



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

Machine learning, waveform preprocessing and feature extraction methods for classification of acoustic startle waveforms



Timothy J. Fawcett^{a,b,c}, Chad S. Cooper^a, Ryan J. Longenecker^a,
Joseph P. Walton^{a,c,d,*}

^a Global Center for Hearing and Speech Research, University of South Florida, Tampa, FL, United States

^b Research Computing, University of South Florida, Tampa, FL, United States

^c Department of Chemical and Biomedical Engineering, University of South Florida, Tampa, FL, United States

^d Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL, United States

A B S T R A C T

The acoustic startle response (ASR) is an involuntary muscle reflex that occurs in response to a transient loud sound and is a highly-utilized method of assessing hearing status in animal models. Currently, a high level of variability exists in the recording and interpretation of ASRs due to the lack of standardization for collecting and analyzing these measures. An ensemble machine learning model was trained to predict whether an ASR waveform is a startle or non-startle using highly-predictive features extracted from normalized ASR waveforms collected from young adult CBA/CaJ mice. Features were extracted from the normalized waveform as well as the power spectral density estimates and continuous wavelet transforms of the normalized waveform. Machine learning models utilizing methods from different families of algorithms were individually trained and then ensemble together, resulting in an extremely robust model.

- ASR waveforms were normalized using the mean and standard deviation computed before the startle elicitor was presented
- 9 machine learning algorithms from 4 different families of algorithms were individually trained using features extracted from the normalized ASR waveforms
- Trained machine learning models were ensemble to produce an extremely robust classifier

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

DOI of original article: [10.1016/j.jneumeth.2020.108853](https://doi.org/10.1016/j.jneumeth.2020.108853)

* Corresponding author.

E-mail address: jwalton1@usf.edu (J.P. Walton).

<https://doi.org/10.1016/j.mex.2020.101166>

2215-0161/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

ARTICLE INFO

Method name: Acoustic Startle Response classification

Keywords: Acoustic startle reflex, Machine learning, Waveform preprocessing, Ensemble modeling

Article history: Received 1 September 2020; Accepted 27 November 2020; Available online 1 December 2020

Specifications Table

Subject Area:	Neuroscience
More specific subject area:	Acoustic Startle Reflex
Method name:	Acoustic Startle Response classification
Name and reference of original method:	Grimsley, C. A., Longenecker, R. J., Rosen, M. J., Young, J. W., Grimsley, J. M., & Galazyuk, A. V. (2015). An improved approach to separating startle data from noise. <i>Journal of Neuroscience Methods</i> , 253, 206–217. https://doi.org/10.1016/j.jneumeth.2015.07.001
Resource availability:	N/A

*Method details

Experimental procedure

Whole-body startle responses were elicited using acoustic startle elicitors from young adult CBA/CaJ mice. The details of the startle waveform acquisition, sampling and procedure are identical to those in Fawcett et al. [4]. Briefly, individual mice were placed in wire cages which rested on platforms

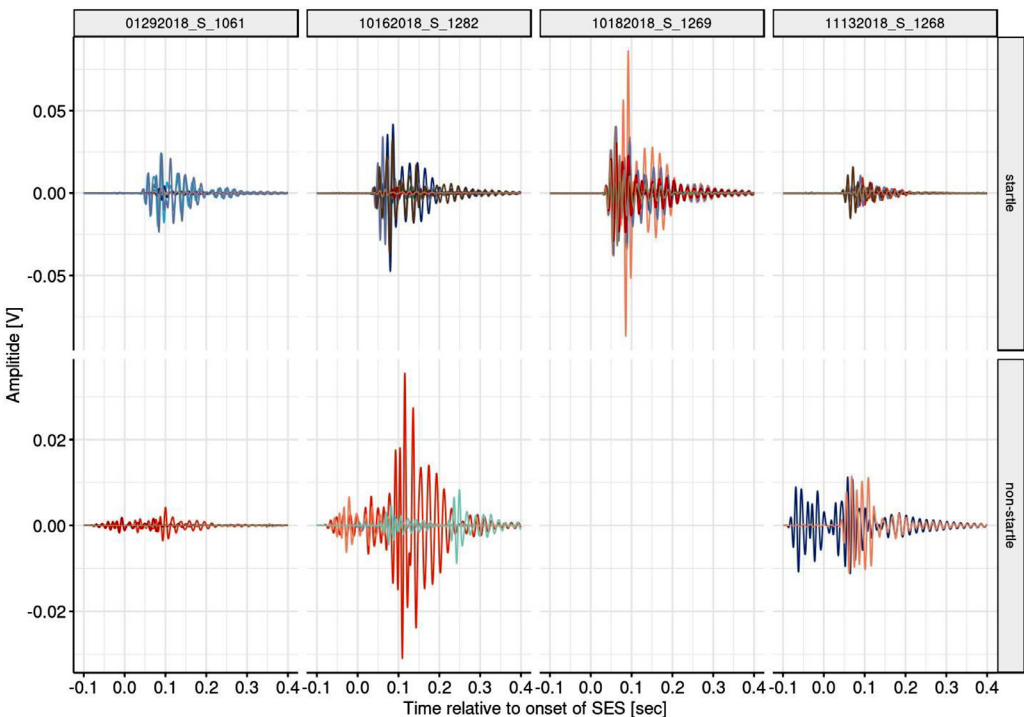


Fig. 1. Representative acoustic startle reflex waveforms from four mice and four randomly selected 85 dB SPL startle I/O sessions manually classified as startles (top) and non-startles (bottom). Data was acquired throughout the 500 msec of time shown on the x-axis. Note colors indicate different trials within the same session and panels without data indicate zero non-startles for that session and/or condition.

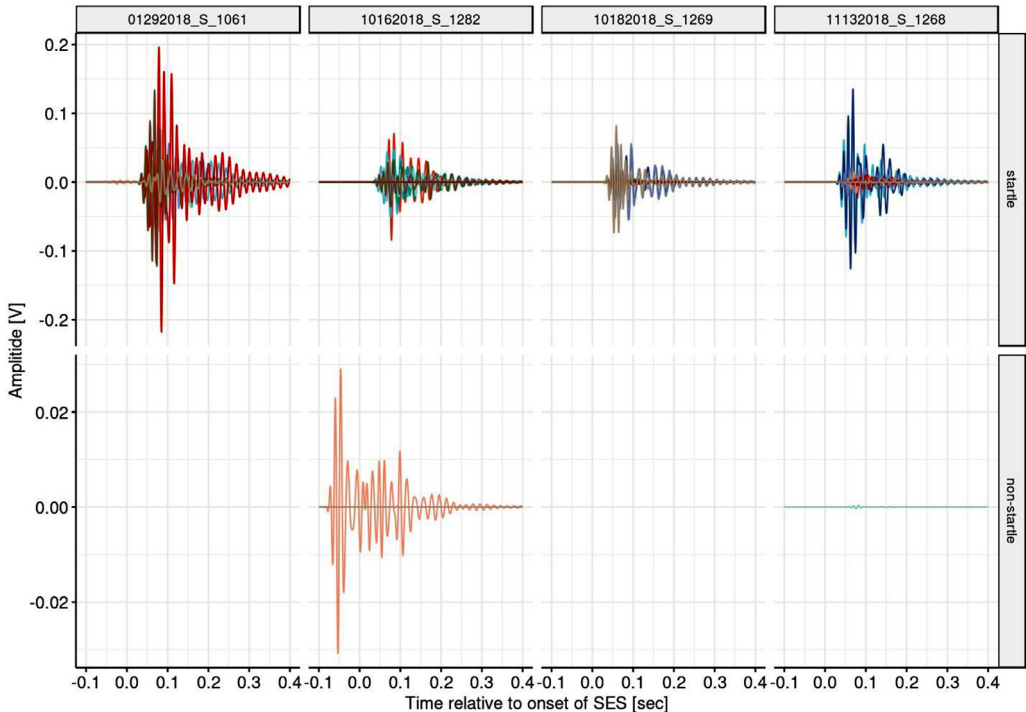


Fig. 2. Representative acoustic startle reflex waveforms from the same four mice and randomly selected startle I/O sessions presented in Fig. 1 with the SES presented at 95 dB SPL manually classified as startles (top) and non-startles (bottom). Note manual classification identified all ASR waveforms as startles for sessions 1061 and 1269 and colors indicate different trials within the same session and panels without data indicate zero non-startles for that session and/or condition.

with embedded piezoelectric transducers. Animals were given 5 min for acclimation inside the sound chamber before testing. Each animal received three testing sessions which included startle input-output that tested sensory motor gating and tone burst used in a pre-pulse inhibition measure (TPP) to test hearing sensitivity. A total of 6911 Acoustic Startle Reflex (ASR) measurements were collected from 40 young CBA/Caj mice (20 male/20 female) ranging in age from 2.3 to 6.2 months (mean age of 3.76 months) over a 14-month period. These ASR measurements were manually classified as startle (4214) or non-startle (2697) by two experienced behavioral neuroscientists.

Exploratory data analysis

The classification of ASR waveforms as startles versus non-startles requires the extraction of features with high predictive capability. To identify relevant features, many ASR waveforms have been visualized and explored. Fig. 1 shows ASR waveforms from four randomly selected startle I/O sessions with the SES presented at 85 dB SPL. All ASR waveforms classified as startles show similar morphology with little activity before the SES ($t = 0$ msec) with extremely large increases in activity after the SES is presented. Fig. 1 also shows the variation in startle amplitudes between mice and/or sessions as well as the variations in the shape of the ASR waveform. The ASR waveforms classified as non-startles generally exhibit much lower amplitudes when compared to startles within the same session. Non-startles also contain mostly noise associated with random animal movement, especially before the SES is presented when little to no activity is expected. This figure also shows only one session (10182018_S_1269) which resulted in all ASR waveforms being classified as a startle, a condition which

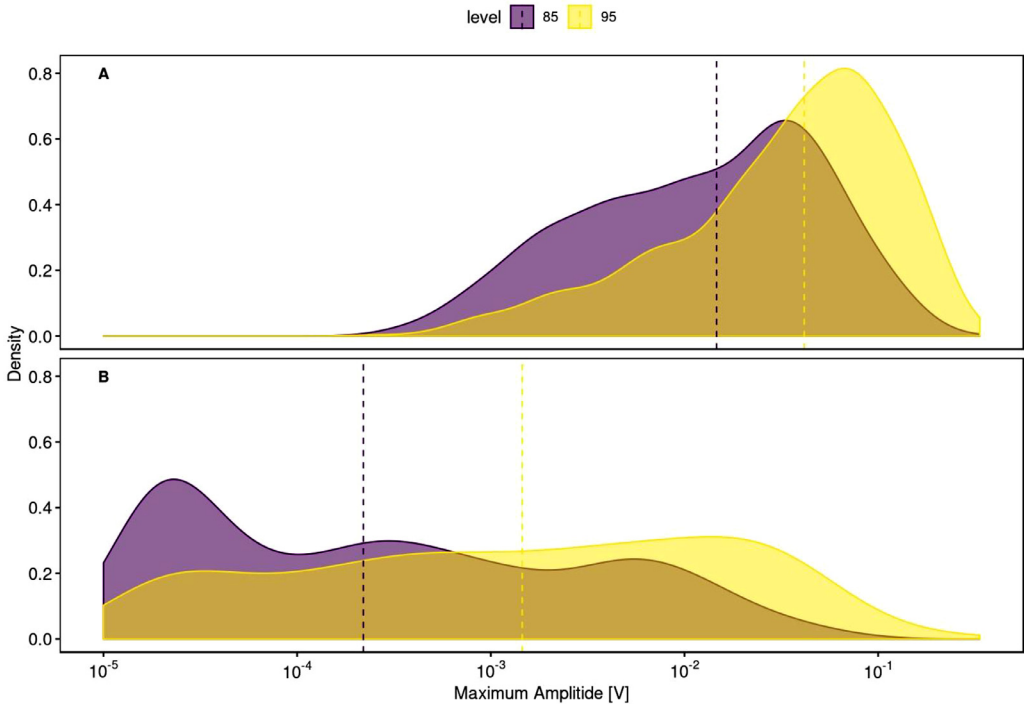


Fig. 3. Distributions of the maximum magnitudes of the Startle I/O ASR waveforms with the SES presented at both 85 dB SPL and 95 dB SPL for startles (panel A) and non-startles (panel B). Vertical dotted lines indicate the median maximum magnitude of each distribution.

occurred in only 14 of the 109 (12.84%) startle I/O ASR waveforms (SES at 85 dB SPL) analyzed in this study.

