the parameter is fixed, $\hat{\sigma}_{JS_U\,\delta}/\sigma$ has a high MSE value and is quite far from 1. $\hat{\sigma}_{JS_U\,\lambda}/\sigma$ reached an infinite value at 5 Hz and has high ratios up to 40 Hz. To explain why the standard deviation estimates are too large for these cases, the variation of the parameters with frequency is examined in Figure 3.7. Since the truncated Normal distribution for 0 Hz is used here, the parameters are shown starting from 5 Hz. For $\hat{\sigma}_{JS_U\,\delta}/\sigma$, δ values in 5 Hz and 10 Hz equal to each other, but γ has lowest value at 5 Hz. Due to the cosh function in Eq. 2.12, large deviations of γ away from 0 cause the computed standard deviation to increase rapidly. This explains the large amplitude of $\hat{\sigma}_{JS_U\,\delta}/\sigma$ at 5 Hz. In order to show why γ changes abruptly at 5 Hz, the truncation threshold values at 5-10 Hz intermediate iterations and the changes of γ and ξ in these iterations are shown in Figure 3.8. Here, from the first iteration, γ_{5Hz} and ξ_{5Hz} are fallen in intermediate iterations and diverged from γ_{10Hz}, and ξ_{10Hz}. The final estimated $\hat{\gamma}$ and $\hat{\xi}$ parameters at 5 Hz are affected from this divergence. On the other hand, for $\hat{\sigma}_{JS_U\,\lambda}/\sigma$, the ratio of γ to δ ($-5.4\times10^5$) at 5 Hz caused the cosh function in Eq. 2.12 to produce infinite values.

In other notable comparisons between the curves in Figure 3.7, JS_U ξ yields the curve closest to 1 among the fixed parameter cases because it has the smallest MSE value. It produced values that are almost identical to the Normal distribution (P=0.17; Kruskal-Wallis test).

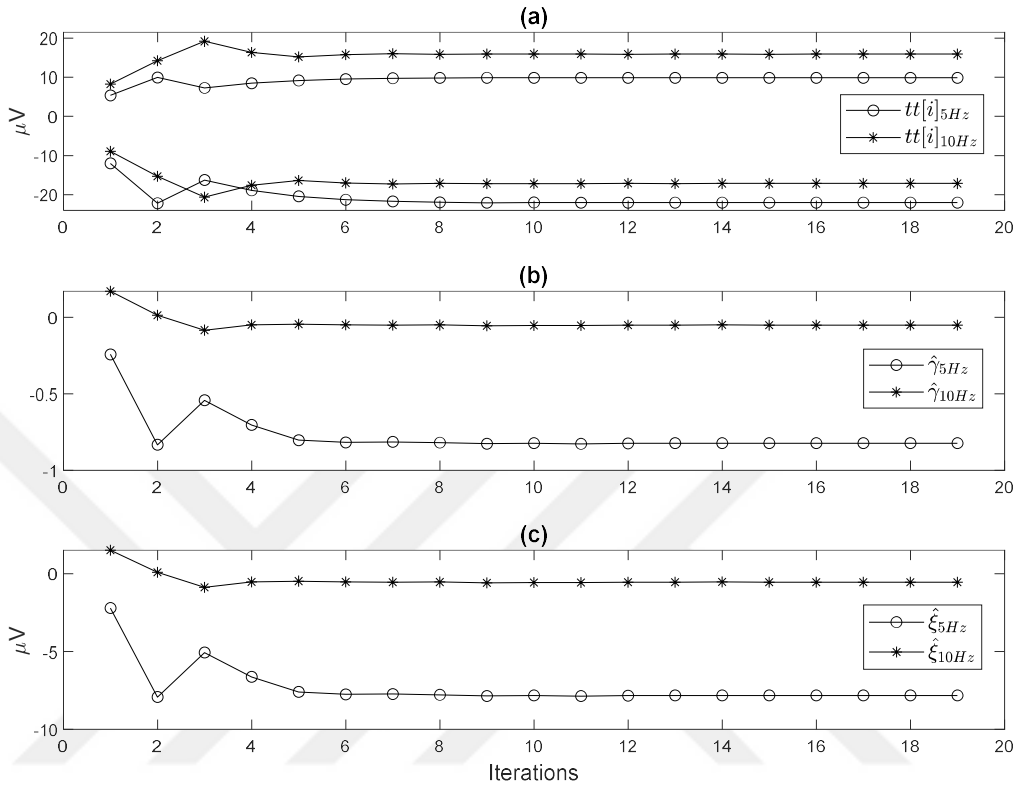

Figure 3.8 : (a) Truncation thresholds in intermediate iterations for $JS_{U\,\delta}$ in 5-10 Hz, (b, c) estimated parameters ($\hat{\gamma}$ and $\hat{\xi}$) at each intermediate iterations respectively.

