A beamforming video recorder for integrated observations of dolphin behavior and vocalizations (L)

Keenan R. Ball[a)]
Woods Hole Oceanographic Institution, Dept 4, MS 18, 86 Waters Street, Woods Hole, Massachusetts 02543

John R. Buck[b)]
Department of Electrical and Computer Engineering & School for Marine Science and Technology, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300

(Received 11 August 2003; revised 26 August 2004; accepted 7 October 2004)

In this Letter we describe a beamforming video recorder consisting of a video camera at the center of a 16 hydrophone array. A broadband frequency-domain beamforming algorithm is used to estimate the azimuth and elevation of each detected sound. These estimates are used to generate a visual cue indicating the location of the sound source within the video recording, which is synchronized to the acoustic data. The system provided accurate results in both lab calibrations and a field test. The system allows researchers to correlate the acoustic and physical behaviors of marine mammals during studies of social interactions. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1831284]

PACS numbers: 43.80.Ev, 43.80.Ka, 43.80.Jz, 43.30.Sf [WA] Pages: 1005–1008

I. INTRODUCTION

Many cetacean species are acoustically active in social contexts. Additionally, these species spend a majority of their lives underwater where it is difficult for human researchers to observe them. These combined factors challenge researchers studying the physical and acoustic behavior of individual animals during social interactions. Bottlenose dolphins (*Tursiops truncatus*) and other smaller cetaceans increase these challenges by interacting in complex fission–fusion social structures, with animals in close physical proximity. Associating acoustic signals with individual dolphins is essential to studies of acoustic repertoires, juvenile acoustic development,[1] social alliances,[2] and the signature whistle hypothesis.[3] Ideally, observations should be made with minimal perturbations to the animals' natural behavior and environment. Moreover, the observations should be archival in the sense that subsequent investigators should be able to see and hear the animals' behaviors directly in their original form, and not have to rely on the original observers' detections and classifications of the behaviors. Data in this form allow both independent confirmation of behavioral hypotheses and reanalysis in light of subsequently proposed alternative hypotheses. Lastly, the system should be portable for situations where the animals do not remain in a single location.

Previous techniques used to associate sounds with individual marine mammals include emitted bubble streams (e.g., Ref. 4) isolating animals (e.g., Refs. 1, 3), tags (e.g., Refs. 5, 6), and hydrophone arrays (e.g., Refs. 7, 8). None of these approaches produced archival video footage, relying on human observers to link the animals to the sounds made. Three recent systems[9–11] integrate hydrophone arrays with video recordings to produce archival observations, but still differ in significant ways from the system described in this Letter. The system in Ref. 10 uses only two hydrophones spaced at five times the human intra-aural distance to provide coarse localization cues (left, right, or both) while reviewing the video recordings. The system provides no vertical resolution cues.

Thomas *et al.*[11] combined an elevated video camera with a distributed array of eight hydrophones around a lagoon perimeter to observe behavior. The location of each sound source was determined by cross-correlating the hydrophone signals to estimate the relative arrival times of the sounds at the array. The resulting location estimate was projected into the video image. Calibration tests determined an accuracy of roughly 2 m for the system. While the system produces large-scale archival overhead video records of the physical and acoustic behaviors of the animals with little or no perturbation of the observed animals' behavior, the limited accuracy makes it impossible to discern acoustic behavior among closely spaced dolphins, e.g., mother–calf pairs, and the need for the manual selection of sounds makes the video post-processing labor intensive.

Au and Herzing[9] developed a two-dimensional Y-shaped four hydrophone array including a video camera to study Atlantic spotted dolphin (*Stenella frontalis*) echolocation clicks. This system used differences in the clicks' arrival times to estimate the range to the echolocating dolphin, but did not estimate the bearing to the animal.[12] Because it was designed solely for click analysis, the system has memory limitations that prevent it from recording and analyzing entire whistles or tracking a whistling dolphin moving across multiple video frames.

