Automated acoustic analysis of 50-kHz ultrasonic vocalizations using template matching and contour analysis

David J. Barker

National Institute on Drug Abuse Intramural Research Program, Neuronal Networks
Section, National Institutes of Health, 251 Bayview Boulevard, Baltimore, Maryland 21224,
USA
djamesbarker@gmail.com

Aaron M. Johnson^{a)}

Department of Otolaryngology–Head and Neck Surgery, New York University Voice
Center, New York University School of Medicine, 345 East 37th Street, Suite 306,
New York, New York 10016, USA
aaron.johnson@nyumc.org

Abstract: Ultrasonic vocalizations are a useful tool for inferring affective states in the rat and have been incorporated in research paradigms modeling important human conditions. While the majority of studies report the quantity or rate of observed ultrasonic vocalizations, growing evidence suggests that critical data may be contained in the acoustic features of individual vocalizations. Thus, the goal of the present study was to develop and validate a method for measuring acoustic parameters of ultrasonic vocalizations that were collected using automatic template detection. Acoustic parameters derived using this method were found to be comparable to those collected using commercially available software.

© 2017 Acoustical Society of America
[CCC]

Date Received: September 30, 2016 **Date Accepted:** February 16, 2017

1. Introduction

Ultrasonic Vocalizations (USVs) are a useful tool for inferring affective states in the rat and have been incorporated in research paradigms modeling important human conditions such as drug abuse,^{1–3} depression, fear, and anxiety disorders,⁴ Parkinson disease,⁵ and aging.⁶ Moreover, USVs are a useful tool for examining the neurobiological substrates of reward.^{7–9}

While the majority of studies report the quantity and/or rate of observed USVs, growing evidence suggests that critical data may be contained in the acoustic features of individual vocalizations. For example, in models of Parkinson disease and neurodegeneration, it has been shown that dopamine depleting lesions and aging produce changes in the acoustic features of USVs, such as decreases in amplitude and reductions in fundamental frequency bandwidth.^{5,10–12} Notably, these changes affect the acoustic *qualities* of vocalizations rather than the *quantity* of observed USVs. Moreover, it has been convincingly shown that repeated training of USVs improves the vocal deficits induced by dopamine depleting lesions.^{11,12} Still, quantification of the magnitude of this improvement is tedious and would benefit from increased automation.

Similarly, drug abuse research has demonstrated that certain abused drugs affect the quantity of USVs, others affect the acoustic features of USVs, and that some drugs may affect both the quantity and acoustic features of USVs (reviewed in Simola¹³ and Barker *et al.*¹⁴). Finally, there is some evidence which suggests that highly salient events may preferentially elicit USVs with higher rates of modulation than events of lower salience.^{15,16} Nevertheless, the quantification of this assertion has remained limited to categorical analysis of USVs.

Thus, it is clear that studies of reward processing and vocal sensorimotor function would benefit from novel, automated, methods for the quantification of the acoustic parameters of USVs. Current automated methods of vocalization detection are based on either intensity threshold or whistle detection. However, non-vocalization noises and rapid frequency modulations, both of which are common in USV

^{a)} Author to whom correspondence should be addressed.

recordings, reduce the reliability of these approaches. Previously, we developed and validated a novel tool for automatically detecting 50-kHz USVs using a template matching procedure.² The goal of the present study was to develop and validate a method for measuring acoustic parameters of USVs following the automatic template detection. Parameters included the minimum, maximum, and mean fundamental frequencies of the entire USV, bandwidth, duration, the minimum, maximum, and mean derivative, number of frequency modulations per second, number of frequency peaks, and number frequency peaks per second. Following automatic detection, our approach involved using USV amplitude as a way to threshold and detect the contour of USVs. Acoustic parameters determined using this method were then compared to the most commonly used parameters automatically measured using commercially available software.

