

RESEARCH ARTICLE

Automatic mouse ultrasound detector (A-MUD): A new tool for processing rodent vocalizations

Sarah M. Zala^{1*}, Doris Reitschmidt¹, Anton Noll², Peter Balazs², Dustin J. Penn¹

1 Konrad Lorenz Institute of Ethology, Department of Integrative Biology and Evolution, University of Veterinary Medicine, Vienna, Austria, **2** Acoustic Research Institute, Austrian Academy of Sciences, Vienna, Austria

* sarah.zala@vetmeduni.ac.at



Abstract

House mice (*Mus musculus*) emit complex ultrasonic vocalizations (USVs) during social and sexual interactions, which have features similar to bird song (i.e., they are composed of several different types of syllables, uttered in succession over time to form a pattern of sequences). Manually processing complex vocalization data is time-consuming and potentially subjective, and therefore, we developed an algorithm that automatically detects mouse ultrasonic vocalizations (*Automatic Mouse Ultrasound Detector* or A-MUD). A-MUD is a script that runs on STx acoustic software (S_TOOLS-STx version 4.2.2), which is free for scientific use. This algorithm improved the efficiency of processing USV files, as it was 4–12 times faster than manual segmentation, depending upon the size of the file. We evaluated A-MUD error rates using manually segmented sound files as a ‘gold standard’ reference, and compared them to a commercially available program. A-MUD had lower error rates than the commercial software, as it detected significantly more correct positives, and fewer false positives and false negatives. The errors generated by A-MUD were mainly false negatives, rather than false positives. This study is the first to systematically compare error rates for automatic ultrasonic vocalization detection methods, and A-MUD and subsequent versions will be made available for the scientific community.

OPEN ACCESS

Citation: Zala SM, Reitschmidt D, Noll A, Balazs P, Penn DJ (2017) Automatic mouse ultrasound detector (A-MUD): A new tool for processing rodent vocalizations. PLoS ONE 12(7): e0181200. <https://doi.org/10.1371/journal.pone.0181200>

Editor: Brenton G. Cooper, Texas Christian University, UNITED STATES

Received: February 14, 2017

Accepted: June 26, 2017

Published: July 20, 2017

Copyright: © 2017 Zala et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information file.

Funding: This research was supported by grants (Austrian Science Fund (AT) P 24711-B21 to DJP and Austrian Science Fund (AT) P 28141-B25 to DJP and SMZ) of the Austrian Science Foundation (<http://www.fwf.ac.at>).

Competing interests: The authors have declared that no competing interests exist.

Introduction

House mice (*Mus musculus*) emit a wide repertoire of vocalizations across five octaves (from ca. 3 to more than 120 kHz), which are mostly ultrasonic vocalizations (USVs), beyond the range of human hearing (>20 kHz). Mice begin emitting USVs as pups, which function as ‘isolation calls’ that induce maternal retrieval [1], and adults of both sexes produce complex USVs during courtship and other social contexts (see reviews [2, 3–9]). Adult males, for example, emit USVs when presented with a female mouse, or their scent, and during courtship and copulation (laboratory mice [10–16]; wild house mice [17]). Vocal production is generated by a ‘glottal jet’ mechanism [18], and the USVs of male mice are innate since they do not use [19]

or require [20, 21] auditory feedback (vocal learning). On the other hand, females acquire their auditory preferences for the USVs of unrelated males [17] at an early age, through auditory learning (familial imprinting) [22]. USV emission is under neuro-endocrine control [2, 9], and varies among individuals depending upon age, sex, genetic background, social status, health, and other factors [2, 8, 23–25]. Individuals can modulate USV emission during social and sexual interactions, depending upon their motivation or behavioral state, target receiver, and other contexts [8]. The signals transmitted in USVs (‘information content’, ‘meaning’ or ‘semantics’) and other possible functions are not well understood, though playback experiments with wild house mice indicate that they are sufficient for mice to recognize and discriminate *Mus* species, close kin, and individuals [17, 26, 27]. There are still few studies on wild house mice [25, 28–31], however, and none in natural or naturalistic ecological or social conditions. Most USV studies have focused on laboratory mice (*Mus laboratorius*) [32], which are used as a model system to investigate the genetic basis of communication and disease phenotypes [33], including human speech, autism and other neuropsychiatric disorders [4, 5]. Mouse vocalizations may also provide useful non-invasive indicators of disease and animal welfare [34–36]. Thus, research on mouse USVs is increasing, and this work has largely been inspired by the discovery of their complexity and similarity to birdsong [15], which was made possible by technical developments in bioacoustic analyses.

Mouse USVs are structurally and temporally complex at multiple levels of acoustic analyses [8]. First, USVs can be classified into different types of *syllables*, which consist of single notes (pure tones) or complex syllables with frequency jumps, and can be classified according to their particular spectrotemporal features, i.e., amplitude, frequency and duration (‘syllable acoustics’). Mice emit as many as 12 different types of syllables, and the number (‘vocal repertoire’) and the frequency in the usage of different syllables (‘prevalence’) can vary within and among mice. Syllable classifications vary among researchers [8, 15, 29, 37–40], and the challenge is to determine whether or how mice perceive and respond to different syllables. Laboratory mice can discriminate at least four types based on their ability to be trained by reinforcement (operant conditioning) [41]. Second, mice vary the number of syllables emitted per minute (‘vocalization rate’). Males emit around 13 to 90 syllables/min upon presentation of a female urine stimulus (in wild mice [29] or laboratory mice [42], respectively), and laboratory mice produce up to 160 to 230 syllables/min when presented with an anesthetized mouse or an intruder [43]. Third, mice emit USV syllables in phrases or bouts, composed of a series of syllables, usually more than one type, in succession over time to form a pattern of sequences [15]. The sequence of different syllables can vary (‘syntax’), as well as the number and type of different bouts (Markovian chains) [30, 31]. Thus, there is a surprisingly amount of complexity in mouse USVs, and better and more efficient methods for data processing and analyzing are needed to improve our understanding of these vocalizations.

