ARL-TR-8205 ● Nov 2017

US Army Research Laboratory

Strategies for Characterizing the Sensory Environment: Objective and Subjective Evaluation Methods using the VisiSonic Real Space 64/5 Audio-Visual Panoramic Camera

By Joseph McArdle, Ashley Foots, Chris Stachowiak, and Kelly Dickerson

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-8205 ● NOV 2017

US Army Research Laboratory

Strategies for Characterizing the Sensory Environment: Objective and Subjective Evaluation Methods using the VisiSonic Real Space 64/5 Audio-Visual Panoramic Camera

by Joseph McArdle, Ashley Foots, Chris Stachowiak, and Kelly Dickerson
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
November 2017

2. REPORT TYPE
Technical Report

3. DATES COVERED (From - To)
July 14, 2017–July 20, 2017

4. TITLE AND SUBTITLE
Strategies for Characterizing the Sensory Environment: Objective and Subjective Evaluation Methods using the VisiSonic Real Space 64/5 Audio-Visual Panoramic Camera

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Joseph McArdle, Ashley Foots, Chris Stachowiak, and Kelly Dickerson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
Human Research and Engineering Directorate (ATTN: RDRL-HRF-D)
Aberdeen Proving Ground, MD 21005-5066

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-8205

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The VisiSonics RealSpace 64/5 Audio-Visual Panoramic (VRAP) Camera captures high-resolution panoramic video and full-dimension audio in laboratory or field environments. Online and offline analysis and visualization tools enable sound source identification localization using, among other sources, sound pressure level dynamics. This report aims to extend the user manual produced by VisiSonics by familiarizing potential users of the VRAP with features, display elements, and procedures for operation of the VRAP hardware and software by using an illustrative use case where the VRAP was deployed in the field. The use case addresses limitations of the ecological frequency (i.e., prevalence) measure, an approach described by Ballas (1993) for determining the representativeness of a set of sounds in a given environment. The use case evaluates the impact of high-quality panoramic audiovisual captures of the environment on the observer's ability to extract ecological frequency information from a scene.

15. SUBJECT TERMS
ecological frequency, audio-visual panoramic camera, VRAP, environmental sound perception, soundscapes, beamforming

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Kelly Dickerson
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified	UU	36	(410) 278-5979

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK.

1. Introduction

There is increasing Army interest in characterizing the sensory aspects of the operational environments, which are often complex and dynamic. Important mission-related information may be conveyed by changes in the ambient auditory and visual information, but extracting meaningful events from background noise is a resource-intensive process for both human observers and technology. Laboratory studies generally offer a very limited approximation of the complexity of real-world operational environments and, consequently, studies of Soldier sensory and perceptual performance conducted under these conditions are often unable to fully characterize the dynamics of sensory and perceptual processing necessary for successful mission performance. Improvements in technology and a push toward developing novel methods of evaluating performance under real-world conditions addresses the potential limitations of traditional laboratory experimentation (ARL 2015); however, with this shift in focus to real-world environments emerges a very real practical issue of capturing and characterizing complex environmental stimuli. With the right tools, the precision in observation of the laboratory can be extended to the real world, potentially extending and validating theories of perceptual performance.

2. Characterizing Environmental Sounds

Generally, approaches to the characterization of environmental sounds have overly relied on the documentation of physical stimulus attributes, such as spectral slope, flux, and amplitude envelope (Gygi et al. 2004). Environmental sounds, however, convey meaningful information that cannot be specified within an array of physical features alone. Specifically, sounds afford strong associations with actions and objects within the environment. This link between objects, action, and perception engages higher-level cognitive/semantic networks that have been difficult to objectively quantify. An emerging literature has demonstrated that these difficult-to-quantify, subjective stimulus parameters can have a profound influence on performance. For example, Dickerson et al. (2015) found that stimulus similarity and identifiability are predictive of participant error rates on change localization tasks. Dickerson et al. (2016) found a similar relationship between pleasantness ratings and cued-recall performance. These findings suggest that semantic and subjective experiences with stimulus events influences performance. Given the clearly important role these subjective parameters play in perceptual performance, there is an evident need for the development of standardized methods of quantifying both the objective contents of the environment and an observer's subjective experience within that environment. A prerequisite to this, however, should be to first determine the representativeness of a set of sounds for a given environment. As subjective stimulus parameters, such as meaningfulness, often develop their roots in the context within which they are experienced, efforts to characterize environmental content should begin with observations within the natural environment. An established approach proposed by Ballas (1993) is to evaluate the ecological frequency for a given sound in a particular environment. The ecological frequency of a stimulus event can convey information about familiarity, identifiability, and signal salience.

Despite clear relevance to understanding perception in the real world, there has not been another comprehensive study of ecological frequency of environmental sounds since the publication of the Ballas paper, which has nearly 400 citations, many that explicitly state the importance of ecological frequency. According to the procedures described by Ballas, observers were provided with a timer, which would cue them at various times of the day to make observations. Upon hearing the cue, the observers were instructed to report on a data sheet the first sound (excluding music and speech) that they heard, the actions and objects involved, and their location at the time of the entry. Once completed, the observers were instructed to reset the timer, based on a provided schedule, if the timing of the next cue would not interfere with their regular activities. Following these procedures, the observers were instructed to log up to 50 sounds over the course of a week. Across the 25 observers who recorded the sounds present in their environment, a total of 1,185 sounds were reported across 8 distinct environments. These sounds were then evaluated by the researchers and binned into categories based on the sound-producing event and the extent to which its cause could be clearly ascribed. While the method used by Ballas is clearly preferable to recall from memory alone, there are several methodological issues that may affect the overall accuracy of ecological frequency measurements. First, Ballas relied on human observers who could have easily missed events, particularly for cases where the events co-occurred. Second, given that the observers were individually deployed to a particular location for a given recording session, there is no way to evaluate the reliability of the report of an observer since the observation conditions were distinct for each of the 25 observers. Finally, the recording time window varied across individual observers, making it impossible to extrapolate frequency data for the reported sounds.

3. The Present Study

The purpose of this technical report is to document the operation and functions of the VisiSonics RealSpace[*] 64/5 Audio-Visual Panoramic (VRAP) camera (Fig. 1). From the perspective of the authors, ecological frequency and real-world perception are critically interrelated and as such, a stable and reliable method for assessing ecological frequency should be developed. Thus, a secondary aim of this report is to present a brief description of an updated ecological frequency measure as a case for demonstrating the functionality and utility of the VRAP. These updated ecological frequency measures were developed for use within a larger research project characterizing common auditory environments (Foots et al. 2016). The VRAP supports the human field observers because it is deployed, along with operators, to the environmental locations. The VRAP camera is capable of not only recording panoramic video, but also 360° sound, from which, following offline processing, sound sources may be accurately discerned. It was intended that with the high-resolution recorded scenes, in conjunction with offline human evaluation, performance in a highly systematic fashion would result in more accurate estimates of ecological frequency (see Data Evaluation Methods, Section 7.2).

