

Universal automated classification of the acoustic startle reflex using machine learning

Timothy J Fawcett^{1,4}, Ryan J Longenecker⁵, Dimitri L. Brunelle¹, Joel I Berger⁶, Mark N Wallace⁷, Alex V Galazyuk⁸, Merri J Rosen⁸, Richard J Salvi⁹, and Joseph P Walton^{1,2,3}

¹Global Center for Hearing and Speech Research, University of South Florida, Tampa, FL, USA

²Department of Medical Engineering, University of South Florida, Tampa, FL, USA

³Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL, USA

⁴Research Computing, University of South Florida, Tampa, FL, USA

⁵Sound Pharmaceuticals Inc, 4010 Stone Way N., Suite 120, Seattle, WA 98103

⁶Department of Neurosurgery, University of Iowa Hospitals and Clinics, Iowa City, IA, USA

⁷Hearing Sciences, School of Medicine, University of Nottingham, Nottingham, UK

⁸Department of Anatomy and Neurobiology, Northeast Ohio Medical University, Rootstown, OH, USA

⁹Center for Hearing and Deafness, University at Buffalo, The State University of New York, Buffalo, USA

These authors contributed equally: Timothy J Fawcett, Ryan J Longenecker. e-mail: tfawcett@usf.edu

Corresponding Author: Joseph P. Walton, PhD. email: jwalton1@usf.edu

Keywords

Acoustic startle response, Pre-pulse inhibition, Machine learning, Ensemble models, Waveform classification

Abstract

The startle reflex (SR), a robust, motor response elicited by an intense auditory, visual, or somatosensory stimulus has been widely used as a tool to assess psychophysiology in humans and animals for almost a century in diverse fields such as schizophrenia, bipolar disorder, hearing loss, and tinnitus. Previously, SR waveforms have been ignored, or assessed with basic statistical techniques and/or simple template matching paradigms. This has led to considerable variability in SR studies from different laboratories, and species. In an effort to standardize SR assessment methods, we developed a machine learning algorithm and workflow to automatically classify SR waveforms in virtually any animal model including mice, rats, guinea pigs, and gerbils obtained with various paradigms and modalities from several laboratories. The universal features common to SR waveforms of various species and paradigms are examined and discussed in the context of each animal model. The procedure describes common results using the SR across species and how to fully implement the open-source R implementation. Since SR is widely used to investigate toxicological or pharmaceutical efficacy, a detailed and universal SR waveform classification protocol should be developed to aid in standardizing SR assessment procedures across different laboratories and species. This machine learning-based method will improve data reliability and translatability between labs that use the startle reflex paradigm.

1. Introduction

The startle reflex (SR), an abrupt motoric response elicited by an intense auditory stimulus, has been used as a tool to assess a subject's overt reaction to various types of acoustic stimuli for over a century (Preyer, 1900; Landis and Hunt, 1939; Davis, 1984). The response is composed of multiple movements involving facial nerve output to the face and ears and spinal nerve output to the neck, back, tail, and limb extensors (Pantoni et al., 2020). Different components of the reflex have been measured using different methods but the main ones in rodents involve measurement of the whole-body response by placing piezoelectric sensors (Dulawa et al., 1997; Longenecker and Galazyuk, 2011; Schilling et al., 2017) or strain gauges/load cell (Berger et al., 2013; Virag et al., 2021) under the base of the cage. Alternatively, the ear flick component (Preyer reflex) can be measured separately by making electromyographic recordings (Cassella and Davis, 1986a; Cassella and Davis, 1986b) or optical tracking of ear movements (Berger et al., 2013). This basic reflex has been used to investigate cognitive, sensory, and mental disorders such as schizophrenia (Geyer et al., 1990; Ahmari et al., 2012; Fendt and Koch, 2013; Buse et al., 2016; Khan and Powell, 2018), bipolar disorder (Mao et al., 2019), hearing loss (Longenecker et al., 2016; Lauer et al., 2017; Wake et al., 2021), and tinnitus (Turner et al., 2006; Berger et al., 2013; Lobarinas et al., 2013; Galazyuk and Hebert, 2015; Zhang et al., 2005). The SR is ubiquitous in the mammalian nervous system (Koch, 1999) and has been used to assess the internal state of rats (Kraus et al., 2011), mice (Longenecker and Galazyuk, 2011; Lowe and Walton, 2015), gerbils (Green et al., 2016), and guinea pigs (Berger et al., 2013). The acoustic startle response waveform has not been extensively characterized, and the lack of a standardized method for quantifying the startle response has led to high variation in startle data across different studies and laboratories. Since SR is widely used to investigate toxicological or pharmaceutical efficacy, a detailed and universal SR waveform classification system should be developed to aid in standardizing SR assessment procedures across different laboratories and species.

The acoustic startle response waveform was first examined in rats many decades ago (Horlington, 1968) and has since been extended to other animal models. Most studies utilizing SR methodologies have not classified individual startle response waveforms to ensure that all force traces were indeed related to an animal's physiological response to the intense stimulus, i.e., a true startle response. Instead, nearly all studies assume that a “true startle” response, as opposed to extraneous motor activity, was evoked by the acoustic stimulus. This assumption was brought into question by a recent study that categorized and detailed the acoustic startle response waveform via template matching (Grimsley et al., 2015) in mice.

Careful analysis indicated that a large proportion of so-called SR waveforms were actually non-startle responses. The Grimsley et al. study used high-speed motion cameras to help characterize the fundamental relationship between the mouse's whole body acoustic startle response and the resulting waveform. From this baseline analysis, they determined that an animal's movement after a positive-startle reaction was extremely stereotyped, to the point that a template (of waveform characteristics) could classify a true-startle from a non-startle almost a 100% of the time when compared to manual identification from trained SR behavioral neuroscientists. Unfortunately, this template analysis was set to assess only one type of waveform. However, the most important aspect of the study demonstrated that animals do not startle at a one-to-one ratio for each presentation presented, e.g., there are non-startle waveforms. Using the concept that not all startle stimuli result in a true-startle response, it was shown that SR input/output curves and PPI tests could predict the% of positive startles based on stimuli and that average startle magnitudes correspond to average

startle probability for a given stimulus set (Longenecker et al., 2016). This fact has also been shown in humans (Blumenthal and Levey, 1989).

Removal of non-startles is important because these trials don't offer information concerning the physiological function of the nervous system and instead should be considered background noise, which importantly, should not be included in the final data analysis (Grimsley et al., 2015; Fawcett et al., 2020). While this was an important first step for SR analysis, this approach had limited effectiveness as it employed a rigid template and model for specific features. In addition, analysis of these waveforms has not been standardized across lab groups, measurement equipment and different animal models. Thus, a more systematic and generalized approach to SR analysis would aid in comparing results from different laboratories and species and advance the discipline.

To address this issue, we have adapted a machine learning model to automatically classify SR waveforms from various species, paradigms, and modalities (Fawcett et al., 2020; 2021). The relative magnitude of neuro-muscular responses varies between species; thus, the stereotypical startle response waveform varies between species. Most mammalian species neurologically favor the whole-body response; however the guinea pig favors the Preyer reflex (see Berger et al., 2013 for a detailed comparison). This is highlighted in the response waveform variation seen in Fig. 2. However, the largest difference between the waveforms is the modality of which it was collected (Fig. 2A [load cell] vs 2F [piezo]) and the type of reflex (Fig. 2C [optical tracking] vs Fig. 2D [whole body response]). The magnitude, but not the overall fine structure of waveforms can be influenced by weight when considering the whole-body responses.

Machine learning is an advanced and powerful method to develop predictive models for category discrimination, while removing measurement bias (Kotsiantis et al., 2006). Even a modicum of manual classification can quickly “teach” a machine learning algorithm to accurately classify waveforms (Fawcett et al., 2020). Importantly, machine learning can work for any type of waveform because it learns the details of waveform shapes, sizes, and latencies unique to that species or setup. A direct comparison of SR waveforms across species can be useful to quantitatively assess the major differences between species and qualitatively determine if certain species are better suited for SR testing.

Here we provide a detailed protocol implemented in the R coding language to assist researchers in classifying SR waveforms as true startles from non-startles. The R language was selected because it is open-source as well as being intuitive and user-friendly, while also being able to design and implement machine learning models due to its rapid prototyping abilities. First, we demonstrate how to load labeled acoustic startle waveforms from a variety of species and modalities into the computational environment. Next, we train an initial machine learning model using a small set of features common to all SR waveforms and use default pre-processing functions to classify SR waveforms into true startles and non-startles. Since the accuracy and generalizability of any machine learning model is highly dependent on the quality of the features, we provide tools to determine the highly predictive features to be extracted from the SR waveforms. We then demonstrate how to train a machine learning algorithm with these extracted features. Finally, we demonstrate advantages of this analysis when applied to typical SR-related paradigms such as prepulse inhibition. This work will help elucidate and standardize the SR field by detailing a breadth of acoustic startle waveforms from most species used in SR research.

The machine learning startle classification method was initially developed to classify true-startles from non-startles using an acoustic SR behavioral paradigm employing CBA/CaJ mice (Fawcett et al., 2020). The goal was to reduce the variability in distinguishing between true-startle and non-startle SR waveforms through waveform normalization, feature

engineering and extraction, and then to combine these features into a robust predictive model that can train a machine learning ensemble to reliably distinguish between a true-startle versus non-startle from SR waveforms with little to no human intervention (Fawcett et al., 2020; Hastie et al., 2009; Kuhn, 2013). Even though prepulse inhibition of the SR has been one of the most widely used diagnostic tools for assessing a laboratory animal's internal state (Fawcett et al., 2020; Davis, 1984; Koch, 1999; Ison and Hoffman, 1983), there is a lack of standardized, systematic methods for distinguishing true-startle from non-startle responses. Due to this variability, the need for the aforementioned methods of systematically classifying true-startles from non-startles is not only necessary for mice, but other laboratory animals as well, such as the rat, gerbil, and guinea pig, as well as different measurement systems. Having successfully developed a machine learning algorithm that correctly predicts true-startles with a high degree of accuracy for the mouse SR collected using a piezoelectric platform, we extended the machine learning workflow to classify SR waveforms across the spectrum of laboratory animals and measurement modalities using data from different laboratories.

2. Materials and methods

Acoustic startle reflex measurements were collected from mouse, rat, guinea pig, and gerbil animal models using piezoelectric transducer, load cell, and optical tracking (Pryer reflex) measurements from five auditory neuroscience laboratories. After SR waveforms were collected, robust machine learning models were separately trained to classify SR waveforms from each species/startle modality data set as startles or non-startles via ensembles of machine learning models utilizing highly predictive feature engineering components from normalized SR waveforms.

For all laboratories, procedures were approved by their university's respective Institutional Animal Care and Use Committee and/or accordance with the Animals (Scientific Procedures) Act 1986. In all instances animals were given food and water ad libitum in temperature and humidity-controlled vivariums with 12-hour light dark cycles. SR waveforms were collected from a variety of hardware systems and were calibrated with either Brüel and Kjaer or Larson Davis microphones/amplifier equipment (for details on hardware specifications and procedural details please see Fawcett et al., 2020; Lowe and Walton, 2015; Longenecker and Galazyuk, 2012; Longenecker et al., 2018; Lobarinas et al., 2004; Berger et al., 2013; Green et al., 2017).