3.3 Conventional Standard Deviation of the Background Activity

In this section, the ratio of the conventional standard deviation of the subthreshold data to the actual standard deviation is calculated to compare the performances of the thresholding methods. The data are truncated with the threshold values indicated in Figure 3.1, and the ratio of conventional standard deviations to actual standard deviation is shown in the Figure 3.9.

In addition, MSE values are shown in Table 3.2 to examine the closeness of the ratios of conventional standard deviations to actual standard deviations to the value of 1.

$s_{JS_U\,\mathrm{all}}/\sigma$, which is obtained by estimating all parameters, moved away from 1 starting from 65 Hz, but still showed better results than $\hat{\sigma}_{JS_U\,all}/\sigma$ according to MSE value.

Table 3.2: The MSE values for conventional standard deviation.

Ratio	MSE with 1
s_K/σ	5.78×10^{-5}
s_N/σ	9.47×10^{-4}
$s_{JS_{U\,\gamma}}/\sigma$	0.0032
$s_{JS_{U\,\xi}}/\sigma$	0.0057
$s_{JS_{U\,\mathrm{all}}}/\sigma$	0.0090
$s_{JS_{U\,\delta}}/\sigma$	0.0588
$s_{JS_{U\,\lambda}}/\sigma$	0.0684

For cases where a parameter is fixed, $s_{JS_{U\,\delta}}/\sigma$ and $s_{JS_{U\,\lambda}}/\sigma$ have high MSE values and are far away from 1, as were $\hat{\sigma}_{JS_{U\,\delta}}/\sigma$ and $\hat{\sigma}_{JS_{U\,\lambda}}/\sigma$. Although $\hat{\sigma}_{JS_{U\,\xi}}/\sigma$ gives the best results in Table 3.1 and is almost the same as the standard deviation estimated by the Normal distribution, $s_{JS_{U\,\xi}}/\sigma$ deviates from s_N/σ (P = 5.55×10^{-7}; Kruskal-Wallis). $s_{JS_{U\,\gamma}}/\sigma$ has a lower MSE value than $\hat{\sigma}_{JS_{U\,\gamma}}/\sigma$, providing the best conventional standard deviation estimate among the fixed parameter cases. In all parametric methods, conventional standard deviation formula yields better MSE values compared to the results in Table 3.1.