2. Materials and methods

2.1 USV recordings

Recordings of USVs were obtained from ten adult male Long-Evans rats (Charles River, Wilmington, MA). Recordings were made in a wire-mesh walled chamber with an ultrasonic microphone (CM16/COMPA, Avisoft Bioacoustics, Berlin, Germany) and digitized using a 250 kHz sampling rate and 16-bit depth (UltraSoundGate 416, Avisoft Bioacoustics, Berlin, Germany). USVs were elicited from the male rats using a mating paradigm, as described in Johnson *et al.*¹¹ In brief, female rats in estrus were introduced to each male rat in the recording chamber. The female was removed after the male rat expressed interest in her, which elicited vocalizations from the male. Acoustic recording began when the female was removed from the room to ensure recordings consisted solely of the male rats' vocalizations.

2.2 SASLab analysis

Recordings were first analyzed using a commercially available software program, SASLab Pro (Avisoft Bioacoustics, Glienicke, Germany). The beginning and end of each USV were manually labeled on a spectrogram with a frequency resolution of 488 Hz and a temporal resolution of 1.024 ms [fast Fourier transform (FFT) window length = 512; 50% overlap]. After manual segmentation, the minimum and maximum fundamental frequencies of each USV were automatically identified and visualized using horizontal lines. SASLab Pro provides a tool for visually erasing portions of the spectrogram. This tool was used to remove any noise surrounding the USV that was interfering with the accuracy of the automatic detection of the minimum and maximum frequencies. No portion of the actual USV was removed. Once USVs were labeled and noise was removed, the automatic parameter measurement tool was used to calculate the minimum, maximum, and mean fundamental frequencies across the entire USV, the bandwidth (calculated as the difference between the minimum and maximum frequency), and the duration. To accomplish this, the acoustic parameters tool identifies the frequency with the greatest amplitude in each spectral bin and then uses these bins to calculate the above-mentioned measurements.

2.3 XBAT analysis

Automatic detection of USVs was conducted as described previously.² Briefly, USVs were detected in XBAT, a MATLAB based graphical user interface and extensible bioacoustics tool developed by the Bioacoustics Research Program at Cornell University. A library of USVs between 35 and 80 kHz were used as templates for the detection of USVs in recorded ".wav" files using spectrogram correlation. Parameters for the FFT were congruent with those described above for Avisoft. In this method, the frequency and amplitude of the template spectrogram (with a specifically defined time and frequency window) is correlated with the frequency and amplitude of recorded data. Correlations at each time point in the recorded data file are accomplished using a sliding window, producing a correlation for each template at each specific time lag in the recorded data file. Correlations that were greater than or equal to 0.3 were considered as putative USVs. Detected elements were post-processed using a custom MATLAB script to group temporally proximal elements into single USVs if they occurred within 50 ms of one another. Finally, detected elements were subjected to a final human review ("Selection Review" graphical user interface) for confirmation or rejection.

In the present study, the Selection Review tool (graphical user interface and custom MATLAB programming) used for human validation of USVs was modified to detect the frequency contours of USVs and calculate numerous acoustic parameters including those included for SASLab analysis (see above). Additionally, measurements

were included to quantify the modulation across the USV contour. This included the maximum instantaneous slope (calculated as the maximum of the absolute value of the first derivative), the minimum instantaneous slope (calculated as the minimum of the absolute value of the first derivative), and the average of all instantaneous slopes (calculated as the mean of the absolute value of all instantaneous slopes across the function fit to the contour). We also quantified modulation based on the number of oscillations in each USV and the rate of oscillation per unit time. This was accomplished by first quantifying the number of local minima and maxima for each USV and then by dividing the number of peaks by the total duration of the USV.

For contour analysis, experimenters were required to delineate a window around detected USVs using a rectangular box, drawn in the graphical user interface. Within this window, the frequency bin with the greatest power was selected for each successive spectral frame (bin of time) across the window. To further separate frequency bins containing USVs from those containing background noise, we further defined the signal by eliminating points that fell below a customizable threshold. The custom threshold was calculated as the ratio of acoustic power within each successive peak of the user-defined window as compared to the average power within the 75–125 kHz range. Values in the 75–125 kHz range were chosen in order to avoid the extraneous noise that is often present in lower frequency bands during recordings in operant chambers or other experimental apparatus. Any peak value within the signal that fell below the user-defined threshold was filtered out and no longer used when fitting a function to each USV. As necessary, experimenters were then able to adjust the minimum threshold individually for each USV to obtain optimal contour detection and control for fluctuations in the signal quality across the recording. Based on initial training the default threshold was set as a ratio of 8:1. Acoustic parameters of each USV were then extracted from the contours detected (Fig. 1).