2.1. Summary of startle modalities

As described above, the acoustic startle reflex is robust and ubiquitous across mammalian species. However, differences in data collection modalities should be noted as startle response waveforms will vary dramatically as a function of how the startle was collected. As such, we will describe the acquisition modalities used in this work, which encompass the most common methodologies currently used in SR animal research. Piezoelectric transducers measure the whole-body animal movements recorded via the voltage generated when the piezoelectric materials are deformed due to animal movement. Piezoelectric transducers do not require a covariate of body weight for data analysis and there is no effect of dynamic range of measurements. Load cell transducers measure whole-body animal movements via an electrical signal whose magnitude is directly proportional to the force being measured. Startle responses from animals of different weights can be compared after body mass is taken into consideration. Load cell systems restrict the dynamic range of response and so that body

mass must be used as a covariate. Optical tracking measures the movement of the pinna reflex, or Preyer reflex, using a series of infrared cameras to track reflective markers placed on each pinna of the animal. It is important to note that Preyer reflex is used preferentially over whole-body startle reaction measurements in guinea pig (Berger et al., 2013).

2.2. Machine learning startle classification

The machine learning approach to classifying SR waveforms as startles or non-startles recently reported by Fawcett, et al. in 2020 was generalized, modified, and extended to operate on multiple species and startle modalities. SR waveforms from each species/startle modality data set collected using the methods described above were separately analyzed to determine the optimum waveform pre-processing procedure, the most predictive features for SR waveform classification, and to train an ensemble of machine learning models to classify SR waveforms as startles or non-startles. After successfully training machine learning models for each species/startle modality data set, the model was then used to classify startle waveforms from a variety of species and data acquisition systems as startles or non-startles. The flowchart shown in Fig. 1 describes the major steps of the machine learning classification.

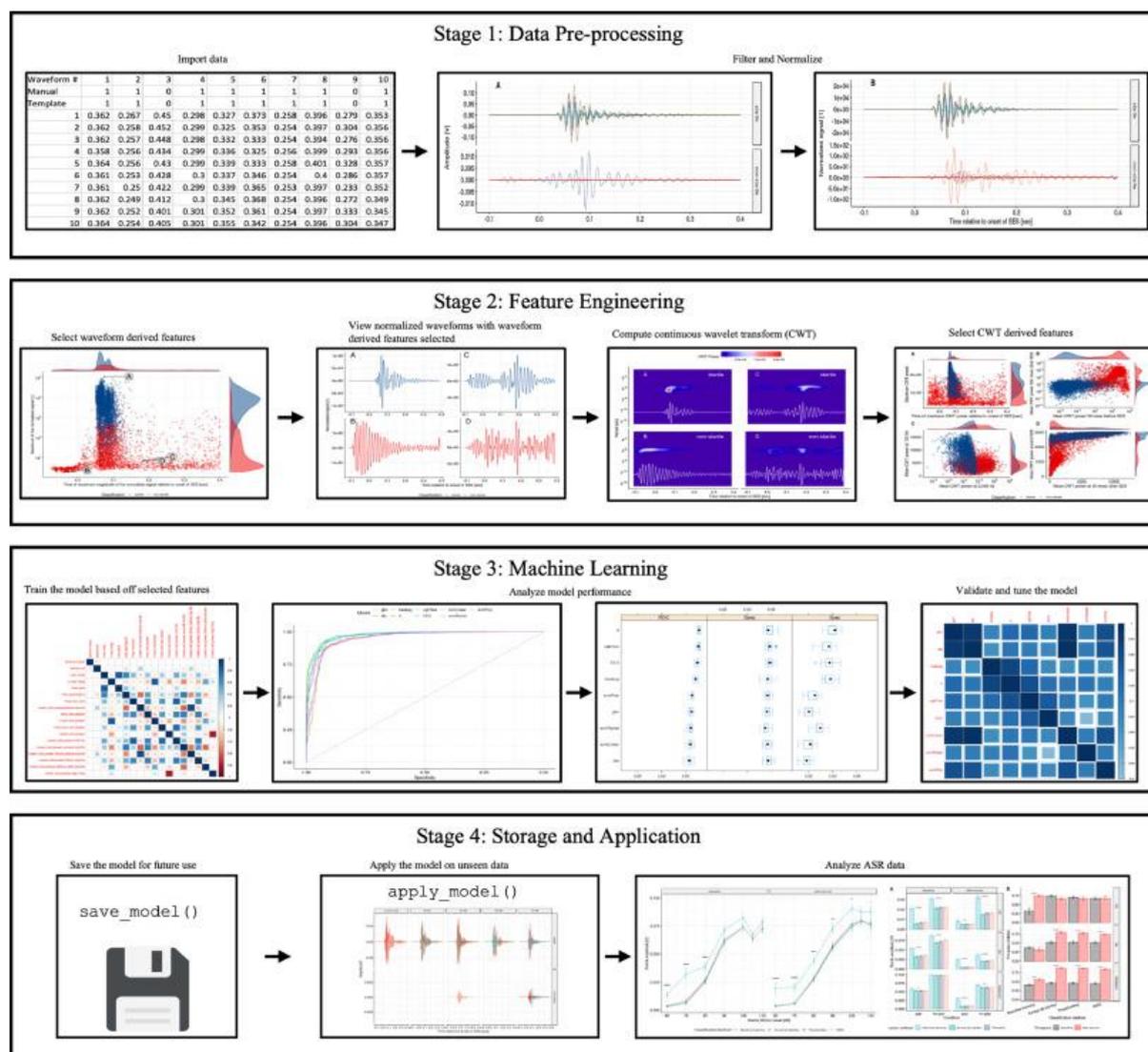


Fig. 1. Diagram of high-level classification stages. The diagram shows the four main stages of the machine learning startle classification workflow: data pre-processing, feature engineering, machine learning, and storage and application. Stage 1 produces filtered and normalized SR waveforms from raw data to the normalized waveforms. Stage 2 involves the selection of waveform derived features, then the subsequent computation and selection of power spectral density (PSD) and continuous wavelet transform (CWT) features. Stage 3 entails training, tuning the fully customizable machine learning model and testing the performance of the model. In stage 4, the model is saved and can be applied to an unseen dataset to analyze the newly classified SR data (described in detail in Fawcett et al., 2020; 2021).

2.2.1. Waveform pre-processing and normalization

In order to derive highly predictive features from SR waveforms which exhibit highly diverse morphological traits in the data, both filtering and normalization are necessary to remove high frequency content not related to the startle reflex response and to ensure that waveforms within a range of amplitudes can be easily compared. The waveforms can be centered (by subtracting the mean value from the SR waveform) and scaled (by dividing the centered waveform by the standard deviation) using the mean and standard deviation measured at various times, including but not limited to before the startle elicitor stimulus (pre-SES), post-SES, or the entire waveform. Additional normalization methods include dividing the SR waveform by the maximum amplitude of the signal to produce a signal with a maximum equal to 1 or subtracting the minimum of the signal then dividing by the range of the signal to produce a normalized signal that varies exactly between ± 1 . All SR waveforms considered in this study were centered using the mean and standard deviation computed before the startle elicitor stimulus (pre-SES) was presented.

2.2.2. Feature engineering

Features consisting of individual voltages at time points in the original SR waveform dataset possess weak predictive ability due to variability in startle response latency, magnitude, and sensor response. Therefore, waveforms must undergo several transformations to extract features which possess stronger predictive capability. Highly predictive features are extracted from the normalized SR waveforms and include features derived from the normalized waveform's characteristics, average frequency content obtained from power spectral density analysis, and temporally-resolved frequency content obtained from continuous wavelet transform (Fawcett et al., 2020).

2.2.2.1. Waveform derived features

Several features were derived from normalized waveforms, including the maximum magnitude and time of the maximum magnitude of the normalized waveform. Typically, the maximum magnitudes of true startles is significantly greater than that of non-startles and tend to occur within a predictable window after the presentation of the startle elicitor, giving these features very strong predictive capability. In our previous work in mice using this technique (Fawcett et al., 2020), we observed overlap between true-startles and non-startles in the maximum magnitudes when the maximum magnitudes occurred around 0.05 s (see Fig. 4 for example of this overlap in Fawcett et al., 2020). The substantial overlap that occurred in the maximum magnitude and the time of the maximum magnitude of the normalized SR true-

startle and non-startle waveforms indicated that additional features are necessary for a machine learning algorithm to accurately classify true SR waveforms.

2.2.2.2. Power spectral density

Because SR waveforms exhibit periodic behavior, analysis of the frequency content of each waveform leads to additional highly-predictive features. In our previous study in mice, the power spectral density (PSD) was estimated from the periodogram of the entire 0.5 s long normalized SR waveform. Most of the energy from the smoothed PSD in the PSD spectrum was concentrated between 10 Hz and 150 Hz. Additionally, this analysis showed a marked difference between the magnitude of the PSD for startle versus non-startle waveforms. Several distinguishing features of the smoothed PSD estimates obtained from the normalized SR waveforms include the maximum PSD power and the frequency of the maximum PSD (see Figs. 5 and 6 for these features in Fawcett et al., 2020). While there are significant differences in the maximum power of the PSD between startles and non-startles, there were no significant differences in the frequency where that maximum occurred. Because of this, a slight overlap in the maximum power of the PSD for startles versus non-startles was observed, indicating a need to identify additional features to accurately classify the SR waveforms.

2.2.2.3. Continuous wavelet transform

Both the magnitude and frequency of the periodic behavior of the SR waveforms are functions of time. Because the PSD estimates represent an average over the entire normalized SR waveform, temporal information is removed from the PSD estimate. In contrast, the continuous wavelet transform (CWT) of the normalized SR waveforms provides an estimate of the frequency content of the waveform as a function of time (Fawcett et al., 2020; Veer and Agarwal, 2015; Allen and Rabiner, 1977; Sejdic et al., 2009). In general, the wavelet power spectra revealed increased maximum wavelet power for true-startle versus non-startle waveforms. The wavelet power was concentrated around the maximum amplitude following the startle elicitor on trials classified as startles whereas the wavelet power was broadly distributed for trials classified as non-startles (Fig. 7 in Fawcett et al., 2020). Therefore, the CWT power spectra from normalized SR waveforms from each species/startle modality data set was used to obtain time and period (frequency) dependent CWT features. However, the exact timing and period (frequency) of the high CWT power is dependent on species and startle modality, requiring exploration of the CWT for several representative SR waveforms to determine the most useful timings and periods of interest to be extracted from the CWT. These features were combined with other metrics to increase the predictive ability (Fawcett et al., 2020; Cai et al., 2018; Vergara and Estevez, 2014).