Developing methods for automatically detecting and classifying types of USVs (syllables) presents a difficult technical challenge, especially for large sound files and recordings of mice in social conditions. Previous studies have usually relied on manually counting the syllables in spectrograms, which is time intensive (e.g., [17, 24, 28, 39, 42]). Other studies have used an in-house script for automated or semi-automated detection (e.g., [15, 30]); or used commercial software to process recordings automatically or semi-automatically, e.g., *Avisoft SAS LabPro* (Avisoft Bioacoustics, Germany) [20, 38, 43] and *Sound Analysis Pro* (SAP) [26, 29]. The error rates of these automated detection methods have never been evaluated to our knowledge, aside from a short paragraph in one study [15], and concerns have been raised about their reliability [17, 24, 28, 30, 39, 42, 44]. One of the main challenges of developing automatic detection methods is that USVs have a very low signal-to-noise ratio (SNR), and broadband interference from ambient, background noise (i.e., sounds other than previously described mouse USVs), which

can generate false positives or mask the signals (false negatives). We developed an algorithm to automatically detect USV syllables (*Automatic Mouse Ultrasound Detector* or A-MUD), which is a script run in STx, S_TOOLS-STx (Acoustic Research Institute, Austria). We measured the error rates of A-MUD using manually segmented files as a ‘gold standard’ reference, and we compared its performance with a commercially available program. This first version of A-MUD is a step towards improving error rates and developing automated syllable classification, and our aim is to provide a tool that will help researchers improve the efficiency of their analyses.

Materials and methods

Subjects and housing

Our study was conducted with wild-derived house mice (*Mus musculus musculus*), the F1 offspring of wild mice caught at the Konrad Lorenz Institute of Ethology (48° 12'38"N, 16° 16'54"E) in Vienna, Austria. Previous studies on this subspecies were conducted only on male USVs elicited by urinary odors, and female responses to male USV playbacks [2, 17, 26–29]. Mice were raised in mixed-sex family groups (standard Type IIL cages, 36.5 x 20 x 14 cm, with stainless steel cover, 1cm mesh width, Tecniplast, Germany) until weaning (21 d of age). Siblings were housed in mixed-sex groups (maximum of four mice per cage) until 5 weeks of age, when the sexes got separated. Males were individually housed to prevent fighting and females were housed in sister pairs. All cages were provided equally with wood shavings (ABEDD, Austria), nesting material (Nestlet, Ehret, Austria), one cardboard paper roll and one nest box (Tecniplast, Germany) for environmental enrichment. Food (rodent diet 1324, Altromin, Germany) and water were provided *ad libitum*. Mice were kept in standard conditions (mean \pm SD room temperature: 22 \pm 2°C, in a 12:12 h light:dark cycle, lights off at 15:00). Red light was used instead of a complete dark period to be able to conduct experiments during the mice active period without disturbing them. We worked with 32 adult mice (mean \pm SD age: 192 \pm 19d; n = 11 males, n = 21 females). We did not sacrifice any of the mice used for this study.

Recording apparatus

We recorded the vocalizations of males (n = 11) and females (n = 3) (‘callers’) in the presence of a female (‘stimulus’) under red light, during the active period of the day for our mice (15:00 to 17:30). The callers were primed (or socially experienced) for 5 min 1 d prior to the recordings by placing a female into their home cage. Some stimulus females were used once as priming animals and once as stimuli, but never for the same caller. The callers were always unfamiliar and unrelated to the priming and stimulus females. The recording apparatus consisted of a Plexiglas cage (36.5 x 21 x 15 cm) divided into two equal compartments, the ‘caller’ and the ‘stimulus compartments’. To ensure that the mice could see and smell each other during the experiment, the two compartments were separated by a 0.5 cm thick Plexiglas divider covered with small holes (0.5 cm diameter). The caller compartment was covered with a metal cage lid (1 cm width mesh), whereas the stimulus compartment was covered with a Plexiglas lid to prevent USVs from being recorded. This design ensured that we recorded vocalizations of the caller (the focal mouse), and not the stimulus mouse, and preliminary tests using USVs playbacks released from an ultrasound speaker (Avisoft Bioacoustics, Germany) positioned into the stimulus compartment, confirmed that the Plexiglas cover was very effective at blocking USVs. The stimulus compartment was also provided with bedding and 2–3 food pellets.

We always used a small plastic cylinder to introduce the mice into their respective compartments. To record, we first placed the stimulus female into the assigned compartment and after 5–10 min habituation time we introduced the focal mouse. The entire cage was then

positioned inside a recording chamber, which was lined with acoustic foam as described in [27]. A condenser ultrasound microphone (Avisoft Bioacoustics/CM16/CMPA with an integrated pre-amplifier and a frequency range from 10 to 200 kHz) and an UltraSoundGate 116–200 (Avisoft Bioacoustics, Germany) were mounted inside the recording chamber, 10 cm above the caller compartment. Before each recording, the microphone was calibrated with a 440 Hz tone of a commercial available tuning fork. Mice were recorded using the RECORDER USGH software with settings at 300 kHz sampling rate, 16 bit format, and 256 Hz FFT size. After positioning the cage inside the recording chamber, we waited for 30 sec and then started recording for 10 min. To avoid and also to standardize any potential estrus status effects of the stimulus female on the caller, we added an additional olfactory stimulus (5 µl of 4 different pooled female urine on a 4 x 4 cm filter paper) into the caller’s compartment. The urine was previously collected in metabolic cages (Techniplast, 600M021) from wild-caught adult females, equally aliquoted and mixed in Eppendorf tubes and stored at -20°C until the recordings. After each recording the entire cage was cleaned with ethanol before reusing.

Development and implementation of the *Automatic Mouse Ultrasound Detector* (A-MUD)

We implemented a segmentation algorithm in a new script (*Automatic Mouse Ultrasound Detector* or A-MUD 1.0) in STx (S_TOOLS-STx version 4.2.2), a software from the Acoustic Research Institute (Austria), which is free for scientific use (<http://www.kfs.oeaw.ac.at/stx>). STx is used for processing large quantities of data in a timely fashion [45, 46], such as for speech analysis [47, 48], noise evaluation [49, 50], and psychoacoustics [51]. It is designed to organize and process large collections of signal and segment data and to implement a large number of signal processing algorithms and elaborate interactive tools. STx has been used for bio-acoustic projects [52], particularly for signal detection and segmentation. It provides a scripting language for extending the software for problem-specific applications. The scripting language can access all the program interfaces (e.g., graphic, file system), the signal and segment data and the extensive internal signal-processing functions. In addition, STx can be extended using C/C++ functions, where it is possible to perform complex methods in a timely fashion. Moreover, STx provides all the necessary functions and tools to interactively check and correct the automatically generated segments, and it also includes flexible and programmable export methods to facilitate segment analysis in other programs, such as R or Microsoft Excel.