2.4 Comparison of results

All USVs detected by the SASLab and XBAT methods were manually compared using spectrograms and time codes to determine the number of vocalizations that were (1) identified using both techniques, (2) identified manually using SASLab but not automatically by XBAT, or (3) identified automatically by XBAT but not manually using SASLab. For the USVs that were identified by both methods, Pearson product moment correlations were used to measure the agreement on acoustic measures of mean frequency, maximum frequency, minimum frequency, bandwidth, and duration. The signal-to-noise ratio (SNR) was calculated for each USV to investigate if recording quality may have contributed to why a USV may have been identified by one method, but not the other. Additionally, spectrograms were manually inspected to investigate sources of measurement discrepancy between the two methods.

3. Results

The newly developed XBAT tools for contour analysis are available at https://github.com/djamesbarker/USV_AutoDetection.¹⁷ A total of 553 unique USVs were manually identified using SASLab. XBAT automatically identified 391 (71%) of those manually identified USVs. Of these mutually identified 391 USVs, XBAT identified 14 as multiple USVs that were manually identified as a single USV in SASLab. XBAT identified only three USVs that were not identified manually in SASLab. Consistent with previous reports, it was observed that USVs were missed by automatic detection in XBAT when the SNR of USVs to background noise was low (Fig. 2). Overall, 91.3% of USVs were detected above the 18 dB SNR, while only 44.5% of USVs below this threshold were automatically detected by XBAT. A logistic regression indicated that SNR significantly predicted whether a USV was or was not detected ($r^2=0.583$, $p<0.0001$). Further, a qualitative analysis of USVs that were missed by the detector suggested that acoustic artifacts were often present in USVs with higher SNRs that were missed by the detector.

The estimated time to obtain contours and derive acoustic measurements using the present XBAT extension was approximately three times faster than the analogous contour detection in SASLab. Combined with the fact that automated detection of USVs is already 6–10× faster than manual detection of USVs, the present method provides a distinct advantage over widely-used methods.²

Acoustic measures of mutually identified USVs were strongly correlated for mean frequency, $r(389)=0.95$, $p<0.001$, maximum frequency, $r(389)=0.82$, $p<0.001$, minimum frequency, $r(389)=0.82$, $p<0.001$, and duration $r(389)=0.81$, $p<0.001$. The correlation for bandwidth was moderate at $r(389)=0.72$, $p<0.001$ (Fig. 3). It is

Fig. 1. (Color online) (A) Visualization of the contour detection algorithm. Portions of the spectrogram exceeding a user-specified SNR are first used to detect each automatically detected USV (top left). From the detected vocalization, the peak amplitude at each time lag is then used to model the contour of the USV (top right). Users are able to visualize USVs and their contours within the graphical user interface (bottom left and right). (B) The parameters of the detected contour are subsequently used to derive various parameters of each USV including the average frequency, duration, and bandwidth.

necessary to note that differences in the algorithms and user-defined settings used for each method of detection and feature extraction will account for some differences in quantified acoustic parameters. In point of fact, the present contour mapping method requires that points exceed a user-defined threshold in order to be included in the

Fig. 2. The percent and raw number of USVs detected using the automatic detection algorithm as a function of the observed SNRs. Signal strength was operationally defined as the peak power (dB) of a given vocalization, while the noise-band was defined as the average power (dB) of a portion of the spectrogram proximal to the vocalization which was free from any USVs. Detection and contour mapping of USVs is most dependent on recording quality; the largest discrepancies between traditional methods and the reported contour mapping were related to SNR.