As the intensity of the startle elicitor increases, the number of true startles as well as the magnitude of the startle response is expected to increase. Fig. 2 shows ASR waveforms from the same four 85 dB SPL startle I/O sessions presented in Fig. 1 but with the SES presented at 95 dB SPL. The increase in SES level to 95 dB SPL resulted not only in an increase in the number of true startles, but also at least an order of magnitude increase in startle magnitude, when compared to the responses at 85 dB SPL. Note that the magnitudes of the ASR waveforms classified as non-startles with an SES of 95 dB SPL presented in Fig. 2 is approximate to the non-startle ASR waveforms presented in Fig. 1 elicited by the 85 dB SPL stimulus. The magnitude of the ASR waveforms is likely a feature with high predictive capability since the magnitude generally increases with SES level for startles but stays about the same for non-startles. The distribution of the magnitude of the ASR waveforms classified as startle (panel A) versus non-startle (panel B) is presented in Fig. 3. This analysis shows the maximum magnitude of startles is significantly higher on average than that of non-startles. In addition, the maximum magnitude for startle is greater for 95 dB SPL startles when compared to 85 dB SPL startles while the maximum magnitude of non-startles is more uniformly distributed.

In addition to the Startle I/O data presented above in Figs. 1 and 2, ASR waveforms measured in a tonal pre-pulse (TPP) protocol were also collected. Fig. 4 shows all trials from a single 16 kHz TPP session with no pre-pulse as well as at tones presented at levels ranging from 40 dB SPL to 70 dB SPL. In this representative example, all trials with no pre-pulse were classified as true startles and also have the largest magnitudes, an expected result given the SES is presented at 110 dB SPL. In

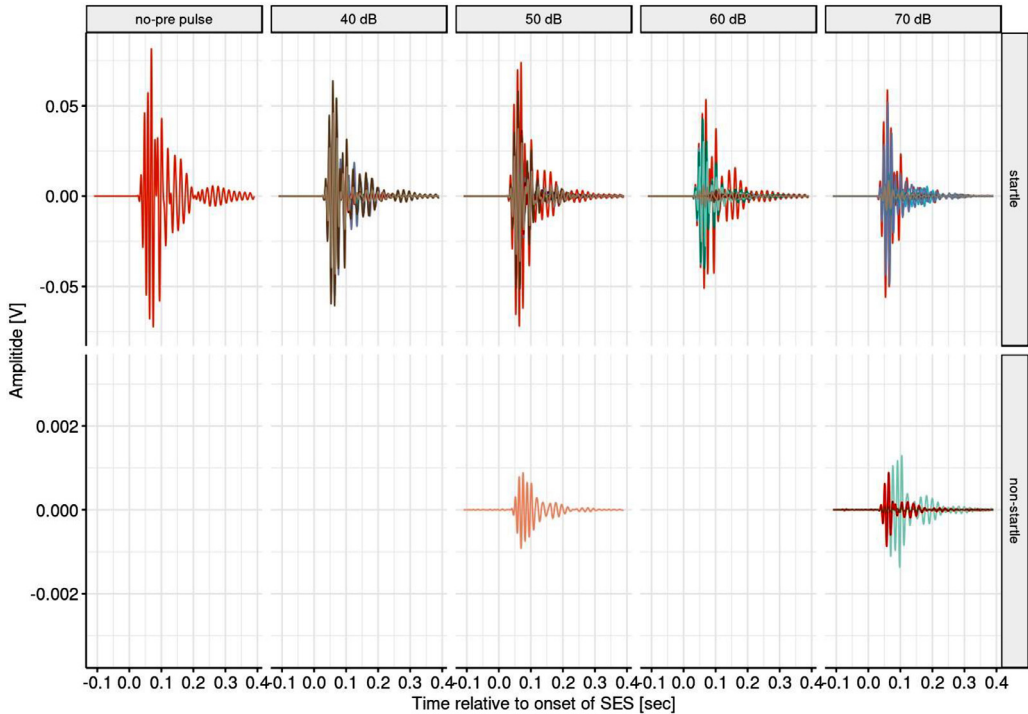


Fig. 4. Representative acoustic startle reflex waveforms from a single Tone Pre-Pulse measurement session with no pre-pulse as well as the 16 kHz tone presented at 40 to 70 dB SPL. All ASR waveforms were manually classified as startles (top) and non-startles (bottom). Note colors indicate different trials within the same session and panels without data indicate zero non-startles for that session and/or condition. Non-startle panels for no-pre pulse, 40 dB and 60 dB conditions had zero non-startle waveforms.

addition, the magnitudes of the ASR waveforms classified as startles decreases with increasing TPP level, demonstrating tone pre-pulse inhibition reported by [6,8,9,15].

Figs. 1,2, and 4 demonstrate these large variations over a variety of sessions conducted on a variety of animals while Fig. 3 shows the distribution of the maximum magnitudes. Fig. 5 shows the ASR waveforms for both the startle I/O and TPP protocols across all SES and pre-pulse levels with the maximum magnitudes (panel A) and minimum magnitudes (panel B) manually classified as startles (top) and non-startles (bottom). It is clear that all startles presented in Fig. 5 follow a similar pattern regardless of the magnitude of the ASR waveform in which there is little to no noise prior to the SES being presented ($t < 0$ msec) followed by large signals due to the involuntary animal movement followed by a period of decreasing magnitude with time and a return to baseline noise levels. This is contrast to ASR waveforms classified as non-startles but with large magnitude where the signal before the SES is presented contains low-level noise followed by increased activity after the SES is presented but does not contain the typical decrease in magnitude back to baseline as seen in those waveforms classified as startles. The ASR waveforms classified as non-startles with minimum maximum magnitudes presented in panel B Fig. 5 contain only low-level noise. Fig. 6 shows the distribution of the maximum magnitudes from all ASR waveforms considered in this study. There is a clear difference in the median maximum magnitudes for startles versus non-startles with the variance associated with startles being much smaller than that of non-startles. Thus, the maximum magnitude will likely be a feature with high predictive power.

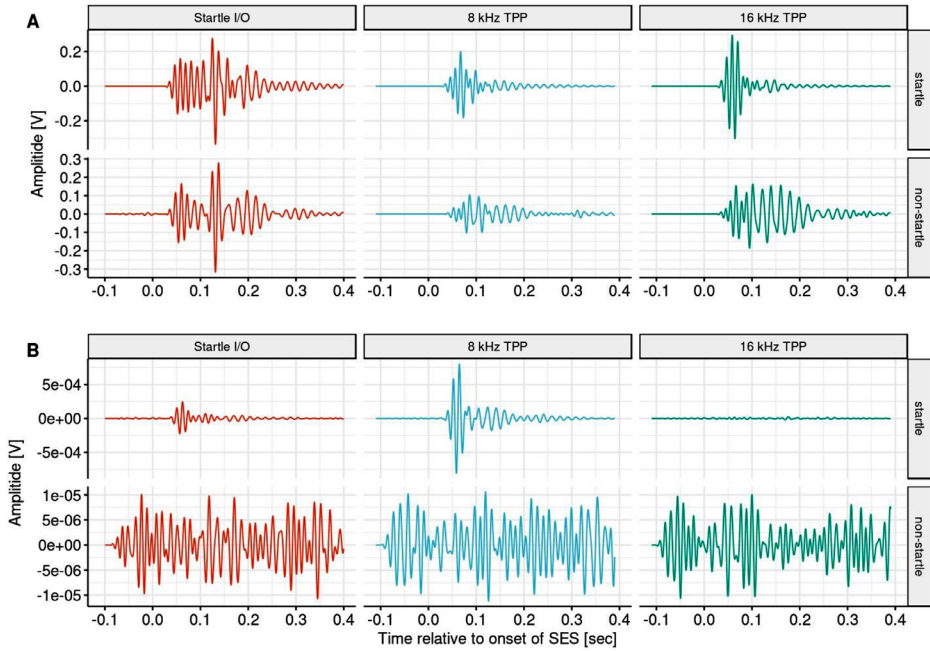


Fig. 5. Acoustic startle response waveforms for measurements producing the highest (panel A) and lowest (panel B) responses across all sessions included in this study for the startle I/O, 8 kHz and 16 kHz tone pre-pulse protocols, manually classified as startle (top) and non-startle (bottom).

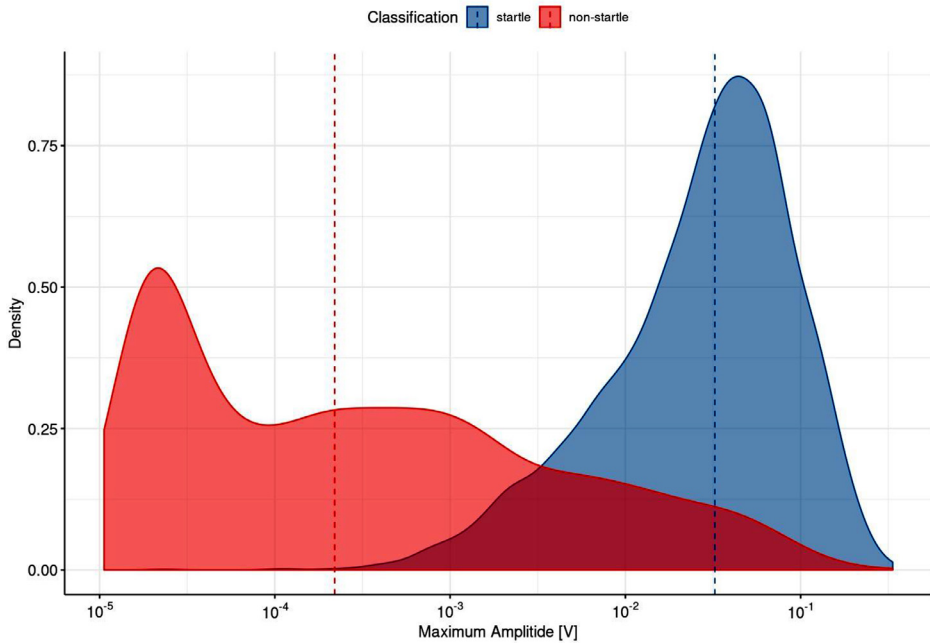


Fig. 6. Distribution of maximum magnitudes of all ASR waveforms for startle (blue) and non-startle (red) with vertical dotted lines indicating the median maximum magnitude of each distribution.