2.2.3. *Machine learning methods*

Machine learning models were trained to classify SR waveforms as startles versus non-startles using a combination of features extracted from the normalized SR waveforms via an ensemble of models from various families of algorithms. The extracted features are first pre-processed to ensure only highly-predictive features that do not contain redundant information are used to train the machine learning models by removing features with near-zero variance, highly correlated features, and features which are linear combinations of other features. The remaining features are then centered and scaled prior to training any machine learning model so the magnitude and variability of any individual feature does not dominate the model.

Individual machine learning models using the Random Forests and Support Vector Machine (SVM) algorithms are then separately trained while also tuning hyperparameters of each machine learning model. After training, individual trained machine learning models are then stacked and combined via a generalized linear model, to provide a robust model which generalizes well to new/previously unseen SR waveforms.

2.2.4. Ground truth

Each of the 1000 SR waveforms per species and measurement paradigm (6000 total) presented in this paper were manually classified as either a startle or non-startle by an experienced neuroscientist. The number of required SR waveforms for successful machine learning model training was examined by generating a learning curve examining model performance metrics, just as accuracy and area under the receiver operating characteristic curve (ROC) as the number of examples used to train the models increases stratified by startle/non-startle classification (Balki et al., 2019). Prior to learning curve generation, 200 examples stratified by startle classification were partitioned from the 1000 SR waveforms and were removed from the training dataset and used for all model performance metric determinations. Next, models were trained using an increasing number of randomly selected examples from the remaining SR waveforms (from 10 to 800) with the variability of model performance metrics evaluated using 150 replicates.

3. Results

3.1. Waveform variability

The variability in SR waveforms both within and across species and startle modality is shown in Fig. 2. All SR waveforms manually classified as startles presented in Fig. 2 demonstrate similar characteristics such as large changes in waveform activity, magnitude, and frequency content immediately after the startle elicitor sound (SES) is presented at $t = 0$. However, the characteristics of this increased waveform activity differs between, but is consistent within, species and startle modalities. For example, the biphasic shape of a startle response for mouse load cell examples (Fig. 2A) is significantly different than the dampened oscillating shape of the mouse piezoelectric examples (Fig. 2F). However, within each set of examples, the general shape and timing of startle responses are consistent with variations in magnitude as see in Fig. 2A.

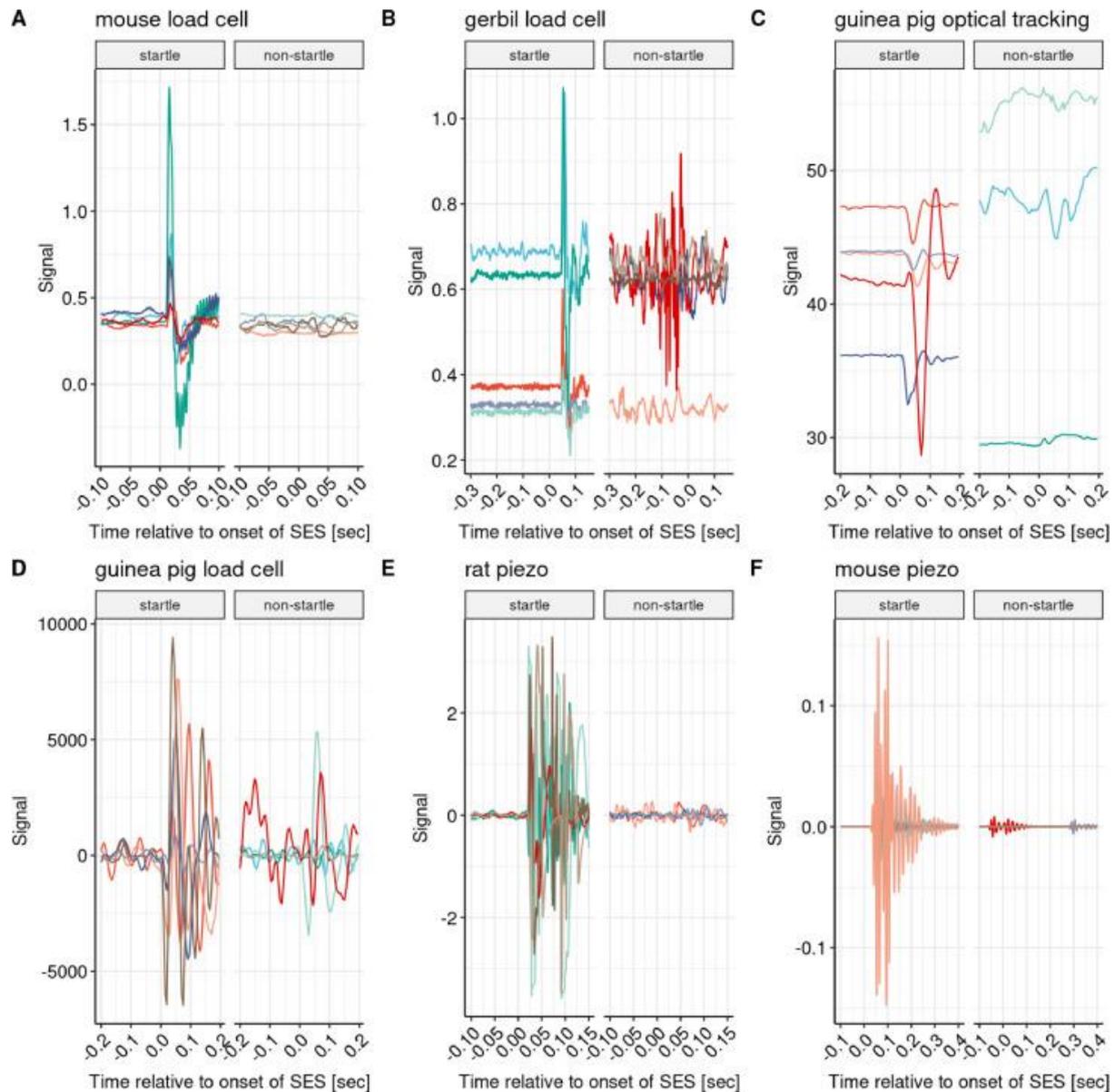


Fig. 2. Examples of startle and non-startle waveforms from four animal models (mouse, rat, gerbil and guinea pig) and three types of startle modalities (piezoelectric, load cell, and optical tracking). Different colors in each plot indicate separate trials. DC offset in panels B and C is due to the baseline offset (B: body mass and C: unit of signal measurement here is distance between two ears in mm).

3.2. Waveform normalization

Fig. 3 shows the application of the various normalization methods described above on the same representative SR waveforms presented in Fig. 2B obtained from gerbils tested on a load cell transducer. Normalization methods which center and scale the waveforms using the mean and standard deviation shown in Fig. 3B and 3E remove any offset in the SR waveform introduced by the measurement technique, allowing SR waveforms from different animals with different weights to be compared using the same features. Normalization methods which scale the waveforms such that the maximum peak amplitude is equal to unity (Fig. 3C) or that the signals range exactly between ± 1 (Fig. 3D) ensure all waveforms, true-startles and non-

startles, have similar maximum peak magnitudes without regard to variability in the pre-SES waveform. All SR waveforms presented in this paper were normalized using the mean and standard deviation from the pre-SES region, which provided extremely robust normalized waveforms for subsequent feature extraction since normalization. This method yields normalized waveforms representing the number of standard deviations from the mean of the pre-SES waveform, significantly magnifying true startles presented well above the pre-SES noise while significantly reducing noisy non-startle signal amplitudes. Each normalization method has strength and weaknesses with the ultimate choice of method governed by trained machine learning model performance and generalization to previously unseen examples.

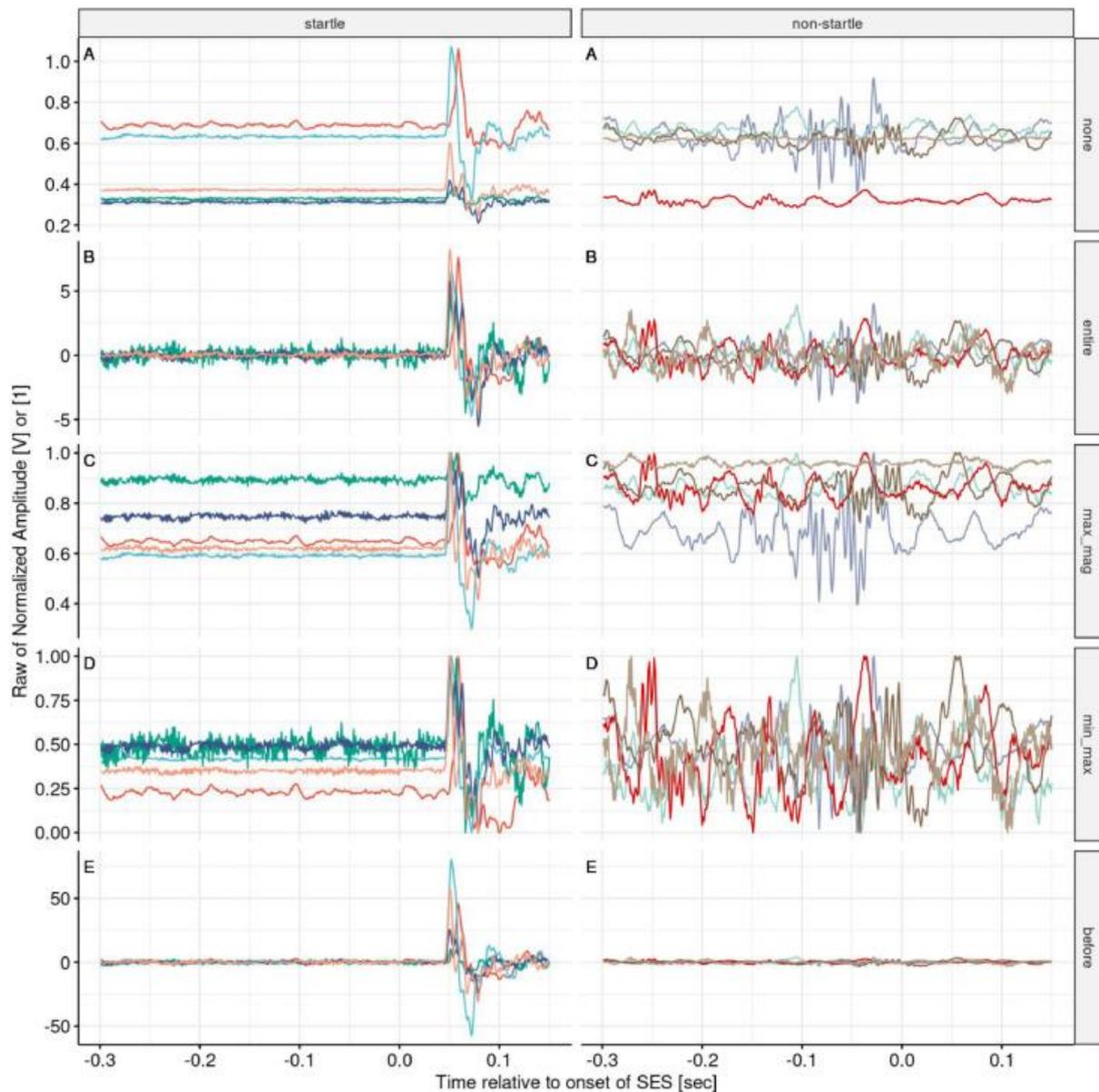


Fig. 3. Raw and normalized startle waveforms collected from the gerbil in Fig. 2B; data obtained with a load cell transducer. True-startle waveforms on the left and non-startle waveforms on the right. Startle and non-startle waveforms, acquired from a startle input-output series. (A) raw startle waveforms, (B) Z-score normalization using the mean and standard deviation of the entire waveform, (C) maximum-magnitude normalization, (D) minimum-maximum normalization, and (E) Z-score normalization

using the mean and standard deviation from before the SES was presented ($t < 0$). Different colors in each plot indicate separate trials.