[*] VisiSonics Corporation, Highland, Maryland 20777

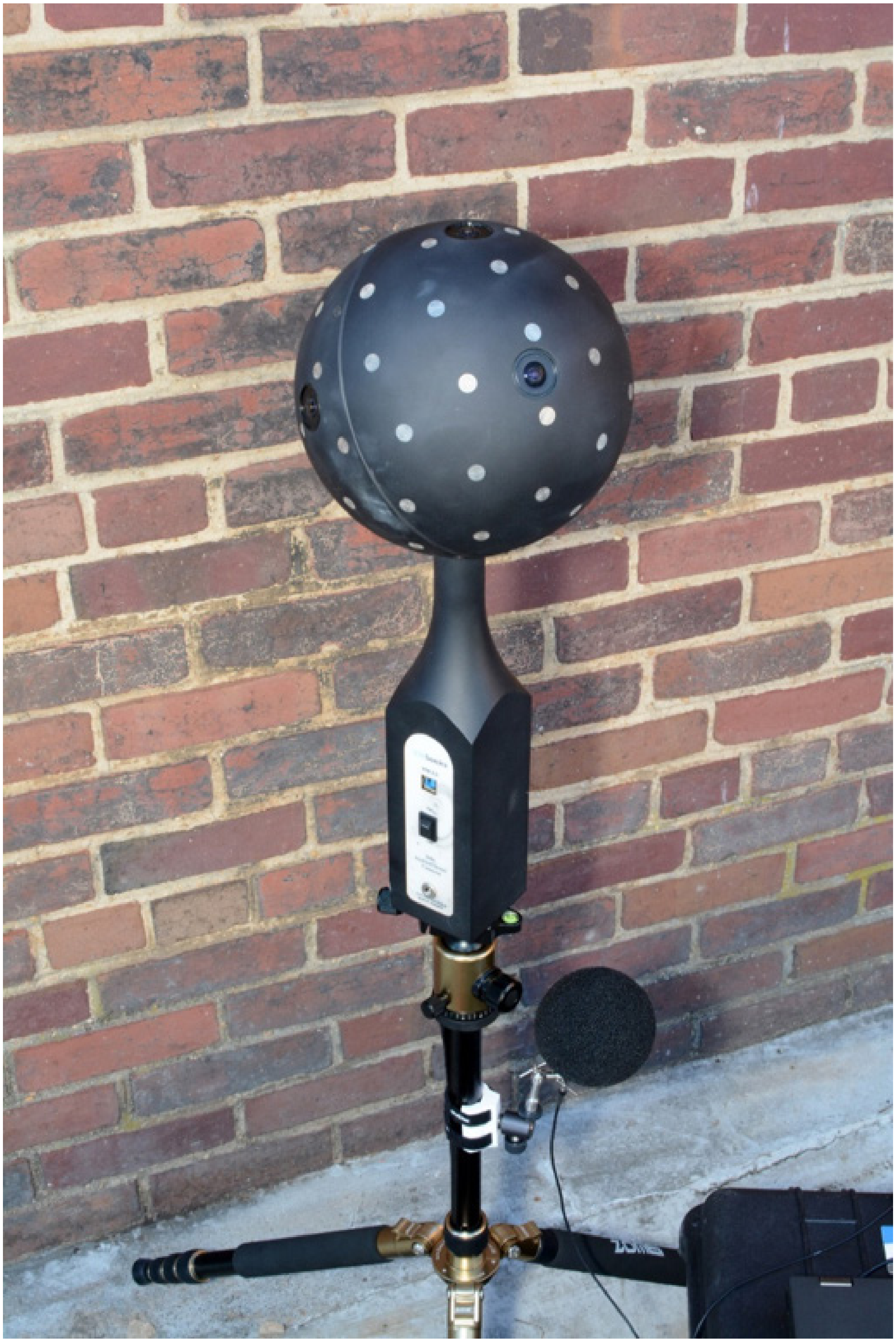


Fig. 1 VisiSonics RealSpace 64/5 Audio-Visual Panoramic (VRAP) camera (photograph courtesy of Ron Carty)

The VRAP camera was deployed to 2 distinct environments (i.e., urban and rural) and in each, several panoramic audio-visual scenes were recorded. To determine what features of the recorded media were necessary for offline ecological frequency rating, the recordings were post-processed to create comparison conditions. For each of the recorded scenes an audio-only comparison file was created. Additionally, to determine if beamforming (a sound localization-filtering method) would affect frequency rating, both the audio-only and audio-visual recordings were beamform-processed and saved separately for comparison. This resulted in 4 comparison conditions for each recording: 1) Audio-only/without beamforming, 2) Audio-only/with beamforming, 3) Audio-visual/without beamforming, and 4) Audio-visual/with beamforming.

4. Capture Methods and Environments

The VRAP camera captured recordings in 2 environments: an urban environment (Fig. 2 top panel) and a rural environment (Fig. 2 bottom panel). The urban environment was a densely populated area where the VRAP was set up roughly 250 ft from the street in an area with light pedestrian foot traffic and on a day where wind was intermittent and light. The rural environment was a loading dock in an industrial area roughly 45 miles from the urban environment. The loading dock area had light and occasional traffic and wind was minimal. At the loading dock there was a heavy vehicle (i.e., construction vehicle or tank) that would occasionally pass by. This is noteworthy because during these events no other sounds could be heard because of the intensity of the heavy vehicle noise.

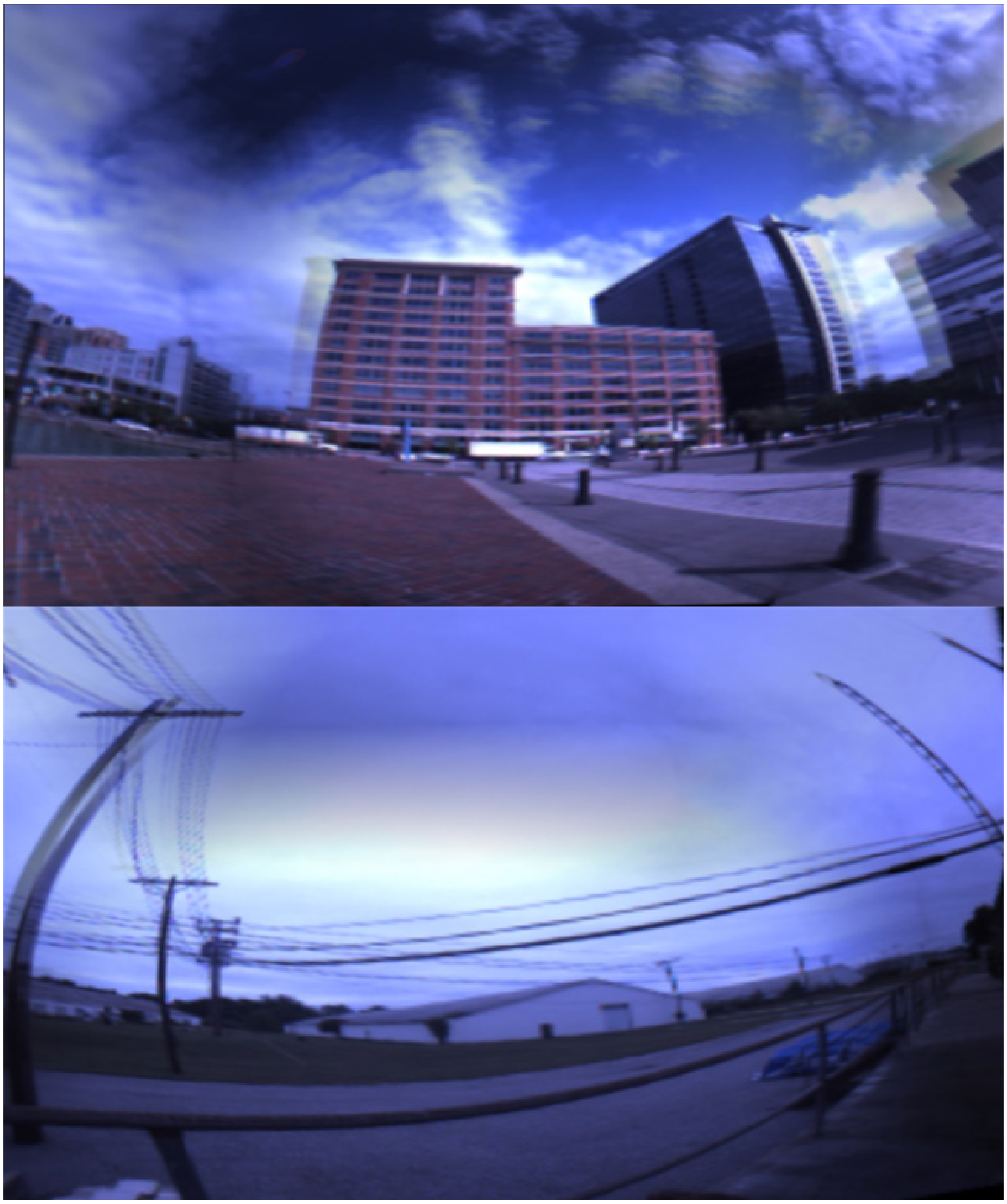


Fig. 2 Top panel: urban environment (city). Bottom panel: rural environment (loading dock).

Samples were recorded for 30 s at 5–10 min intervals for 1 h. There were 4 usable samples generated from each of the environments. In each recording session, up to 8 recordings were captured; however, some of the samples were not useable for an ecological frequency analysis. Samples were excluded from consideration if there was wind noise masking other sounds from the environment or if a curious bystander asked questions during the recording process. Some recordings were incomplete due to equipment failure (i.e., camera or laptop), or human error (e.g., forgot to enable turbo boost, accidentally bumping the camera tripod, or noise from the human operators). A separate single-channel, continuous audio recording was captured during each of the 1-h recording sessions. These recordings will serve as an auditory baseline for future comparisons between the VRAP and more traditional recording arrangements.

5. Hardware

Two recording systems were used simultaneously during the field sessions: a single-channel portable audio recorder and the VRAP camera. A Duracell[*] Powerpack 600 served as a power source for both recording systems.

5.1 VRAP

The VRAP camera consists of 5 USB 3.0 cameras and 64 omnidirectional electret microphones. The microphones are mounted to the surface of an aluminum sphere (d = 8 inches) attached to an 18-inch base resembling a typical camera tripod. Each camera captures video images at 742 × 480 dpi, which are combined in real time to produce a high-definition (HD), panoramic video that can be displayed as a Mercator projection[†] or spherical scene. The 64 microphones are gain matched to within 0.1-dB sound pressure level (SPL) and are synchronized by a common clock and recorded at a 48-kHz sampling rate with 24-bit accuracy and 110-dB dynamic range. While 110 dB of dynamic range provides significant flexibility in where the VRAP can be deployed, for louder environments it is possible to use the gain control settings within the VRAP interface to shift the range upward and prevent signal saturation.

Audio acquisition is handled by a field programmable gate array (FPGA) processor, which formats and converts the data into a single USB 3.0 data stream. The VRAP camera is interfaced through a Lenovo Thinkpad[‡] W540 laptop computer operating with an Intel Core[§] i7-4800MQ CPU and a NVIDIA Quadro[**] K2100M GPU (Figs. 3 and 4).