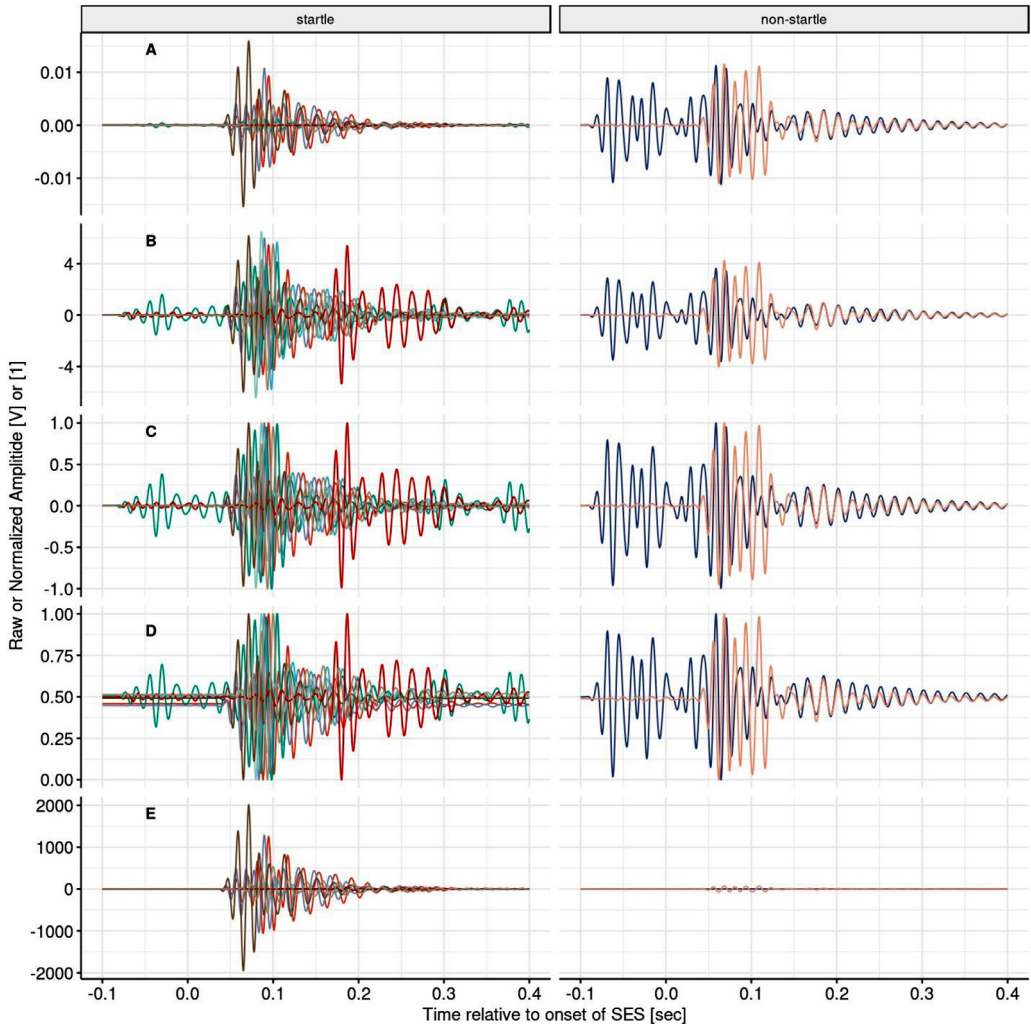


Fig. 7. Raw and normalized representative ASR waveforms collected using the Startle I/O protocol classified as startles (left) and non-startles (right) with the SES presented at 85 dB SPL from session 1268 presented in Fig. 1 including the A: raw waveform as well as several normalization methods including B: Z-score normalization using the entire waveform for mean and standard deviation, C: maximum magnitude normalization, D: minimum-maximum normalization, and E: Z-score normalization using the waveform before the SES was presented to for the mean and standard deviation. Note colors indicate different trials within the same session.

Waveform pre-processing

The ASR waveforms presented in Section 1 show the diversity typically found in acoustic startle response data. In order to derive highly predictive features from ASR waveforms, the waveforms must be normalized such that the waveforms having a range of amplitudes can be easily compared. Several methods of ASR waveform normalization were explored, such as a Z-score normalization by subtracting the mean and dividing by the standard deviation. In this case the mean and standard deviation was computed from both the entire waveform as well as the waveform before the SES was presented. The Z-score normalization essentially transforms the ASR waveforms from voltage to the number of

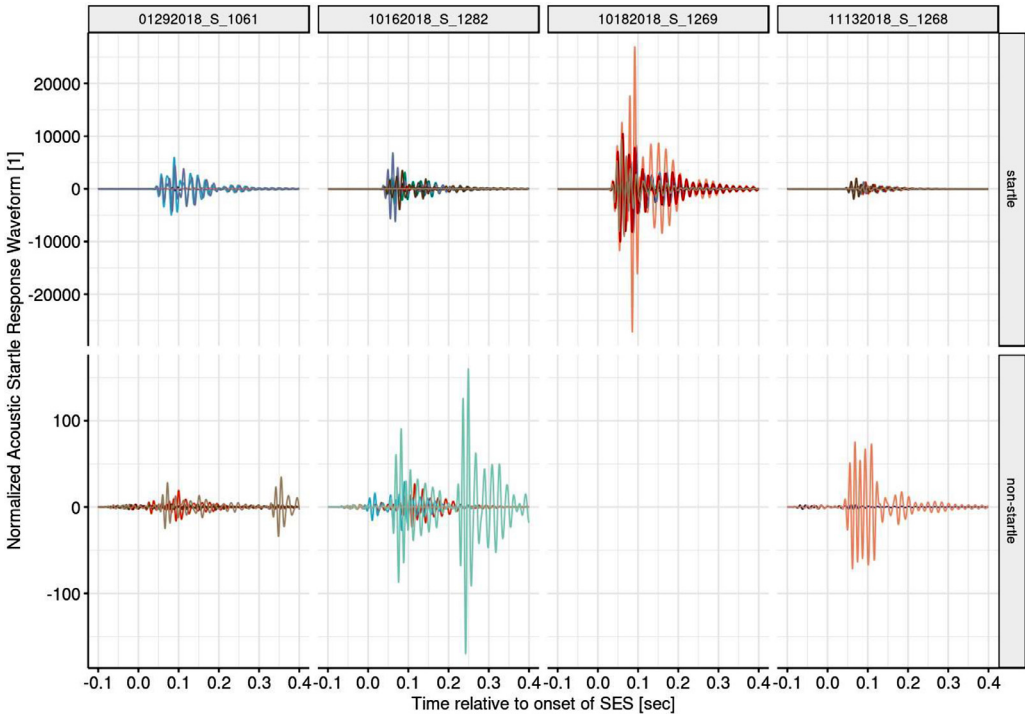


Fig. 8. Normalized ASR waveforms (using Z-score normalization with mean and standard deviation taken before the SES is presented) from the same four randomly selected 85 dB SPL startle I/O sessions presented in Fig. 1. Note colors indicate different trials within the same session as well as the different scale for startles versus non-startles and panels without data indicate zero non-startles for that session and/or condition.

standard deviations away from the mean used to compute the normalized waveform. In addition, ASR waveforms were normalized by dividing by their maximum magnitudes, scaling to the largest value to be equal to one, as well as min-max normalization such that every minimum waveform is transformed to a value of zero and maximum amplitude into a value of one.

Fig. 7 shows the effect these normalization techniques have on the same representative set of representative ASR waveforms subject to the startle I/O protocol with an elicitor presented at 85 dB SPL presented in Fig. 1. Most normalizations techniques produce very similar results for both startles and non-startles but with different scales. However, Z-score normalization using the mean and standard deviation taken before the SES, see plot E in Fig. 7, shows significant differences in scale between startles compared to non-startles. Fig. 8 shows the Z-score normalization with the mean and standard deviation computed using the waveform before the SES. This produces normalized waveforms that have extremely different scales for ASR waveforms classified as startles compared to non-startle. This analysis shows that ASR waveforms classified as startles are up to 2000 standard deviations away from the pre-SES mean (rather $2,000\sigma$) whereas non-startles are up to 80σ . Thus, Z-score normalization using the mean and standard deviation will be applied to ASR waveforms used in this study before feature extraction.

Feature engineering

Machine learning algorithms use features associated with the original ASR waveforms to build models to classify ASR waveforms as startle or non-startle. Features must be extracted from the normalized ASR waveforms as the timeseries (each voltage in time) representing the raw ASR

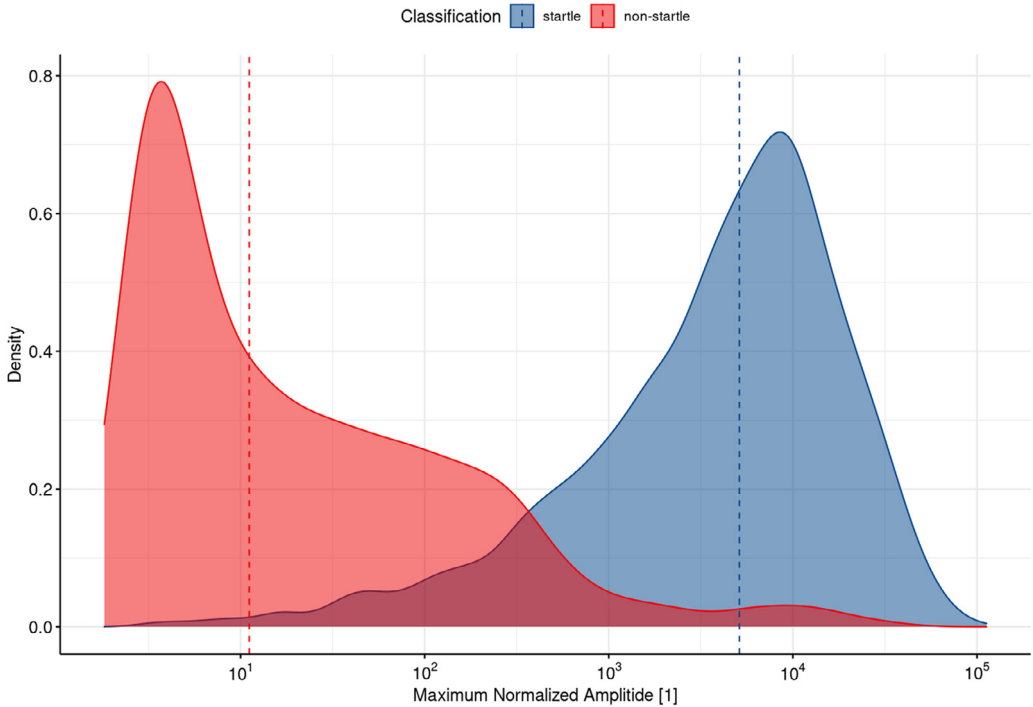


Fig. 9. Distribution of the maximum magnitudes of normalized ASR waveforms for startle (blue) and non-startle (red) with vertical dotted lines indicating the median maximum magnitude of each distribution.

waveforms are generally not good predictors due to the variation of exact timing and shape of startle responses. Thus, features extracted from the normalized ASR waveforms as well as those derived from various transformations of the normalized ASR waveforms will have higher predictive capability than the original waveforms themselves.