Development of A-MUD and its segmentation algorithm. As previously mentioned, one of the main challenges of automatically detecting USVs is the very low signal-to-noise ratio (SNR), as USV signals often have broadband interference or ambient noise, which can partially mask the signals. Mouse USVs occur between 20–120 kHz, and are nearly mono-frequency signals, similar to a whistle. They have a narrow bandwidth, and modulate over a relatively large frequency range, and their duration is 5–100 ms (though some researchers also include < 5 ms sounds, e.g., [38, 39, 44]). Here, we propose a new method that can cope with low SNR. To segment the signal and separate the USV from the background, the script uses the narrow bandwidth, which is a characteristic property not shared by the other part of the recorded signal. All the necessary calculations for the signal detection are done in the time-frequency domain. A short-time Fourier transformation (STFT) [53, 54] with window w of length N_{win} is applied to the signal f with hop size H and FFT-length N_{FFT} :

$$STFT_w(f)[k, l] = \sum_{j=0}^{N_{FFT}-1} f[j] \cdot w[j - kH] \cdot e^{-2\pi i (j-l) N_{FFT}}$$

A Hanning window was chosen for this task. The high sampling rate (250 or 300 kHz) and the shortness and frequency dynamics of the signal require the window length to be between 2.5 and 3.3 ms (choosing $N_{FFT} = N_{win}$ this leads to a frequency resolution being $\Delta f = 300\text{--}400$ Hz) and an overlap of between 75% and 85% (hop size $H \sim 0.5$ ms), which was determined empirically to provide a good time-frequency representation of the USVs. All further calculations are restricted to the power spectra (PS) in the relevant frequency range between 30–120 kHz, assumed to result in N frequency bins. We performed following three steps:

Step 1: Noise reduction (pre-whitening). For de-noising we used a pre-whitening step (see e.g., [55]). The signal energy (*rms*) of the power spectra was calculated for each frame. As an estimation of the maximum level of the background noise, we set $L_{noise} = L_{95} + 2\text{dB}$, where L_{95} is the 95% quantile of all *rms* values. The background noise spectrum was estimated by averaging the spectra of all frames with a level less than L_{noise} . All power spectra were then multiplied with the inverse background spectrum, for the removal of stationary noise. This procedure of pre-whitening has great effects when applied in methods using amplitude as a criterion [56].

Step 2: Compute the segmentation parameter track and set thresholds. The narrowness of the signal's bandwidth (*ebw*) used for segmentation, was defined as the number of frequency bins for which a certain ratio R_{ebw} of the total spectrum energy was achieved, i.e., *ebw* is the smallest number for which

$$R_{ebw} \cdot \sum_{i=1}^N PS_i \leq \sum_{i=1}^{ebw} PS_i.$$

Here, the power spectrum is sorted by descending amplitude. This is done for every time step. This ratio was chosen between $0.3 < R_{ebw} < 0.6$. The inverse of *ebw* is used as the main segmentation criterion. We define the energy concentration by $ec = \frac{N}{ebw}$. Two threshold values for the segmentation algorithm are calculated using the *ec* function: $ec_{on} = 10\%$ quantile of the *ec* function and $ec_{off} = 90\%$ quantile of the *ec* function. Windows with high *ec* can be considered as containing a signal and those with low *ec* as noise.

Step 3: Detect segments and apply time corrections. The *ec* function is then used to detect segments. The *ec* function is searched for a local maximum greater than the ec_{on} threshold. The *ec* function is then applied from this position forward and backwards, and until a point where the *ec* value falls below the ec_{off} threshold. These points are used as the beginning and end positions of the segment. This procedure is repeated until no more matching local maxima are found.

Two correctional steps are applied to the identified segments. First, any sequential segments closer than the minimal distance $t_{mindist}$ (5–10 ms) are merged. Second, any segments shorter than the set minimum length t_{mindur} (5–10 ms) or longer than the maximum length t_{maxdur} (150–200 ms) are removed. The temporal thresholds $t_{mindist}$, t_{mindur} and t_{maxdur} , and all other parameters of the algorithm are derived from a heuristic pre-test and analysis of a small set of manually segmented test signals. The method was then evaluated using a much larger signal set.

Implementing and testing A-MUD. Although the USV detection algorithm is straightforward, it requires high computational effort due to the high sampling rate (up to 300 kHz) and the overlapping necessary for the short length of the calls. The script automatically detects USVs in the recorded sound file. In addition, A-MUD also provides the spectrographic analyses of each detected *element* (i.e., these are candidate syllables) providing its frequency, amplitude and time parameters (Table 1). The resulting segments and their extracted frequency progression parameters are stored as signal metadata in an XML-file. Computation is taking 1.5 to 2.5 times longer than the signal (file) length.

Table 1. USV parameters and definitions in A-MUD.

Parameter	Definition of the parameter (measurement unit)
begin	start of the element (s)
length	length of the element (s)
fmean	mean frequency of the element (Hz)
fband	frequency bandwidth (fmax—fmin) (Hz)
amean	mean amplitude of the element (dB)
t1	start point of the element (ms)
f1	frequency at start point of the element (Hz)
a1	amplitude at start point of the element (dB)
tn	end point of the element (ms)
fn	frequency at endpoint of the element (Hz)
an	amplitude at endpoint of the element (dB)
tfmin	time point of lowest frequency of the element (ms)
fmin	lowest frequency of the element (Hz)
afmin	amplitude at the point of lowest frequency (dB)
tfmax	time point of highest frequency of the element (ms)

Spectrographic parameters detected by A-MUD.

<https://doi.org/10.1371/journal.pone.0181200.t001>

Initially, recordings from four male callers were used to develop the A-MUD algorithm and to choose the appropriate values for the given parameters. We then evaluated error rates of A-MUD and a commercially available software using 10 additional recordings (7 male and 3 female callers). The 10 recordings cover the range of calling rates typical for wild-derived mice ([17, 27, 29]; Zala et al. unpublished data). These 10 recordings were manually segmented three different times to obtain a gold standard reference (see below). We used the gold standard reference first to assess the inter-observer reliability within STx and the intra-observer reliability between both programs, and then to estimate the error rates of the two automatic call detection methods. Thus, two independent observers manually segmented the files in STx, and in addition one manual segmentation was performed in the commercially available software (same observer for both programs). The observers marked the beginning and the end of each syllable and classified 12 different syllable types depending on their duration, frequency and frequency modulation, according to previous classification [27, 29, 40]. We then compared the error rates of A-MUD and the commercially available program. The automatic analysis in the commercially available software was conducted using two different settings. First, using the setting ‘whistle tracking’, which is based on detecting steady signals without rapid frequency modulations and is recommended for analyzing soft whistle like sounds in noisy background such as short syllables of rodent USVs. Second, using the setting ‘single threshold’, which detects all elements above a specified amplitude threshold.