3.3. Feature engineering

Highly predictive features must be extracted from the normalized SR waveforms to train robust machine learning models. The maximum magnitude and time of the maximum magnitude were extracted from the normalized SR waveforms for each species and startle modality dataset. The distributions of these waveform derived features from gerbil SR waveforms measured via a load cell transducer are presented in Fig. 4A. The maximum normalized magnitudes are significantly higher for startles versus non-startles, most of which occur at 0.05 s after the SES. However, there are still example normalized SR waveforms where both true-startles and non-startles have similar maximum magnitudes, less than $10^{0.75}$ and occur prior to the majority of the startle examples. Thus, although these waveform derived features are highly predictive, additional features are required to separate startles from non-startles which have similar maximum magnitude and time of maximum magnitude.

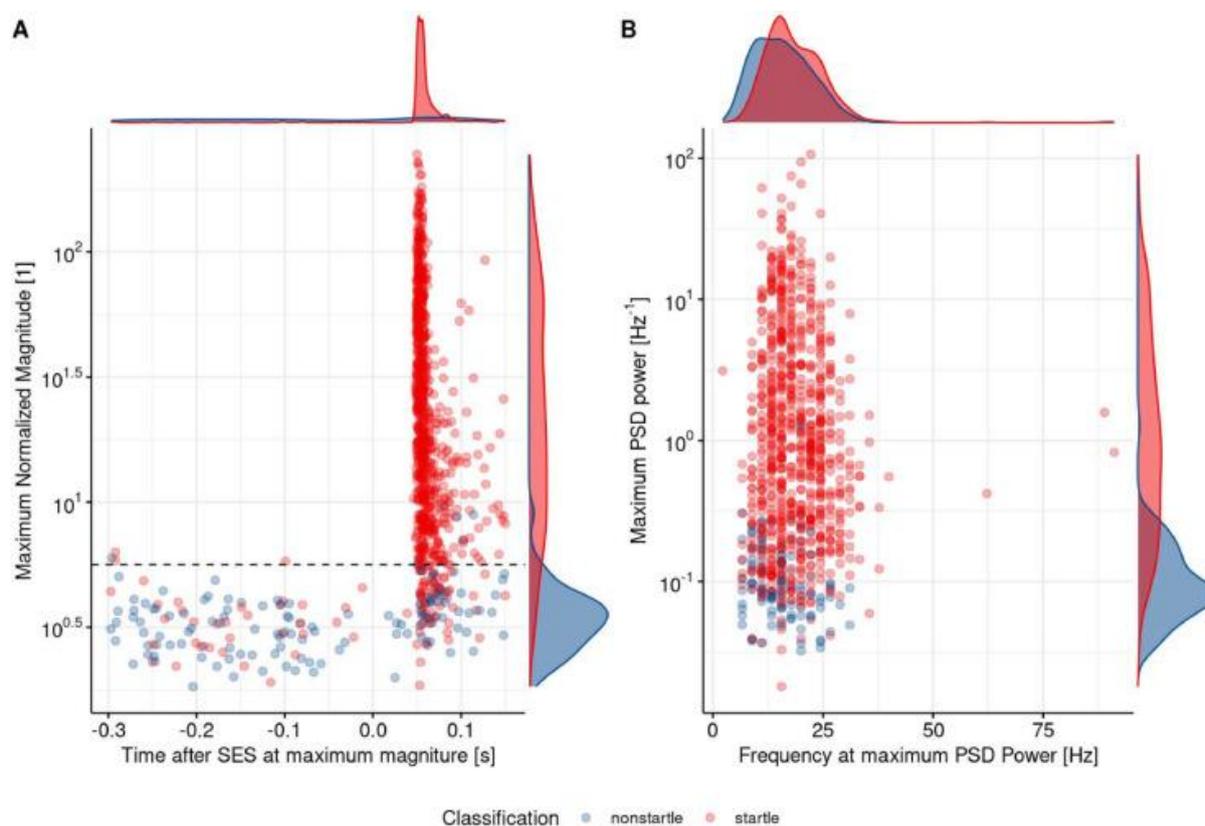


Fig. 4. Distribution of (A) waveform and (B) power spectral density features for all 1000 gerbil startle waveforms measured using a load cell transducer; data normalized using the mean and standard deviation before the startle eliciting sound was presented. Note that there is a strong distinction between startles and non-startles based on the maximum normalized magnitude of the waveform, with the majority of startles tightly clustered at 0.05 s after the SES and with magnitudes greater than $10^{0.75}$ (above the black dotted line). The PSD features (panel B) depict a significant separation between startles and non-startles in regards to maximum power (y-axis), but not in the frequency domain (x-axis).

All example SR waveforms presented in Fig. 2 show periodic temporal behavior which increases in magnitude immediately after the SES is presented for true startles. Thus, highly predictive features can be extracted from the power spectral density estimate (PSD) of the normalized SR waveforms including the maximum PSD power and frequency of the maximum PSD power. Fig. 4B shows the distributions of the maximum PSD magnitudes and the frequency at which they occurred for the same normalized SR waveforms obtained from gerbil via a load cell transducer whose waveform derived features are presented in Fig. 4A. Although most frequencies at maximum PSD power occur between 5 and 35 Hz with no discernible separation between startles and non-startles, there is significant separation in maximum PSD power with most non-startles being less than $10^{-0.5}\text{Hz}^{-1}$ and most startles being greater than 10^{-1}Hz^{-1} .

The power spectral density computes the distribution of power across the frequency components that make up the signal under test across the entire duration of the signal. Thus, the PSD provides no temporal information about the time course of the power computed from the PSD. However, the continuous wavelet transform allows for temporally resolved frequency content to be computed for all SR waveforms as a function of time. Continuous wavelet transform spectra from six representative SR waveforms, three true-startles and three non-startles, from each species/startle modality data set are shown in Fig. 5. The CWT spectra for all species/startle modalities presented in Fig. 5 show maximum CWT power for true-startles a short period of time after the SES is presented, while the magnitude of the maximum CWT power is generally lower for non-startles and the timing of maximum CWT power is highly variable.

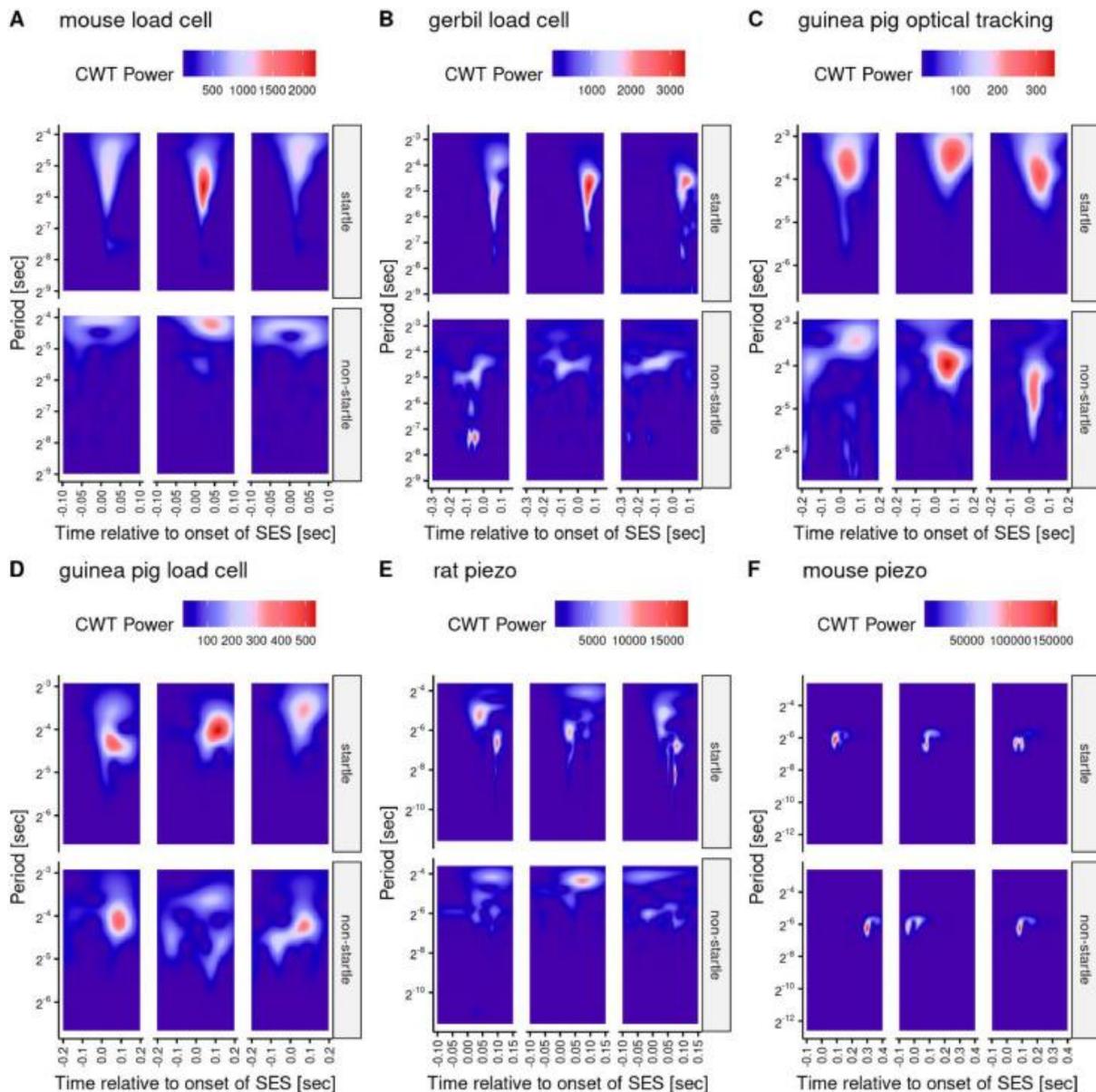


Fig. 5. Examples of continuous wavelet transform (CWT) power spectra of the startle waveforms from the six data sets, (A) mouse load cell, (B) gerbil load cell, (C) guinea pig optical tracking, (D) guinea pig load cell, (E) rat piezo, and (F) mouse piezo. Each panel contains six randomly selected trials, three true-startles and three non-startles, across multiple animals, sessions, and startle elicitor intensities over a 400 ms of acquisition time. Note the scale bar of the normalized power varies across the data sets.