[*] Duracell, Wilmington, Delaware 19801

[†] The Mercator projection is a cylindrical map projection presented by the Flemish geographer and cartographer Gerardus Mercator in 1569.

[‡] Lenovo PC International, Hong Kong, China

[§] Intel Corporation, Santa Clara, California 95054

[**] NVIDIA Corporation, Santa Clara, California 95050

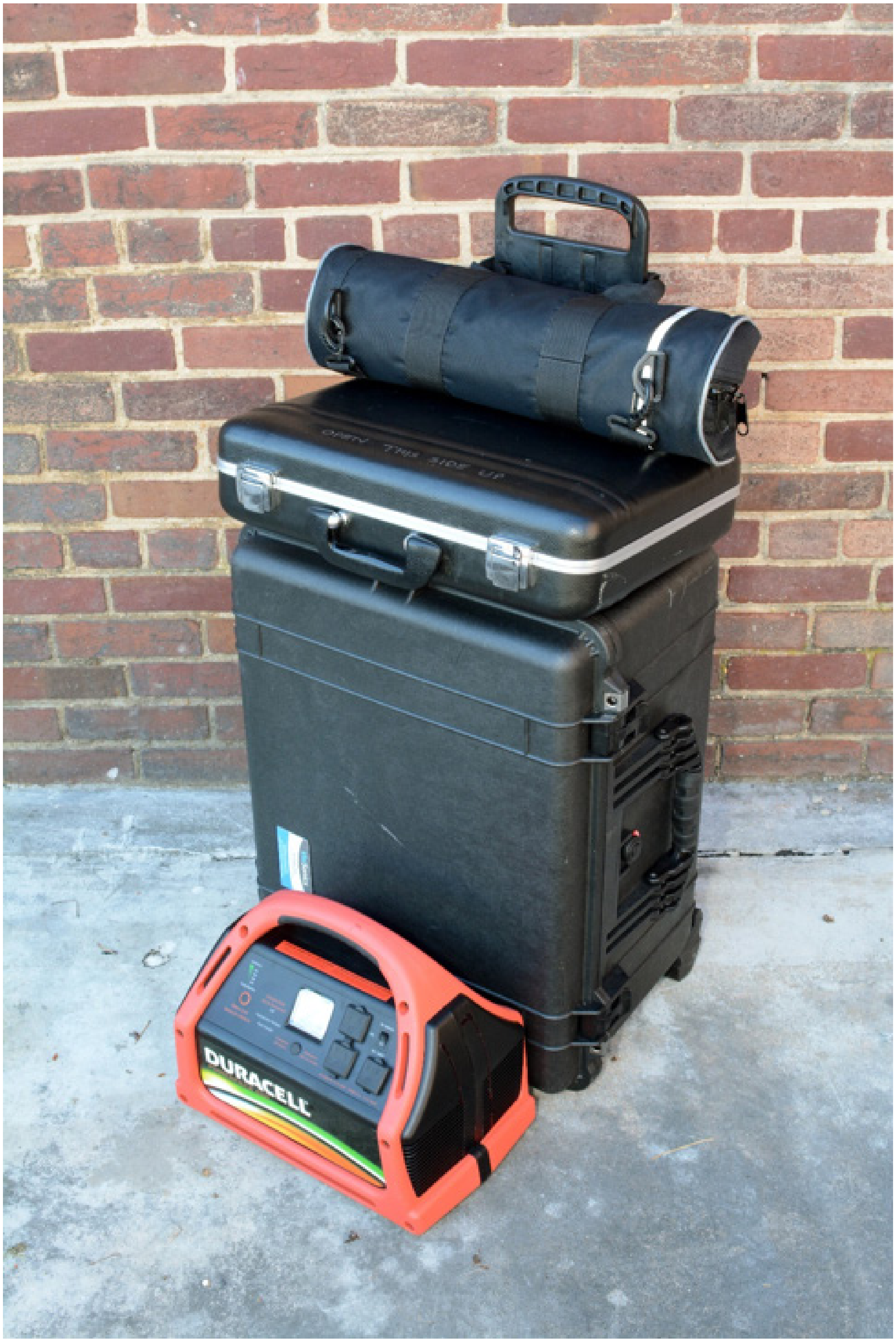


Fig. 3 VRAP camera system packaged for deployment (photograph courtesy of Ron Carty)

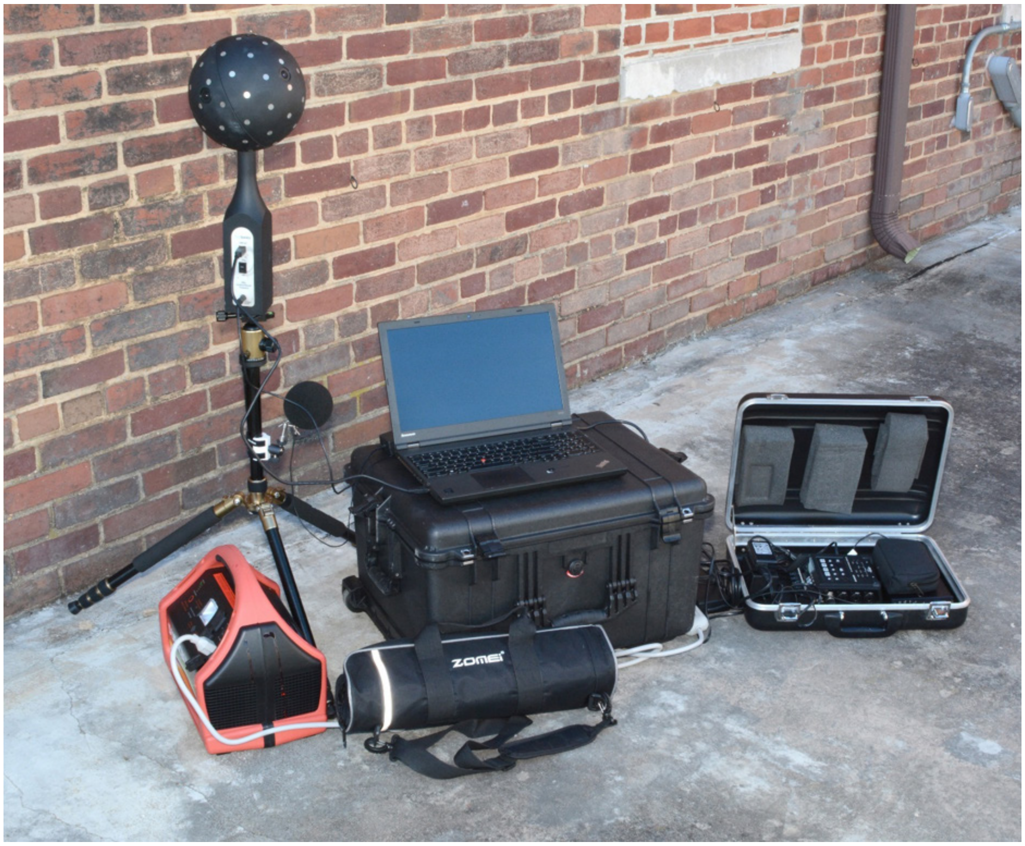


Fig. 4 VRAP camera system deployed (photograph courtesy of Ron Carty)

5.2 Single-Channel Recorder

A single-channel recording system was deployed with the VRAP camera. This additional system was used to record baseline audio. Baseline recordings served as a validation check since the ecological frequency project was the first full-field deployment of the VRAP. By using this auxiliary system the operators would have a record of the auditory environment in the event that the VRAP malfunctioned or the signal saturated due to high environmental noise intensity.

The single-channel recording system consisted of a G.R.A.S. Sound & Vibration* 40AF free-field microphone affixed to the VRAP camera tripod (Fig. 5). The microphone is driven by a G.R.A.S. power module type 12AK. The output is captured by a Roland Edirol† R-44-E digital recorder capable of 16-bit or 24-bit resolution at sampling frequencies of 44.1kHz/48kHz/88.2kHz/96kHz/192kHz. The power module, recorder, and their power transformers are secured in a small transit case for easy deployment (Fig. 6). The particular system described here was

* G.R.A.S. Sound & Vibration, Holte, Denmark

† Roland Corporation, Shizuoka, Japan

optimized for portability making it ideal for use with the VRAP; however, other users interested in a single-channel validation/baseline recording could use any microphone and recorder.