Fig. 3. Correlation plots between the two measurement procedures for the spectral and temporal measurements. The solid line is the line of best fit and the dashed line represents a perfect correlation. Each data point is a single USV.

mapped contour. Qualitatively, this approach is beneficial for eliminating the influence of spurious background noise on the USV contour. Nevertheless, manual spectrogram inspection revealed intensity often varied considerably during single vocalizations, particularly at the beginning and end of vocalizations and in those USVs with a high degree of frequency modulation. Thus, variation in the signal quality across individual USVs is likely to account for differences in the minimum or maximum frequency reported between detection methods or for differences in the beginning and endpoints of individual USVs.

4. Discussion

The purpose of this study was to develop and validate a novel method for automatically measuring acoustic parameters of rat USVs. The results from our new method, based on template matching for USV detection and frequency contour analysis for acoustic analysis, were highly correlated to the commonly-used USV analysis technique using SASLab. Further, the development of a contour mapping approach has allowed for the acquisition of additional measurements that might be of use to USV researchers.

The primary advantage of this newly developed method is the robust acoustic analysis based on the frequency contour of automatically detected USVs. This approach allows users to visualize the data points that are included for contour analysis and adjust contours to improve signal quality. Furthermore, once the frequency contour has been identified, further acoustic and statistical analyses can be performed within MATLAB. For example, this frequency contour analysis could be used to provide a detailed temporal analysis of frequency modulation, a potentially important variable in disease models that impact vocal sensorimotor control such as aging and Parkinson disease. This automated analysis technique could also aid in automatic classification based on acoustic parameters or in developing new classifications by allowing for automated analysis of large numbers of USVs. Current USV classification schemes range from very simple divisions based on frequency changes (frequency modulated vs constant frequency) to complex classifications with up to 14 categories requiring manual inspection of individual USVs.¹

The present data illustrate that the primary advantage of manual scoring methods is the ability to detect USVs with a low SNR. Undoubtedly, this is a limitation of automated detection that one must be aware of when processing USV data. Still, it may be difficult to accurately acoustically analyze these low-SNR USVs, even when manually identified; measurements of acoustic parameters rely on the quality of the signal and are therefore influenced by high background noise or low signal quality. Thus, the present data also illustrate the need to adjust recording parameters in order to optimize signal acquisition, when possible. This could be achieved by using a microphone array or by limiting the movement of the animal. The ultrasonic microphones

used to collect our data are highly directional, with a 5 to 13 dB amplitude decrease at 30° off axis for frequencies at 40 and 60 kHz, respectively. Finally, this emphasizes the assertion that a subset of files should always be scored manually in order to produce and report inter-rater reliability between manual and automated scoring.²

Our future goal is to further automate USV detection and analysis to remove human review and, therefore, significantly increase speed of analysis and improve repeatability and reliability. The method presented here adds a powerful *post hoc*/post-detection acoustic analysis method that both increases analysis speed and keeps the data in the MATLAB environment for continued analysis.

Acknowledgments

This work was made possible due to the financial support from the National Institute on Deafness and Other Communication Disorders (K23DC014517, PI: J.) and the Intramural Research Program (IRP) of the National Institute on Drug Abuse (PI: Marisela Morales). Funding sources were not involved in study design, data collection and interpretation, or the decision to submit this research for publication. We thank Christopher Herrera, Elizabeth Banaszak, and Anna Fischer for their excellent assistance.