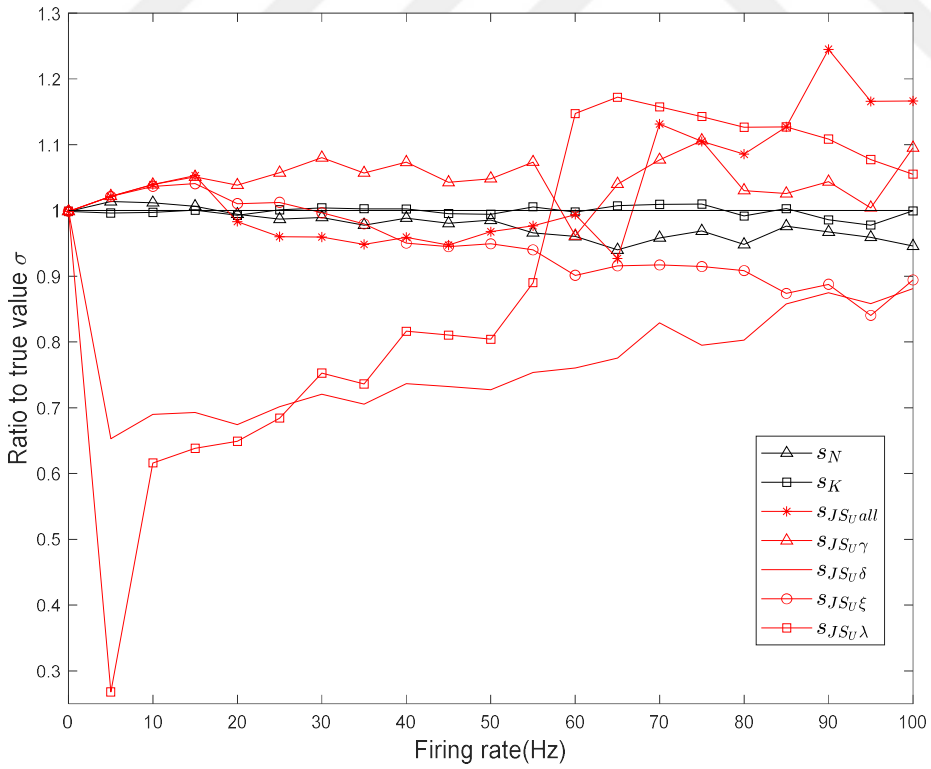

Figure 3.9 : Ratio of conventional standard deviation estimates to σ.

3.4 Distribution of Subthreshold Data Under the Otsu-Based Thresholds

The KS test is passed at all frequencies when the data between the Otsu-based threshold values are fit using the $JS_{U\ all}$ distribution [21]. The same data are fit with $JS_{U\ \gamma}$, $JS_{U\ \delta}$, $JS_{U\ \xi}$ and $JS_{U\ \lambda}$ using parameter fixation, and the KS test results are shown in Figure 3.10. Since the P values obtained with $JS_{U\ \delta}$ and $JS_{U\ \lambda}$ are very close to 0 at low f, they are shown in Figure 3.10 starting from from 70 Hz and 45 Hz, respectively. The KS test fails for $JS_{U\ \xi}$ beyond f=35 Hz. The best P values for the fixed parameter case are obtained with $JS_{U\ \gamma}$, but that too fails the KS test at 60, 65, 70 and 80Hz.

