2019

Who Said That? Towards a Machine-Prediction-Based Approach to Tursiops Truncatus Whistle Localization and Attribution in a Reverberant Dolphinarium

Sean Fisher Woodward

WHO SAID THAT? TOWARDS A MACHINE-PREDICTION-BASED APPROACH TO *TURSIOPS TRUNCATUS* WHISTLE LOCALIZATION AND ATTRIBUTION IN A REVERBERANT DOLPHINARIUM

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Sean Fisher Woodward

June 2019

WHO SAID THAT? TOWARDS A MACHINE-PREDICTION-BASED APPROACH TO *TURSIOPS TRUNCATUS* WHISTLE LOCALIZATION AND ATTRIBUTION IN A REVERBERANT DOLPHINARIUM

Sean Fisher Woodward, Ph.D.

The Rockefeller University 2019

Dolphin communication research is an active period of growth. Many researchers expect to find significant communicative capacity in dolphins given their known sociality and large and complex brains. Moreover, given dolphins' known acoustic sensitivity, serving their well-studied echolocation ability, some researchers have speculated that dolphin communication is mediated in large part by a sophisticated "vocal" language. However, evidence supporting this belief is scarce.

Among most dolphin species, a particular tonal class of call, termed the *whistle*, has been identified as socially important. In particular, for the common bottlenose dolphin, *Tursiops truncatus* – arguably the focal species of most dolphin cognitive and communication research – research has fixated on "signature whistles," individually-distinctive whistles that seem to convey an individual's identity to conspecifics, can be mimicked, and can be modulated under certain circumstances in ways that may or may not be communicative.

Apart from signature whistles, most studies of dolphin calls concern group-based repertoires of whistles and other, pulse-form call types. However, studies of individual repertoires of non-signature whistles, and the phenomenon of combined signature and non-signature vocal exchanges among dolphins, are conspicuously rare in the literature, tending to be limited by either extreme subject confinement or sparse attributions of vocalizer identity. Nevertheless, such studies constitute a logical prerequisite to an understanding of the communicative potential of whistles.

This absence can be explained by a methodological limitation in the way in which dolphin sounds are recorded. In particular, no established method exists for recording

the whistles of an entire social group of dolphins so as to reliably attribute them to their vocalizers.

This thesis proposes a dolphinarium-based system for achieving audio recording with whistle attribution, as well as visual behavioral tracking. Towards achieving the proposed system, I present foundational work involving the installation of permanent hydrophone arrays and cameras in a dolphinarium that enforces strict animal safety regulations.

Attributing tonal sounds via the process of sound localization – estimation of a sound's point of origin based on the physical properties of its propagation – in a highly reverberant environment is a notoriously difficult problem, resistant to many conventional signal processing techniques. This thesis will provide evidence of this difficulty, and also a demonstration of a highly effective machine-learning-based solution to the problem.

This thesis also provides miscellaneous hardware and the pieces of a computational pipeline towards completion of the full proposed, automated system.

Once completed, the proposed system will provide an enormous data stream that will lend itself to large-scale studies of individual repertoires of non-signature whistles and combined signature and non-signature vocal exchanges among an invariant group of socializing dolphins, representing a unique and necessary achievement in dolphin communication research.

Acknowledgements

This project was big. Not only was it big in its physical dimensions – the equipment was big, the equipment was distributed and connected in a big, dolphinarium-size network, and it was all located 200 miles from where I lived – but it was big in the breadth of the technical fields it covered – it involved concepts from mechanical and electrical engineering, physics and acoustics, machine learning, and animal behavior. Moreover, the project was delicate, much of it occurring in plain sight of the thousands of people that visit the National Aquarium's Dolphin Discovery every week to view our irreplaceable dolphin subjects. Entering the project as a solitary, somewhat bumbling student with vague backgrounds in physiology and physics, and from a small theoretical neuroscience lab, I owe whatever success I have achieved these past years to a large support network that has been absolutely essential to the project's growth and survival, as well as my own.

Apart from me, this project impacted the day-to-day responsibilities of no one more than the collective membership of the National Aquarium Dolphin Discovery training and veterinary staffs. For over two years, hydrophone arrays and other apparatuses have been installed in the habitat of the animals that this membership alone is ultimately responsible for keeping safe, and I am grateful to have received its trust and time. Concerning the training staff in particular, the work hours it has dedicated to installing and maintaining the hydrophone arrays are unmatched by anyone. Every hydrophone array installation or removal required the scuba support of at least one trainer, and once installed a hydrophone array required ongoing monitoring and upkeep, executed primarily by the training staff. Moreover, the training staff's cooperation and assistance was essential to every other pool-related task I organized, such as making underwater measurements, performing sound calibrations, and even installing overhead cameras. The staff is quite large and dynamic and I cannot thank everyone individually here, but I would like to highlight and thank Kerry Diehl, Allison Ginsburg, Susie Walker, April Martin and head veterinarian Leigh Clayton for their participation in this work.