Detailed analysis of the CWT spectra of each SR waveform results in many CWT derived features with the extraction of seventeen features reported in our initial report (see Table 1 and Fig. 9 in Fawcett et al., 2020). These features can be separated into general CWT features which apply to all CWT spectra of SR waveforms without regard to the timing and/or period of high CWT power (McKearney and MacKinnon, 2019). Fig. 6 shows the distribution of six general CWT features from the gerbil/load cell SR waveform data set. Fig. 6A shows the maximum CWT power after the SES is generally higher for startles versus non-startles with no differences in mean CWT power over the entire CWT spectra. The time of maximum CWT power after the SES is presented occurs between 0.05 and 0.1 s

for most startles and appears to be uniformly distributed across all post-SES times in Fig. 6B. Fig. 6C shows that most startles have lower mean CWT power before the SES compared to non-startles while also having higher mean CWT power over the time period before the SES to the maximum CWT power.

Table 1. Continuous wavelet transform features of interest for the gerbil startle waveforms presented in Fig. 7.

CWT Feature of Interest	Time Range [s]	Period [s^{-1}]	CWT Measure
Well before startle	(-0.26, -0.24)	NA	mean
During startle	(0.04, 0.06)	NA	mean
Around startle	(-0.02, 0.125)	NA	mean
After startle	(0.14, 0.149)	NA	mean
Low frequency	NA	$2^{-4.5}$	mean
High frequency	NA	$2^{-8.5}$	mean

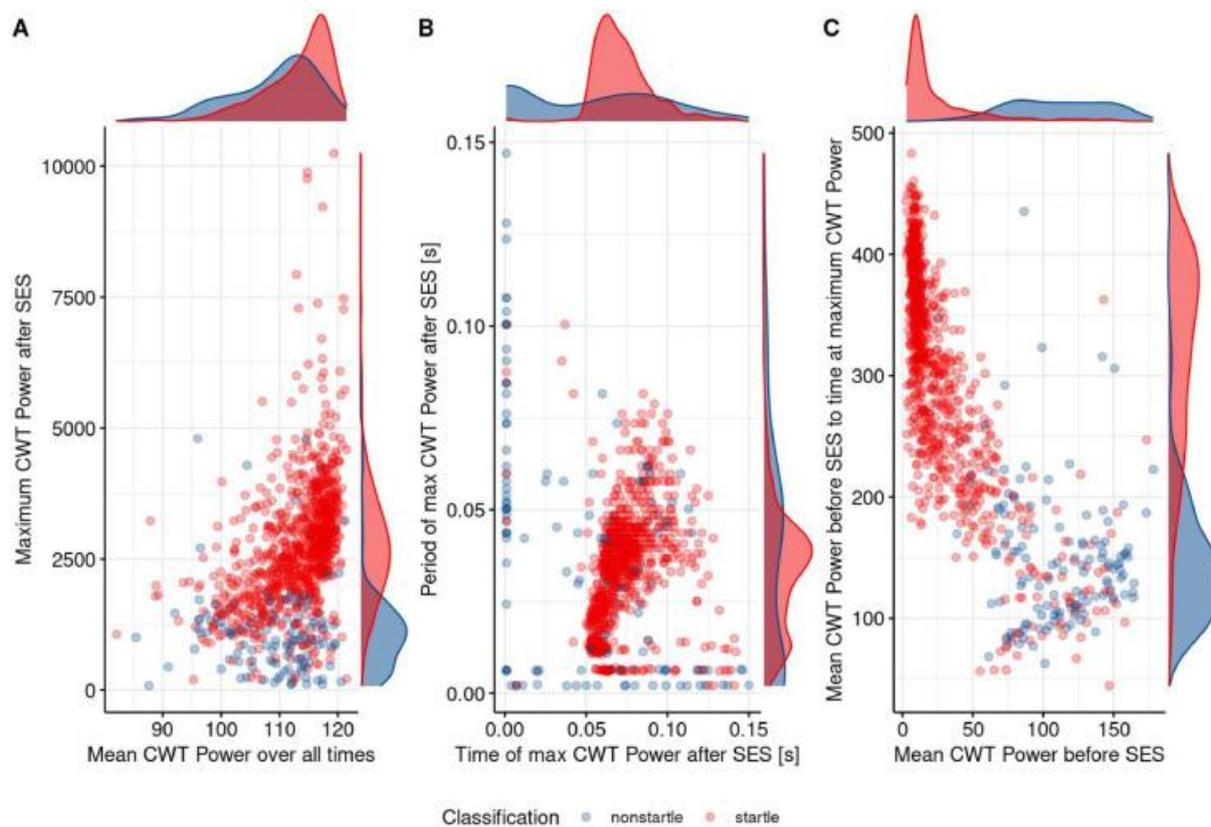


Fig. 6. Distribution of general continuous wavelet transform (CWT) features for all 1000 gerbil startle waveforms measured using a load cell transducer. Panel A depicts the maximum CWT power after the SES, with startles exhibiting a larger maximum power and a tighter spread towards greater time elapsed relative to the SES than non-

startles. Panel B portrays the relationship between the duration of a maximum in the CWT occurring as a function of time elapsed after the SES. Panel C illustrates the association between the difference of CWT power before the SES versus the maximum CWT power as a function of the mean CWT power before the SES.

The general continuous wavelet transform features described above and presented in Fig. 6 provide only initial and general insights related to the time-resolved frequency content of SR waveforms. Analysis of each panel in Fig. 5 for each species/startle modality reveals that the location, both the timing and period (frequency) of the high-power regions of the CWT spectra, is dependent on species and measurement modality. A detailed exploratory analysis of the CWT of representative startle and non-startle SR waveforms is required to determine the times (or time ranges) and period (or period ranges) of regions of high (or low) CWT power. Fig. 7 presents the CWT of six of the representative gerbil startle waveforms from Fig. 2B whose CWT spectra are also presented in Fig. 5B. Inspection of many example CWTs of the gerbil startle waveforms allowed the selection of several CWT features of interest with time ranges determined for: 1) well before the startle when any SR waveform activity is due to animal movement, 2) around the peak amplitude of the startle encompassing all high power CWT activity due to the SES, 3) during the startle which captures the initial response to the SES and for the frequency domain: 4) a high frequency component capturing all high power CWT activity due to the SES, and 5) a low frequency component where CWT power is lower for startles versus noisy non-startles (required a \log_{10} transform to visualize). A measure of the CWT power (mean and/or max) was extracted from the CWT spectra of each SR waveform. The CWT specialized features of interest shown in Fig. 7 are also summarized in Table 1.

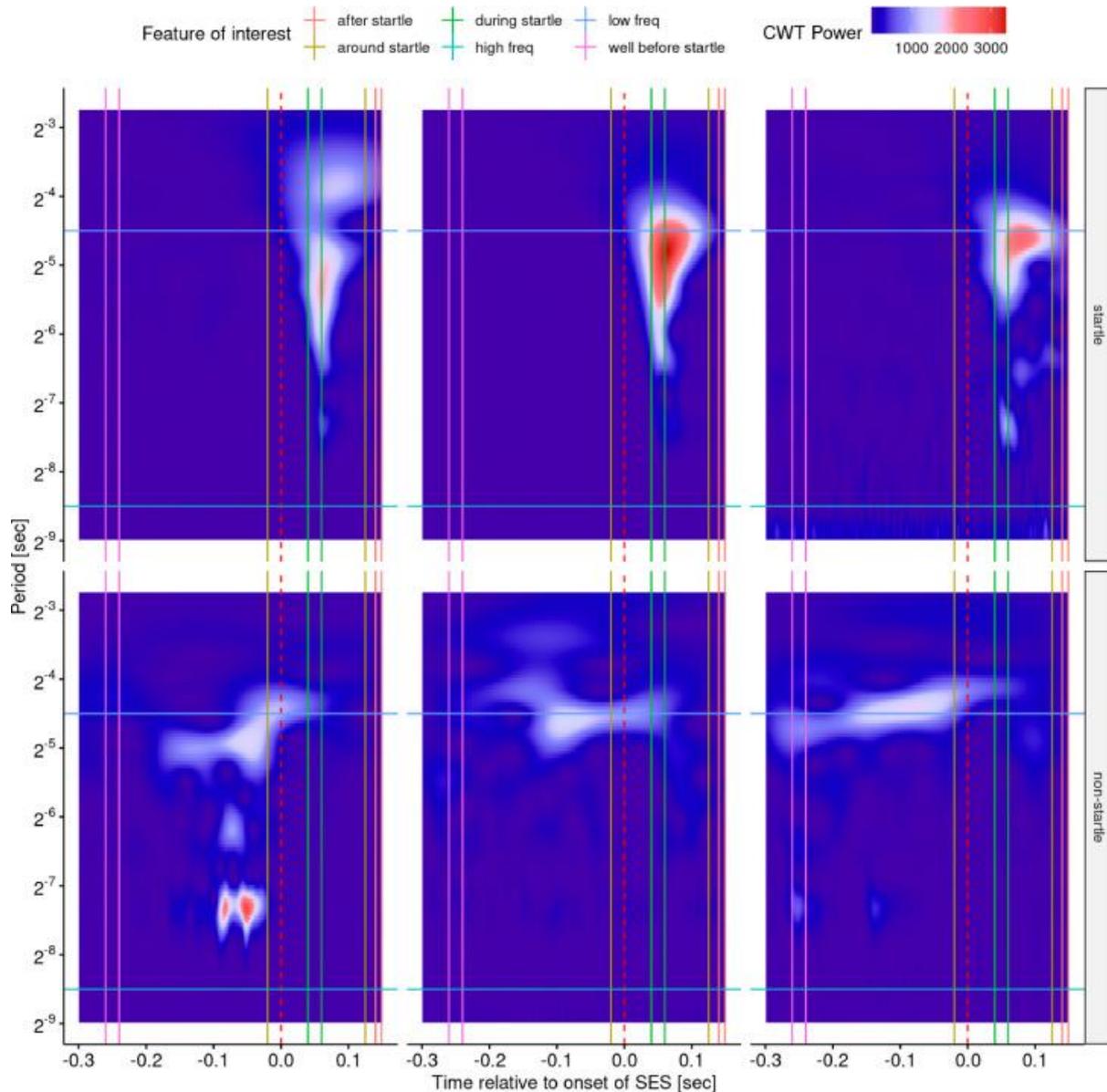


Fig. 7. CWT of six of the representative startle waveforms from gerbil using a load cell transducer presented in Fig. 2B with raw CWTs presented in Fig. 5B. The red dashed line shows when the SES is presented ($t = 0$), the vertical lines show time ranges of interest while the horizontal lines show periods of interest in which the mean CWT power at or within these regions are highly-predictive features.