Waveform derived features

Fig. 6 showed distribution of maximum magnitudes from the ASR waveforms collected prior to waveform normalization with large differences between startles versus non-startles. However, Fig. 8 showed Z-score normalization using the mean and standard deviation of the ASR waveform prior to the SES being presented produces even larger differences between startles versus non-startles. The distribution of the maximum magnitude of normalized ASR waveforms is presented in Fig. 9. Although Fig. 9 is similar to Fig. 6 for non-normalized ASR waveforms, the shape of the non-startle distribution for normalized ASR waveforms favors lower magnitude startle responses when the ASR waveforms are not normalized. In addition, the difference between the median maximum magnitudes for startle versus non-startles is about 2.5 orders of magnitude as compared to normalized ASR waveforms.

The normalized ASR waveforms presented in Fig. 8 show variation in the time in which the maximum magnitude within about 0.1 s after the SES is presented at $t = 0$ s. Fig. 10 shows the distribution of the time the maximum magnitude occurred for the normalized ASR waveforms for both startles and non-startles with the maximum magnitude located between 0.05 s and 0.15 s for both startles and non-startles. Almost all startles waveforms occurring within this time-range have a significant number of waveforms with the time of maximum magnitude outside of this time range as compared to non-startle waveforms.

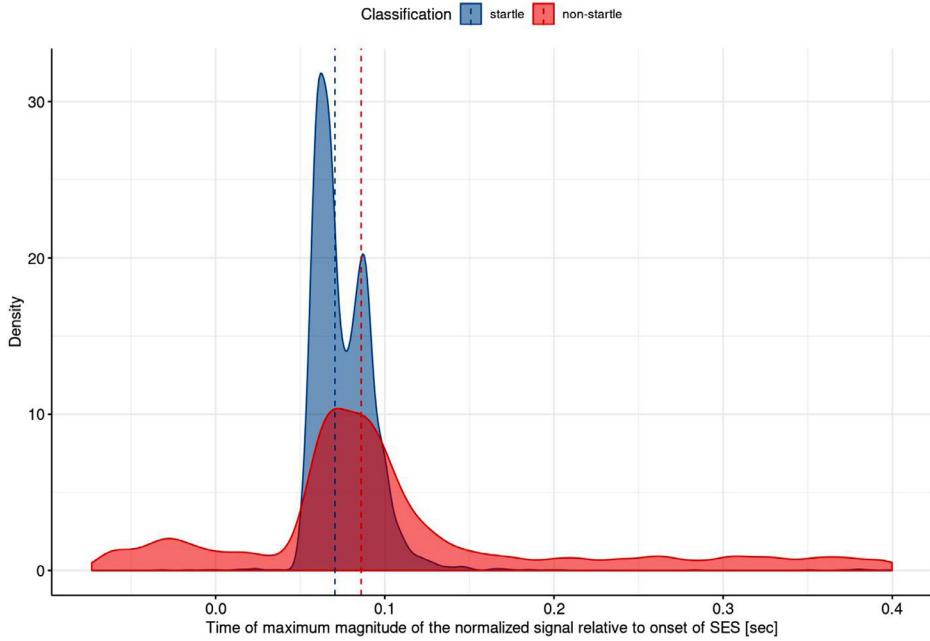


Fig. 10. Distribution of the time of the maximum magnitude of normalized ASR waveforms manually classified as startle (blue) and non-startle (red) with vertical dotted lines indicating the median maximum magnitude of each distribution.

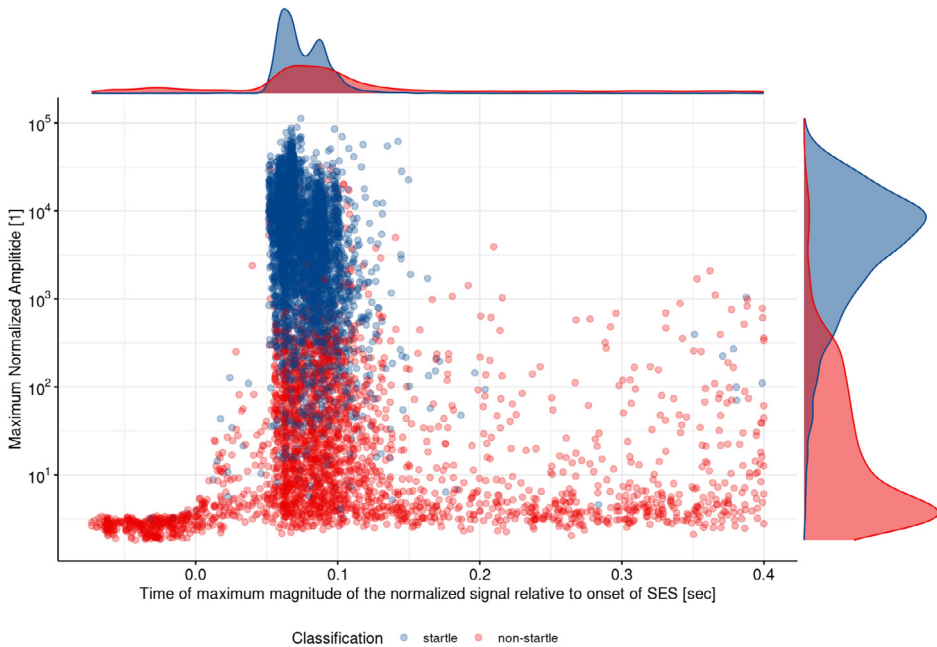


Fig. 11. The maximum magnitude as a function of the time the maximum magnitude occurred extracted from each normalized ASR waveform manually classified to be a startle (blue) or a non-startle (red). The distribution of each feature is above (time of maximum magnitude) and to the right (maximum magnitude) of the individual data.

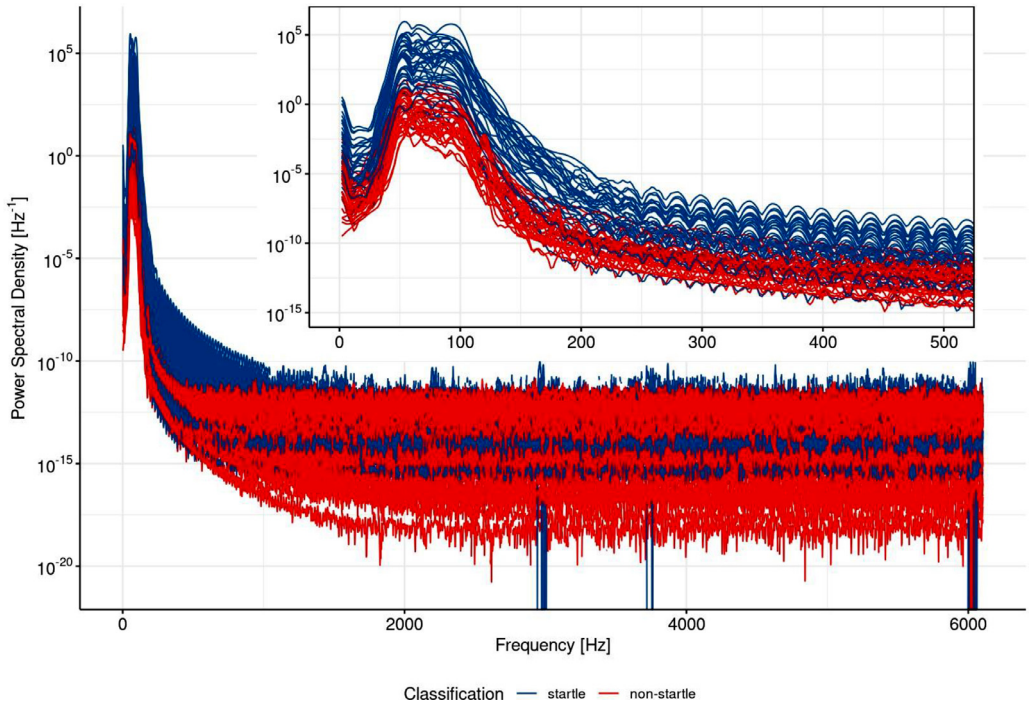


Fig. 12. Power spectral density estimate of the normalized ASR waveforms from the same four randomly selected 85 dB SPL startle I/O measurements presented in Fig. 8 manually classified as startle (blue) and non-startle (red). Inset shows the PSD for frequencies from 0 to 525 Hz.

Fig. 11 shows the individual maximum magnitudes for the normalized waveforms as a function of the time of the maximum magnitude with the marginal distributions on each axis. This analysis clearly shows most startle waveforms present large maximum magnitudes ($> 5 \times 10^2$) at times concentrated around $t = 0.1$ s. However, there is some overlap in maximum magnitude and time of the maximum magnitude for startles and non-startles, suggesting additional features must be extracted from the normalized ASR waveforms for successful classification via machine learning algorithms.

Power spectrum density derived features

Figs. 1, 2, and 4 display periodic behavior in the ASR waveforms with the magnitude, and possibly the frequency content, varying between startles versus non-startles. These figures as well as the normalized ASR waveforms presented in Fig. 8 and distributions shown in Figs. 9 and 11 show the magnitude of the periodic behavior also varies between startles versus non-startles. Thus, the power contained within a spectral density (PSD) estimate from each normalized ASR waveform is expected to also vary both in the magnitude of the power as well as which frequency contains the maximum power. Fig. 12 show the power spectral density estimates from the same 85 dB SPL startle I/O normalized ASR waveforms presented in Fig. 8. The frequency content is clearly concentrated between 50 Hz to 100 Hz with startles exhibiting about 5 orders of magnitude higher power on average compared to non-startles.

Features extracted from the PSD estimate of the normalized ASR waveforms include the maximum power and the frequency at maximum power. Fig. 13 shows the distribution of the maximum PSD power for all normalized ASR waveforms considered in this study. This PSD analysis shows the median

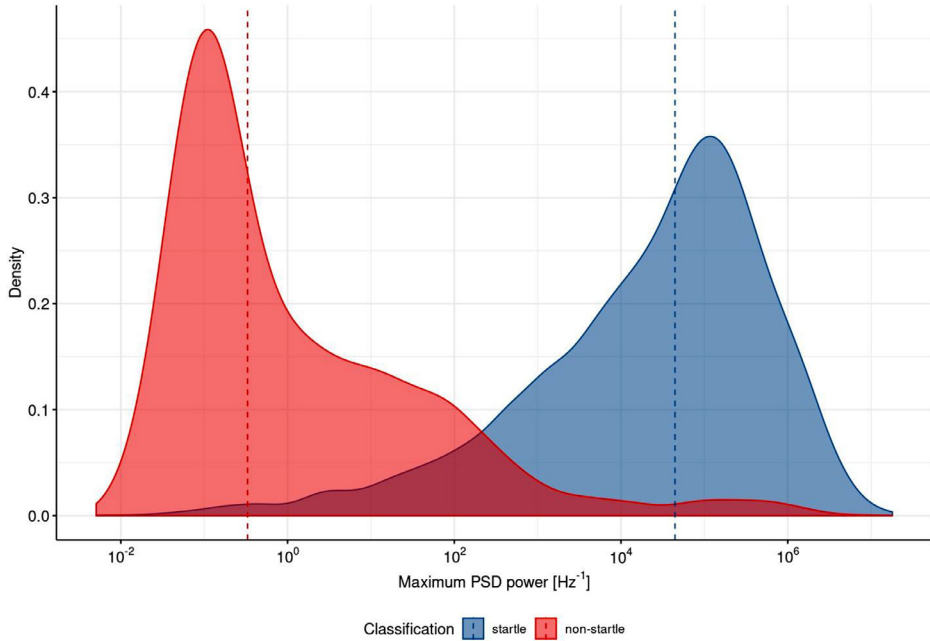


Fig. 13. Distribution of the maximum power spectral density estimate for each normalized ASR waveforms manually classified as startle (blue) or non-startle (red) with vertical dotted lines indicating the median maximum magnitude of each distribution.