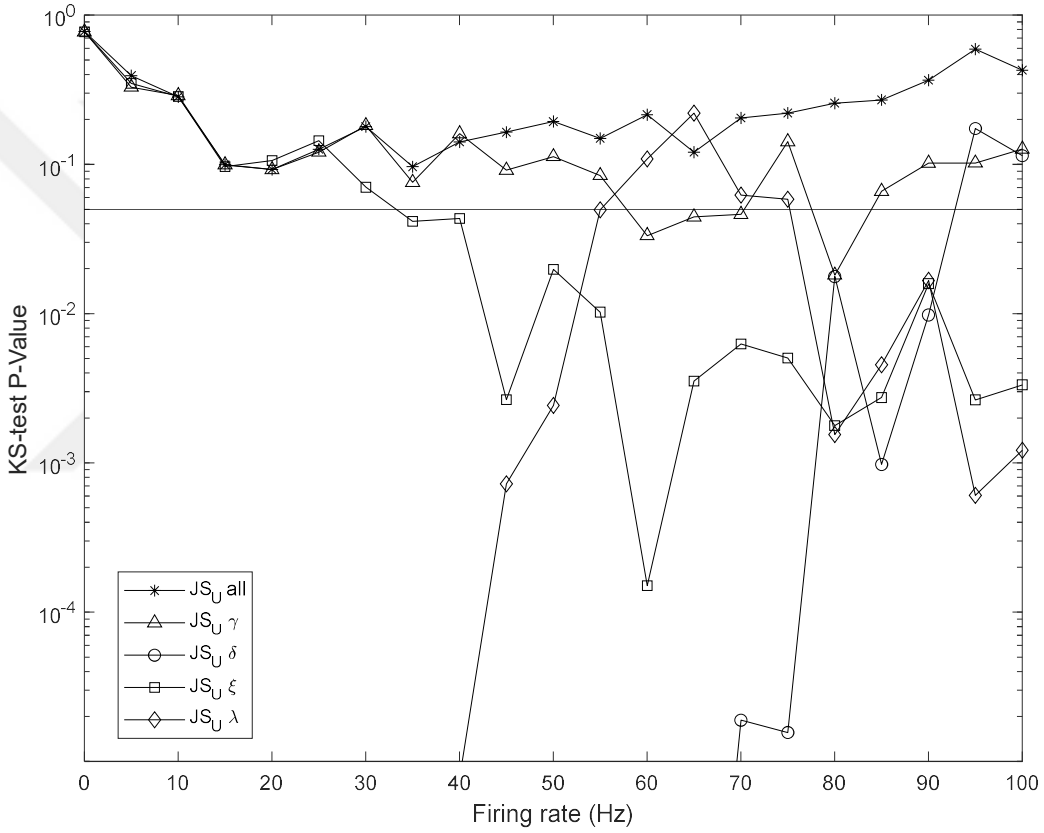

Figure 3.10 : Goodness-of-fit of truncated JS_U all, $JS_U\ \gamma$, $JS_U\ \delta$, $JS_U\ \xi$ and $JS_U\ \lambda$ to the subthreshold data under the Otsu-based thresholds. Horizontal line indicates the level 0.05.

3.5 Analysis with Real Data

So far, all analyzes have been performed using simulated extracellular neural recordings of [20]. Although simulated data contain characteristics of real data such as spike overlapping and realistic background noise power, measuring the performance of thresholding methods on real data is critical for the acceptability of thresholding

methods. Therefore, the extracellular neural recording shown in Figure 2.1 is analyzed using the same methods. The threshold values and, standard deviation values computed using the real data are shown in Figure 3.11. along with P values obtained by fitting the data encompassed by the Otsu-based thresholds with truncated Normal and truncated JS_U distributions.

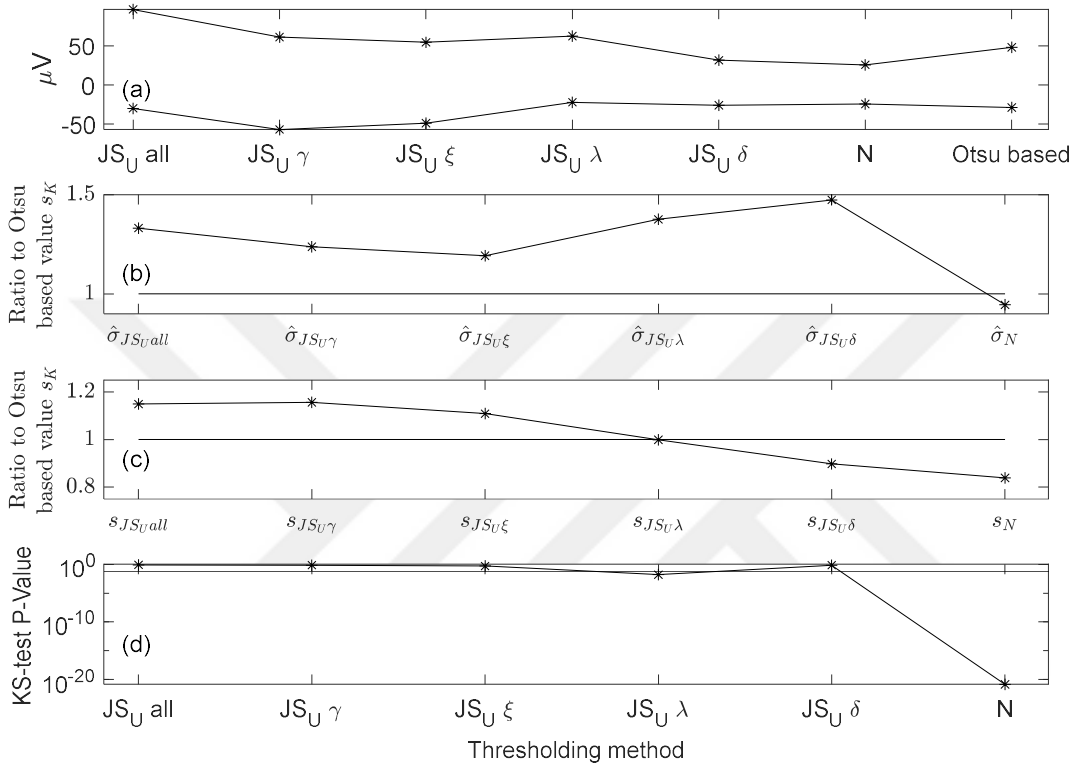

Figure 3.11 : (a) Low and high truncation thresholds, in the order of decreasing threshold separation (except for the Otsu-based thresholds), (b) Ratio of estimated standard deviation to Otsu-based conventional standard deviation, (c) Ratio of conventional standard deviation to Otsu-based conventional standard deviation, (d) Goodness of fit results for the truncated data between Otsu-based truncation thresholds.