The specialized CWT features using the time range (or period) described in [Table 1](#) are shown in Fig. 8. Panel A plots the mean CWT power before the startle onset as a function of the maximum CWT during the startle and indicates a significant difference between startles and non-startles, with non-startles exhibiting a higher overall CWT power well before the SES and during the startle response. Panel B highlights an increase in mean CWT power around the startle for startles compared to non-startles. Panel C shows that the mean CWT power at both high and low frequencies between startles and non-startles is similar.

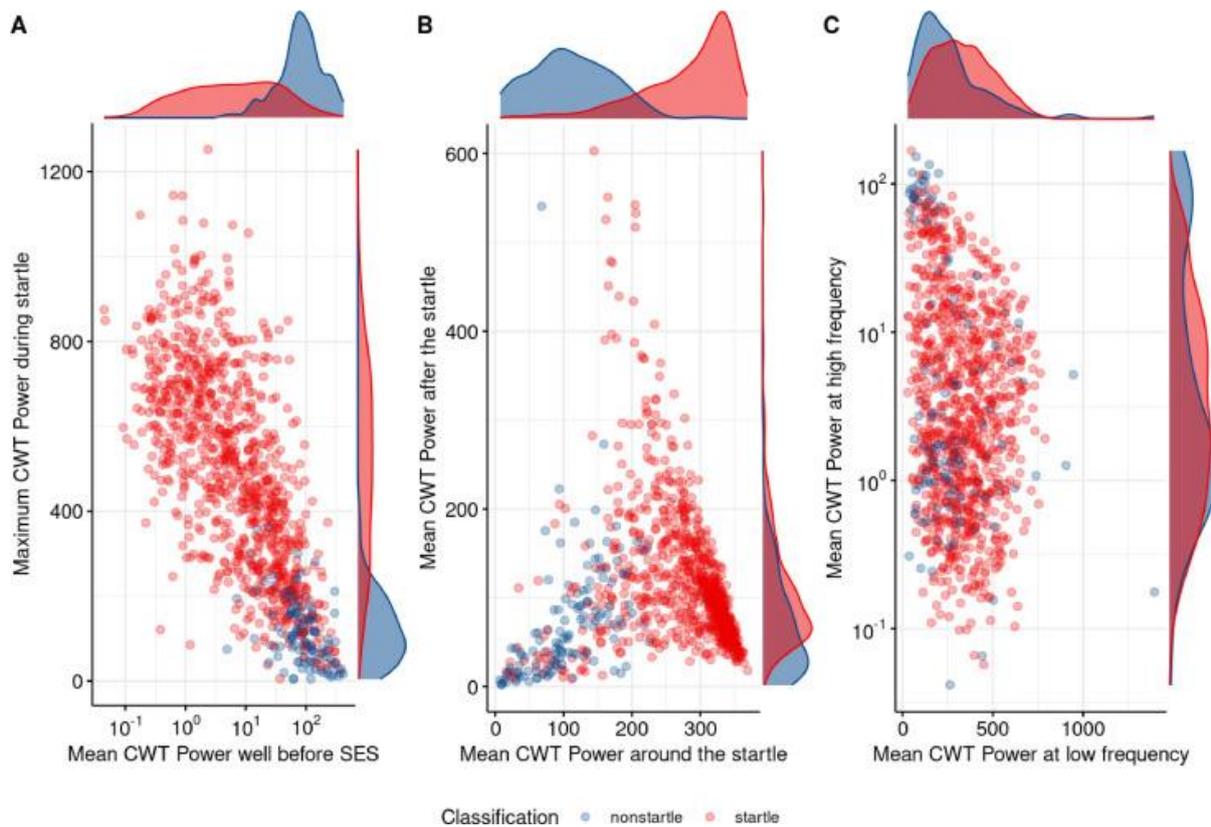


Fig. 8. Distribution of the specialized continuous wavelet transform features for all 1000 gerbil startle waveforms measured using a load cell transducer with time ranges and periods of interest described in Table 1. (A) the maximum CWT power during the startle (highest frequency content) versus the mean CWT power well before the SES (only animal movement), (B) mean CWT power after the startle (back to baseline activity) versus around the startle (contains all startle activity), and (C) mean CWT power at high frequency (where large and fast changes occur) versus at low frequency.

3.4. Machine learning model training and performance

After feature engineering, all features extracted from the SR waveforms including the waveform derived, PSD derived and CWT derived features were partitioned into 80% training samples and 20% test samples. Several machine learning models from various families, Random Forests and Support Vector Machine, are then trained using the training samples. The individual machine learning models are then combined via a generalized linear model to create a robust startle classifier (Fawcett, et al. 2020). Once the ensembled model is trained, new SR waveforms were classified as startles or non-startles using the exact same features as the trained ensembled model. The prediction accuracies of the ensembled models for each species/startle modality are presented in Table 2. The prediction accuracy for most species/startle modalities were well over 0.9 and greater than 0.95 for mouse/load cell, rat/piezoelectric, and mouse/piezoelectric. These high prediction accuracies for machine learning models separately trained for each species/startle modality data set demonstrate the robust and general nature of the machine learning startle classification procedure presented in this paper.

Table 2. Prediction accuracies of ensembled machine learning models for each species/startle modality data set.

Species	Startle Modality	Test Accuracy
Mouse	load cell	0.955
Gerbil	load cell	0.920
Rat	piezoelectric	0.960
Mouse	piezoelectric	0.980

Learning curves were evaluated to investigate the effect the number of training examples on trained machine learning model performance metrics. The number of training examples varied from 10 to 700 with model metrics evaluated using 200 previously unseen test samples. All training and test samples were stratified by startle label so the same proportion of startles/non-startles is present in all samplings. Fig. 9 shows the test accuracy (Panel A) and area under the received operating characteristic curve (AUC) (Panel B) as the number of training examples increases for each species and measurement method. In general, both test accuracy and AUC increases with the number of training examples for all species and measurements methods. This is expected as machine learning models demonstrate improved learning as the number of training examples increases and generalize better to unseen examples as the variety of those training also increases (Balki et al., 2019). Although 1000 SR waveforms for each species and measurement method were manually classified with 800 used to train the machine learning models presented in Table 2, the learning curves shown in Fig. 9 show that significantly less samples are required to obtain satisfactory model performance. The AUC must be greater than 0.5, when the machine learning model is no better than random guessing and not able to adequately distinguish between startles and non-startles, with higher AUC leading to better machine learning model performance (Fawcett, 2006). For example, the gerbil load cell median AUC is 0.5 until greater than 200 training examples are used in which the AUC increases to over 0.95, allowing for 200 versus 800 used to train successfully machine learning models if a test accuracy of slightly less than 0.9 is acceptable. However, training machine learning models with minimal examples can lead to increased variability in model performance as well as significantly decreased model generalization. Thus, we recommend using as many training examples as possible to reduce model performance variability and ensure those training examples represent as many types of examples as possible to ensure the model generalizes to previously unseen examples.

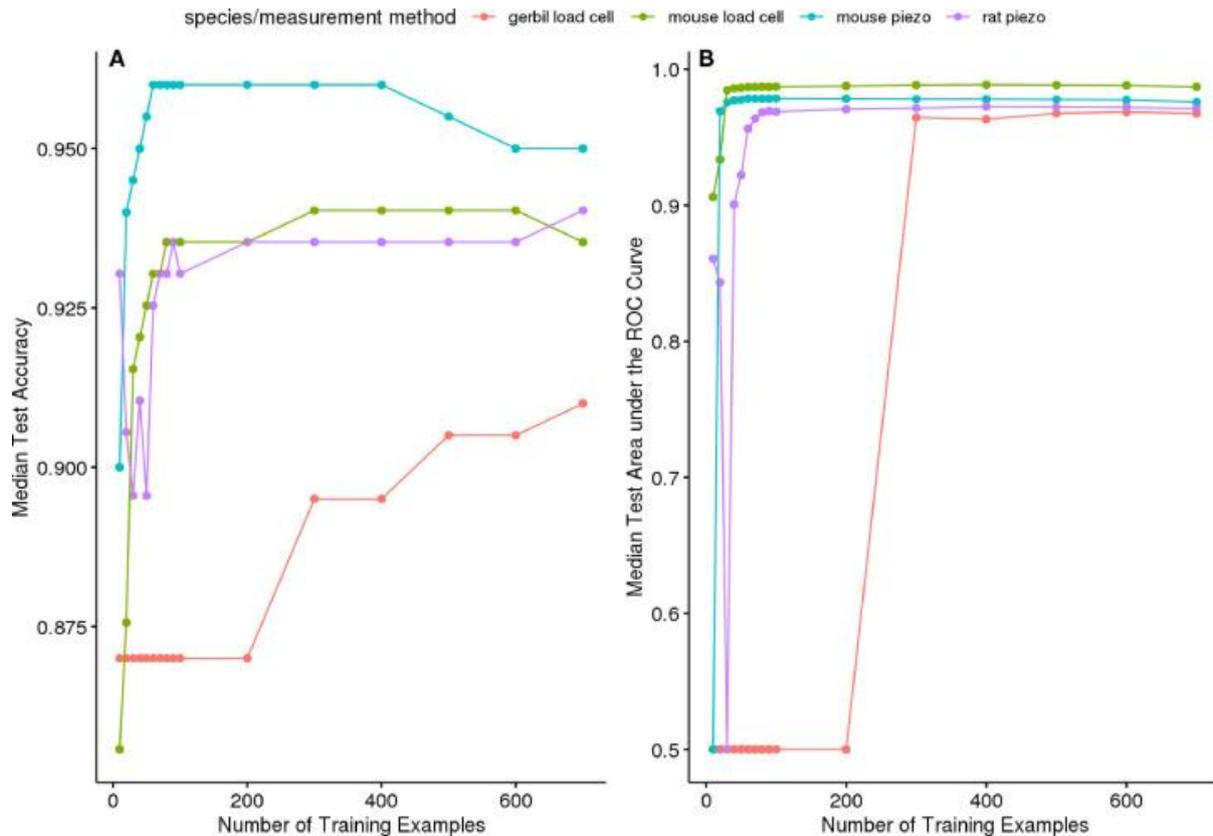


Fig. 9. Learning curve showing the (A) accuracy and (B) area under the receiver operating characteristics (ROC) curve for 200 holdout test examples stratified by startle label for each species and measurement method.

4. Discussion

There are several existing methods for SR classification including thresholding, RMS-based, and template matching classification methods presented in Grimsley et al., 2015 and the invalid trials method presented in Schilling et al. 2017. These methods typically classify SR waveforms as true startles at a much higher rate than manual classification (see Table 3 in Fawcett et al., 2020) resulting in significant inaccuracies in mean amplitude computations in the majority of datasets using these classification methods due to the large number of low amplitude non-startles being classified as startles, thus lowering mean startle amplitudes used for important behavioral measures such as pre-pulse inhibition (see Figs. 11, 12, and 13 in Fawcett et al., 2020). The approach described in this report allows machine learning algorithms to combine the mathematical definitions of human interpretable features to classify SR waveforms as startles or non-startles. Machine learning models are trained with many examples of both startle and non-startle waveforms demonstrating the required flexibility to use all information, via features, when classifying waveforms versus the rigid thresholding or template requirements imposed by the other SR classification methods discussed above. Another recent publication in the field of auditory assessments used a similar machine learning approach to objectively classify auditory brainstem responses (McKearney and MacKinnon, 2019). Taken together, these works demonstrate the utility of using machine learning to quickly and accurately (SR: up to 98%; ABR: up to 92.6%) analyze large sets of data. Data analysis will also benefit from reduced bias, especially for near-threshold waveform analysis. For these reasons, machine learning models have and will continue to help standardize data across laboratories and fields.