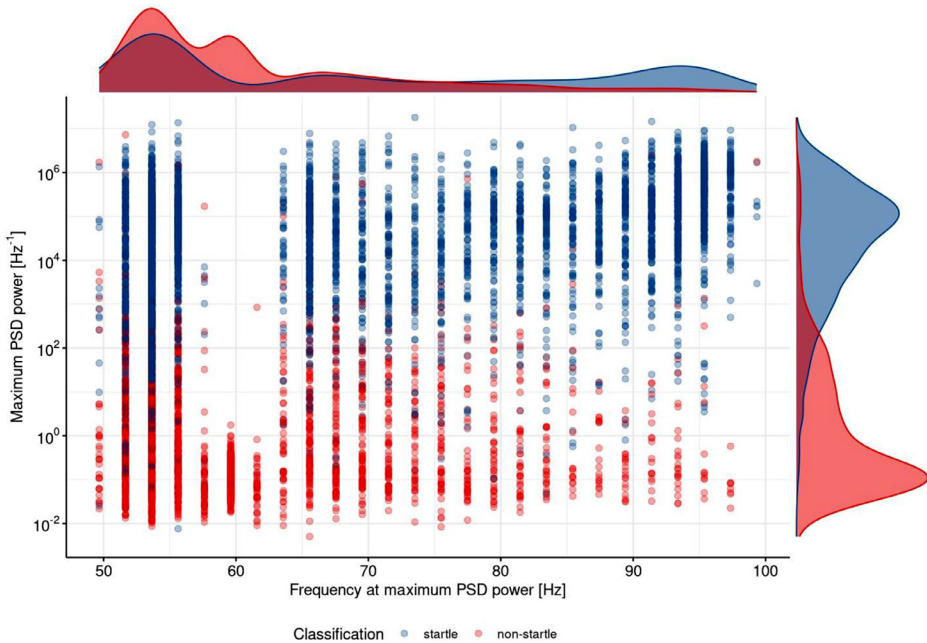


Fig. 14. The maximum power as a function of the frequency at which the maximum power occurred extracted from the power spectral density estimates from each normalized ASR waveform manually classified to be a startle (blue) or a non-startle (red). The distribution of each feature is above (frequency of maximum power) and to the right (maximum power) of the individual data.

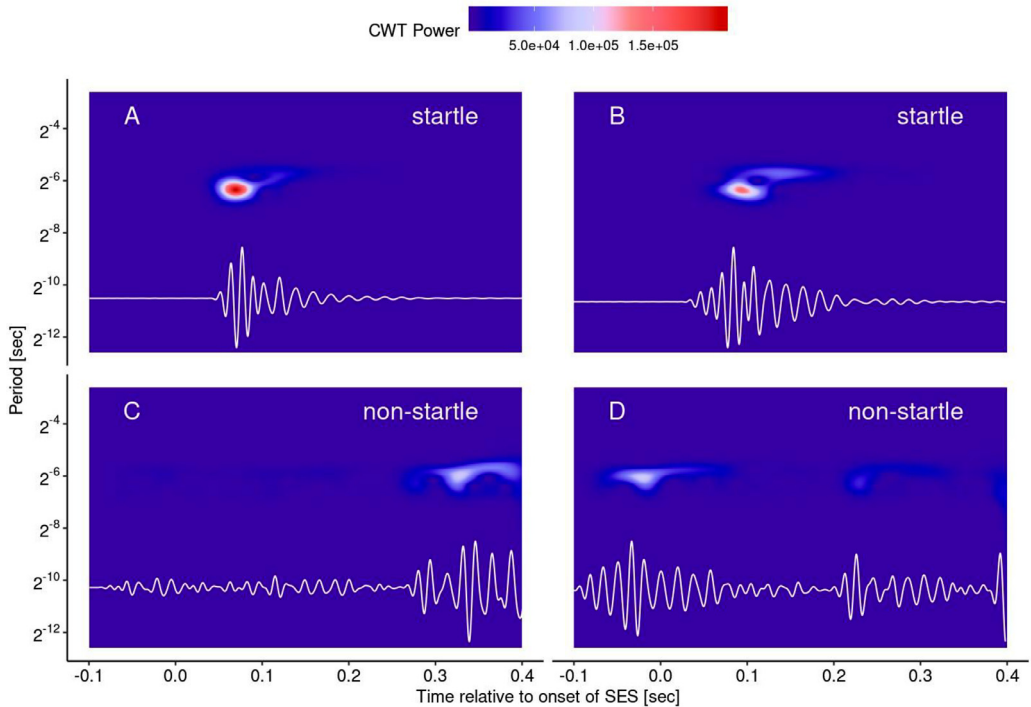


Fig. 15. Continuous wavelet transform power spectra of four randomly selected normalized ASR waveforms from a single 85 dB SPL startle I/O session whose normalized ASR waveforms (shown in white) are presented in Fig. 1. Note: the scale of the normalized ASR waveforms is not constant for each panel.

maximum PSD power for startles is about 5 orders of magnitude higher than that of non-startles, indicating higher magnitude periodic behavior for startles versus non-startles. The maximum PSD estimates presented in Fig. 13 will occur at slightly different frequencies. Fig. 14 shows the individual maximum PSD power as a function of the frequency of the maximum PSD power for each normalized ASR waveform. This analysis shows the distributions of the frequency of the maximum PSD power for startles is similar to that for non-startles.

Continuous wavelet transform derived features

Close inspection of the ASR waveforms and normalized ASR waveforms presented in Figs. 1, 2, 4, and 8 shows the timing of the maximum high magnitude periodic activity is drastically different between startles and non-startles. Startles generally exhibit high magnitude periodic behavior within 0.05 s and 0.15 s after the SES is presented. Many non-startles also exhibit a maximum magnitude within the same time frame, but many also exhibit high magnitude periodic behavior throughout the waveform. The power spectral density estimate is based on the Fourier Transform and computed power as a function of frequency based on the entire waveform, not any one period of time. Thus, the PSD estimate weights the frequency content the same across all times. This equal weighting does not allow for the computation of the power spectral density as a function of both frequency and time in order to analyze the temporal changes in the normalized ASR waveforms. One alternative would be to use the short time Fourier Transform (STFT) for the analysis of frequency content as a function of time. However, Zhang et al. [18] provided evidence for the advantageous use of a continuous wavelet transform, which exhibited good time-frequency resolution, as well as accurate time-frequency power estimation.

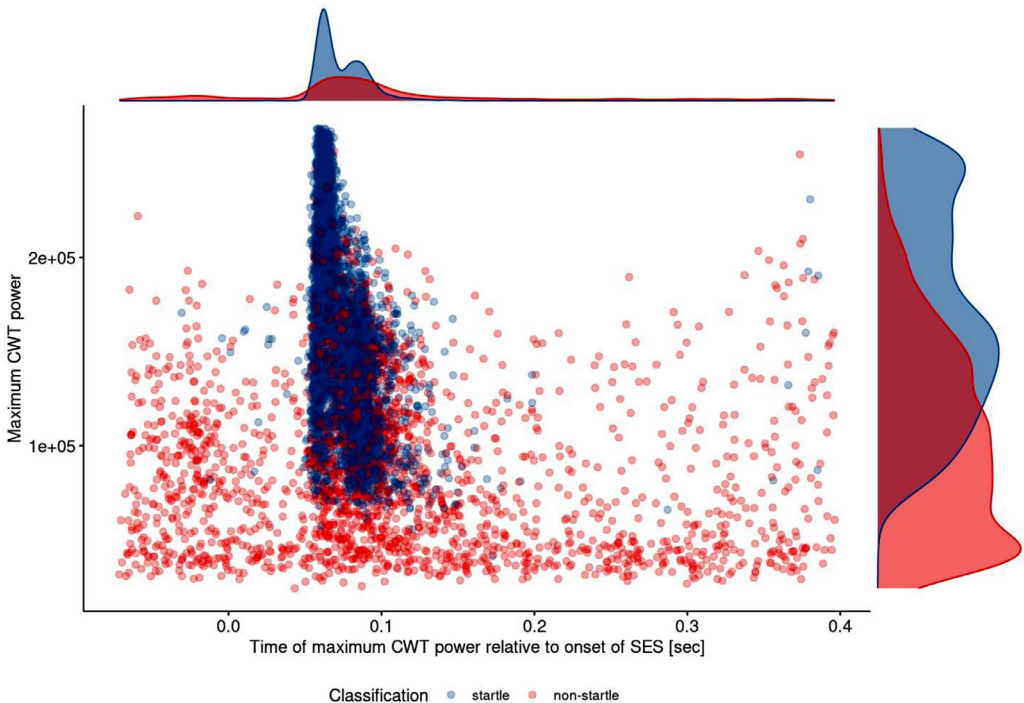


Fig. 16. Distribution of the maximum continuous wavelet transform (CWT) power versus the time of the maximum CWT power for normalized ASR waveforms manually classified as startle (blue) or a non-startle (red).

The continuous wavelet transform (CWT) power spectra of normalized ASR waveforms from four randomly selected trials from a single 85 dB SPL session (whose normalized ASR waveforms are presented in Fig. 8) are presented in Fig. 15. These CWT power spectra show large magnitudes when the associated normalized ASR waveform contains high-magnitude frequency content. Specifically, the CWT power spectra of the startles presented, panels A and B, show high CWT power in a small time window (0.05 to 0.15 s) as well as a small period range (2^{-6} to 2^{-7} Hz) whereas the non-startles contain CWT power at inconsistent times. The CWT power spectra for the non-startles shown in panels C and D of Fig. 15 show high-magnitude frequency content well after (panel C) and well before (panel D) the SES is presented.

Characteristics of the CWT power spectra of each normalized ASR waveform provide insight into their time-dependent frequency content, allowing for the extraction of features with high predictive capability. Fig. 16 shows the distribution of maximum CWT power of the normalized ASR waveforms as a function of the time in which the maximum CWT power occurred. This analysis shows the maximum CWT power to be elevated for startles when compared to non-startles. In addition, the time of the maximum CWT power is between 0.05 and 0.15 s after the startle elicitor for startles versus non-startles whose time of maximum CWT power spans the entire recording time. Fig. 17 shows the mean CWT power (over all times and periods) as a function of period of the maximum CWT power. The period at maximum CWT power is generally lower for startles when compared to non-startles. However, the mean CWT power is generally equal across almost all normalized ASR waveforms and is not a highly predictive feature.