In this Letter we describe a beamforming video recorder (BVR) designed to produce archival video recordings of complicated social interactions including whistles and

[a)]Formerly at the Department of Electrical and Computer Engineering, University of Massachusetts, Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300; electronic mail: kball@whoi.edu
[b)]Electronic mail: johnbuck@ieee.org

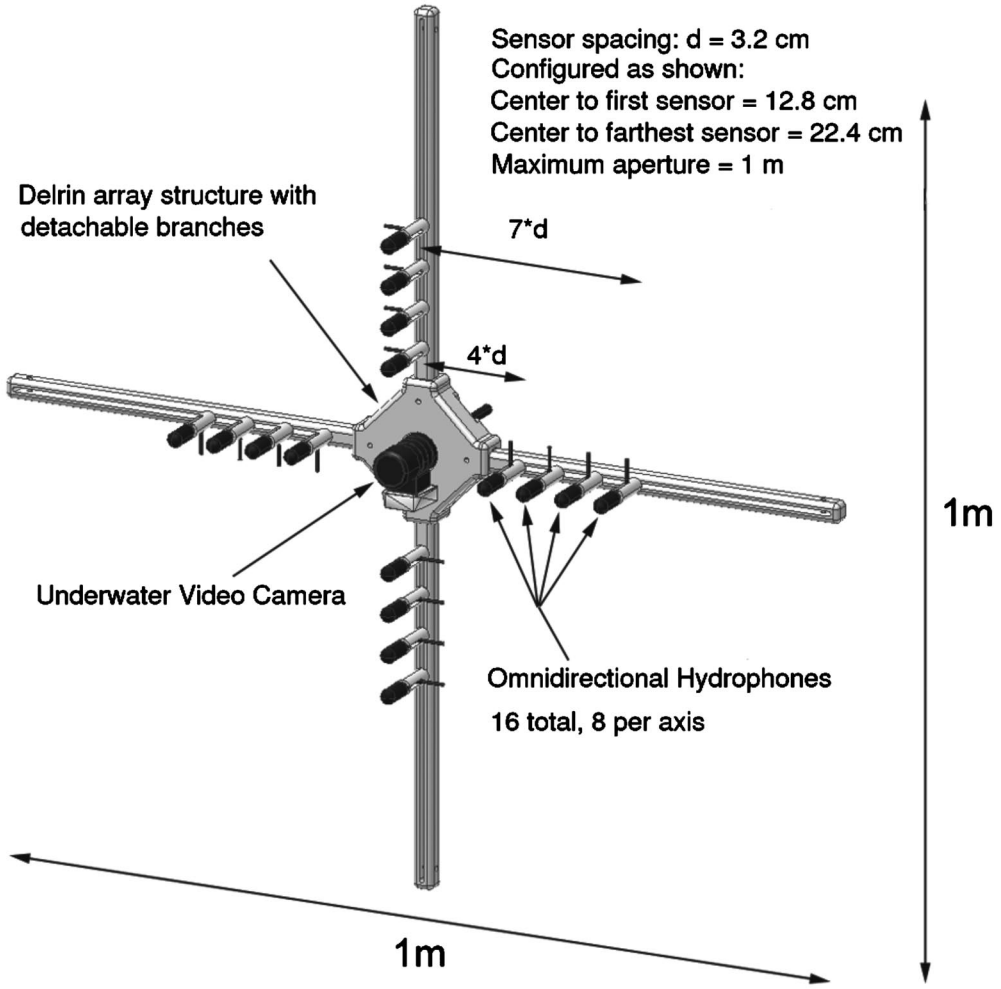

FIG. 1. Two-dimensional handheld audio/video array with detachable branches. The array consists of 16 hydrophones, 8 per axis, and an underwater video camera located at the center. The diver hand holds, mounting brackets, and tether are not shown.

echolocation clicks. The video recordings include visual cues indicating the location of each sound observed, tracking across video frames where necessary. The BVR provides short range underwater observations with a high resolution beamformer that can potentially distinguish among several vocally active dolphins in view of the camera.

II. SYSTEM DESCRIPTION

The BVR consists of a short-baseline Mills cross 16 hydrophone array incorporating a video camera at the array center and associated audio and video recording hardware. Placing the video camera at the array center gives a common reference frame for the bearing estimates and the video image, removing the need for field recalibration and registration as in Ref. 11. The system is largely comprised of off-the-shelf parts, minimizing the need for custom hardware. The acoustic signals are processed using broadband frequency–domain beamforming algorithms, allowing for automated processing and independent position estimates for each video frame. The visual cues incorporated into the video record can track a whistling animal as it moves across the image. The resulting video stream is an archival record linking the acoustic and physical behaviors of small cetaceans in dense social contexts.