Each of the 10 sound files was thus processed six times using both programs and as follows: (1) commercially available software, manual observer 1 (n = 10), (2) commercially available software, setting ‘whistle tracking’ (n = 10), (3) commercially available software, setting ‘single threshold’ (n = 10), (4) STx, manual observer 1 (n = 10), (5) STx, manual observer 2 (n = 10), (6) STx, A-MUD (n = 10). The settings chosen to create the spectrograms for manual segmentation were the best time and frequency resolutions for visually analyzing mouse USVs in both programs. The spectrograms for the manual segmentations in the commercially available software were generated using following settings: FFT length = 512, frame size = 100% (flat top window) and overlap = 50%. The settings of the automatic processing in the commercially

available software (single threshold) were: threshold = -50 dB and hold time = 20 ms; and the settings for the same software (whistle tracking) were: tolerated maximum change of frequency modulation 10 253 Hz, hold time = 20 ms. For all automatic analyses in the commercially available software only frequencies > 25 kHz and only elements \geq 10 ms were included. All the settings were empirically chosen, to maximize the quality of USV detection. For the manual segmentation in STx we scrolled through the whole spectrogram in steps of 2 seconds with the overlap of 25%. Spectrograms were generated with a range of 50 dB, frame = 4 ms and an overlap of 75%. The spectrograms used a Hanning window and displayed frequencies between 25–150 kHz. For automatic processing in STx, we ran the script A-MUD 1.0. This script's threshold of element duration was set at 10 ms.

Calculating error rates. We first confirmed the reliability of our manual segmentation by comparing the total number of manually detected elements between individual observers and within the same observer using the two detection programs (as their spectrograms present visually different patterns). After confirming the reliability of our manual detection (see Results), we used these manually segmented data as our gold standard reference to calculate the percentage of elements that were detected correctly (correct positives, 'correct'), missed incorrectly ('false negatives'), and detected incorrectly ('false positives') by the two automatic call detection programs. Thus, we compared the 'manual reference elements' (Nref) with the 'corresponding elements' (Ncorr). Nref was calculated using only the elements that were detected by all three manual segmentations, i.e. the gold standard. Ncorr identified the elements that were detected by both the manual (Nref) and the automatic processing methods (Nauto) within each software separately. We used following formulas:

$$\text{Correct} = (100 * \text{Ncorr}) / \text{Nauto}$$

$$\text{False positives} = [100 * (\text{Nauto} - \text{Ncorr})] / \text{Nauto}$$

$$\text{False negatives} = [100 * (\text{Nref} - \text{Ncorr})] / \text{Nref}$$

Error estimates were calculated for all 10 files first by using the entire manually segmented dataset Nref, which also contained short USV elements < 10 ms (Fig 1), and then again after excluding these short elements from the gold standard (Fig 2), and here we explain the reason for this second analysis. Short sounds in USV recordings pose a technical challenge for automatic (and sometimes even manual) detection because they are very difficult to distinguish from background noise. We therefore set a threshold in A-MUD and in the commercial software to detect only sounds \geq 10 ms, as described above, to reduce the number of false positives [6]. The downside of this threshold is that both methods will fail to detect very short elements that mice emit (false negatives). This is an acceptable tradeoff [57], as long as most of these short sounds are background noise and not USVs. However, to evaluate the error rate for calls \geq 10 ms, which both methods can detect, we repeated the analysis after omitting the short elements from the manual gold standard (Fig 2). For this second error estimate, five files with < 50 manually detected elements were excluded, as larger files likely have more reliable percentages of error rates. Thus, this second analysis provides estimates of error rates for USV detection within the threshold, and it is based on files containing the largest number of calls.

Results are reported as mean \pm 1 standard deviation, and statistical analyses were conducted in IBM SPSS Statistics 22. To analyze amount of USV emission, we used the total number of elements detected per file ('number of elements'). We performed non-parametric statistics as the assumptions of parametric statistics were not met, used two-tailed tests, and results are considered statistically significant at $\alpha \leq 0.05$.

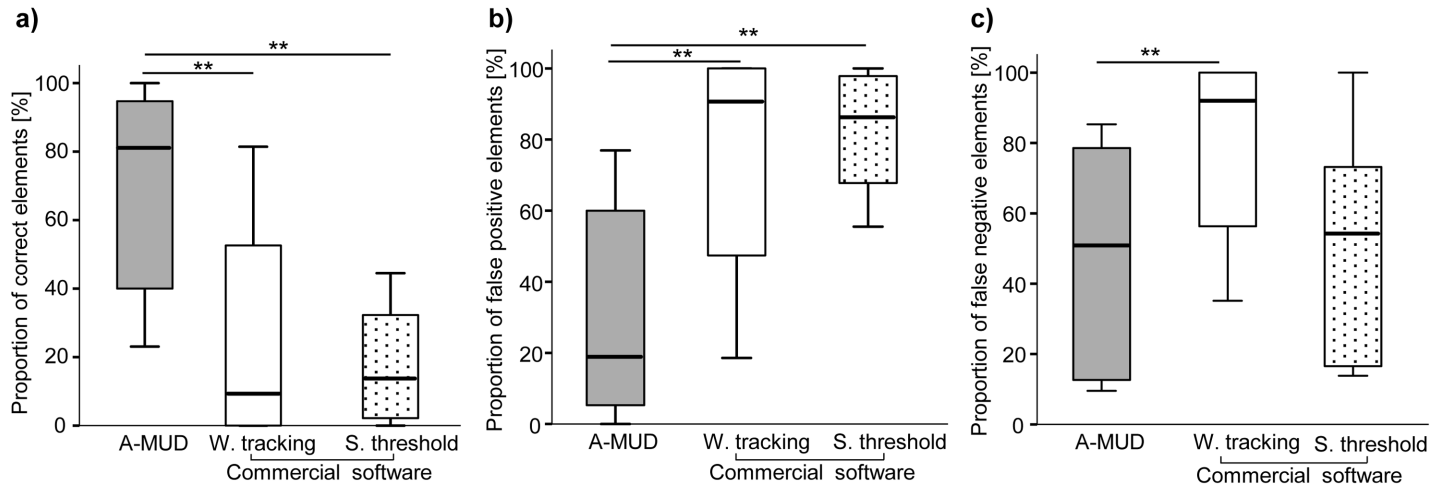


Fig 1. Evaluation of error rates among automatic USV detection methods (all recordings). Boxplots showing the percentage of correct positives (a), false positives (b) and false negatives (c) comparing three automatic processing methods: A-MUD (grey) and a commercially available software using ‘whistle tracking’ (white) or ‘single threshold’ (stippled) settings. The graph shows median \pm 95% CI, including the 25th and the 75th percentiles. ** = $p \leq 0.01$.

<https://doi.org/10.1371/journal.pone.0181200.g001>

Ethical statement

This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All the experiments were conducted at the Konrad Lorenz Institute of Ethology, Austria and the protocols have been approved and were in accordance with ethical standards and guidelines in the care and use of experimental animals of the Ethical and Animal Welfare Commission of the University of Veterinary Medicine Vienna (Austria).