Inspection of the CWT power spectra presented in Fig. 15 shows that the combination of time, period, and CWT power allows for the extraction of features with high predictive capability. The CWT power before the SES was presented ($t < 0$ s) is essentially zero across all periods for startles (panels A and B) but can be non-zero for non-startles (panel D) while the CWT power around the time the SES

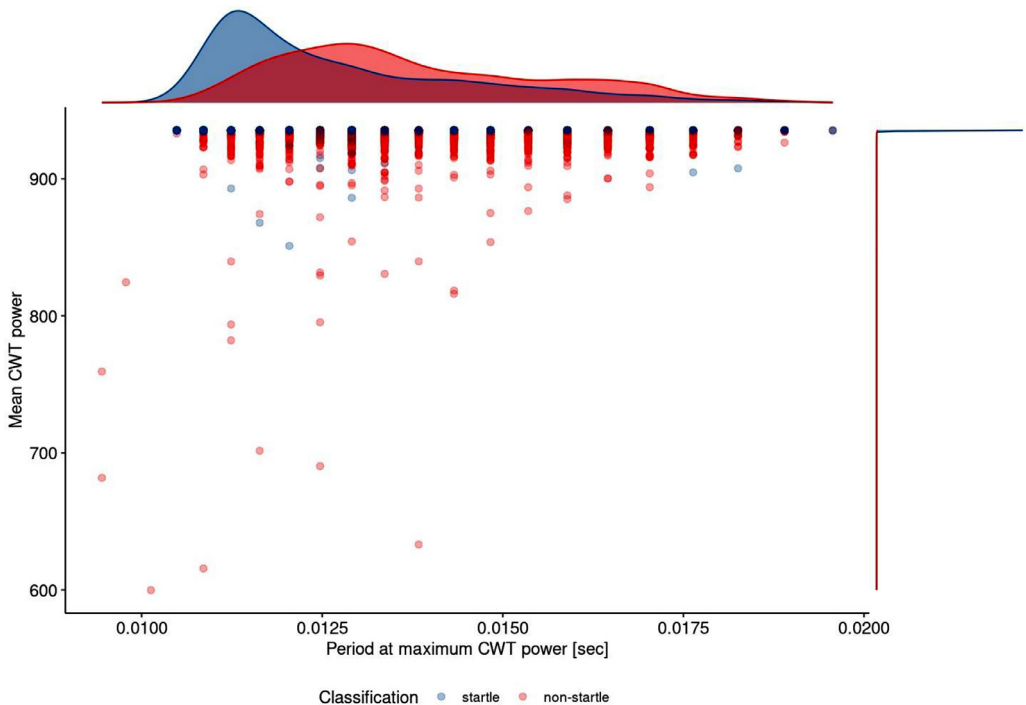


Fig. 17. Distribution of the mean continuous wavelet transform (CWT) power across all periods and times as a function of the period of maximum CWT power for normalized ASR waveforms manually classified as startle (blue) or a non-startle (red).

is presented ($t > 0.005$ and $t < 0.15$ s) is high for a subset of periods for startles for the representative CWT power spectra shown in Fig. 15. Fig. 18 shows the mean CWT power (across all periods) around the time the SES is presented as a function of the mean CWT power before the SES is presented (all times before $t = 0$ s) for all normalized ASR waveforms sampled for this study.

The CWT power spectra across all periods 50 msec before the SES is shown in Fig. 15. The CWT power spectra are essentially zero for both startles (panels A and B) as well as the non-startle in panel C but is not zero for the non-startle in panel D. Fig. 19 shows the distribution of the mean CWT power (across all periods) at 50msec before the SES is presented. This analysis shows significant separation between startles and non-startles in the distribution of mean CWT power 50msec before the SES is presented, making the mean CWT power 50 msec before the SES a feature with high predictive capability.

Periods of high magnitude activity and recovery to baseline are represented by the CWT power at 50 msec and 350msec after the SES is presented. Fig. 20 displays the mean CWT power at 350 msec after SES compared to 50msec after SES. Generally, startle responses show a decreased mean CWT power 350 msec after SES as compared to the non-startle responses. This illustrates that activity returns to baseline for classified startles as compared to non-startles, when using the normalized ASR waveforms. However, the mean CWT power 50 msec after the SES is presented is generally low for non-startles but is almost uniformly distributed for startles. This is due to the variable latency in the animal's startle response following the startle elicitor.

The variable latency of the high magnitude activity for normalized ASR waveforms classified as startles shown in Fig. 8 and embedded with the corresponding CWT power spectra in Fig. 15 highlights the difficulty in choosing any specific time in which to extract CWT power. However, inspection of the representative CWT power spectra presented in Fig. 15 show the high magnitude

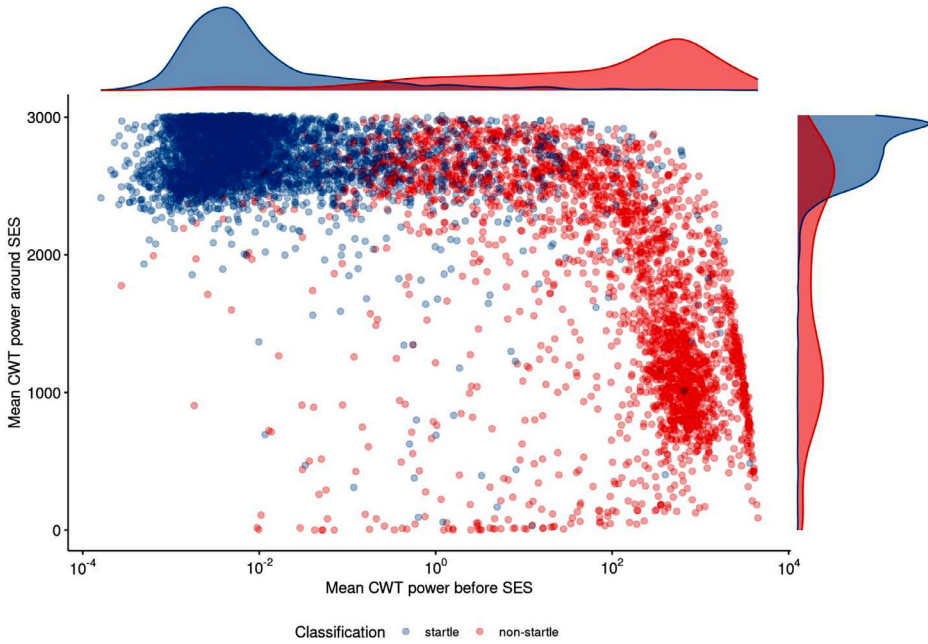


Fig. 18. Distribution of the mean continuous wavelet transform (CWT) power across all periods around the SES was presented ($t > -0.005$ and $t < 0.15$ s) as a function of the mean CWT power before the SES was presented ($t < 0$ s) for normalized ASR waveforms manually classified as startle (blue) or a non-startle (red).

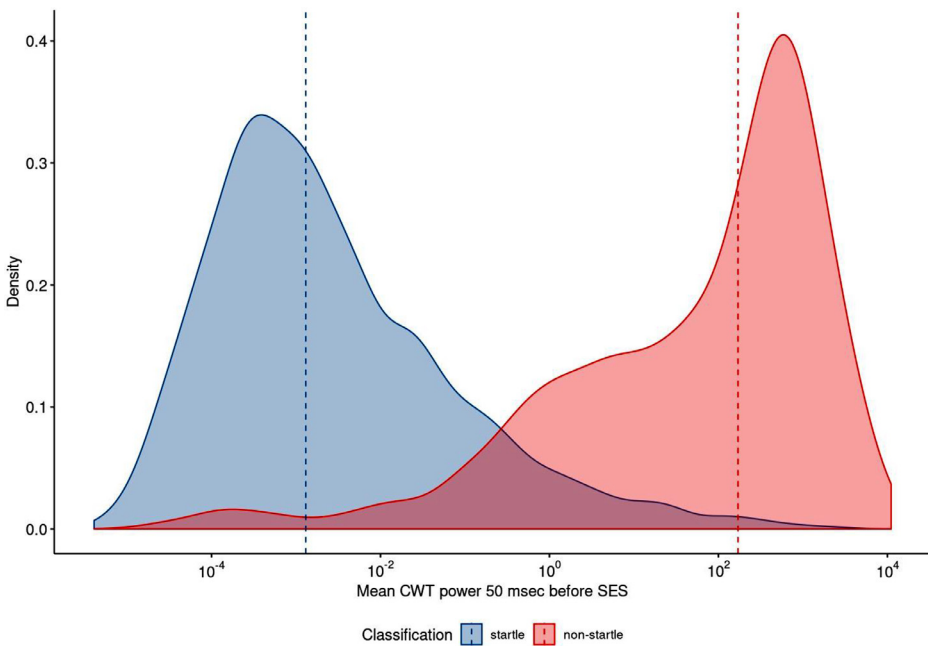


Fig. 19. Distribution of the mean continuous wavelet transform (CWT) power across all periods at $t = 50$ msec before the SES was presented for normalized ASR waveforms manually classified as startle (blue) or a non-startle (red).

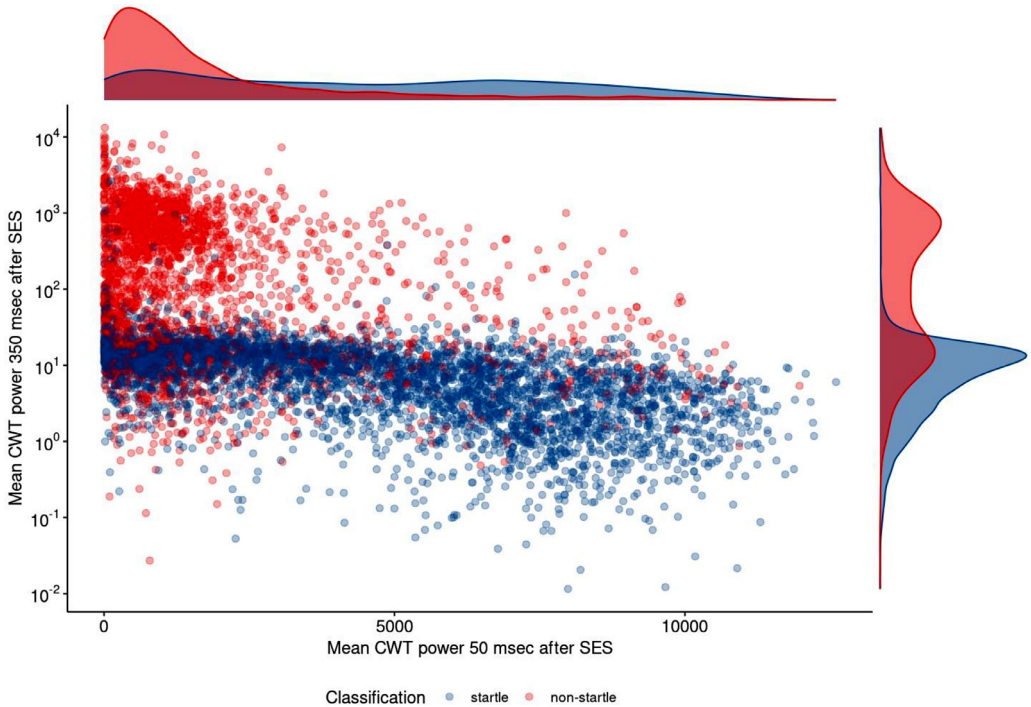


Fig. 20. Distribution of the mean continuous wavelet transform (CWT) power across all periods at $t = 50$ msec after the SES was presented as a function of the mean CWT power at $t = 350$ msec after the SES was presented for normalized ASR waveforms manually classified as startle (blue) or a non-startle (red).

activity occurs at a frequency near the resonant frequency of the piezoelectric transducers of 100 Hz or a period of $2^{6.64}$ s. Fig. 21 shows the log transform of the same representative CWT power spectra presented in Fig. 15. Analysis of the log transformed CWT power spectra show significant differences in high frequency (low period) activity for startles with an approximately 5 orders of magnitude greater CWT power at 2048 Hz across all times for non-startles versus startles. Fig. 22 shows the mean CWT power across all times at 2048 Hz versus that at 100 Hz.