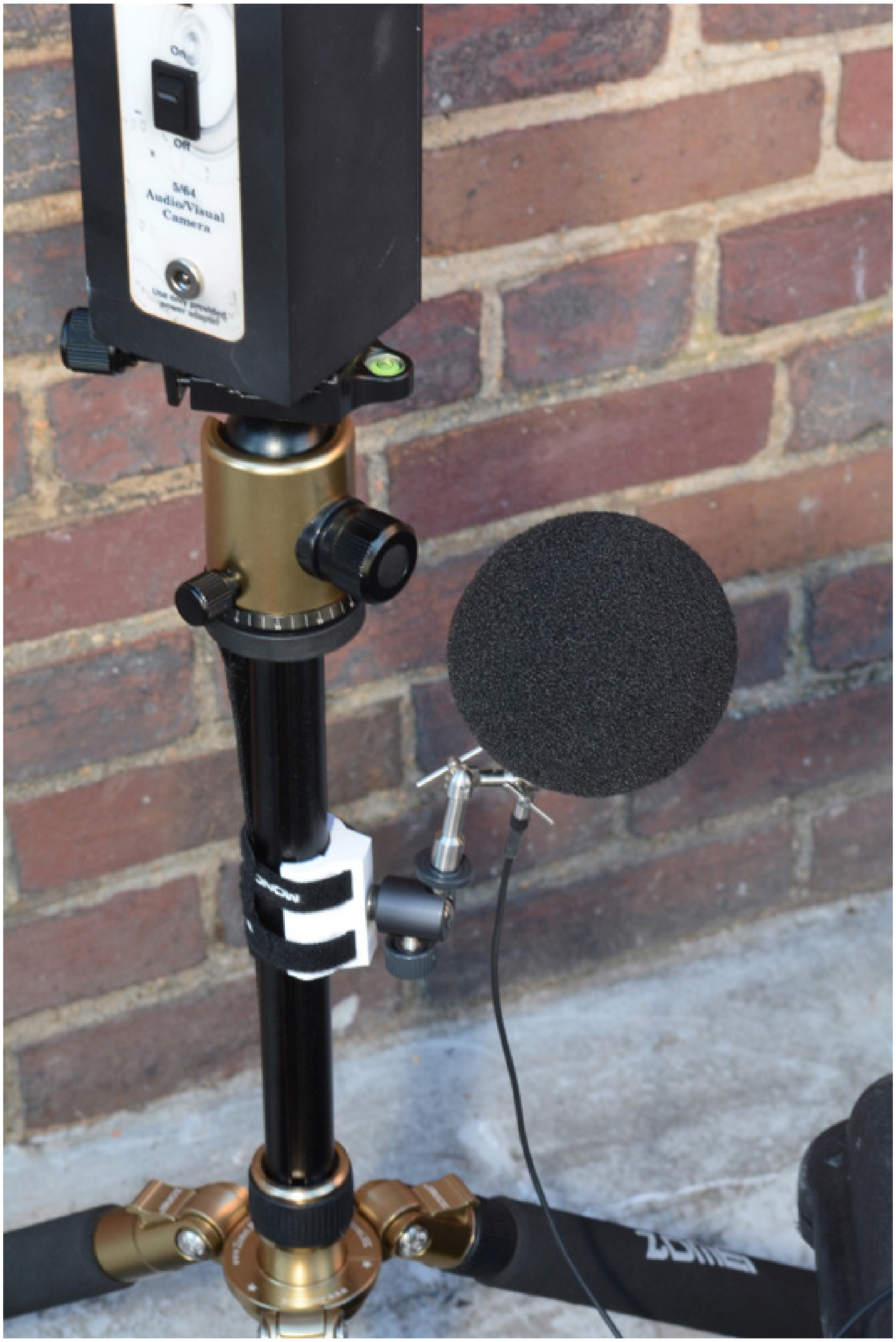


Fig. 5 G.R.A.S. Sound & Vibration 40AF free-field microphone affixed to the VRAP camera tripod (photograph courtesy of Ron Carty)

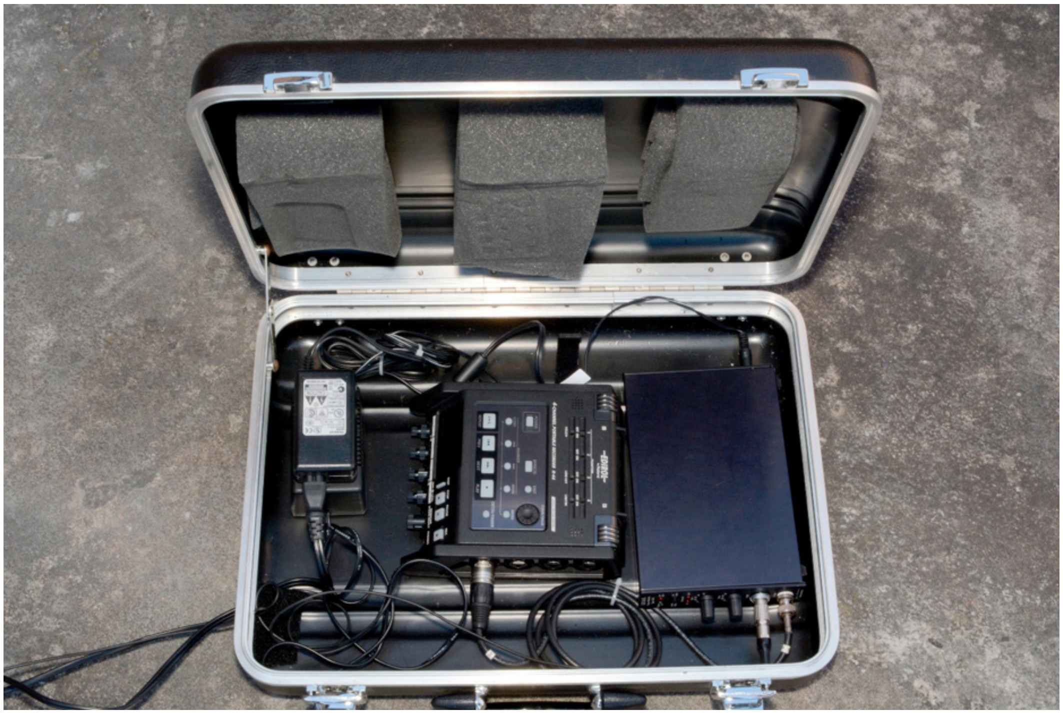
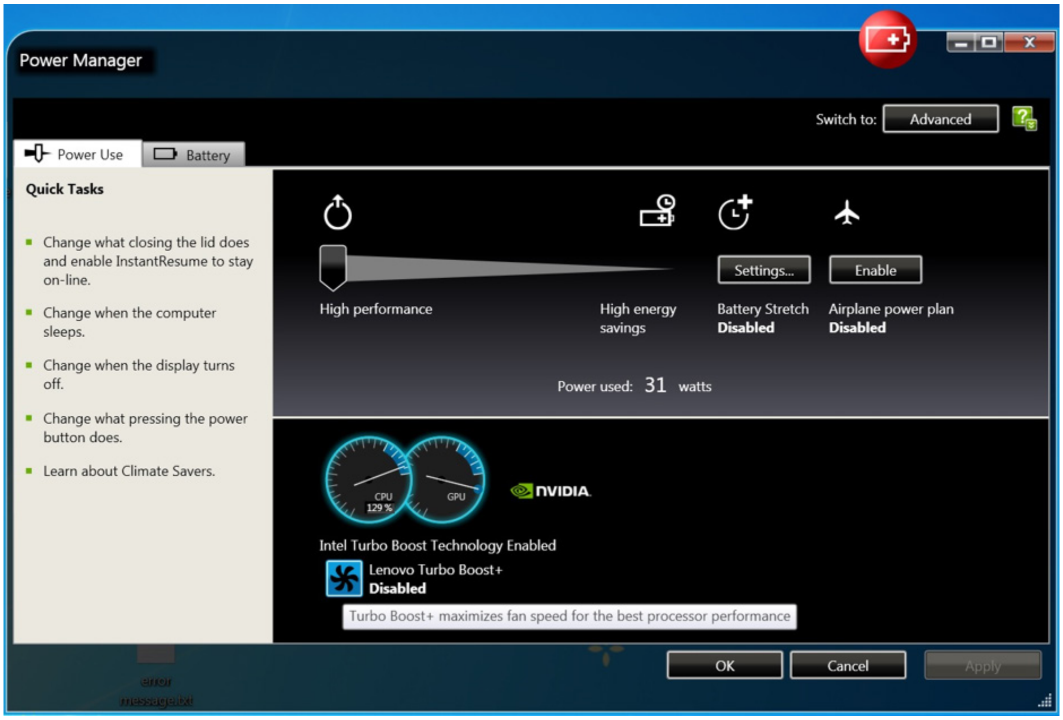


Fig. 6 Auxiliary audio-only recording system (photograph courtesy of Ron Carty)

5.3 Processer Demands and System Resources

The demands of capturing environmental events using 5 HD video and 64 audio channels are such that standard PC power settings are inadequate to handle the load. The VisiSonics technical staff recommend that users activate the “turbo boost+” mode of the Lenovo Thinkpad laptop prior to initiation of data acquisition. Turbo boost+ can be activated from the basic settings menu of the Lenovo Power Manager tool bar (Fig. 7). Activating turbo boost+ increases the speed of the laptop’s system fan to its maximum. Increasing fan speed manages the heat generated by the processors during data acquisition and reduces issues with poor system performance directly related to overheating. The Lenovo W540 laptop is designed to automatically conserve system resources when not connected to an external power source. Users should always connect the power adapter, provided with the Lenovo laptop, to a reliable AC power source during data acquisition. Recording sessions reliably failed in all instances when the laptop was powered by its own internal battery, which seems counterintuitive given that the VRAP system was specifically designed for use in field settings where AC power is typically not available.