A. Hydrophone array and camera

Figure 1 is a diagram of the BVR array and camera with the hydrophones configured in the reduced aperture as used for the tests described in Sec. III B. The array is constructed of Delrin with detachable branches. Each of the four arms is 45.7 cm long by 3.2 cm square, holding four hydrophones with integrated 10 dB preamplifiers (High Tech Inc., HTI-96-MIN). The hydrophones can be reconfigured to suit recording conditions using a slot in each arm. Calibrations confirmed that the phase variations among the hydrophones produce a negligible bias on the beamforming results. The array's center piece is 17.8 cm tall by 18.4 cm wide by 4.4 cm deep. A 5 cm diameter hole in this piece mounts a DeepSea 2050 (DeepSea Power and Light, San Diego, CA) underwater camera at the array center. The maximum total aperture of the array is roughly 1 m, making it convenient to deploy from a dock or moving boat in an inverted periscope configuration, or to be maneuvered by a swimmer. Due to the low visibility during the field test described in Sec. III B, the hydrophones were positioned at ± 13.3, ± 16.5, ± 19.7, and ± 22.9 cm on each axis relative to the origin at the array center, giving a total aperture of about 46 cm.

B. Recording system

The 16 hydrophones are connected through custom breakout boxes to a Tascam MX-2424 multitrack hard disk recorder. The MX-2424 synchronously records all 16 channels at 44.1 kHz with 24 bit resolution directly to its internal SCSI hard disk. An additional channel on the MX-2424 records a sync signal that is also recorded on the left channel of the video camera. This allows the video and acoustic array recordings to be synchronized with a precision of 1/44.1 kHz $=22.7\ \mu$s. The video signal from the DeepSea 2050 underwater camera is recorded by a Sony DCR-TRV-530 Digital 8 camcorder. To simplify processing, the video is converted from the NTSC 29.97 frames-per-second DF (Drop Frame) video standard to the 30 frames-per-second ND (Non-Drop Frame) standard. At the 30 Hz frame rate, there are exactly 1470 samples of 44.1 kHz acoustic data per video frame.

C. Beamforming algorithm

The recorded data is processed in Matlab (The Mathworks, Natick, MA) to synthesize video recordings with localization cues. Whistles and echolocation clicks were detected by comparing the average acoustic energy in each 1470 sample frame against an empirically determined threshold value. The azimuth and elevation of each detected sound are estimated using the broadband frequency–domain algorithm for sparse arrays in Ref. 13. Equivalent angular resolution could be obtained using an array of only two hydrophones, however, the multiple hydrophone array geometry employed reduces the number of grating lobes and attenuates the amplitudes of the sidelobes in the frequency–wave number response or beam pattern (Ref. 14, Sec. 2.2). Consequently, the BVR rejects noise from undesired bearings more robustly than could a two hydrophone per axis array. The system beamforms the array data at the peak frequency in each block, and also at harmonics of the peak frequency. Frequency–domain beamforming allows automated processing of the acoustic data, updating the position estimate 30 times a second, which is not possible for systems that use time–domain autocorrelations on an entire whistle such as Ref. 11. The estimated elevation and azimuth angles from the beamformer were converted to pixels in the video image using conversions of 11.083 pixels/degree in azimuth and 8.727 pixels/deg in elevation established during calibration tests in the SMAST Acousto-Optic tank. A localization cue of a small + sign was spliced into each frame of video based on these pixel coordinates. Elevations or azimuths just offscreen were indicated by highlighting the nearest border of the video image.

III. RESULTS

A. Calibration tests

The BVR was calibrated in both indoor tank tests and a free swimming outdoor test. The tank tests transmitted five cycles of a 5 kHz sinusoid with three harmonics (10, 15, and 20 kHz) from a source at a known fixed range and bearing. These tests established that for the array geometry in Sec. II A, any source at a range of 1 m or more was sufficiently in the farfield to produce an accurate bearing estimation. The tests varied the elevation and azimuth of the source over the camera's field of view, and the BVR nearly always placed the + cue on the source or within five pixels of the source. This is an error of less than 4 cm even at a 4 m range. It is also possible to use the BVR to triangulate the range to a source using the two outermost sensors on each arm to obtain four separate bearing estimates. When the hydrophones are set to provide the maximum aperture (about 1 m, twice that used in Sec. III B), this triangulation approach produced range estimates accurate to $\pm 10\%$ for ranges of 1–4 m. The free swimming test was conducted off the UMass Dartmouth SMAST pier in water roughly 3–4 m deep over a bottom with a mix of rocks and sand. The source played recorded dolphin whistles with three to four harmonics. The video cue tracked the source closely as the current and the swimmer's motion caused the source to move within the video frame.