Feature pre-processing

A total of 17 features were extracted from the ASR waveforms as described in Fawcett et al. [4], Table 1] with their distributions presented in Figs. 11,14,16–20, and 22. Feature variability was assessed to ascertain which, if any, features have little to no variability. Features with little to no variability would be considered poor candidates for features use in machine learning. Table 1 shows the majority of features had over 99% unique value. The four features that contained less than 99% unique values also maintained a ratio of less than three when placing the most prevalent value over the second most prevalent value. A ratio value of three is smaller than the ratio of 20 described by Kuhn and Johnson [12] which was used to classify predictors as having variance that approximates zero. As a result of all the features having a near-zero variance, none were removed from the feature space.

Next, the correlation between features was analyzed to determine if any pairs of features are highly correlated and thus contain redundant information. Features containing redundant information must be removed from the feature space prior to the use of machine learning algorithms [5,17]. Fig. 23

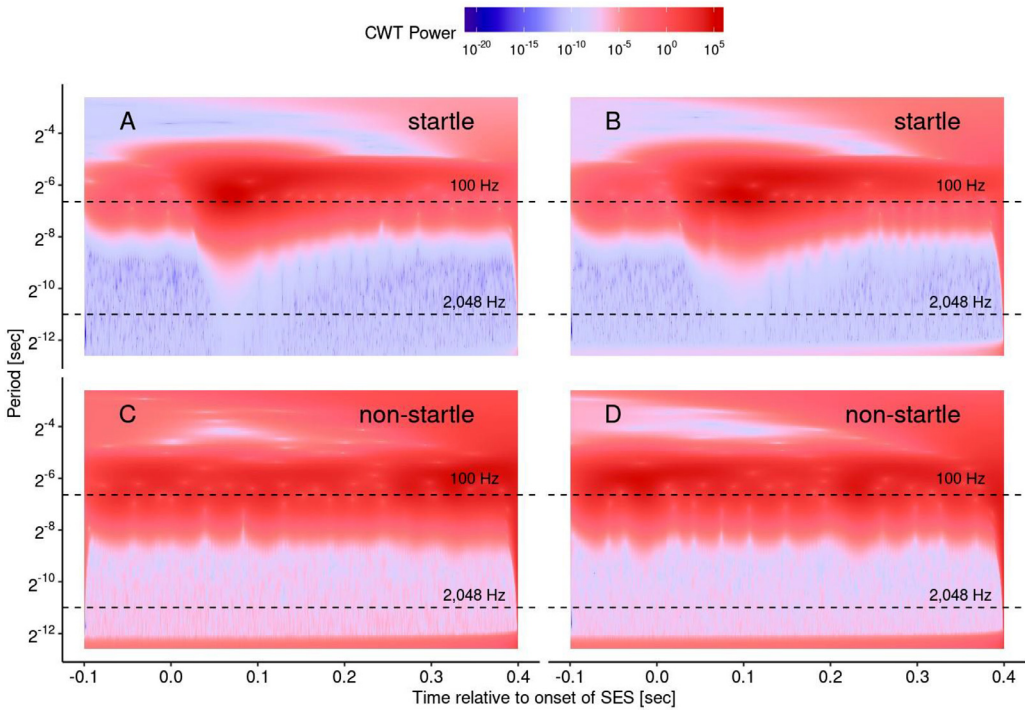


Fig. 21. Log transform of the same continuous wavelet transform power spectra of four randomly selected normalized ASR waveforms from a single 85 dB SPL startle I/O session presented in Fig. 15. Dotted lines represent the period corresponding to 100 Hz and 2048 Hz.

Table 1
Feature variability.

Feature	Frequency Ratio	Percent Unique
before.mean	1.0000	96.4385
before.sd	1.0000	96.5084
max.mag	1.0000	96.5084
t.max.mag	1.0870	35.5726
max.psd	1.0000	96.5084
max.psd.log10	1.0000	96.5084
freq.max.psd	2.6388	0.3771
mean.cwt.power.before.startle	1.0000	96.5084
max.cwt.power	1.0000	96.5084
t.max.cwt.power	1.1071	34.8324
freq.max.cwt.power	1.2391	0.3073
mean.cwt.power	1.0000	96.5084
mean.cwt.power.100.Hz	1.0000	96.5084
mean.cwt.power.around.startle	1.0000	96.5084
mean.cwt.power.50 ms.before.startle	1.0000	96.5084
mean.cwt.power.50 ms.startle	1.0000	96.5084
mean.cwt.power.350 ms.after.startle	1.0000	96.5084
mean.cwt.power.high.freq	1.0000	96.5084

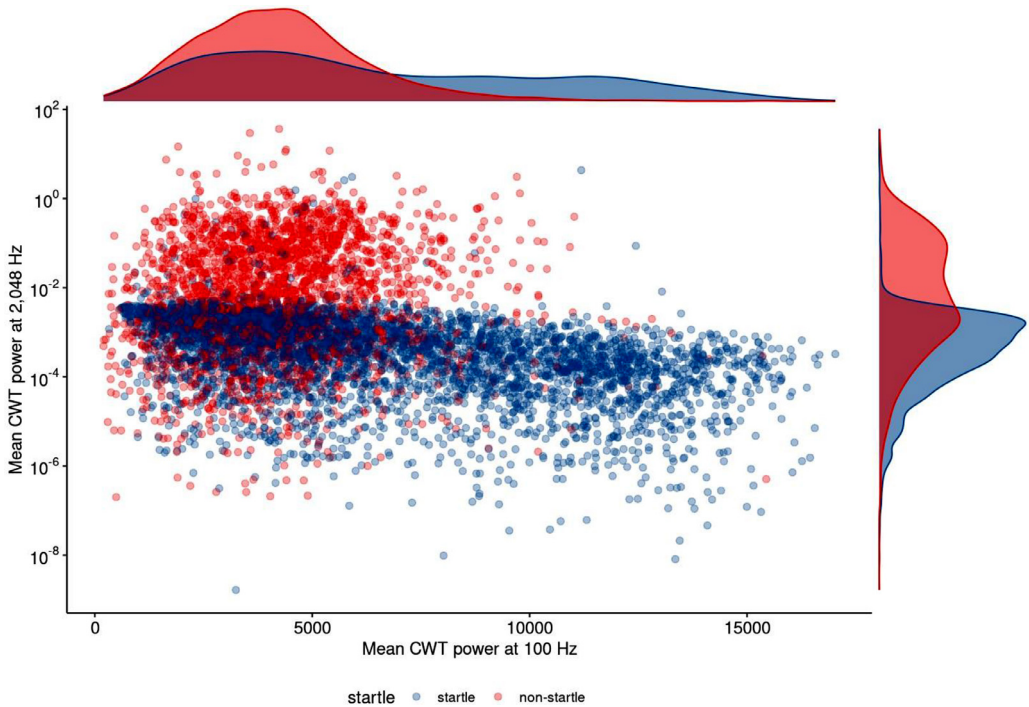


Fig. 22. Distribution of the mean continuous wavelet transform (CWT) power across all times at 100 Hz as a function of the mean CWT power at 2048 Hz for normalized ASR waveforms manually classified as *startle* (blue) or a *non-startle* (red).

shows the correlation matrix of all pairs of features extracted from the ASR waveforms with most pairs of features having small correlation coefficients. Although there are several pairs of features that are somewhat correlated, this analysis shows that no pair of features exceed the highly correlated threshold of $r > \pm 0.9$ [1], allowing all features to be included in the feature space use for the machine learning model development described below.

Similar to the correlation analysis described above, any feature which is a linear combination of other features contains redundant information and must be removed from the feature set [12]. No features were removed from the feature set since no feature was determined to be a linear combination of any other features.

Data partitioning

Features were extracted from 6911 ASR waveforms and randomly split into two sections: 80% was used for training and 20% was used for testing and validation. Additionally, all model training included 10-fold cross validation to help reduce training biases across the specific subsets of data [10,13,14]. As previously mentioned, all features were centered and scaled within their data partitions to use via machine learning methods as outlined in the next section.

Machine learning methods

No combination of single features or feature pairs shown in the previous data can differentiate a *startle* from a *non-startle*. In order for machine learning algorithms to distinguish *startles* and *non-startle* responses with little error, machine learning techniques must be applied to

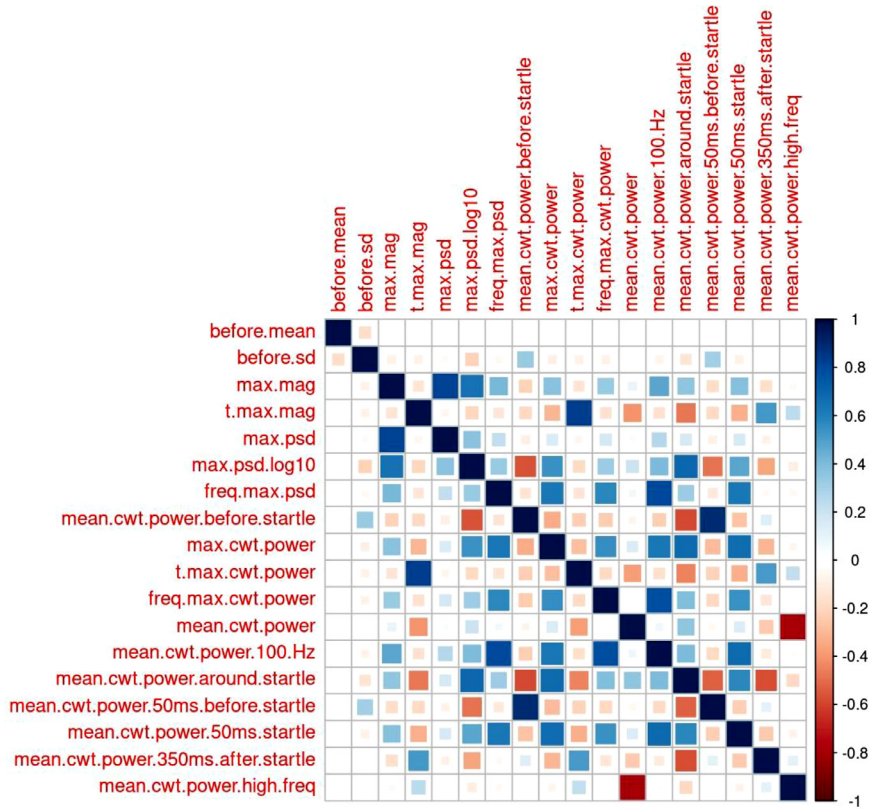


Fig. 23. Feature correlation plot.

combine the features described previously [7,12]. In an attempt to employ algorithms to find optimal feature combinations, various techniques from diverse algorithm families including Linear classification (logistic regression and linear discriminant analysis), Bagged classification (bagged tree and random forests), Boosted classification (extreme gradient boosted and C5.0), and Discriminative classification via kernels (support vector machine) were employed as reported in Fawcett et al. [4]. All machine learning algorithms were implemented in the caret package [11] in the R programming language.

Individual model performance

Fig. 24 shows the Receiver Operating Characteristic (ROC) curve for the prediction of ASR waveforms being a startle or non-startle for all machine learning algorithms listed above. The Random Forest (rf) algorithm produced the best classification (up to 100% training accuracy) and demonstrates the ROC curve farthest from the 45° line. The specificity, sensitivity, and ROC when predicting 25 randomly selected training examples for all machine learning algorithms tested is presented in Fig. 25, again showing Random Forests (rf) is the highest performing algorithm followed closely by eXtreme Gradient Boosting Tree (xgbTree).