Quick Tasks

- Change what closing the lid does and enable InstantResume to stay on-line.
- Change when the computer sleeps.
- Change when the display turns off.
- Change what pressing the power button does.
- Learn about Climate Savers.

High performance

High energy savings

Battery Stretch **Disabled**

Airplane power plan **Disabled**

Power used: 31 watts

CPU 129 %

Intel Turbo Boost Technology Enabled

Lenovo Turbo Boost+ **Disabled**

Turbo Boost+ maximizes fan speed for the best processor performance

Fig. 7 Lenovo ThinkPad power management window

5.4 Data Collection

In preparation for recording, the VRAP camera was attached to the center column of the tripod by means of a quick-release mounting plate threaded into the camera base. To connect the VRAP camera to the laptop, the type-B connector of the provided USB 3.0 cable is plugged into the corresponding marked port on the VRAP camera and the type-A connector is plugged into the marked USB port on the left side of the laptop. Both the VRAP and laptop AC adapters were plugged into available power ports on the Duracell Powerpack. “Turbo boost+” mode was turned on, as depicted in Fig. 7. VisiSonics recommends enabling Lenovo Turbo Boost prior to running the RealSpace acquisition software. Turbo boost is enabled by toggling the blue, fan-shaped icon in the lower-left-hand corner of the power management window.

The VRAP camera was switched on and individual checks of each of the 5 camera elements were performed using the Point Grey FlyCapture 2.0[*] software. This program is accessed through the “Fly Cap 2” icon located on the Lenovo desktop. The FlyCapture 2.0 software displays a list of cameras currently connected to the computer. The cameras are listed by their individual serial numbers and IP addresses. Starting from the top of the list, a camera was selected with the left

[*] Point Grey Research, Inc., Richmond, BC Canada

mouse button followed by clicking the “OK” button. A new window opened, containing a display of the live video stream from the corresponding camera. Once satisfied that a camera was operational, the new window was closed and the next camera on the list was checked in the same manner. Following the check of the final camera, the FlyCapture 2.0 program was closed by left clicking the cancel button in the lower-right corner. Next, the RealSpace Capture program was opened. The program is accessible through 1 of 2 shortcuts on the Lenovo desktop. For routine data collection, the standard program can be accessed through the shortcut labeled “RealSpace Capture”. To aid in troubleshooting connectivity issues between the VRAP and software, the second icon, labeled “Debug RealSpace Capture”, should be selected. In addition to opening the RealSpace Capture program, the shortcut also opens a terminal window with camera status updates. In the main window of the RealSpace Capture program, a round icon, located in the bottom-left corner of the program window, is provided to indicate the status of the connection between the laptop and the VRAP camera. When the icon is green, the camera is correctly communicating with the laptop. If the icon is red, the connection should be checked. A final check of the VRAP camera was performed by selecting the “Real-Time Display” button in the lower right of the program window. Following the software initialization, a new window opened, with a Mercator projection composite display of the 5 camera video feeds. The stitch distance between images in the real-time display can be adjusted by pressing the “o” and “p” keyboard keys; however, this may also be done during post-processing. Several other keyboard shortcuts are available for adjusting the audio and video features of the real-time display (Table 1).

Table 1 Real-time display shortcut keys and respective functions

Keyboard input	Function	Remarks
ESC	Quit	Cleanly exits the program
“a”	Uncertain	Do not press the “a” key. It appears to disrupt the timing of the audio and visual.
, OR <	Decrease sensitivity	Only makes loud sounds stand out in the visual panorama
. OR >	Increase sensitivity	Fainter sounds stand out in the visual image
“–“	Increase persistence filter	More spatially persistent sounds are shown
“=“	Decrease persistence filter	All sounds are shown
o OR p	Decrease/increase camera stitch distance	Panoramic viewer can be set for closer or farther environments
“ c “	Toggle between single-band color mapped output and tri-band red, green, blue (RGB) mapped output	Sound pressure mapping—single band displays as “heat map” Tri-band—displays analogous to video camera with low, middle, and high frequency mapped to RGB, respectively
“[”	Zoom in when the zoom view is enabled	In the lower-left zoom panel the zoom level will be changed
“]”	Zoom out when the zoom window is enabled	In the lower-left zoom panel the zoom level will be changed
“9–0”	Decreases/increases volume of audio output	. . .

Table 1 Real-time display shortcut keys and respective functions (continued)

Keyboard input	Function	Remarks
“b”	Toggle beamformer/3-D sound rendering	Beamformer mode will render the isolated audio that is in the direction clicked on the screen and under the red cursor. 3-D sound mode will render spatialized sound assuming that the red cursor is the look direction.
“h”	Change display to mono head-mounted display (HMD) mode	Will render a view for mono-rendered HMDs
“r”	Enable rift rendering mode Press the sequence “v,h,r” to enable render to Oculus Rift	v enables zoom view, h enables HMD mode, r enables rift mode
“v”	Brings up zoom view window	Unwarped zoomed view window. Steerable by mouse click in main image. “[“ and “]” zooms in and out.

To record a session, the “Real-Time Display” window is closed by pressing the “Esc” button on the keyboard. In the Real-Space Capture window, the “File” menu item is selected, and in the submenu, “New Session” is selected. Sessions are automatically named according to the date and time of creation using the naming convention “vsrsSession-YY-MM-DD-HH-MM-SS”, and by default, stored in the “Sessions” folder of the RealSpace program. Signal gain and the number of seconds in which to record can be set using the controls in the upper left of the window. To begin a recording session, the “Capture” button, just below these controls, is selected. There are no apparent indications to show that the session is recording, and once the recording time has elapsed the system may require additional processing time before the session is complete. Users should refrain from pressing keys or using the mouse until the session processing is complete as doing this may cause the system to become unstable and the Capture program to crash. Following session completion, a new window opens with the captured video scene displayed as a Mercator projection. The captured scene plays and, once completed, the window closes automatically. Again, users should refrain from attempting to enter commands until this process is completed. The next session, and all following sessions, are recorded by returning to the file menu and once again selecting the “New Session” option from the submenu. A new session must be created before selecting the capture button, otherwise the previously recorded session will be overwritten. During a recording session, audio data from the 64 omnidirectional microphone arrays are streamed to the laptop and stored as a single interleaved binary. Video data from the 5 cameras are transferred as separate AVI files. Once all data are transferred to the laptop and the session is complete, single-channel audio from the 64 microphones are automatically extracted from the binary file and stored as individual mono recordings in WAV format, sampled at 44.1 kHz. Video from the 5 camera sources is automatically overlaid as a single Mercator projection and stored as a single, 24 fps, 1280 × 720 dpi, AVI-formatted video. All generated files are retained in the session’s folder.

6. Data Preparation

All of the data captured were analyzed using the RealSpace Acoustic Analysis Tool software, Version 1.8 (VisiSonics 2014). All sessions were processed individually using the software’s graphical user interface (GUI). Optionally, the software supports batch processing from the Windows command line interface (Fig. 8).

6.1 Batch Processing of Beamformers

To execute batch commands under the Windows operating system, first open a command prompt by running cmd.exe from the start menu. At the prompt, change the path to the VisiSonics directory by typing:

“cd v:\visisonics”.

Beamformed sources can be generated using the command structure:

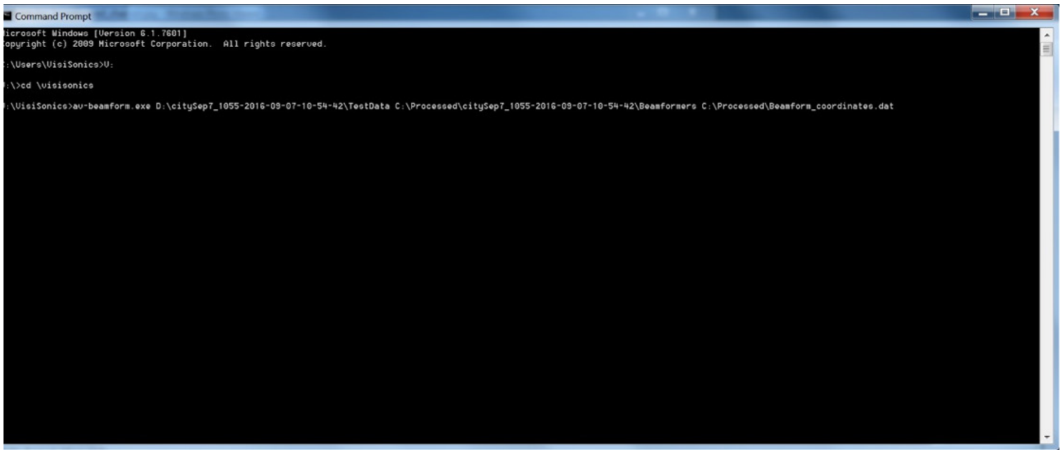
“av-beamform.exe <input .wav prefix> <output .wav prefix> <beam form data list>”

Where:

<input .wav prefix> is the path and prefix of the sessions WAV files

<output .wav prefix> is the path and prefix of the beamformer to be generated, and