B. Field test

The BVR field test took place at the Dolphin Connection at Hawk's Cay Resort in Duck Key, Florida. The Dolphin Connection was chosen as the testing site because of its natural lagoon setting, the number of well-trained dolphins, and the representative acoustic conditions including snapping shrimp and nearby boat traffic. The sand and coral bottom of the lagoon provide realistic boundary conditions for acoustic propagation. The seaward edges of the lagoon are formed by plastic fencing secured to wooden posts embedded in the bottom. The dolphins generally acclimatize well to the novel activities because the Dolphin Connection frequently hosts research projects.

The dolphin training staff suggested that a dummy array be used to desensitize the dolphins during the months preceding the experiments using the actual array. The dummy array, constructed out of Delrin to closely resemble the actual array, was installed in March 2002. The array was mounted to the dock with a clamping mount holding a 1.2 m PVC tube with a 0.61 m horizontal tube glued on the top end. A buckled strap closed the clamp mount while allowing the rapid release and removal of the array in the event of an adverse reaction. The array could be rotated in order to aim the camera at any activity in the lagoon.

The field test took place during the last week of May and the first week of June 2002. Most testing took place during 1 h sessions in the morning before the trainers began their daily morning cleaning and feeding routine. We obtained useful data on five of seven days of deployment. In total, over 145 minutes of data were processed, including over 57 distinct whistles. The narrowband signal-to-noise ratio (SNR) at the fundamental frequency was at least 80 dB in all of the data processed. For all 57 whistles, the cue was placed on or very close to (two to three pixels) the dolphin's head, indicating a maximum error of less than 0.5 deg of elevation or azimuth. Figure 2 shows a still frame from the video recorded on 28 May 2002 observing a adolescent male dolphin named Kai when he was the only dolphin in the lagoon. The contrast of this image was increased from the original color image for a black and white publication. More video data including the sample frame presented in Fig. 2 are available

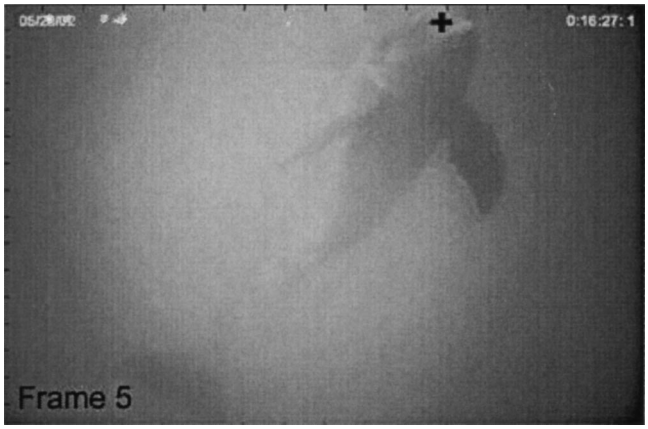

FIG. 2. Representative still frame from the BVR output, recorded on 28 May 2002 showing Kai whistling. The cue accurately followed Kai's head throughout a 3 s whistle including many echolocation clicks. The original color video frame has been digitally retouched to improve the contrast for a black and white publication.

at the website indicated in Ref. 15. In all trials, the BVR performed well through interference such as snapping shrimp, boat noise, and echolocation clicks. The high SNRs in the recordings meant that the limit on the BVR's performance was generally underwater visibility (4–5 m), and not the beamforming algorithm.

IV. DISCUSSION

The BVR performed very well monitoring dolphins in a natural lagoon. In almost every test, the cue was placed on the head of a dolphin in the video image. In most tests, the few erroneous cues, if any, were single frame transients and did not generate confusion about the identity of the whistling animal. The only whistle series that consistently confounded the BVR was when the whistling dolphin swam beneath the dock along the fence, causing a complicated multipath arrival structure. Although we never observed a whistle exchange between two dolphins, the BVR's consistent precision placing the cue on the dolphin's head gives us confidence that the system will perform well when such an event is observed.