Firstly, the results in Figure 3.11 show that all threshold, standard deviation and P-value types considered here are computable for the real data. Threshold values obtained with $JS_{U\,\delta}$ and N produced values close to each other. The N method showed the narrowest range compared to the other methods.

Otsu-based thresholds were the best-performing thresholds for simulated data, and the ratio of the obtained standard deviation to the actual standard deviation was close to 1 at each frequency. Therefore, the standard deviation value obtained by the Otsu-based

method is assumed as the true standard deviation and the ratios of the standard deviations estimated by the other methods to it are shown in Figure 3.11 (b). According to this analysis, the best standard deviation estimate is obtained with the threshold values obtained under N. For the conventional standard deviations, the data were truncated with the threshold values obtained in Figure 3.11 (a), and the conventional standard deviation of the data in between was calculated. According to this analysis $s_{JS_{U\ \lambda}}/s_K$ (0.99) gave the best standard deviation estimate. Since the MSE values were better when the conventional standard deviation formula was used in simulated data analysis (Table 3.2 versus Table 3.1), it may be argued that $JS_{U\,\lambda}$ may be preferred for estimating the standard deviation of the background activity in real data using the JS_U approach. However more real data analysis results are needed before reaching a reliable conclusion.

The samples between the threshold values obtained by the Otsu-based method are fit successfully by the $JS_{U\,\gamma}$, $JS_{U\,\delta}$, $JS_{U\,\xi}$ and $JS_{U\,all}$ distributions, as shown in Figure 3.11 (d). This was achieved in simulated that only by the $JS_{U\,all}$ distribution. This result suggests that the real data encompassed by the Otsu-based thresholds may be fit by some of the fixed-parameter JS_U distributions. On the other hand, the truncated Normal distribution could pass the KS test only at 0 and 5 Hz in the simulated data and it failed this test for the real data too.

4. CONCLUSION

In this study, truncated Normal [18], Otsu-based method [19], and truncated JS_U [21] methods are compared in their ability to separate the noise and non-noise components of extracellular neural recordings and parameter-fixed modeling is proposed for the JS_U-based approach. In the analysis, simulated data with 21 different firing rates (0-100 Hz with 5 Hz increments) were used [20]. In addition, analyzes were performed on a real 1-second long recording [20]. Among all the methods tested, the standard deviation of the simulated background activity was estimated most accurately by the Otsu-based method. In addition, the simulated data between the Otsu-based thresholds was successfully modeled with the truncated JS_U distribution and not with the truncated Normal distribution. However, among these methods, the processing time of the truncated JS_U method is longer than the others. In order to reduce this processing time, four different parameter-fixed JS_U methods are proposed by fixing one parameter of this distribution at a time. Except for fixing λ, parameter fixing reduces the computation time of the thresholds under the truncated JS_U distribution. But the reduced times are not better than the time it takes to compute the thresholds under the truncated Normal distribution or the Otsu-based method. The latter is the fastest among the methods considered here.

Calculated truncation thresholds are compared in Figure 3.1. It is noted that, in the lower threshold values, the $JS_{U\ all}$ has a high variation in the 15-20 Hz and 65-70 Hz. The intermediate iterations are examined in Figure 3.3 for JS_U all. Intermediate threshold values move up or down from one iteration to the next depending on the KS test result during the operation of the bisection search algorithm. There may be many upper and lower threshold pairs that can pass the KS test. Since the algorithm tests only those threshold pairs that are selected by the bisection algorithm, the ultimate solution is not guaranteed to be the global optimum. Thus, the threshold pairs obtained at adjacent frequencies may occasionally be very different from each other. In other words, the large frequency-dependent changes in the threshold values may be due to different local optima discovered at those firing rates. Such large differences in the calculated thresholds at adjacent firing rate values can be eliminated with a higher

resolution search that tests many more threshold pairs. However, searching for threshold values with high resolution will be computationally more costly. This problem may be examined further in future studies.