References and links

- ¹J. M. Wright, J. C. Gourdon, and P. B. Clarke, "Identification of multiple call categories within the rich repertoire of adult rat 50-kHz ultrasonic vocalizations: Effects of amphetamine and social context," *Psychopharmacology (Berl.)* **211**(1), 1–13 (2010).
- ²D. J. Barker, C. Herrera, and M. O. West, "Automated detection of 50-kHz ultrasonic vocalizations using template matching in XBAT," *J. Neurosci. Methods* **236**(0), 68–75 (2014).
- ³D. J. Barker, S. J. Simmons, L. C. Servilio, D. Bercovicz, S. Ma, D. H. Root, A. P. Pawlak, and M. O. West, "Ultrasonic vocalizations: Evidence for an affective opponent process during cocaine self-administration," *Psychopharmacology (Berl.)* **231**(5), 909–918 (2014).
- ⁴A. Borta, M. Wöhr, and R. K. Schwarting, "Rat ultrasonic vocalization in aversively motivated situations and the role of individual differences in anxiety-related behavior," *Behav. Brain Res.* **166**(2), 271–280 (2006).
- ⁵L. M. Grant, C. A. Kelm-Nelson, B. L. Hilby, K. V. Blue, E. S. Paul Rajamanickam, J. D. Pultorak, S. M. Fleming, and M. R. Ciucci, "Evidence for early and progressive ultrasonic vocalization and oromotor deficits in a pink1 gene knockout rat model of Parkinson's disease," *J. Neurosci. Res.* **93**(11), 1713–1727 (2015).
- ⁶A. M. Johnson, M. R. Ciucci, and N. P. Connor, "Vocal training mitigates age-related changes within the vocal mechanism in old rats," *J. Gerontol. A Biol. Sci. Med. Sci.* **68**(12), 1458–1468 (2013).
- ⁷J. Burgdorf, P. L. Wood, R. A. Kroes, J. R. Moskal, and J. Panksepp, "Neurobiology of 50-kHz ultrasonic vocalizations in rats: Electrode mapping, lesion, and pharmacology studies," *Behav. Brain Res.* **182**(2), 274–283 (2007).
- ⁸A. J. Wintink and S. M. Brudzynski, "The related roles of dopamine and glutamate in the initiation of 50-kHz ultrasonic calls in adult rats," *Pharmacol. Biochem. Behav.* **70**(2–3), 317–323 (2001).
- ⁹T. Scardocho, I. Trujillo-Pisanty, K. Conover, P. Shizgal, and P. B. Clarke, "The effects of electrical and optical stimulation of midbrain dopaminergic neurons on rat 50-kHz ultrasonic vocalizations," *Front. Behav. Neurosci.* **9**, 331 (2015).
- ¹⁰M. R. Ciucci, L. Vinney, E. J. Wahoske, and N. P. Connor, "A translational approach to vocalization deficits and neural recovery after behavioral treatment in Parkinson disease," *J. Commun. Disord.* **43**(4), 319–326 (2010).
- ¹¹A. M. Johnson, E. J. Doll, L. M. Grant, L. Ringel, J. N. Shier, and M. R. Ciucci, "Targeted training of ultrasonic vocalizations in aged and parkinsonian rats," *J. Vis. Exp.* **54**, e2835 (2011).
- ¹²A. M. Johnson, L. M. Grant, T. Schallert, and M. R. Ciucci, "Changes in rat 50-kHz ultrasonic vocalizations during dopamine denervation and aging: Relevance to neurodegeneration," *Curr. Neuropharmacol.* **13**(2), 211–219 (2015).
- ¹³N. Simola, "Rat ultrasonic vocalizations and behavioral neuropharmacology: From the screening of drugs to the study of disease," *Curr. Neuropharmacol.* **13**(2), 164–179 (2015).
- ¹⁴D. J. Barker, S. J. Simmons, and M. O. West, "Ultrasonic vocalizations as a measure of affect in preclinical models of drug abuse: A review of current findings," *Curr. Neuropharmacol.* **13**(2), 193–210 (2015).
- ¹⁵A. M. Ahrens, S. T. Ma, E. Y. Maier, C. L. Duvauchelle, and T. Schallert, "Repeated intravenous amphetamine exposure: Rapid and persistent sensitization of 50-kHz ultrasonic trill calls in rats," *Behav. Brain Res.* **197**(1), 205–209 (2009).
- ¹⁶S. M. Brudzynski, M. Silkstone, M. Komadoski, K. Scullion, S. Duffus, J. Burgdorf, R. A. Kroes, J. R. Moskal, and J. Panksepp, "Effects of intraaccumbens amphetamine on production of 50 kHz vocalizations in three lines of selectively bred long-evans rats," *Behav. Brain Res.* **217**(1), 32–40 (2011).
- ¹⁷D. J. Barker (2017). "Automatic detection of ultrasonic vocalizations and contour analysis" [Computer software] (Last viewed 3/6/2017), available at https://github.com/djamesbarker/USV_AutoDetection.