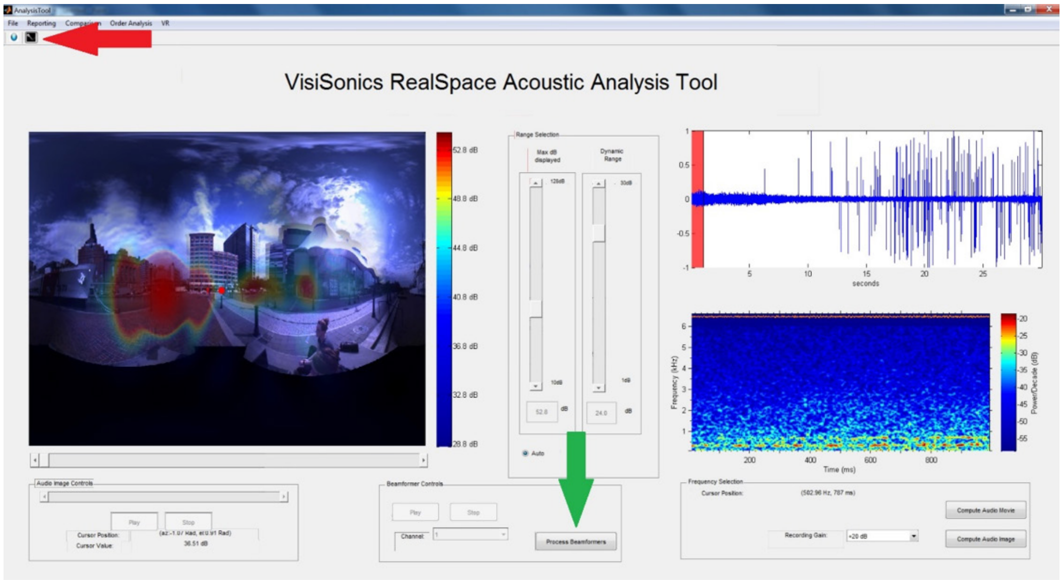
<beam form data list> is the path and file name of a text file, containing the spherical coordinates for the beamformer projection.



```
icrosoft Windows [Version 6.1.7601]
opyright (c) 2009 Microsoft Corporation.  All rights reserved.

:\Users\VisiSonics>V:

:\>cd \visisonics

:\VisiSonics>av-beamform.exe D:\citySep7_1055-2016-09-07-10-54-42\TestData C:\Processed\citySep7_1055-2016-09-07-10-54-42\Beamformers C:\Processed\Beamform_coordinates.dat
```

Fig. 8 Batch processing command entered at the Microsoft Windows command prompt

The analysis tool software is provided for visualization and post-processing of the audio-visual data contained within the session’s folder. The software GUI is sectioned in 3 windows. The session’s time-domain audio signal is depicted in the upper-right section of the interface (Fig. 9). A spectrogram, representing the frequency features of the audio recording over time, is featured in the lower-right section. The left section of the interface contained a Mercator projection of the 5 video streams with a visualized overlay of the acoustic image. Alternately, the recording environment may be viewed as a spherical projection by selecting the first of 2 icons in the far upper-left-hand corner of the interface (Fig. 10). The spherical projection will open in a new window. Orientation of the sphere can be controlled by placing the mouse cursor on the sphere, holding the left mouse button down, and dragging the mouse within the window.



Max dB displayed: 52.8 dB

Dynamic Range: 24.0 dB

Cursor Position: (az:-1.07 Rad, el:0.91 Rad)

Cursor Value: 36.51 dB

Channel: 1

Cursor Position: (502.96 Hz, 787 ms)

Recording Gain: +20 dB

Fig. 9 GUI for the VisiSonics RealSpace Acoustic Analysis Tool. The red arrow pointing to the microphone icon initializes the virtual microphone. The green arrow denotes the button for initializing the beamformers process; this should be selected following coordinate selection in the Mercator projection.

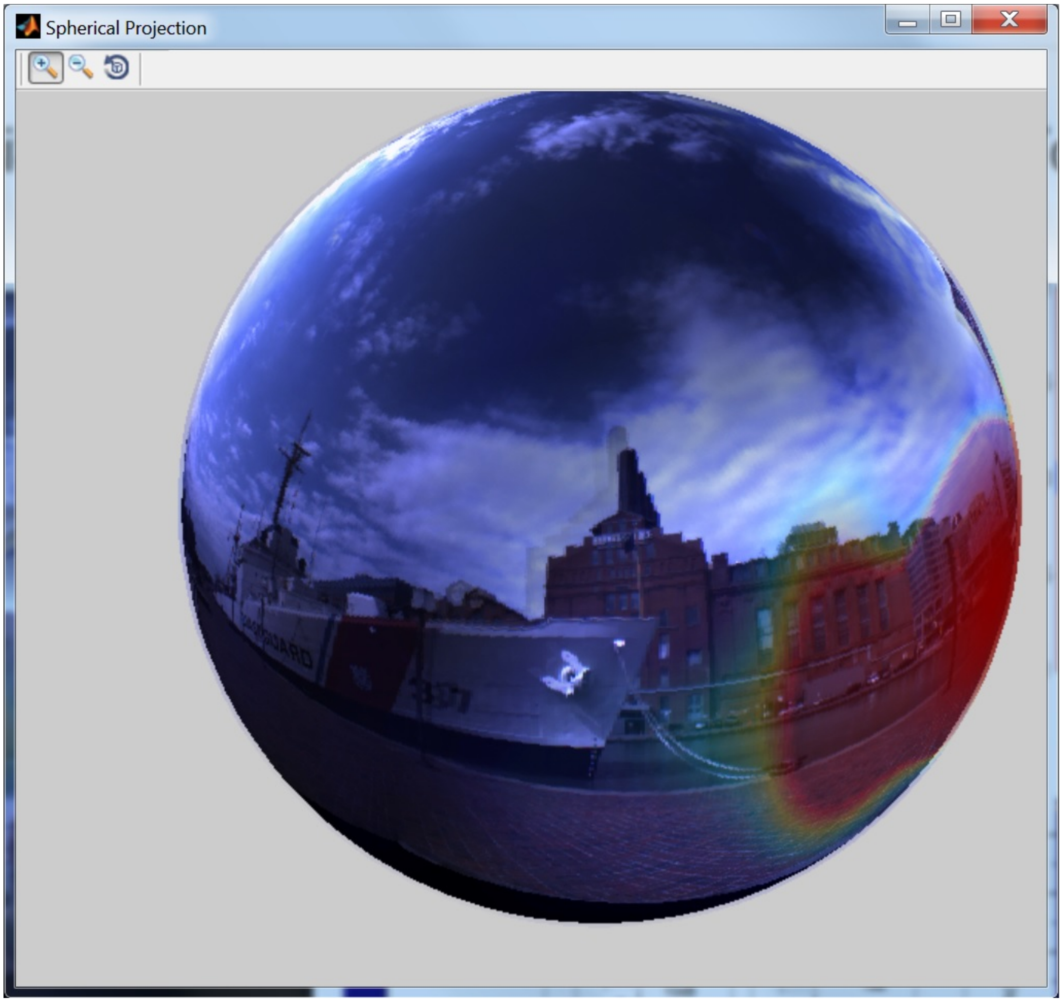


Fig. 10 Spherical projection of the recording environment. The displayed scene was recorded at the location described in the report as an urban setting. A sound source is represented in the highlighted lower-right region, with red indicating the highest sound pressure.

6.2 Audio Data

To compare ecological frequency rating performance between source localized (i.e., beamformed audio) and unedited, single-channel audio, 8 recording sessions (4 from each of the respective locations) were selected for post-processing. Source localization was performed using the beamform function of the audio analysis tool and single-channel audio comparisons were created by retaining the microphone channel closest in spatial approximation to the location selected for the beamformed audio channel and removing the remaining 63 microphone channels.

6.3 Beamforming

Beamforming, or spatial filtering, utilizes an array of microphones to isolate localized sound sources. This technique, either alone or in combination with panoramic visualization of the recording environment, may prove beneficial to raters in estimating the ecological frequencies of sounds. Beamformed audio channels were generated using the RealSpace Acoustic Analysis Tool software, Version 1.8 (VisiSonics 2014). To generate a beamformed audio, a location is first selected within the recorded scene. This is done by toggling on the virtual microphone icon, located in the top left corner of the display (Fig. 9). A location within the Mercator projection is selected using the left mouse button. A red dot and subscript number "1" marked the selected coordinates. This marker remains on the screen after a session is closed and another opened, and can be used as a visual aid for approximating consistent beamformer coordinates between recording sessions. Following the selection of a beamform location, the "Process Beamformer" button (located in the lower, central region of the window) is selected. Beamformed sources were saved in WAV format within a subfolder labeled "Beamform" under the session heading. These steps were followed to create beamformed audio sources for each of the 8 recording sessions.

6.4 Channel Section: Single-Channel Data

For comparison to beamformed audio, a single unprocessed audio channel was selected from the 64-microphone array. Single-channel audio can be exported from any of the 64 microphone channels by copying its corresponding WAV file from the session folder. For the current example, audio data were extracted from microphone channel 12, which was determined to be spatially the nearest physical channel to the spherical coordinates of the beamformed channel. The WAV files from the beamformers and the physical channel served as the audio-only comparison samples.