The same high variation problem that occurred in $JS_{U\,all}$ also exists for $JS_{U\,\lambda}$ at 55-60 Hz. In order to explain this variation, intermediate iterations of $JS_{U\,\lambda}$ are examined in Figure 3.5. While samples between the intermediate threshold values could be fitted with $JS_{U\,\lambda}$ in the first iteration at 55 Hz, they could not be fitted in the first iteration at 60 Hz. Therefore, the ϕ values differ in Eq. 2.8 (see Section 2.3.1.2) and lead to a high variation in truncation thresholds between 55 and 60 Hz.

The ratios of the estimated standard deviations to the actual standard deviations are compared in Figure 3.6. Among the parameter-fixed methods, the lowest MSE value is obtained with $JS_{U\,\xi}$. Also, $\hat{\sigma}_{JS_{U\,\xi}}/\sigma$ and $\hat{\sigma}_{\mathrm{N}}/\sigma$ cannot be distinguished from each other (P=0.17; Kruskal-Wallis test). Furthermore, the estimated standard deviation for $JS_{U\,\delta}$ at 5 Hz is calculated to be infinity. In order to examine this problem in detail, the estimated parameters of the $JS_{U\,\delta}$ are examined, where it was observed that the γ and δ parameters had a sudden change at 5 Hz. Intermediate iterations to examine this sudden change at 5 Hz and the estimated γ and ξ parameters in these iterations are shown in Figure 3.8. Similar local optima problem occurred in $JS_{U\,all}$ (15-20Hz and 65-70Hz) is observed here as well.

In Figure 3.9, the ratios of conventional standard deviations of subthreshold data to actual standard deviations are compared and the lowest MSE value among the parameter-fixed methods is obtained with $JS_U\,\gamma$. These methods produced lower MSE values than JS_U all, which shows that fixing one of the parameters instead of estimating all parameters can improve the estimation of the standard deviation of the background activity. When the distribution of the samples encompassed by the Otsu-based thresholds in Figure 3.10 is modeled with parameter-fixed methods, the best P values are obtained with $JS_U\,\gamma$, but it does not pass the KS test for 60, 65, 70, and 80 Hz. The only model that successfully fits the distribution of the samples encompassed by the Otsu-based thresholds at all firing rates is the truncated $JS_{U\,all}$ distribution.

Finally, the same analyzes were repeated in Figure 3.11 for the 1 second-long real data. Due to the success of the Otsu-based method in simulated data, the standard deviation estimate computed by that method has been taken as ground truth for the real data.

According to this analysis, the conventional standard deviation of the background activity is best estimated with JS_U λ. While this method was the parameter-fixed method with the highest MSE value in simulated data, it gave results close to the Otsu-based threshold value in real data. In addition, the distribution of the samples encompassed by the Otsu-based thresholds could be fit successfully with truncated JS_U all, JS_U γ, JS_U δ, and JS_U ξ distributions.

Overall, our results show that the truncated $JS_{U\ all}$ distribution is the only distribution, among the ones considered here, that explains the distribution of the data encompassed by the Otsu-based thresholds in both simulated and real data with a KS P value of at least 0.05 at all firing rates. We also found that the Otsu-based thresholding method is the fastest and most accurate method, among the ones considered here, in estimating the standard deviation of the background activity in simulated data. Our results also suggest that, when the truncated JS_U distribution is used for modeling the distribution of the subthreshold data, fixing one of its four parameters may substantially reduce the computation time while yielding comparable threshold or standard deviation estimates, especially when the fixed parameter is γ or ξ.

REFERENCES

[1] **Squire, L., Berg, D., Bloom, F. E., Du Lac, S., Ghosh, A., and Spitzer, N. C.** (2012). Fundamental neuroscience. Academic press.

[2] **Kettenmann H, Ransom BR**. (1995). Neuroglia. New York: Oxford University Press.

[3] **Buzsáki, G.** (2004). Large-scale recording of neuronal ensembles. *Nat Neurosci.*, *7(5),* 446-451.

[4] **Musk, E.** (2019). An integrated brain-machine interface platform with thousands of channels. *Journal of medical Internet research, 21.10, e16194*.