An analysis of the classification accuracy of the machine learning method approach presented in this paper compared to the threshold, RMS, and invalid trials methods for each species and measurement method is presented in Fig. 10 via receiver operating characteristic (ROC) curves (Obuchowski and Bullen, 2018). Panels A through E show ROC curves for each classification method across each species and measurement method via a cutoff analysis (Thiele and Hirschfeld, 2020; NCSS Statistical Software, 2022) with the independent variables used to generate each curve being the startle prediction probability for the machine learning method. Here we show the difference in maximum post- and pre-SES magnitudes for the threshold method (> 0 indicating a startle), the difference in root mean square of up to 100 ms of the post- and pre-SES signals for the RMS method (> 0 indicating a startle), and the maximum pre-SES magnitude for the invalid trials method ($< \text{threshold}$ in a startle). These ROC curves show the machine learning method is the most robust classifier with the largest area under the curve with the invalid trials method being the least accurate with many ROCs close to the chance line. To explore the overall performance of the various methods independent of species and measurement method, the area under the ROC curve (AUC) was bootstrapped with 100 replicates of 150 randomly selected samples and presented in Panel F. Fig. 10F shows the AUC of the machine learning method being notably higher with substantially less variability than the other classification methods.

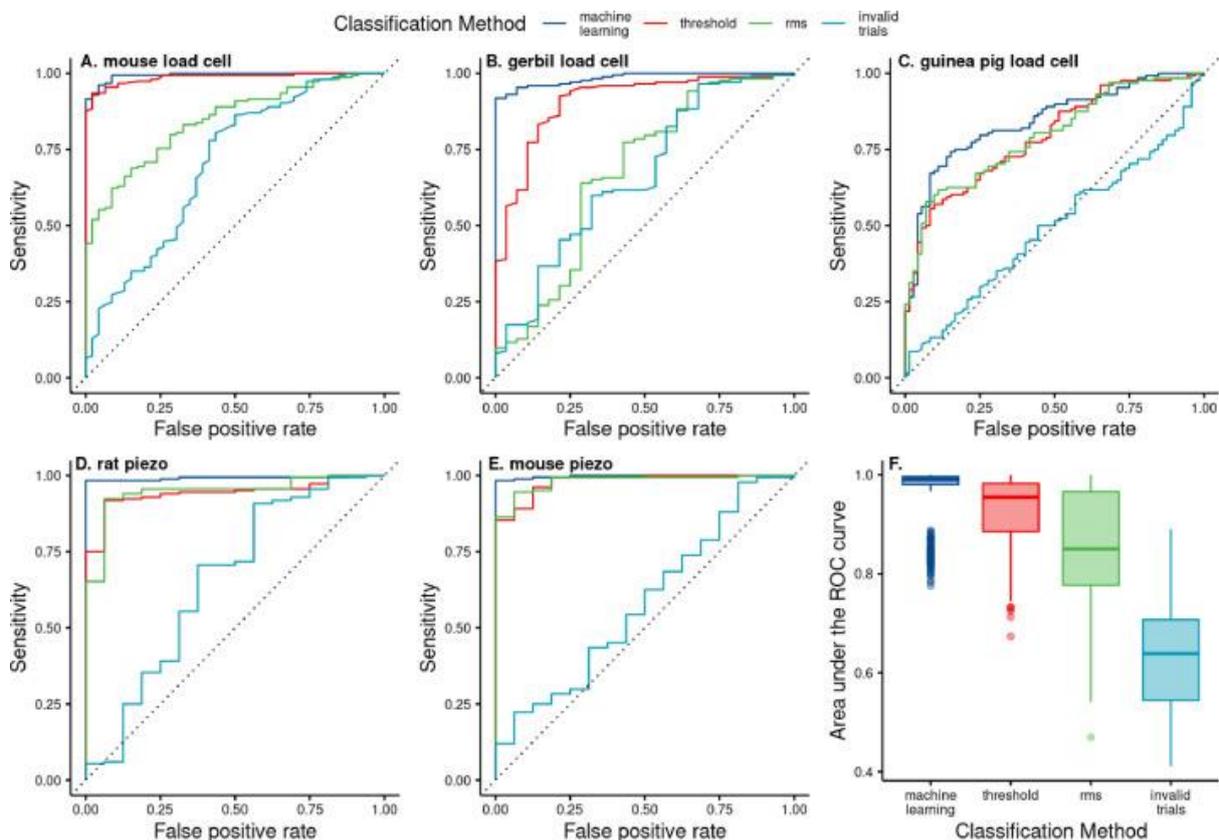


Fig. 10. Receiver operating characteristic (ROC) curves for the machine learning method presented in this paper, the threshold method and RMS method presented in Grimsley et al. (2015) and the invalid trials method presented in Schilling et al. (2017) for several species and measurement methods: (A) mouse load cell, (B) gerbil load cell, (C) guinea pig load cell, (D) rat piezoelectric transducer, and (E) mouse piezoelectric transducer with (F) showing the bootstrapped area under the ROC curve (AUC) for each method independent of species and measurement method.

The machine learning approach presented in this paper can be easily adapted for any startle reflex data through the use of the *asrclassify* R package. The *asrclassify* package contains functions to load SR data from a Matlab.mat file containing a StartleData structure or a variety of comma separate value (csv) file formats including the Kinder Scientific format or generic csv file formats with either one waveform per row or per column. SR waveforms required for feature engineering and machine learning model training must have manual classification labels such as 1 for startle and 0 for non-startle or similar classification. After SR data has been loaded, the raw and/or normalized waveforms can be explored to determine the optimum waveform normalization method. Next, the continuous wavelet transform (CWT) of representative startles and non-startles can be explored to determine the time ranges, periods, and CWT measures for CWT features of interest (see Fig. 7 and Table 1 for examples). Then, features can be extracted from raw and/or normalized SR waveforms as well as their power spectral density estimate and CWT. After the feature dataset is collected and partitioned into training and testing datasets, machine learning models are trained on the training set with hyperparameter tuning with the best model from each machine learning method stacked to form a robust ensemble model. After verification of the performance of the ensemble machine learning model using the testing dataset, the model can then be used to predict whether new SR waveforms are startles or non-startles. Specific details about implementing the *asrclassify* package to build a machine learning model and classify new SR waveforms are thoroughly described in the vignettes of the *asrclassify* package. The SR is ubiquitous in behavioral assays of the mammalian nervous system and has been used in the fields of pharmacology, psychology, and neuroscience to assess the internal state of rats, mice, gerbil and guinea pig. However, the acoustic startle response waveform has not been extensively characterized across species, and the lack of a standardized method for quantifying the startle response has led to high variation in startle data across different studies and laboratories. The automated classification method using machine learning reported by Fawcett et al. (2020) was specifically developed for CBA/CaJ mice with SR waveforms collected on piezoelectric transducers. In this report, the machine learning approach to classify SR waveforms as startles or non-startles was generalized for use with SR waveforms from virtually any species and startle modality using the *asrclassify* R package. Using this machine learning approach on startle reflex waveforms measured from any species or startle modality will allow unified startle classification as well as significantly improve data reliability and translatability between laboratories conducting startle reflex behavioral research.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

The *asrclassify* source code is available at <https://gitlab.com/waltonlab/asrclassify>. The code can be accessed and used by readers without restriction.

Funding

This work was supported by the National Institutes of Health AG009524 to JPW, NIDCD R01 DC000937 to AVG, MRC U135097126 to JIB, RNID grant TRIH 2018 to MNW and NIDCD R01 DC013314 to MJR.

CRediT authorship contribution statement

Timothy J. Fawcett: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Ryan J. Longenecker:** Conceptualization, Validation, Resources, Writing – original draft, Writing – review & editing. **Dimitri L. Brunelle:** Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Joel I. Berger:** Resources, Writing – review & editing. **Mark N. Wallace:** Resources, Writing – review & editing. **Alex V. Galazyuk:** Resources, Writing – review & editing. **Merri J. Rosen:** Resources, Writing – review & editing. **Richard J. Salvi:** Resources, Writing – review & editing. **Joseph P. Walton:** Conceptualization, Validation, Resources, Supervision, Funding acquisition, Resources, Writing – review & editing, Project administration, Funding acquisition.

Declaration of Competing Interests

The authors declare that they have no competing financial interests.

Data availability

Data will be made available on request.

References

- S.E. Ahmari, V.B. Risbrough, M.A. Geyer, H.B. Simpson **Impaired sensorimotor gating in unmedicated adults with obsessive-compulsive disorder.** *Neuropsychopharmacology*, 37 (5) (2012), pp. 1216-1223, [10.1038/npp.2011.308](https://doi.org/10.1038/npp.2011.308)
- J.B. Allen, L.R. Rabiner **A unified approach to short-time Fourier analysis and synthesis.** *Proc. IEEE*, 65 (11) (1977), pp. 1558-1564
- Balki, A. Amirabadi, J. Levman, A.L. Martel, Z. Emersic, B. Meden, A. GarciaPedrero, S.C. Ramirez, D. K ong, A.R. Moody, P.N. Tyrrell. **Sample-size determination methodologies for machine learning in medical imaging research: a systematic review.** *Can. Assoc. Radiol. J.*, 70 (4) (2019), pp. 344-353, [10.1016/j.carj.2019.06.002](https://doi.org/10.1016/j.carj.2019.06.002)
- J.I. Berger, B. Coomber, T.M. Shackleton, A.R. Palmer, M.N. Wallace. **A novel behavioural approach to detecting tinnitus in the guinea pig.** *J. Neurosci. Methods*, 213 (2) (2013), pp. 188-195, [10.1016/j.jneumeth.2012.12.023](https://doi.org/10.1016/j.jneumeth.2012.12.023)
- J. Buse, C. Beste, E. Herrmann, V. Roessner. **Neural correlates of altered sensorimotor gating in boys with Tourette Syndrome: a combined EMG/fMRI study** *World J. Biol. Psychiatry*, 17 (3) (2016), pp. 187