6.5 Video Data

Audio-visual scenes were generated using OpenShot*, an open-source video editor. The video editor GUI is arranged into 3 windows (Fig. 11). The upper-left window, Project Files, displays the input files provided by the user, and the upper-right window displays the Video Preview of the processed outputs. The bottom, untitled window displays the timeline of the video and audio tracks to be included in the

* OpenShot Studios, LLC, Arlington, Texas 76016

output file. The timeline window is further subdivided into individual track windows.

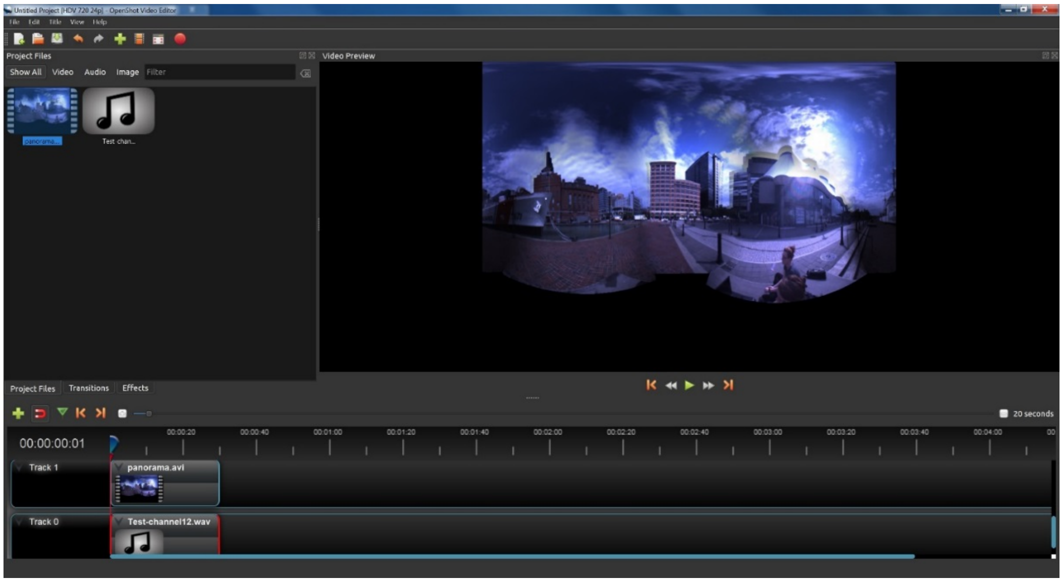
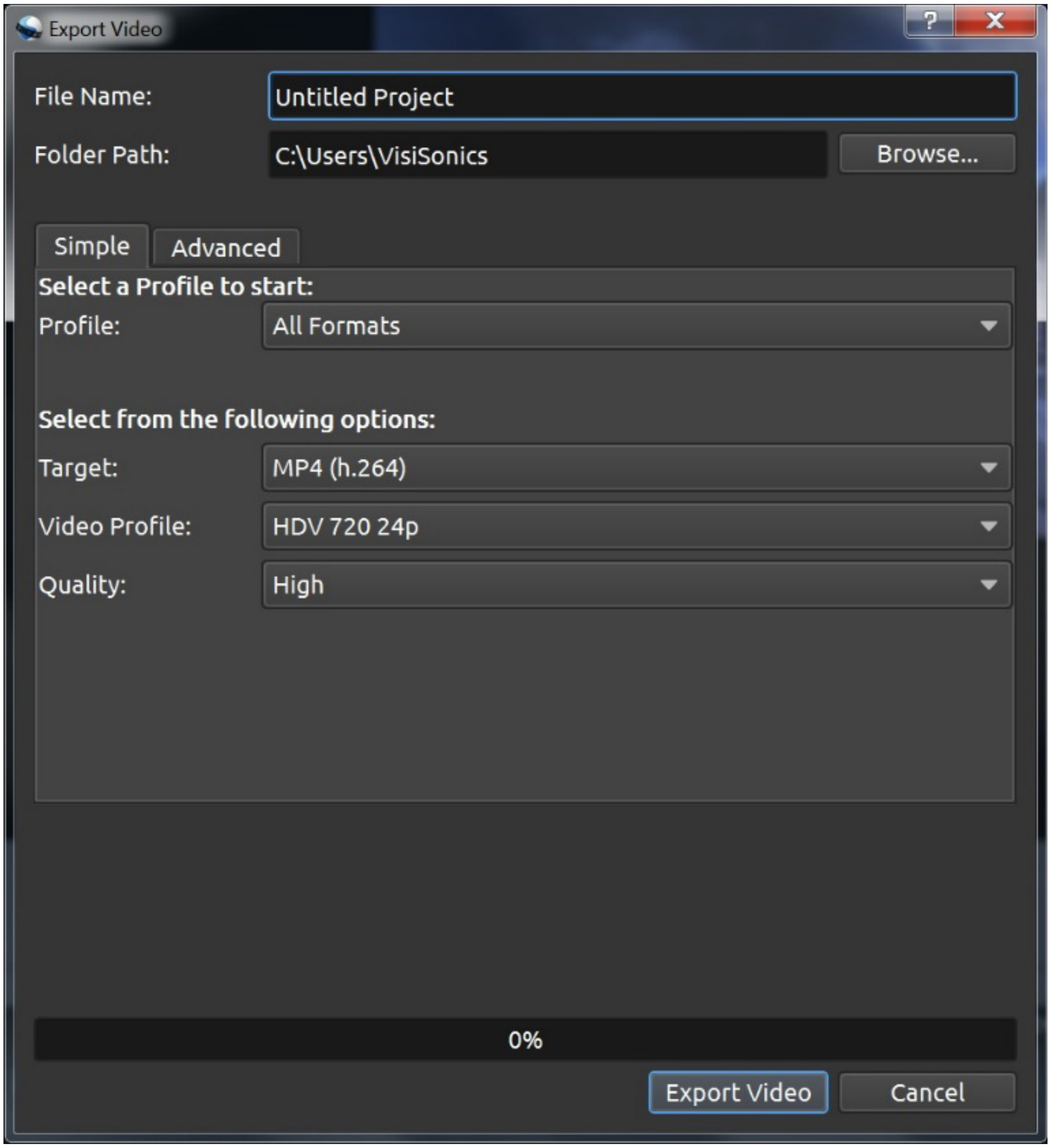


Fig. 11 OpenShot, an open-source video editor, may be used to combine audio and video tracks into a single, exportable file

To create video files with beamformed or single-channel audio for an individual session, the session's panoramic video file is first imported, using the "drag and drop" interface, into the Project Files window. Next, the beamformed audio or single-channel WAV file is imported using the same method. The panoramic video file and the audio track are then drag-and-dropped from the Project Files window to Tracks 0 and 1, respectively, of the timeline window. From the File menu, Export Video is selected to merge the audio and video tracks. Once selected, a new window opens in which the name and folder destination of the output file is specified (Fig. 12). Tabs, in the lower portion of the window, allow users to choose between "Simple" and "Advanced" export options. Three export options are available from the Simple interface: Target, the first option, indicates the video format and codec of the output file. The second option, Video Profile, specifies frame rate and resolution. Quality, the final option of the Simple interface, allows the user to choose between High, Medium, or Low and corresponds to the bit rate of the outputted file. All exported files were saved in MP4 format using the h.264 codec at 24 fps. The High Quality option was selected, indicating a bit rate of 15 Mb/s for all outputted files. The original audio sampling rate of 44.1 kHz was maintained for all files.



File Name: Untitled Project

Folder Path: C:\Users\VisiSonics

Select a Profile to start:

Profile: All Formats

Select from the following options:

Target: MP4 (h.264)