Results

We first confirmed the reliability of our gold standard, and we compared manual segmentation results between observers, who both visualized spectrograms using the same program

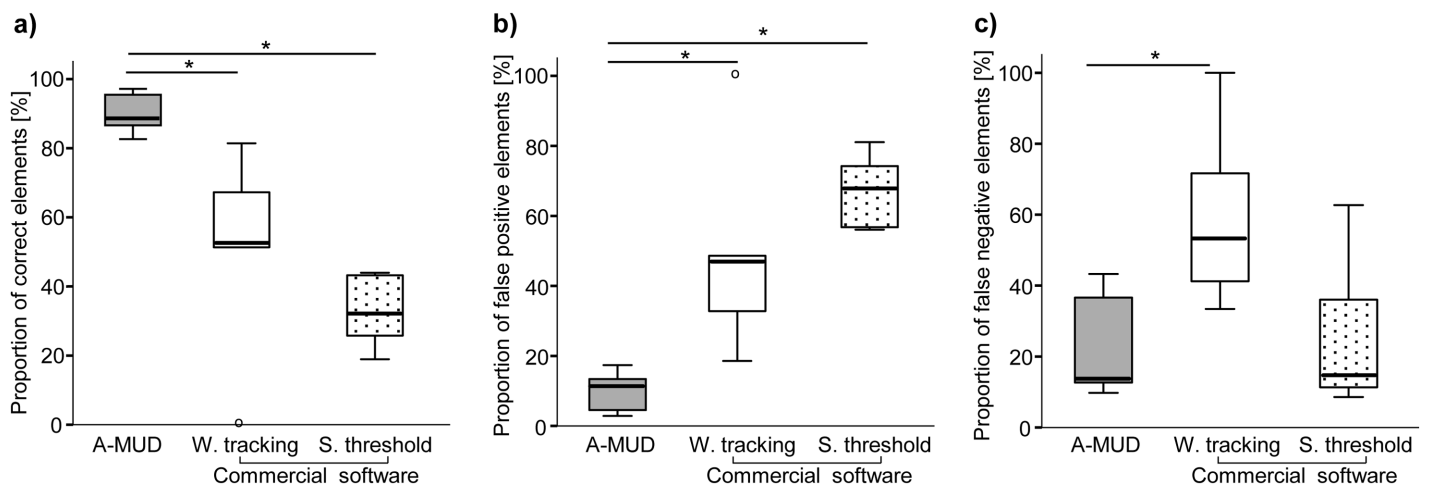


Fig 2. Evaluation of error rates among automatic USV detection methods (selected recordings). Boxplots showing the percentage of correct positives (a), false positives (b) and false negatives (c) with three automatic processing methods: A-MUD (grey) and a commercially available software using ‘whistle tracking’ (white) or ‘single threshold’ (stippled) settings. The graph shows median \pm 95% CI, including the 25th and the 75th percentiles. * = $p \leq 0.05$. ° = outliers.

<https://doi.org/10.1371/journal.pone.0181200.g002>

(STx), and between different programs, evaluated by the same observer. The number of elements detected was highly correlated between observers and between programs (Spearman's ρ , $n = 10$, $\rho = 0.99$, $p < 0.001$ for both correlations). Thus, manual detection was highly repeatable regardless of the observer or the software used for visualizing spectrograms, suggesting that any potential influence of different parameters used to generate the spectrograms within the two programs is negligible. We then used our generated gold standard (Nref) to calculate and compare error rates of A-MUD with a widely used commercial software. First, we calculated the proportion of correct positives, false positive, and false negative elements (candidate syllables; see Methods) for both A-MUD and the commercially available program (with the two different settings) (Fig 1). We found that A-MUD was significantly more reliable for detecting (a) correct positives (A-MUD vs. commercial software, whistle tracking: Wilcoxon signed ranks test, $Z = -2.8$, $n = 10$, $p = 0.005$ and A-MUD vs. commercial software, single threshold: Wilcoxon signed ranks test, $Z = -2.8$, $n = 10$, $p = 0.005$; Fig 1A); (b) avoiding false positives (A-MUD vs. commercial software, whistle tracking: Wilcoxon signed ranks test, $Z = -2.8$, $n = 10$, $p = 0.005$ and A-MUD vs. commercial software, single threshold: Wilcoxon signed ranks test, $Z = -2.8$, $n = 10$, $p = 0.005$; Fig 1B); and (c) avoiding false negatives (A-MUD vs. commercial software, whistle tracking: Wilcoxon signed ranks test, $Z = -2.8$, $n = 10$, $p = 0.005$ and A-MUD vs. commercial software, single threshold: Wilcoxon signed ranks test, $Z = -0.56$, $n = 10$, $p = 0.58$; Fig 1C). Thus, A-MUD had lower error rates than the commercial software for most comparisons.

Second, we re-calculated error rates after removing the short elements (<10 ms) from the gold standard, and excluding small files with only few (<50) manually detected elements. Again, we found that A-MUD was significantly more reliable than the commercial software (Fig 2) for most comparisons: (a) correct positives (A-MUD vs. commercial software, whistle tracking: Wilcoxon signed ranks test, $Z = -2.02$, $n = 5$, $p = 0.04$ and A-MUD vs. commercial software, single threshold: Wilcoxon signed ranks test, $Z = -2.02$, $n = 5$, $p = 0.04$; Fig 2A); (b) false positives (A-MUD vs. commercial software, whistle tracking: Wilcoxon signed ranks test, $Z = -2.02$, $n = 5$, $p = 0.04$ and A-MUD vs. commercial software, single threshold: Wilcoxon signed ranks test, $Z = -2.02$, $n = 5$, $p = 0.04$; Fig 2B); and (c) false negatives (A-MUD vs. commercial software, whistle tracking: Wilcoxon signed ranks test, $Z = -2.02$, $n = 5$, $p = 0.04$ and A-MUD vs. commercial software, single threshold: Wilcoxon signed ranks test, $Z = -0.14$, $n = 5$, $p = 0.89$; Fig 2C).

Thus, A-MUD had significantly lower error rates than the commercial software in both approaches of evaluation. The second method arguably provides the more informative assessment of A-MUD and the commercial software for calls ≥ 10 ms, as it excluded sounds below the defined threshold, which is a known constraint. The results indicate that A-MUD had more correct positives, fewer false positives, and fewer false negatives compared than the commercial software (Fig 2). The errors for A-MUD were mainly false negatives (mean: $23 \pm 16\%$), rather than false positives ($10 \pm 6\%$).