[5] **Engel, A.K., Moll, C.K., Fried, I. and Ojemann, G.A.** (2005). Invasive recordings from the human brain: clinical insights and beyond. *Nat Rev Neurosci., 6(1)*, 35-47.

[6] **Wang, S. Gao, J., Su, X. L. and Guo, Y. F.** (2013). Different changes of firing activities between ventral and dorsal medial prefrontal cortex in PD rats. *J Neurological Sciences (Turkish), 30(4),* 717-723.

[7] **Berke, J.D., Okatan, M., Skurski, J. and Eichenbaum, H.B.** (2004). Oscillatory entrainment of striatal neurons in freely moving rats. *Neuron, 43(6)*, 883-896.

[8] **Cinel, C., Valeriani, D., and Poli, R.** (2019). Neurotechnologies for human cognitive augmentation: current state of the art and future prospects. *Front. Hum. Neurosci., 13:13*.

[9] **Quiroga, R.Q., Nadasdy, Z. and Ben-Shaul, Y.** (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput, vol.16,* pp. 1661-1687.

[10] **Yang, Z., Liu, W., Keshtkaran, M. R., Zhou, Y., Xu, J., Pikov, V., Guan, C. and Lian, Y.** (2012). A new EC–PC threshold estimation method for in vivo neural spike detection. *Journal of neural engineering*, *9(4),* 046017.

[11] **Vargas-Irwin, C. and Donoghue, J.P.** (2007). Automated spike sorting using density grid contour clustering and subtractive waveform decomposition. J*ournal of neuroscience methods, vol. 164,* pp. 1-18.

[12] **Takekawa, T., Isomura, Y. and Fukai, T.** (2010). Accurate spike sorting for multi-unit recordings. *European Journal of Neuroscience, vol. 31,* pp. 263-272.

[13] **Jäckel, D., Frey, U., Fiscella, M., Franke, F. and Hierlemann, A**. (2012). Applicability of independent component analysis on high-density microelectrode array recordings. *Journal of Neurophysiology, vol 108,* pp. 334-348.

[14] **Todorova, S., Sadtler, P., Batista, A., Chase, S. and Ventura, V.** (2014). To sort or not to sort: the impact of spike-sorting on neural decoding performance. *Journal of neural engineering, vol. 11*.

[15] **Caro-Martín, C. R., Delgado-García, J. M., Gruart, A., and Sánchez-Campusano, R.** (2018). Spike sorting based on shape, phase, and distribution features, and K-TOPS clustering with validity and error indices. *Scientific reports*, *8*(1), 1-28.

[16] **Oby, E. R., Perel, S., Sadtler, P. T., Ruff, D. A., Mischel, J. L., Montez, D. F. and Chase, S. M.** (2016). Extracellular voltage threshold settings can be tuned for optimal encoding of movement and stimulus parameters. *Journal of neural engineering*, *13*(3), 036009.

[17] **Dağdevir, E., Kocatürk, M., and Okatan, M.** (2019). Likelihood-Based Amplitude Thresholding in Extracellular Neural Recordings. *27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, IEEE.

[18] **Okatan, M., and Kocatürk, M.** (2017). Truncation thresholds: a pair of spike detection thresholds computed using truncated probability distributions. *Turkish Journal of Electrical Engineering and Computer Sciences*, *25*(2), 1436-1447.

[19] **Kabakçi, K. A., Töreyin, B. U., and Okatan, M**. (2020). Estimation of Noise Standard Deviation Using an Otsu-Based Approach in Extracellular Neural Recordings. *2020 28th Signal Processing and Communications Applications Conference (SIU),* pp. 1-4, IEEE.

[20] **Okatan, M. and Kocatürk, M.** (2020). Simulated extracellular neural recordings, with white noise of physiologically relevant variance, and firing rates in the range of 0-100 Hz, generated using the action potential of a putative pyramidal neuron recorded from the primary motor cortex of an awake behaving rat. http://dx.doi.org/10.6080/K0S180QW, The data are available at https://web.itu.edu.tr/okatan/veri/ (online available).