Our efforts to desensitize the dolphins to the array's presence were only partially successful, as the animals often focused their behavior on the array, scientists, and trainers. The animals' focus on the array was useful for our engineering test, providing a substantial dataset in a short period, but would be a drawback in any behavioral experiment. In such a test, the animals must be more thoroughly desensitized to the array's presence, or perhaps the array should be physically isolated by fencing and deployed unattended to minimize the animal's interest.

In conclusion, video recordings incorporating visual cues indicating the origins of sounds are a powerful tool for studies linking physical and acoustic behavior in cetaceans. The BVR system presented in this paper is a portable system suitable for observing highly active social contexts and producing archival recordings of these observations. The BVR holds promise for studying open questions in bottlenose dolphin behavior.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Gina Wood, the head trainer at the Dolphin Connection. Gina generously and creatively contributed to the planning and execution of the field test. This work would not have been possible without the cooperation and contributions of the training staff at the Dolphin Connection. Doug and Cheryl Messinger generously allowed us access to the animals at their facility. Jason Davis, Lisa Davis, Amy Dobelle, Adrien Domske, and Sylvia Rickett patiently came in early and stayed late to help us with our field test. Conversations with Peter Tyack inspired the BVR. Larry Reinhart and Rob Fisher at UMass Dartmouth gave valuable advice and assistance in the design and calibration of the BVR. Will Moore machined the BVR. Two reviewers made numerous helpful suggestions to improve this Letter. This research was funded by NSF Ocean Sciences CAREER award 9733391. This Letter is UMass Dartmouth School for Marine Science and Technology Contribution No. 040301.

[1] L. S. Sayigh, P. L. Tyack, R. S. Wells, and M. D. Scott, "Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: stability and mother-offspring comparisons," Behav. Ecol. Sociobiol. **26**, 247–260 (1990).

[2] R. C. Connor, M. R. Heithaus, and L. M. Barre, "Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'," Proc. R. Soc. London **268**, 263–267 (2001).

[3] M. C. Caldwell and D. K. Caldwell, "Individualized whistle contours in bottle-nosed dolphins (*Tursiops truncatus*)," Nature (London) **207**, 434–435 (1965).

[4] M. E. Dahlheim and F. Awbrey, "A classification and comparison of vocalizations of captive killer whales (*Orcinus orca*)," J. Acoust. Soc. Am. **72**, 661–670 (1982).

[5] M. Johnson and P. L. Tyack, "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12 (2003).

[6] P. L. Tyack, "An optical telemetry device to identify which dolphin produces a sound," J. Acoust. Soc. Am. **78**, 1892–1895 (1985).

[7] W. A. Watkins and W. E. Schevill, "Sound source location by arrival times on a non-rigid 3-dimensional hydrophone array," Deep-Sea Res. **19**, 691–706 (1972).

[8] C. W. Clark, "A real-time direction finding device for determining the bearing to the underwater sounds of southern right whales, *Eubalaena australis*," J. Acoust. Soc. Am. **68**, 508–511 (1980).

[9] W. W. L. Au and D. L. Herzing, "Echolocation signals of wild atlantic spotted dolphin (*Stenella frontalis*)," J. Acoust. Soc. Am. **113**, 598–604 (2003).

[10] K. M. Dudzinski, C. W. Clark, and B. Wursig, "A mobile video/acoustic system for simultaneous underwater recording of dolphin interactions," Aq. Mam. **21**, 187–193 (1995).

[11] R. E. Thomas, K. M. Fristrup, and P. L. Tyack, "Linking the sounds ofdolphins to their locations and behavior using video and multichannel acoustic recordings," J. Acoust. Soc. Am. **112**, 1692–1701 (2002).

[12] W. W. L. Au (personal communication, 2003).

[13] A. Thode, T. Norris, and J. Barlow, "Frequency beamforming of dolphin whistles using a sparse three-element towed array," J. Acoust. Soc. Am. **107**, 3581–3584 (2000).

[14] H. L. Van Trees, "Optimum array processing," Part IV of *Detection, Estimation and Modulation Theory* (Wiley, New York, 2002).

[15] See EPAPS Document No. EPAPS-JASMAN-117-705501 for downloadable files containing the video files showing the beamforming video recorder output as it tracks whistles and clicks from a dolphin and places a + cue on the video image based on the beamformed array data. A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps. See the EPAPS homepage for more information.