Discussion

Our aims were to develop an algorithm for automatically detecting mouse USVs, evaluate its performance using manual segmentation (gold standard), and compare error rates with a commercially available program. The main challenge for developing automatic USV detection methods, as for any signal detection task, is to minimize false positives from background noise. We developed an algorithm for automatic USV detection (A-MUD 1.0), and in this validation, we set a threshold so that it would not detect sounds with a very short (<10 ms) duration. This approach probably reduces false positives, but misses very short vocalizations. We considered this to be an acceptable trade-off since most USVs are ≥ 10 ms, and the problematic

background noise is often <10 ms. We recorded USVs of wild-derived mice, and used these recordings to evaluate the performance of A-MUD and one commercially available software program. We evaluated error rates using manual segmentation (gold standard reference), which we confirmed to be highly repeatable (between observers and different programs used for manual segmentation), and then we compared the error rates of A-MUD and the commercial program using the same 10 sound files. We first estimated error rates using all of the candidate syllables (elements) in the manual segmentation. However, since short elements in the manual reference were not detected by A-MUD (or the commercial software) due to the threshold, we also evaluated error rates for elements ≥ 10 ms. Thus, in our second evaluation, we removed short elements from the manual segmentation files, and in addition, we only used files with a large number of USVs (>50 manually detected elements). Again, we found that A-MUD was significantly more reliable for detecting USV elements (correct positives, i.e. confirmed syllables) and also for avoiding false positives and false negatives compared to the commercial software (tested with two different settings) (Figs 1 and 2). Thus, both evaluation methods are consistent and lead to the same conclusion. Errors in A-MUD were mostly false negatives (mean \pm SD: $23\pm 16\%$) rather than false positives ($10\pm 6\%$). The false negatives can mainly be attributed to short elements, which can be reduced by increasing the overlap (reducing the hop size), but this change would also increase the computational effort. In addition, we found that low amplitude elements (with a low signal-to-noise-ratio) were often not detected (false negatives).

As expected, A-MUD greatly improved the efficiency of processing USV sound files (i.e., the processing speed was 4–12 times faster than manual segmentation, depending upon the size of the file). A-MUD required ca. 3.5 h for processing all 14 sound files used to develop and validate it, whereas manual segmentation required ca. 30 h. For processing each 10 min recording, A-MUD required 12–15 min per file (the duration depends on the PC processor speed and we used an Intel[®] Core™ i5-3470 Processor and 8 GB of RAM), whereas manual segmentation required 1–3 h per file, depending on the number of syllables.

Thus, A-MUD provides a fast and reliable method for processing USV data, and it outperformed at least one other method, which is often used in USV studies. More such comparisons of methods would be highly valuable for the field, though it would require publishing the codes and other information necessary for the exact re-implementation (e.g., exact window information for the sonogram, noise thresholds, filtering methods, etc.). We are making this first version of A-MUD available for the scientific community (non-commercial use) at <https://www.kfs.oeaw.ac.at/doc/amud/AMUD1b.sts> (Script); Readme: <https://www.kfs.oeaw.ac.at/doc/amud/AMUD1b-Readme.odt>. This link also provides the code for A-MUD 1.0. Additional comparisons, such as analyzing A-MUD with sound files generated by other mice or under different recording conditions are now more feasible with the recent development of an online database for uploading mouse vocalizations (*mouseTube*) [58]. It is unclear whether the error rates we observed in our study will apply to other mice or conditions. Reliable automatic USV detection is especially challenging for recording animals during direct interactions due to the increased background noise from activity. The main challenge is to develop simple and accurate techniques that make it possible to distinguish and identify an individual's USVs during social interactions [16]. Under such conditions, we find that automatic call detection is error-prone. We are currently developing an improved version of A-MUD to reliably detect and classify syllables when mice are physically interacting.

Supporting information

S1 Table. Data: Automatic Mouse Ultrasound Detector (A-MUD).
(XLSX)

Acknowledgments

We thank H. Sasse and U. Madlspurger for animal care, J. Nicod for encouragement to develop a method for automatic call detection, M. Kalcounis-Rüppell, M. A. Marconi and an anonymous reviewer for helpful comments on a previous version of the manuscript and R. Hengsberger for helping with formatting the manuscript.

Author Contributions

Conceptualization: DJP.

Data curation: SMZ DR AN PB DJP.

Formal analysis: SMZ DR AN PB DJP.

Funding acquisition: DJP PB.

Investigation: SMZ DR.

Methodology: SMZ DR AN PB DJP.

Project administration: DJP.

Resources: DJP PB.

Software: AN PB.

Supervision: DJP PB.

Validation: SMZ DR AN.

Visualization: SMZ DR.

Writing – original draft: SMZ DR PB DJP.

Writing – review & editing: SMZ DR AN PB DJP.

References

1. Ehret G. Infant Rodent Ultrasounds—A Gate to the Understanding of Sound Communication. *Behav Genet.* 2005; 35(1):19–29. <https://doi.org/10.1007/s10519-004-0853-8> PMID: 15674530
2. Musolf K, Penn DJ. Ultrasonic vocalizations in house mice: a cryptic mode of acoustic communication. In: Macholán M, Baird SJE, Munclinger P, Piálek J, editors. *Evolution of the House Mouse*. Cambridge University Press; 2012. p. 253–77.
3. Lahvis GP, Alleva E, Scattoni ML. Translating mouse vocalizations: prosody and frequency modulation. *Genes Brain Behav.* 2011; 10:4–16. <https://doi.org/10.1111/j.1601-183X.2010.00603.x> PMID: 20497235
4. Scattoni ML, Crawley J, Ricceri L. Ultrasonic vocalizations: a tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. *Neurosci Biobehav Rev.* 2009; 33(4):508–15. <https://doi.org/10.1016/j.neubiorev.2008.08.003> PMID: 18771687
5. Fischer J, Hammerschmidt K. Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav.* 2011; 10(1):17–27. <https://doi.org/10.1111/j.1601-183X.2010.00610.x> PMID: 20579107
6. Arriaga G, Zhou EP, Jarvis ED. Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. *PLoS ONE.* 2012; 7(10):e46610. <https://doi.org/10.1371/journal.pone.0046610> PMID: 23071596
7. Portfors CV, Perkel DJ. The role of ultrasonic vocalizations in mouse communication. *Curr Opin Neurobiol.* 2014; 28:115–20. <https://doi.org/10.1016/j.conb.2014.07.002> PMID: 25062471
8. Heckman J, McGuinness B, Celikel T, Englitz B. Determinants of the mouse ultrasonic vocal structure and repertoire. *Neurosci Biobehav Rev.* 2016; 65:313–25. <https://doi.org/10.1016/j.neubiorev.2016.03.029> PMID: 27060755