[21] **Ögütcen, M. Y., Kocatürk, M., and Okatan, M**. (2021). Using Johnson's SU Distribution for Modeling the Background Activity in Extracellular Neural Recordings. *2021 Medical Technologies Congress (TIPTEKNO)*, pp. 1-4, IEEE.

[22] **Johnson, N. L.** (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149-176.

[23] **Donoho, D. L., and Johnstone, J. M.** (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika*, *81*(3), 425-455.

[24] **Okatan, M., and Kocatürk, M.** (2015). Firing rate dependence of" truncation thresholds". In *2015 Medical Technologies National Conference (TIPTEKNO)*, pp. 1-4, IEEE.

[25] **Otsu, N.** (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics, 9(1),* 62-66.

[26] **Lewicki, M. S.** (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, *9*(4), R53.

[27] **Hossain, M. M., Imtiaz, I., and Hasan, M. R.** (2019). Forecasting the general index of Dhaka Stock Exchange. *Int. Res. J. Finan. Econ*, *171*, 20-37.

[28] **Choi, P., and Min, I. S.** (2008). Further Applications of Johnson's S U-normal Distribution to Various Regression Models. *Communications for Statistical Applications and Methods*, *15*(2), 161-171.

[29] **Jones, D. L.** (2014). The Johnson Curve Toolbox for Matlab: analysis of non-normal data using the Johnson system of distributions. The tool available at: https://www.usf.edu/marine-science/research/matlab-resources/johnson-system-of-distributions-johnson-curves.aspx (online available).

[30] **Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.** (2007). *Numerical recipes 2nd edition: The art of scientific computing*. Cambridge university press.

[31] **Pearson, K.** (1895). X. Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London.(A.)*, (186), 343-414.

[32] **Pearson, K.** (1916). IX. Mathematical contributions to the theory of evolution.—XIX. Second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *216*(538-548), 429-457.

APPENDICES

APPENDIX A : Details of experiments

APPENDIX A

All computations performed on a system with Intel(R) Core (TM) i7-10750H CPU @ 2.60GHz, 12-core processor with 16 GB RAM under Windows 11 Pro (64 bit) operating system.

All developments are carried on MATLAB Version 9.8.0.1323502 (R2020a), Natick, Massachusetts, USA: The MathWorks Inc.

In MATLAB, maximum likelihood estimation is computed with mle.m function and this function uses the Nelder-Mead optimizer by default.

Maximum number of iterations and maximum number of objective function evaluations of optimizer are set to 400000 and 800000 respectively.

CURRICULUM VITAE

Name Surname : Melih Yılmaz ÖGÜTCEN

EDUCATION :

- **B.Sc. :** 2019, Suleyman Demirel University, Engineering Faculty, Electronics and Communication Engineering Department

PROFESSIONAL EXPERIENCE AND REWARDS:

- Data Scientist / ML Engineer at CicekSepeti Internet Services Inc. (2021-…)
- Data Scientist at Linktera Information Technologies Inc. (2020-2021)

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Ögütcen, M. Y.,** Kocatürk, M., and Okatan, M. (2021). Using Johnson's SU Distribution for Modeling the Background Activity in Extracellular Neural Recordings. *2021 Medical Technologies Congress (TIPTEKNO)*, pp. 1-4, IEEE.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Akca, E., Özçay, T., Dinç, Y., Yalçı, N., Taşabat, S. E., Varol, M. A., Kayı, B**, Ögütcen, M. Y.** and Öztürk, B. (2022). A CNN Based Ensemble Approach for Malfunction Detection from Machine Sounds. *The European Journal of Research and Development*, *2*(2), 411-420.
- **Ögütcen, M. Y.,** Kocatürk, M., and Okatan, M. (2021). A python code for maximum likelihood estimation of the location and scale parameters of the truncated normal distribution. *2021 Medical Technologies Congress (TIPTEKNO)*, pp. 1-4, IEEE.
- **Ögütcen, M. Y.**, and Koç, T. (2020). Segmentation of Compressed High Speed Video Images of Vocal Folds. *28th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, IEEE.
- **Ögütcen, M. Y.**, and Koc, T. (2019). Yüksek Hızlı Ses Telleri Görüntülerinin Düzlemsel Aydınlatma Modeli ile Aktif Kontur Tabanlı Segmentasyonu, EEMKON 2019, pp. 427-431 (in Turkish).