Response to: Making WAAVES in the vocalization community: How big is the splash?

James M. Reno[a,c] and **Christine L. Duvauchelle**[b,c,*]

[a] The University of Texas at Austin, Department of Psychology, Behavioral Neuroscience, 108 E. Dean Keeton Avenue, Stop A8000, Austin, TX 78712-1043, USA

[b] The University of Texas at Austin, College of Pharmacy, Division of Pharmacology and Toxicology, 2409 University Avenue, Stop A1915, Austin, TX 78712, USA

[c] Waggoner Center for Alcohol & Addiction Research, The University of Texas at Austin, 2500 Speedway, Stop A4800, Austin, TX 78712, USA

Commentary response

We are pleased to address Dr. Barker's responses and concerns regarding our WAAVES program (Reno et al., 2013) and hope that this forum can further clarify the utility of customized USV analyses programs and prerequisite procedures necessary for every laboratory to obtain reliable USV data assessment.

An important concern regarding accuracy in signal detection was raised in reference to Fig. 2 that displays a scatterplot of the relationship between human detection and the WAAVES program, along with the associated correlations. We understand the confusion caused by this comparison because the differences in counts between WAAVES and human could be interpreted as inadequate USV detection by WAAVES. We should have pointed out in the original report that inherent differences in Human versus WAAVES detection criteria could largely account for discrepancies in USV counts. For instance, as a rule-based algorithm, WAAVES requires characteristics of USVs and noise to be rigorously defined. One critical characteristic is USV duration, set at 5 ms (or 20 ms in separate analyses). When processing sound files WAAVES omitted USVs < 5 ms, just as intended. On the other hand, when USV sound files were slowed to 4% of original speed to enable auditory and visual USV assessment, human analyzers had been instructed to count all USVs regardless of duration. As reported in the original publication, the number of 5 ms USV assessed using WAAVES turned out to be 92% of the calls detected by human analysts (692/752 total USV counts in 50 one-minute files). We were satisfied with that outcome, knowing that many 50–55 kHz USVs (actual parameters set at 30–90 kHz) are less than 5 ms.

Yet, in terms of signal detection, as pointed out by Barker, the WAAVES paper did not provide exact rates of true positives, false negatives and false positives contributing to the reported USV counts. We realize that in the absence of this data, it cannot be assumed that lower USV counts by WAAVES are solely due to duration-based omissions. To address this issue, a subset (20%) of sound files having the largest WAAVES/Human USV count discrepancies were chosen from the files used in the WAAVES paper.

* Corresponding author at: The University of Texas at Austin, College of Pharmacy, Division of Pharmacology and Toxicology, 2409 University Avenue, Stop A1915, Austin, TX 78712, USA. Tel.: +1 512 471 1090; fax: +1 512 475 6088.
duvauchelle@mail.utexas.edu, christyduvauchelle@gmail.com (C.L. Duvauchelle)..

Using the 5 ms USV duration criterion, WAAVES USV count for this subset of files was 370. At this point, USVs were re-counted (at a much slower pace than initial analyses) and each file was closely examined to determine which USVs were omitted by WAAVES. The re-counted number of USVs was 458, making the original Human USV count of 422, 92.1% the total calls. Of the 458 USVs, 408 were of durations 5 ms, making the 370 USVs counted by WAAVES 90.7% of the potentially detectable calls. See below chart for WAAVES rates of true positives, false negatives and false positives (refer to Reno et al., Fig. 1 WAAVES Program Flowchart as indicated).

It should be noted that of the false positives, all 5 were the result of longer USVs appearing as two USVs because of segmentation by vertical noise objects (see criterion 3). Though no false positives were the result of noise being counted as USVs, all false negative calls were due to some aspect of noise. The current version of WAAVES cannot separate out USVs occurring in the midst of non-USV noise, and most false negatives (28/38) were due to their location within solid patches of noise. The remaining 10 false negatives appeared to be omitted because of shared characteristics with a defined type of noise (e.g., some harmonic calls were likely omitted because of similarities with reverberating noise, as per criterion 4).

In response to the query regarding types of USVs detectable by the algorithm, the current version of WAAVES distinguishes between flat and frequency-modulated (FM) USVs based on defined changes in USV frequency. Though we do not currently have an algorithm in place for categorizing trill calls an algorithm could be added to WAAVES after determining the characteristics of these calls within the test environment.

Barker also comments that the "USV community as a whole might contribute in order to improve the results of the detector should the authors choose to publish the source code alongside the article." Although we concur with this sentiment, at this stage we deem it premature. One critical aspect of the WAAVES algorithm is to accurately characterize the "noise", and to formalize an algorithm for identifying signal embedded in this noise. For example, within our WAAVES program, we must specify the detailed characteristics of non-USV signals (e.g., length of vertical objects, co-occurrence of USVs and USV reverberations, sound objects less than 2.5 ms) emanating from within our test chambers during experimental sessions. These detailed measurements are necessary to define "noise" that will be filtered out by the algorithm. Analogously, we must specify the detailed characteristics of USV signals (e.g., frequency range, duration, inter-USV intervals). These detailed measurements are necessary to define "signal". The key is to strike an acceptable balance so as to optimize signal detection within the specific experimental environment. Should we set these parameters as constants in our algorithm and make this available to the broader USV community we run the risk that the algorithm will not be accurately calibrated for alternate experimental environments. In future work we hope to develop a calibration protocol that would allow the USV community to easily calibrate the WAAVES algorithm for their specific experimental context.