9. Egnor SER, Seagraves KM. The contribution of ultrasonic vocalizations to mouse courtship. *Curr Opin Neurobiol.* 2016; 38:1–5. <https://doi.org/10.1016/j.conb.2015.12.009> PMID: 26789140
10. Nyby J, Wysocki CJ, Whitney G, Dizinno G. Pheromonal regulation of male mouse ultrasonic courtship (*Mus musculus*). *Anim Behav.* 1977; 25(2):333–41. [https://doi.org/10.1016/0003-3472\(77\)90009-4](https://doi.org/10.1016/0003-3472(77)90009-4) PMID: 889149
11. Maggio JC, Whitney G. Ultrasonic Vocalizing by Adult Female Mice (*Mus musculus*). *J Comp Psychol.* 1985; 99(4):420–36. <https://doi.org/10.1037/0735-7036.99.4.420> PMID: 4075780
12. Warburton VL, Sales GD, Milligan SR. The Emission and Elicitation of Mouse Ultrasonic Vocalizations: The Effects of Age, Sex and Gonadal Status. *Physiol Behav.* 1989; 45(1):41–7. [https://doi.org/10.1016/0031-9384\(89\)90164-9](https://doi.org/10.1016/0031-9384(89)90164-9) PMID: 2727141
13. James PJ, Nyby JG, Saviolakis GA. Sexually stimulated testosterone release in male mice (*Mus musculus*): roles of genotype and sexual arousal. *Horm Behav.* 2006; 50(3):424–31. <https://doi.org/10.1016/j.yhbeh.2006.05.004> PMID: 16828762
14. Wang H, Liang S, Burgdorf J, Wess J, Yeomans J. Ultrasonic Vocalizations Induced by Sex and Amphetamine in M2, M4, M5 Muscarinic and D2 Dopamine Receptor Knockout Mice. *PLoS ONE.* 2008; 3(4):e1893. <https://doi.org/10.1371/journal.pone.0001893> PMID: 18382674
15. Holy TE, Guo Z. Ultrasonic Songs of Male Mice. *PLoS Biol.* 2005; 3(12):e386. <https://doi.org/10.1371/journal.pbio.0030386> PMID: 16248680
16. Neunuebel JP, Taylor AL, Arthur BJ, Egnor SER. Female mice ultrasonically interact with males during courtship displays. *eLife.* 2015; 4. <https://doi.org/10.7554/eLife.06203> PMID: 26020291
17. Musolf K, Hoffmann F, Penn DJ. Ultrasonic courtship vocalizations in wild house mice, *Mus musculus*. *Anim Behav.* 2010; 79(3):757–64. <https://doi.org/10.1016/j.anbehav.2009.12.034>
18. Mahrt E, Agarwal A, Perkel D, Portfors C, Elemans CPH. Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. *Curr Biol.* 2016; 26(19):R880–R1. <https://doi.org/10.1016/j.cub.2016.08.032> PMID: 27728788
19. Kikusui T, Nakanishi K, Nakagawa R, Nagasawa M, Mogi K, Okanoya K. Cross Fostering Experiments Suggest That Mice Songs Are Innate. *PLoS ONE.* 2011; 6(3):e17721. <https://doi.org/10.1371/journal.pone.0017721> PMID: 21408017
20. Hammerschmidt K, Reisinger E, Westekemper K, Ehrenreich L, Strenzke N, Fischer J. Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci.* 2012; 13(1):40. <https://doi.org/10.1186/1471-2202-13-40> PMID: 22533376
21. Mahrt EJ, Perkel DJ, Tong L, Rubel EW, Portfors CV. Engineered Deafness Reveals That Mouse Courtship Vocalizations Do Not Require Auditory Experience. *J Neurosci.* 2013; 33(13):5573–83. <https://doi.org/10.1523/JNEUROSCI.5054-12.2013> PMID: 23536072
22. Asaba A, Okabe S, Nagasawa M, Kato M, Koshida N, Osakada T, et al. Developmental Social Environment Imprints Female Preference for Male Song in Mice. *PLoS ONE.* 2014; 9(2):e87186. <https://doi.org/10.1371/journal.pone.0087186> PMID: 24505280
23. Nyby J, Dizinno GA, Whitney G. Social Status and Ultrasonic Vocalizations of Male Mice. *Behav Biol.* 1976; 18(2):285–9. [https://doi.org/10.1016/S0091-6773\(76\)92198-2](https://doi.org/10.1016/S0091-6773(76)92198-2) PMID: 999582
24. D'Amato FR. Courtship ultrasonic vocalizations and social status in mice. *Anim Behav.* 1991; 41:875–85. [https://doi.org/10.1016/S0003-3472\(05\)80354-9](https://doi.org/10.1016/S0003-3472(05)80354-9)
25. Lopes PC, König B. Choosing a healthy mate: sexually attractive traits as reliable indicators of current disease status in house mice. *Anim Behav.* 2016; 111:119–26. <https://doi.org/10.1016/j.anbehav.2015.10.011>
26. Hoffmann F, Musolf K, Penn DJ. Spectrographic analyses reveal signals of individuality and kinship in the ultrasonic courtship vocalizations of wild house mice. *Physiol Behav.* 2012; 105(3):766–71. <https://doi.org/10.1016/j.physbeh.2011.10.011> PMID: 22037196
27. Musolf K, Meindl S, Larsen AL, Kalcounis-Rueppell MC, Penn DJ. Ultrasonic Vocalizations of Male Mice Differ among Species and Females Show Assortative Preferences for Male Calls. *PLoS ONE.* 2015; 10(8):e0134123. <https://doi.org/10.1371/journal.pone.0134123> PMID: 26309246
28. Hoffmann F, Musolf K, Penn DJ. Freezing urine reduces its efficacy for eliciting ultrasonic vocalizations from male mice. *Physiol Behav.* 2009; 96(4–5):602–5. <https://doi.org/10.1016/j.physbeh.2008.12.014> PMID: 19150619
29. Hoffmann F, Musolf K, Penn DJ. Ultrasonic courtship vocalizations in wild house mice: spectrographic analyses. *J Ethol.* 2012; 30(1):173–80. <https://doi.org/10.1007/s10164-011-0312-y>
30. von Merten S, Hoier S, Pfeifle C, Tautz D. A Role for Ultrasonic Vocalisation in Social Communication and Divergence of Natural Populations of the House Mouse (*Mus musculus domesticus*). *PLoS ONE.* 2014; 9(5):e97244. <https://doi.org/10.1371/journal.pone.0097244> PMID: 24816836