- J. Cai, J. Luo, S. Wang, S. Yang. **Feature selection in machine learning: a new perspective.** Neurocomputing, 300 (2018), pp. 70-79
- J.V. Cassella, M. Davis. **The design and calibration of a startle measurement system.** Physiol. Behav., 36 (2) (1986), pp. 377-383, [10.1016/0031-9384\(86\)90032-6](https://doi.org/10.1016/0031-9384(86)90032-6)
- J.V. Cassella, M. Davis. **Habituation, prepulse inhibition, fear conditioning, and drug modulation of the acoustically elicited pinna reflex in rats.** Behav. Neurosci., 100 (1) (1986), pp. 39-44, [10.1037//0735-7044.100.1.39](https://doi.org/10.1037//0735-7044.100.1.39)
- M. Davis. **The Mammalian Startle response, Neural mechanisms of Startle Behavior.** Springer (1984), pp. 287-351
- S.C. Dulawa, R. Hen, K. Scarce-Levie, M.A. Geyer. **Serotonin1B receptor modulation of startle reactivity, habituation, and prepulse inhibition in wild-type and serotonin1B knockout mice.** Psychopharmacology, 132 (2) (1997), pp. 125-134, [10.1007/s002130050328](https://doi.org/10.1007/s002130050328)
- T.J. Fawcett, C.S. Cooper, R.J. Longenecker, J.P. Walton. **Automated classification of acoustic startle reflex waveforms in young CBA/CaJ mice using machine learning.** J. Neurosci. Methods, 344 (2020), Article
- T.J. Fawcett, C.S. Cooper, R.J. Longenecker, J.P. Walton. **Machine learning, waveform preprocessing and feature extraction methods for classification of acoustic startle waveforms.** MethodsX, 8 (2021), Article 101166, [10.1016/j.mex.2020.101166](https://doi.org/10.1016/j.mex.2020.101166)
- M. Fendt, M. Koch. **Translational value of startle modulations.** Cell Tissue Res., 354 (1) (2013), pp. 287-295, [10.1007/s00441-013-1599-5](https://doi.org/10.1007/s00441-013-1599-5)
- A. Galazyuk, S. Hebert. **Gap-prepulse inhibition of the acoustic startle reflex (GPIAS) for tinnitus assessment: current status and future directions.** Front. Neurol, 6 (2015), p. 88, [10.3389/fneur.2015.00088](https://doi.org/10.3389/fneur.2015.00088)
- M.A. Geyer, N.R. Swerdlow, R.S. Mansbach, D.L. Braff. **Startle response models of sensorimotor gating and habituation deficits in schizophrenia.** Brain Res. Bull., 25 (3) (1990), pp. 485-498, [10.1016/0361-9230\(90\)90241-q](https://doi.org/10.1016/0361-9230(90)90241-q)
- D.B. Green, M.M. Mattingly, Y. Ye, J.D. Gay, M.J. Rosen. **Brief stimulus exposure fully remediates temporal processing deficits induced by early hearing loss.** J. Neurosci., 37 (32) (2017), pp. 7759-7771, [10.1523/JNEUROSCI.0916-17.2017](https://doi.org/10.1523/JNEUROSCI.0916-17.2017)
- D.B. Green, J. Ohlemacher, M.J. Rosen. **Benefits of stimulus exposure: developmental learning independent of task performance.** Front. Neurosci., 10 (2016), p. 263, [10.3389/fnins.2016.00263](https://doi.org/10.3389/fnins.2016.00263)
- C.A. Grimsley, R.J. Longenecker, M.J. Rosen, J.W. Young, J.M. Grimsley, A.V. Galazyuk. **An improved approach to separating startle data from noise.** J. Neurosci. Methods, 253 (2015), pp. 206-217, [10.1016/j.jneumeth.2015.07.001](https://doi.org/10.1016/j.jneumeth.2015.07.001)
- T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman. **The Elements of Statistical learning: Data mining, inference, and Prediction.** Springer (2009)

- M. Horlington. **A method for measuring acoustic startle response latency and magnitude in rats: detection of a single stimulus effect using latency measurements.** *Physiol. Behav.*, 3 (6) (1968), pp. 839-844
- J.R. Ison, H.S. Hoffman. **Reflex modification in the domain of startle: II. The anomalous history of a robust and ubiquitous phenomenon.** *Psychol. Bull.*, 94 (1) (1983), pp. 3-17
- A. Khan, S.B. Powell. **Sensorimotor gating deficits in "two-hit" models of schizophrenia risk factors.** *Schizophr. Res.*, 198 (2018), pp. 68-83, [10.1016/j.schres.2017.10.009](https://doi.org/10.1016/j.schres.2017.10.009)
- M. Koch. **The neurobiology of startle.** *Prog. Neurobiol.*, 59 (2) (1999), pp. 107-128, [10.1016/s0301-0082\(98\)00098-7](https://doi.org/10.1016/s0301-0082(98)00098-7)
- S.B. Kotsiantis, I.D. Zaharakis, P.E. Pintelas. **Machine learning: a review of classification and combining techniques.** *Artif. Intell. Rev.*, 26 (3) (2006), pp. 159-190
- K. Kraus, D. Ding, H. Jiang, E. Lobarinas, W. Sun, R. Salvi. **Relationship between noise-induced hearing-loss, persistent tinnitus and growth-associated protein-43 expression in the rat cochlear nucleus: does synaptic plasticity in ventral cochlear nucleus suppress tinnitus?** *Neuroscience*, 194 (2011), pp. 309-325
- M.J., K. Kuhn. **Applied Predictive Modeling.** Springer Science + Business Media, New York (2013)
- A.M. Lauer, D. Behrens, G. Klump. **Acoustic startle modification as a tool for evaluating auditory function of the mouse: progress, pitfalls, and potential.** *Neurosc. Biobehav. Rev.*, 77 (2017), pp. 194-208.
- E. Lobarinas, S.H. Hayes, B.L. Allman. **The gap-startle paradigm for tinnitus screening in animal models: limitations and optimization.** *Hear. Res.*, 295 (2013), pp. 150-160, [10.1016/j.heares.2012.06.001](https://doi.org/10.1016/j.heares.2012.06.001)
- E. Lobarinas, W. Sun, R. Cushing, R. Salvi. **A novel behavioral paradigm for assessing tinnitus using schedule-induced polydipsia avoidance conditioning (SIP-AC).** *Hear. Res.*, 190 (1-2) (2004), pp. 109-114.
- R. Longenecker, A. Galazyuk. **Methodological optimization of tinnitus assessment using prepulse inhibition of the acoustic startle reflex.** *Brain Res.*, 1485 (2012), pp. 54-62.
- R.J. Longenecker, F. Alghamdi, M.J. Rosen, A.V. Galazyuk. **Prepulse inhibition of the acoustic startle reflex vs. auditory brainstem response for hearing assessment.** *Hear. Res.*, 339 (2016), pp. 80-93, [10.1016/j.heares.2016.06.006](https://doi.org/10.1016/j.heares.2016.06.006)
- R.J. Longenecker, A.V. Galazyuk. **Development of tinnitus in CBA/CaJ mice following sound exposure.** *J. Assoc. Res. Otolaryng.*, 12 (5) (2011), pp. 647-658, [10.1007/s10162-011-0276-1](https://doi.org/10.1007/s10162-011-0276-1)
- R.J. Longenecker, I. Kristaponyte, G.L. Nelson, J.W. Young, A.V. Galazyuk. **Addressing variability in the acoustic startle reflex for accurate gap detection assessment.** *Hear. Res.*, 363 (2018), pp. 119-135, [10.1016/j.heares.2018.03.013](https://doi.org/10.1016/j.heares.2018.03.013)

A.S. Lowe, J.P. Walton. **Alterations in peripheral and central components of the auditory brainstem response: a neural assay of tinnitus.** PLoS One, 10 (2) (2015), Article e0117228, [10.1371/journal.pone.0117228](https://doi.org/10.1371/journal.pone.0117228)

R.M. McKearney, R.C. MacKinnon. **Objective auditory brainstem response classification using machine learning.** Int. J. Audiol., 58 (4) (2019), pp. 224-230

NCSS Statistical Software. 2022. One ROC curve and cutoff analysis.

N.A. Obuchowski, J.A. Bullen. **Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine.** Phys. Med. Biol., 63 (7) (2018), p. 07TR01, [10.1088/1361-6560/aab4b1](https://doi.org/10.1088/1361-6560/aab4b1)

M.M. Pantoni, G.M. Herrera, K.R. Van Alstyne, S.G. Anagnostaras. **Quantifying the acoustic startle response in mice using standard digital video.** Front. Behav. Neurosci., 14 (2020), p. 83

W.T. Preyer. **Die Seele des Kindes: Beobachtungen über Die Geistige Entwicklung des Menschen in Den Ersten Lebensjahren.** T. Grieben (1900).

A. Schilling, P. Krauss, R. Gerum, C. Metzner, K. Tziridis, H. Schulze. **A new statistical approach for the evaluation of gap-prepulse inhibition of the acoustic startle reflex (GPIAS) for tinnitus assessment.** Front. Behav. Neurosci., 11 (2017), p. 198, [10.3389/fnbeh.2017.00198](https://doi.org/10.3389/fnbeh.2017.00198)

E. Sejdić, I. Djurović, J. Jiang. **Time–frequency feature representation using energy concentration: an overview of recent advances.** Digit Signal Process, 19 (1) (2009), pp. 153-183

Thiele, C., Hirschfeld, G., 2020. cutpointr: improved estimation and validation of optimal cutpoints in R. arXiv preprint arXiv:[2002.09209](https://arxiv.org/abs/2002.09209).

J.G. Turner, T.J. Brozoski, C.A. Bauer, J.L. Parrish, K. Myers, L.F. Hughes, D.M. Caspary. **Gap detection deficits in rats with tinnitus: a potential novel screening tool.** Behav. Neurosci., 120 (1) (2006), pp. 188-195, [10.1037/0735-7044.120.1.188](https://doi.org/10.1037/0735-7044.120.1.188)

K. Veer, R. Agarwal. **Wavelet and short-time Fourier transform comparison-based analysis of myoelectric signals.** J. Appl. Stat., 42 (7) (2015), pp. 1591-1601.

J.R. Vergara, P.A. Estévez. **A review of feature selection methods based on mutual information.** Neural Comput. Applic., 24 (1) (2014), pp. 175-186.

D. Virag, J. Homolak, I. Kodvanj, A. Babic Perhoc, A. Knezovic, J. Osmanovic Barilar, M. Salkovic-Petrisic. **Repurposing a digital kitchen scale for neuroscience research: a complete hardware and software cookbook for PASTA.** Sci. Rep., 11 (1) (2021), p. 2963, [10.1038/s41598-021-82710-6](https://doi.org/10.1038/s41598-021-82710-6).

N. Wake, K. Ishizu, T. Abe, H. Takahashi. **Prepulse inhibition predicts subjective hearing in rats.** Sci. Rep., 11 (1) (2021), p. 18902, [10.1038/s41598-021-98167-6](https://doi.org/10.1038/s41598-021-98167-6)

T.Y. Zhang, P. Chretien, M.J. Meaney, A. Gratton. **Influence of naturally occurring variations in maternal care on prepulse inhibition of acoustic startle and the medial prefrontal cortical dopamine response to stress in adult rats.** J. Neurosci., 25 (6) (2005), pp. 1493-1502, [10.1523/JNEUROSCI.3293-04.2005](https://doi.org/10.1523/JNEUROSCI.3293-04.2005)