In the meantime, by offering detailed descriptions of USVs and noise from the testing chambers that we used in the WAAVES paper, we are providing a detailed step-by-step flow chart that other USV laboratories can use to expand their USV studies. Having experienced the process, we realize it is not trivial to develop an algorithm based on unique acoustic environments. Yet working in conjunction with computer scientists and developing the necessary software has proven extremely beneficial to our research program by increasing the speed and efficiency of data analysis. Although we have not developed a fully automated USV detection software application that can be utilized "as is" in any laboratory, we have solved many of the problems associated with automated USV detection. Without the limitations in data collection and analyses associated with non-automated means of USV

assessment, the USV community can exponentially expand USV studies of all types within the field of drug abuse and beyond.

Reference

Reno JM, Marker B, Cormack LK, Schallert T, Duvauchelle CL. Automating ultrasonic vocalization analyses: the WAAVES program. J Neurosci Methods. 2013; 219:155–61. [PubMed: 23832016]
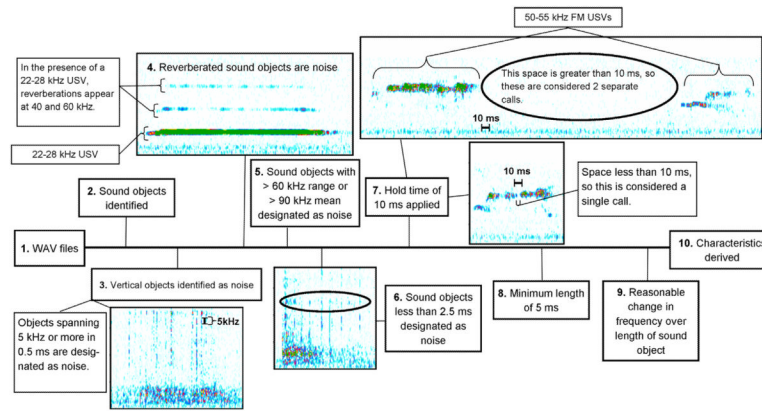
Fig. 1.

WAAVES Program Flowchart. The order of separation criteria is an important factor when developing an automated analysis program such as WAAVES. (1) Sound files (.wav) are read into the program. (2) Sound objects are identified within the files for closer inspection. (3) Sound objects spanning 5 kHz and above consecutive frequencies in 0.5 ms are considered noise. (4) Within the experimental apparatus, noise generated by reverberations from high decibel 22–28 kHz USVs often appears at 40 and 60 kHz. (5) Sound objects defined as noise if frequency range > 60 kHz and mean frequency > 90 kHz. (6) Sound objects defined as noise if <2.5 ms. (7) Inter-call interval at least 10 ms for call count > 1. (8) Call duration setting (e.g., 5 or 20 ms). (9) Median value of change in mean frequency in 0.5 ms steps was required to be >0 and <2 kHz. (10) At this point calls have been successfully separated from noise and further USV subcategories can be created (e.g. flat versus frequency-modulated (FM)).
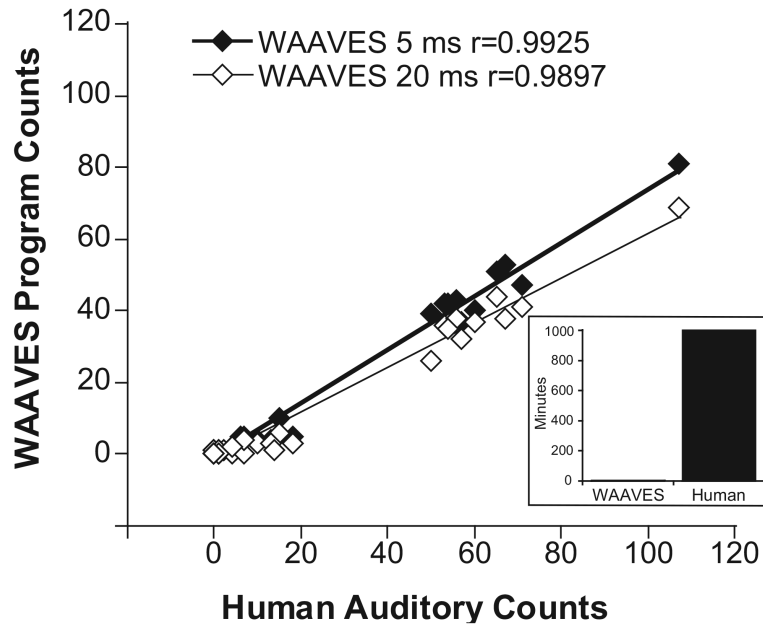
Fig. 2.
Comparison between WAAVES output and human auditory counts. The WAAVES outputs using either 5 or 20 ms USV duration criteria were significantly correlated with the number of USVs tabulated using human visual and auditory assessment (e.g., during playback of USV files at 4% speed of the original recordings; p < 0.001 for both). Inset bar graph displays the average time (min) to analyze 50 1-min USV data files data using WAAVES (≈3.75 min) compared to the amount of time needed for human visual and auditory detection during playback (≈1000 min).

USVs	5 ms = 408	WAAVES output = 370	Circumstances
True positives		365	Meeting all definitions of USVs
False positives		5	USV segmented by noise (criterion 3), counted twice
False negatives		38	Embedded in noise or having characteristics resembling noise (criteria 3, 4, 5)