Fig. 25 shows the distributions of several performance metrics include the Receiver Operating Characteristics (ROC), Sensitivity (Sens), and Specificity (Spec) when predicting a ASR waveform to be a startle or non-startle with the training dataset as well as the accuracy of predicting the correct

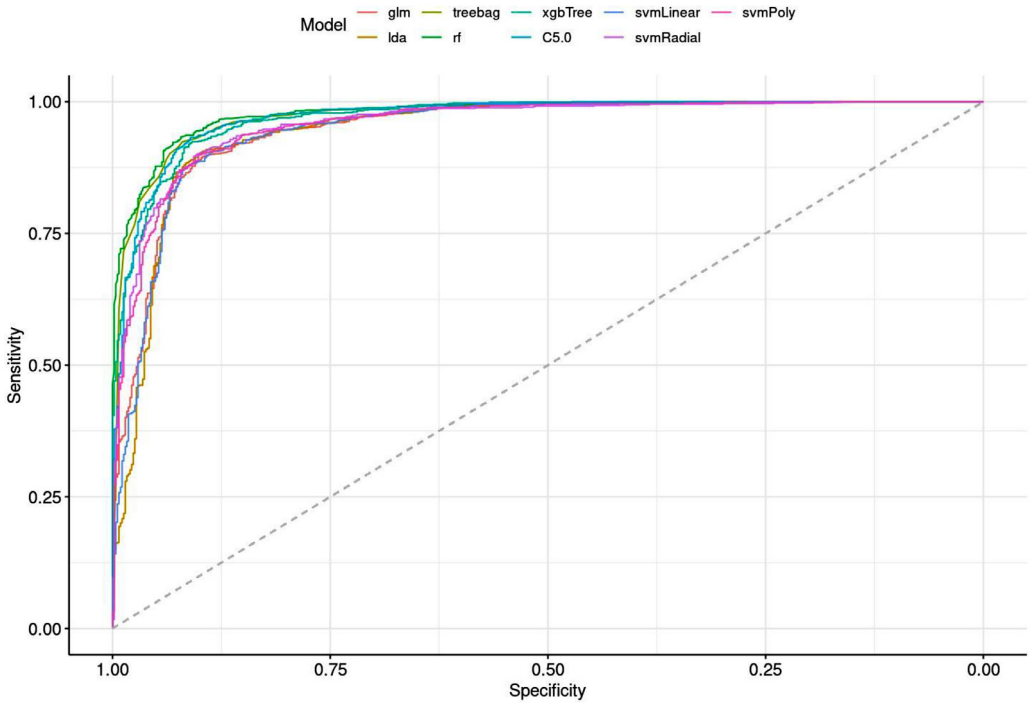


Fig. 24. Receiver Operating Characteristic (ROC) Curve for all individual machine learning used in this study.

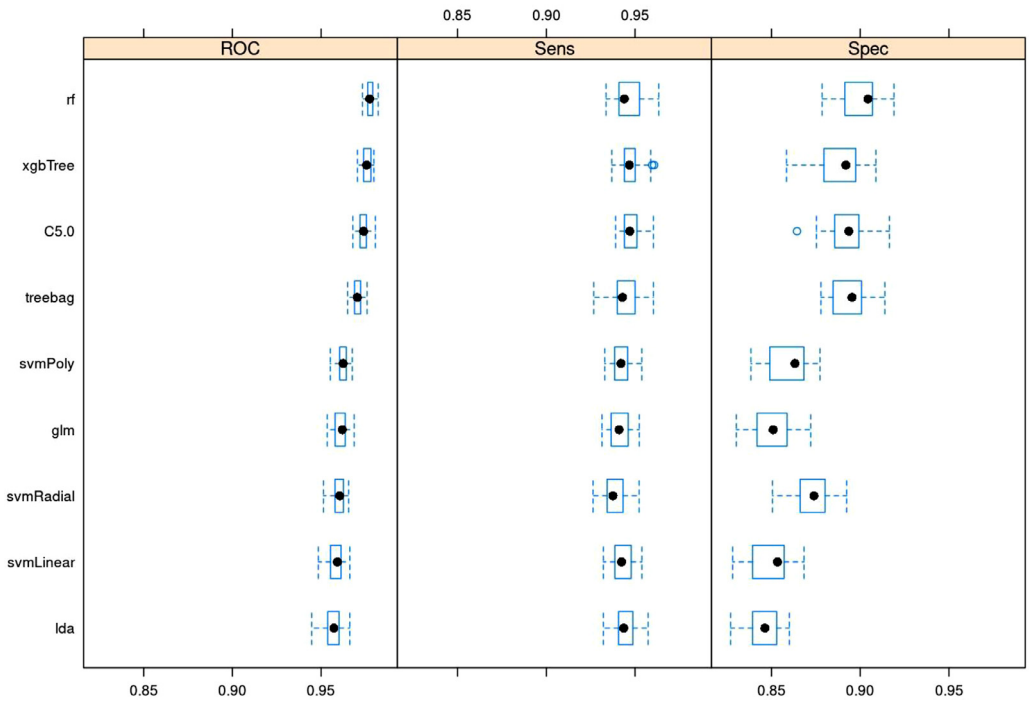


Fig. 25. Individual machine learning model performance.

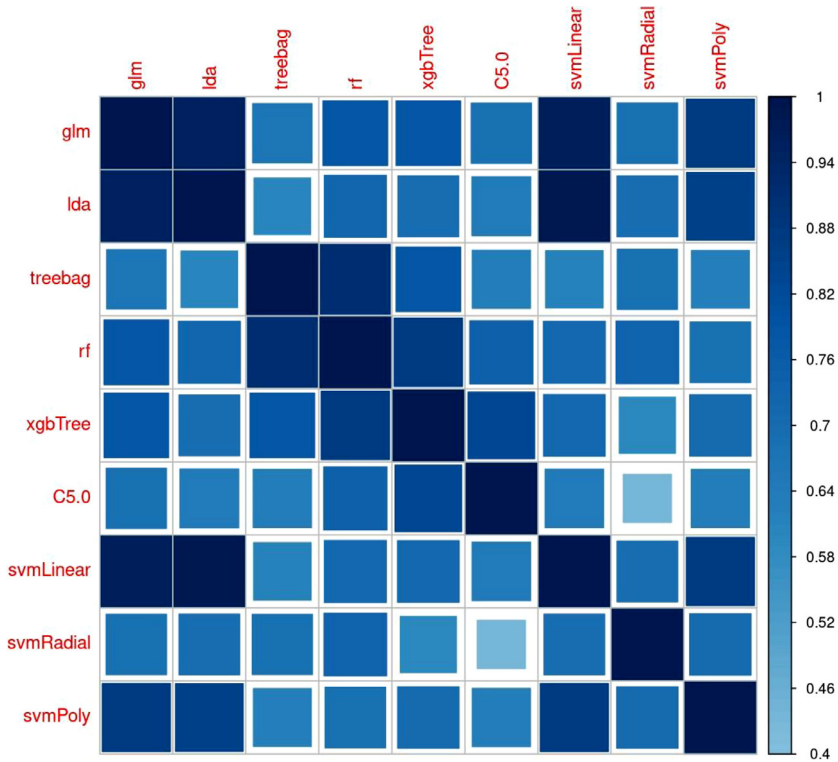


Fig. 26. Machine learning model correlation plot.

classification on the testing dataset for each of the individual machine learning methods described above. All mean ROCs were over 0.95 with Random Forests demonstrating the greatest performance within the individual models with a ROC of 0.9779 and a testing accuracy of 0.9266 [4].

Ensemble model performance

In addition to the individual machine learning algorithms listed above, multiple models have been ensemble together via stacking [2,16] in order to produce an extremely robust model that generalizes well to test and validation data. However, similar to correlated features, highly correlated models should not be included in an ensemble model and thus removed. Fig. 26 shows the correlation between the predictions between pairs of individual machine learning models. The ensemble machine learning workflow was implemented in R using the caretEnsemble package [3].

A Generalized Linear Model (GLM) was used to determine the optimum weight of each model. The ensemble model performance using all models is presented in demonstrated a training ROC of 0.9779 and testing accuracy of 0.9301. The Bagged tree (treebag) as well as the Support Vector Machine with Linear and Polynomial kernels were removed from the final ensemble model as they possessed statistically insignificant weights, indicating these models are highly correlated with other models. Removing models based on the statistical significance of their coefficient versus correlation coefficient alone in order to decide which model in the pair of highly correlated models to keep.

Acknowledgments

This work was supported by the [National Institutes of Health \[NIH-NIA 490 AG00954\]](#). The authors would like to acknowledge the use of the services provided by Research Computing at the [University of South Florida](#). We thank Demitri Brunnell and Mary Reith for their technical assistance and oversight of behavioral experiments. Rachal Love for oversight of animal care and Kendra Stebbins for expert copy editing.

Declaration of Competing Interest

None.

References

- [1] J. Biesiada, W. Duch, Feature selection for high-dimensional data - a pearson redundancy based filter, *Adv. Soft Comput.* (2006) 242–249.
- [2] R. Caruana, A. Niculescu-Mizil, G. Crew, A. Ksikes, Ensemble selection from libraries of models, *Twenty-First International Conference on Machine Learning - ICML '04*, ACM Press, 2002.
- [3] Deane-Mayer, Z.A., Knowles, J.E., 2016. caretEnsemble: ensembles of caret models.
- [4] 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) 108853.
- [5] I. Guyon, J. Weston, An introduction to variable and feature selection, *Mach. Learn. Res.* 3 (2003) 1157–1182.
- [6] J. Halonen, S.A. Hinton, D.R. Frisina, B. Ding, X. Zhu, P.J. Walton, Long-term treatment with aldosterone slows the progression of age-related hearing loss, *Hear. Res.* 336 (2016) 63–71.
- [7] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York, New York, NY, 2009.
- [8] J.E. Jeskey, J.F. Willott, Modulation of prepulse inhibition by an augmented acoustic environment in DBA/2J mice, *Behav. Neurosci.* 114 (5) (2000) 991–997.
- [9] R. Joobar, J.M. Zarate, G.A. Rouleau, E. Skamene, P. Boksa, Provisional mapping of quantitative trait loci modulating the acoustic startle response and prepulse inhibition of acoustic startle, *Neuropsychopharmacology* 27 (5) (2002) 765–781.
- [10] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, Morgan Kaufmann Publishers Inc, 1992, pp. 1137–1143.
- [11] M. Kuhn, caret: classification and regression training, *Tech. Rep.* (2018).
- [12] M. Kuhn, K. Johnson, *Applied Predictive Modeling*, Springer New York, New York, NY, 2013.
- [13] M. Michailidis, *Investigating Machine Learning Methods in Recommender Systems* Ph.D. thesis, 2015.
- [14] J.D. Rodriguez, A. Perez, J.A. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* (2010).
- [15] J.G. Turner, J.F. Willott, Exposure to an augmented acoustic environment alters auditory function in hearing-impaired DBA/2J mice, *Hear. Res.* 118 (1–2) (1998) 101–113.
- [16] H.D. Wolpert, Stacked generalization, *Neural Netw.* 5 (2) (1992) 241–259.
- [17] L. Yu, H. Liu, Feature selection for high-dimensional data: a fast correlation-based filter solution, in: *Vol. 2 of Proceedings, Twentieth International Conference on Machine Learning, 2003*, pp. 856–863.
- [18] Y. Zhang, Z. Guo, W. Wang, S. He, T. Lee, M. Loew, A comparison of the wavelet and short-time fourier transforms for Doppler spectral analysis, *Med. Eng. Phys.* 25 (7) (2003) 547–557.