31. Hoier S, Pfeifle C, von Merten S, Linnenbrink M. Communication at the Garden Fence—Context Dependent Vocalization in Female House Mice. *PLoS ONE*. 2016; 11(3):e0152255. <https://doi.org/10.1371/journal.pone.0152255> PMID: 27022749
32. Guénet J-L, Bonhomme F. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet*. 2003; 19(1):24–31. [https://doi.org/10.1016/S0168-9525\(02\)00007-0](https://doi.org/10.1016/S0168-9525(02)00007-0) PMID: 12493245
33. Crawley JN. What's wrong with my mouse? Behavioral phenotyping of transgenic and knockout mice. New York: Wiley; 2000. XIII, 329 p.
34. Manteuffel G, Puppe B, Schön PC. Vocalization of farm animals as a measure of welfare. *Appl Anim Behav Sci*. 2004; 88(1–2):163–82. <https://doi.org/10.1016/j.applanim.2004.02.012>
35. Kurejova M, Nattenmuller U, Hildebrandt U, Selvaraj D, Stösser S, Kuner R. An improved behavioural assay demonstrates that ultrasound vocalizations constitute a reliable indicator of chronic cancer pain and neuropathic pain. *Mol Pain*. 2010; 6:18. <https://doi.org/10.1186/1744-8069-6-18> PMID: 20346101
36. Cuomo V, De Salvia MA, Maselli MA, Santo L, Cagiano R. Ultrasonic Calling in Rodents: A New Experimental Approach in Behavioural Toxicology. *Neurotoxicol Teratol*. 1987; 9(2):157–60. [https://doi.org/10.1016/0892-0362\(87\)90093-6](https://doi.org/10.1016/0892-0362(87)90093-6) PMID: 3309610
37. Panksepp JB, Jochman KA, Kim JU, Koy JJ, Wilson ED, Chen Q, et al. Affiliative Behavior, Ultrasonic Communication and Social Reward Are Influenced by Genetic Variation in Adolescent Mice. *PLoS ONE*. 2007; 2(4):e351. <https://doi.org/10.1371/journal.pone.0000351> PMID: 17406675
38. Scattoni ML, Gandhi SU, Ricceri L, Crawley JN. Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS ONE*. 2008; 3(8):e3067. <https://doi.org/10.1371/journal.pone.0003067> PMID: 18728777
39. Grimsley JMS, Monaghan JJM, Wenstrup JJ. Development of Social Vocalizations in Mice. *PLoS ONE*. 2011; 6(3):e17460. <https://doi.org/10.1371/journal.pone.0017460> PMID: 21408007
40. Hanson JL, Hurley LM. Female Presence and Estrous State Influence Mouse Ultrasonic Courtship Vocalizations. *PLoS ONE*. 2012; 7(7):e40782. <https://doi.org/10.1371/journal.pone.0040782> PMID: 22815817
41. Neilans EG, Holfoth DP, Radziwon KE, Portfors CV, Dent ML. Discrimination of Ultrasonic Vocalizations by CBA/CAJ Mice (*Mus musculus*) Is Related to Spectrotemporal Dissimilarity of Vocalizations. *PLoS ONE*. 2014; 9(1):e85405. <https://doi.org/10.1371/journal.pone.0085405> PMID: 24416405
42. Rouillet FI, Woehr M, Crawley JN. Female urine-induced male mice ultrasonic vocalizations, but not scent-marking, is modulated by social experience. *Behav Brain Res*. 2011; 216(1):19–28. <https://doi.org/10.1016/j.bbr.2010.06.004> PMID: 20540967
43. Hammerschmidt K, Radyushkin K, Ehrenreich H, Fischer J. The Structure and Usage of Female and Male Mouse Ultrasonic Vocalizations Reveal only Minor Differences. *PLoS ONE*. 2012; 7(7):e41133. <https://doi.org/10.1371/journal.pone.0041133> PMID: 22815941
44. Scattoni ML, Ricceri L, Crawley JN. Unusual repertoire of vocalizations in adult BTBR T+tf/J mice during three types of social encounters. *Genes Brain Behav*. 2011; 10(1):44–56. <https://doi.org/10.1111/j.1601-183X.2010.00623.x> PMID: 20618443
45. Balazs P, Noll A, Deutsch WA, Laback B, editors. [Poster] Concept of the integrated signal analysis software system STX. Proceedings of the Austrian Physical Society Meeting 2000: Österreichische Physikalische Gesellschaft (ÖPG).
46. Balazs P, Noll A. [Presentation] Masking Filter, Phase Vocoder and STX—an introduction. SAMPTA03-Sampling Theory and Applications 20032003. p. 12.
47. Moosmüller S, Brandstätter J. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language Sciences*. 2014; 46(A):84–95. <https://doi.org/10.1016/j.langsci.2014.06.016> PMID: 25400323
48. Moosmüller S. Evaluation of the speech behaviour of reference speakers. *Int J Speech, Lang Law*. 2012; 18(2):179–205. <https://doi.org/10.1558/ijssl.v18i2.179>
49. Gygi B, Deutsch W, Noll A, Becker T, Hirner A, editors. NOIDESc: Incorporating feature descriptors into a novel railway noise evaluation scheme. 19th International Congress on Acoustics; 2007 2–7 Sept. 2007; Madrid.
50. Kasess C, Noll A, Majdak P, Waubke H. Effect of train type on annoyance and acoustic features of the rolling noise. *J Acoust Soc Am*. 2013; 134(2):1071–81. <https://doi.org/10.1121/1.4812771> PMID: 23927107
51. Balazs P, Laback B, Eckel G, Deutsch WA. Introducing Time-Frequency Sparsity by Removing Perceptually Irrelevant Components Using a Simple Model of Simultaneous Masking. *IEEE/ACM Trans Audio, Speech, Language Process*. 2010; 18(1):34–49. <https://doi.org/10.1109/TASL.2009.2023164>
52. Kratochvil H, Noll A, Bolldorf J, Parker AG. Acoustic method for testing the quality of sterilized male tsetse flies (*Glossina pallidipes*). In: IAEA, editor. Quality Control for Expanded Tsetse Production,

- Sterilization and Field Application. IAEA-TECDOC. 1683. Vienna: International Atomic Energy Agency (IAEA); 2012. p. 101–8.
53. Gröchenig K. Foundations of Time-Frequency Analysis. Basel: Springer Science+Business Media; 2001. p. 360.
 54. Oppenheim AV, Schafer R, Buck JR. Discrete-Time Signal Processing 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall; 1999. 870 p.
 55. Carmona R, Hwang W-L, Torr sani B. Practical time-frequency analysis. Wavelet and Gabor Transforms with an Implementation in S. San Diego: Academic Press; 1998. 490 p.
 56. Blackman RB, Tukey JW. The Measurement of Power Spectra. From the Point of View of Communications Engineering. New York: Dover Publications; 1959. 190 p.
 57. Wiley RH. Noise Matters. The evolution of communication. Cambridge, Massachusetts: Harvard University Press; 2015. xi, 502 pages p.
 58. Torquet N, de Chaumont F, Faure P, Bourgeron T, Ey E. mouseTube—a database to collaboratively unravel mouse ultrasonic communication [version 1; referees: 2 approved]. F1000Research [Internet]. 2016; 5:[2332 p.]. Available from: <http://f1000r.es/7uf>.