Video Profile: HDV 720 24p

Quality: High

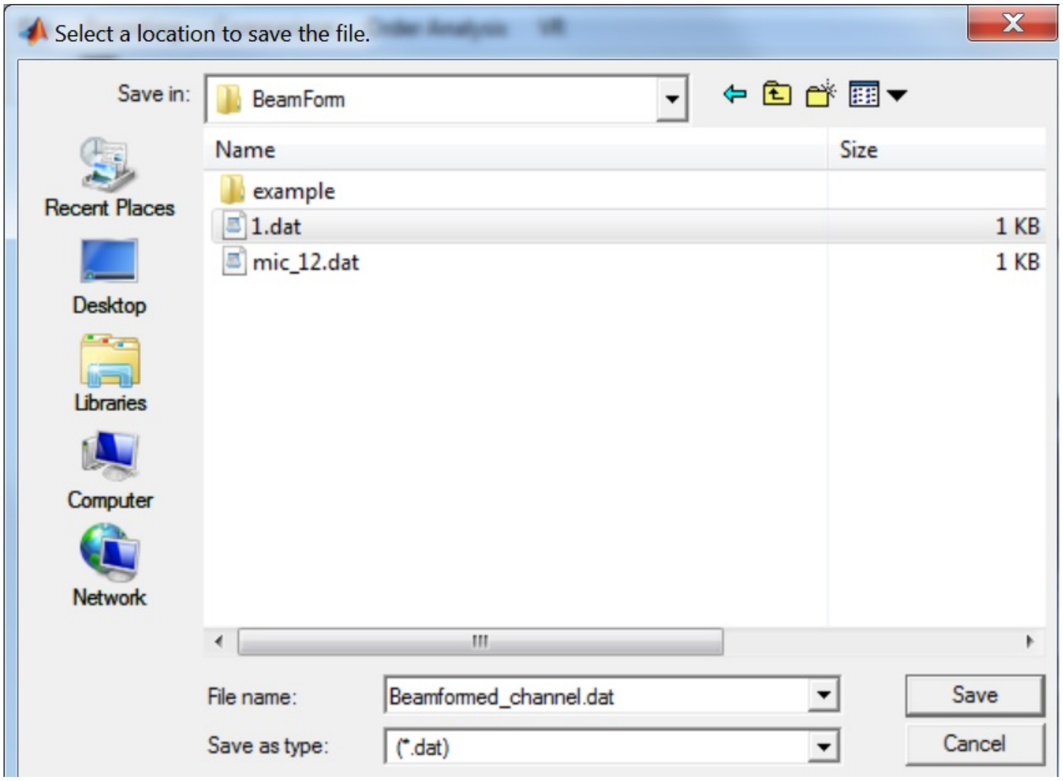
0%

Fig. 12 OpenShot export options window. This window provides options for formatting the exported AV file. Under the advanced tab are options for manipulating frame rate and bit rate.

7. Results and Discussion

7.1 Accuracy of Beamformer Coordinates between Sessions

As noted in Section 6.2, the location of the beamforming coordinates was approximated between sessions by overlaying the beamforming marker with the marker from the previously processed session, which remained on the display after the file had been closed. The authors assumed that the developers of the software intended for the marker to remain on the active display for this purpose; however, this appears to be a programming error rather than an intended feature. Following the completion of the study it was found that the precise coordinates of a beamformed source could be saved and retrieved between sessions. To save beamformer coordinates for later use, follow the procedures outlined in Section 6.3 to generate a beamformer source. Then, select "save beamformer file" from the file menu (Fig. 13). When prompted, select a file location, type a name for the file, and click the "save" button. The coordinates are saved in ASCII format and can be retrieved for later use. To retrieve the coordinates, select "open beamformer file" from the file menu and when prompted, select the file previously saved.



Save in: BeamForm

Name	Size
example	
1.dat	1 KB
mic_12.dat	1 KB

File name: Beamformed_channel.dat

Save as type: (*.dat)

Fig. 13 Coordinates, selected for beamforming within the Mercator projection, can be saved to a text file and retrieved for use with other recorded sessions

7.2 Data Evaluation Methods

To compute ecological frequency, 2 human raters evaluated 4 samples from 2 recording locations (urban and rural) both with and without the video data. Each recording was reviewed twice to ensure the human rater was able to accurately and fully assess each sample. The 2 raters' evaluations of the samples were compared and reconciled to the extent that reconciliation was possible. After reconciliation, inter-rater reliability (IRR) was calculated based on the percentage of cases where the raters disagreed. A detailed discussion of the ecological frequency results are provided in Foots et al. (2016), and will be published in a proceedings paper in the near future.

7.3 Inter-Rater Reliability

Prior to evaluating IRR, the 2 raters checked each disagreement to determine if it was a reconcilable coding error or a genuine difference in scores. Out of 465 detected sound events, the raters disagreed on the presence or labeling of only 91 (19%) of cases. Disagreements between raters were more likely to occur in the urban than the rural environment, which may be attributed directly to the overall level of activity in the urban environment. All urban sessions contained intervals where traffic noise was present to such an extent that uncertainty about the presence of other sources of sound was a noteworthy issue. This informational masking is present in many complex auditory environments; however, during ecological frequency assessment, either live as in Ballas (1993) or from a recording (as in Foots et al. 2016), informational masking is a significant challenge that needs to be addressed.

When the proportion of disagreements were considered as a function of video availability (audio only, audio-video), no significant differences (p >.05) were observed, suggesting that for evaluation of the presence or identity of a sound in a complex scene the addition of video information does not significantly aid in this process. Therefore, it is the opinion of these authors that the VRAP could be used with or without video and provide reliable representation of most complex environments.

8. Conclusions and Lessons Learned

The preliminary results from evaluating the scenes captured using the VRAP camera demonstrate it is unlikely that humans in the environment evaluating a complex scene in real time would be able to produce a reliable estimate of the contents of that environment. Even with the opportunity to pause the recorded session and review the information twice there were numerous discrepancies between the 2 human raters. The authors are currently working to develop additional analysis strategies to better evaluate the rich and complex output from the VRAP camera, both in terms of ecological frequency and other measures to evaluate the contents of the scene captures using the VRAP camera.

The VRAP camera has great potential to capture and play back faithful reproductions of environmental scenes; however, the technology is not without its limitations. The data acquisition using the capture interface built into the VRAP is clunky and has a somewhat steep learning curve for troubleshooting in the field. Practically speaking, the tendency of the VRAP to be highly sensitive to wind noise and overheating makes this device less than ideal for recording in outdoor environments. That is not to say it is impossible to capture outdoor environments, just that the user must take weather conditions into account. The post-processing and suite of analysis tools are not well integrated, and to accomplish basic tasks, such as combining audio and video output streams, additional software was required. Additionally, the provided analysis tools have limited functionality for visualizing the VRAP camera data. To generate plots of scenes, or aspects of scenes, additional software applications will be required; however, for the complexity of data produced, the equipment setup process for acquisition is remarkably simple. The VRAP enables significant improvement in ecological frequency estimate methodology and will have potential utility to any project involving the characterization of a realistically complex and dynamic sensory environment.

9. References

[ARL] ARL Human Sciences Campaign Plan. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015 Feb 5 [accessed 2017 Jan 20]. http://www.arl.army.mil/www/default.cfm?page=2519.

Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. J Exp Psychol Hum Percept Perform. 1993;19(2):250–267.

Carty, R. 2017. Photographs provided courtesy of Ron Carty and used with permission. Weapon System Interface Dynamics Team, Ukpeaġvik Iñupiat Corporation, Bowhead Total Enterprise Solutions.

Dickerson K, Gaston JR, Perelman BS, Mermagen T, Foots AN. Sound source similarity influences change perception in complex scenes. 169th Meeting of the Acoustical Society of America; 2015 May 18–22; Pittsburgh (PA). Proc Mtgs Acoust. 2015;23(1):5–12. doi: 10.1121/2.0000152.

Dickerson K, Sherry L, Gaston J. The relationship between perceived pleasantness and memory for environmental sounds. J Acoust Soc Am. 2016;140(4):3390.

Foots A, Dickerson K, Gaston J. Characterizing real-world auditory scenes using 360° audio-visual capture. J Acoust Soc Am. 2016;140(4):3276.

Gygi B, Kidd GR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. J Acoust Soc Am. 2004;115(3):1252–1265.

[VisiSonics] RealSpace Acoustic Analysis Tool software. Version 1.8. VisiSonics Corporation. 2014 [2017 Oct 31].

List of Symbols, Abbreviations, and Acronyms

3-D	3-dimensional
AC	alternating current
ARL	US Army Research Laboratory
ASCII	American Standard Code for Information Interchange
AVI	audiovideo interleave
CPU	computer processing unit
dB	decibel
dpi	dots per inch
FPGA	field programmable gate array
fps	frames per second
GPU	graphics processing unit
GUI	graphical user interface
HD	high-definition
HMD	head-mounted display
IP	internet protocol
IRR	inter-rater reliability
kHz	kilohertz
Mb/s	megabytes per second
MP4	MPEG-4 Part 14 digital multimedia container format
PC	personal computer
RGB	red, green, blue
SPL	sound pressure level
USB	Universal Serial Bus
VRAP	VisiSonics RealSpace 64/5 Audio-Visual Panoramic
WAV	waveform audio file format

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIR ARL
(PDF) IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 ARMY RSCH LAB – HRED
(PDF) RDRL HRB B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

8 ARMY RSCH LAB – HRED
(PDF) SFC PAUL RAY SMITH
CENTER
RDRL HRO COL H BUHL
RDRL HRF J CHEN
RDRL HRA I MARTINEZ
RDRL HRR R SOTTILARE
RDRL HRA C A RODRIGUEZ
RDRL HRA B G GOODWIN
RDRL HRA A C METEVIER
RDRL HRA D B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 USA ARMY G1
(PDF) DAPE HSI B KNAPP
300 ARMY PENTAGON
RM 2C489
WASHINGTON DC 20310-0300

1 USAF 711 HPW
(PDF) 711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

ABERDEEN PROVING GROUND

15 DIR USARL
(PDF) RDRL HR
J LOCKETT
P FRANASZCZUK
K MCDOWELL
K OIE
RDRL HRB
D HEADLEY
RDRL HRB C
J GRYNOVICKI
RDRL HRB D
C PAULILLO
RDRL HRF A
A DECOSTANZA
RDRL HRF B
A EVANS
RDRL HRF C
J GASTON
RDRL HRF D
A MARATHE
J MCARDLE
A FOOTS
C STACHOWIAK
K DICKERSON