Whenever I visited the National Aquarium, a member from the resident audiovisual group was invariably at my side providing a reliable second pair of eyes and hands. This person was often Jim Tunney; together we lifted hydrophone arrays into the water, hung off of high catwalks to secure cameras, pulled cables through meters of conduit, and pulled calibration buoys across the pool. Manny Chico also filled in for these duties, but more regularly shared his trove of information about audio tech and hand tools; Richard Snader, head of the team, provided broad organizational support while never hesitating to be involved in the manual work as well. As this project comes to a close, I count the members of the AV team among my friends. Along with the AV team I would also like to thank the National Aquarium's network administrator, Allan Consolati, who was quick to share his network as well as his expertise, as well as Mark Kennedy for his contribution of his aesthetic sense in the design of the panels.

From the National Aquarium administration, I am particularly thankful to Jill Arnold for helping me to navigate the aquarium's bureaucratic machinery.

Otherwise, many more members of the National Aquarium staff provided assistance at various points over the years, from providing cots to life support information, and so I thank the entire staff of the National Aquarium.

Of course, without the optimism and enthusiasm of Diana Reiss and Marcelo Magnasco for this project, it certainly would not have made it this far. This project bears the marks of Diana and Marcelo's tenacity, sense of innovation, and passion for dolphins. Additionally, their positivity has been inspirational and I hope to carry a portion with me onward.

Apart from Diana and Marcelo, several colleagues and friends assisted with hydrophone array installations and removals, overnight observations of arrays, catwalk measurements, sound calibration sessions, and various logistics. Megan McGrath, Ana Hočevar, and Stephanie Bousseau provided especially frequent and significant support, and I also thank Alla Villafana, Raymond Van Steyn, Miranda Trapani, Kristi Collom, Brigid Maloney, Robert Dutchen, Kevin Justus, Elias Ohrstrom, Adrienne Koepke, Jennifer Savoie, Dan Firester, and Jason Manley. Also, Eric Ramos and Daisy Kaplan

from the Reiss lab, who I accompanied to field sites to get a taste for wild dolphin research, have been invaluable sources of dolphin information and data.

The hydrophone arrays (of which there are two sets) are intended to be identical assemblies consisting of many pieces, which were sometimes constructed of stubborn materials and to surprisingly tight tolerances. Virtually all of the credit for machining the arrays goes to Vadim Sherman; a small batch machinist of his quality can be hard to find, and I was fortunate that Vadim had set up shop one building away from my office. Sanjee Abeytunge is responsible for much of the design and all of the fabrication of custom voltage fan-out circuit boards that are important in the hydrophone recording circuit. Vadim and Sanjee not only executed the jobs commissioned of them but freely acted as mentors to me in their respective fields, which allowed me to adapt a degree of their expertise to my work with them as well as to other tasks. Pierre-David Letourneau, Paul Mountcastle, Jim Petrillo, Kunal Shah, Carl Modes, Christoph Kirst, Joe Olsen, Mark Turner, Ofer Tchernichovski, and Fernando Nottebohm, all experts in different areas, also freely offered mentorship at different points throughout this and related projects. Together with the dolphin, exhibition, AV and IT staffs at the National Aquarium, these people helped me to navigate technical fields I was deeply unfamiliar with.

When home, my dog Kepler provided constant emotional support. During my frequent and sometimes long stays in Baltimore, a subset of my friends were indispensable to Kepler's care. At one time Tommy Hsiao and Dana Mamriev took care of him so regularly as to have their own toys for him, and Michael Mitchell and Sara Liu, Dimitrios Moirogiannis, Frank Tejera and Daniela Pla Jauregui, Javier Marquina, and Anna Yoney have also stepped in when I have been in need.

I have been lucky to have many friends supporting me, but in particular I thank Michael Mitchell for our frequent work coffee discussions. I also thank Ann Viteri-Jackson and Melanie Lee from The Rockefeller University for logistical support.

This project was not easy to preside over given the many random variables on which its success depended, many reflecting the unpredictability of the dolphins as

well as the institution responsible for them. I shared the stress of this duty in part with my committee. Its chair, Jim Hudspeth, demonstrated a much needed mixture of enthusiasm for the project and concern for my graduate experience, as well as responsibility to the committee and a sense of pragmatism. I trusted him immensely. Alipasha Vaziri generously agreed to join the committee late, and quickly contributed sharp suggestions and criticisms tempered by an understanding of the unique challenges of working at the aquarium. Both Jim and Alipasha were pleasures to consult with one-on-one. Together with Marcelo, Diana, and Sarah Woolley, who generously agreed to join the committee for the defense, I thank the committee.

For funding, I thank the National Science Foundation, Division of Physics, the Eric and Wendy Schmidt Fund for Strategic Innovation, and the Rockefeller University.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Humans and Dolphins

Members of the family Delphinidae, in particular those belonging to the common bottlenose dolphin species (*Tursiops truncatus*), have been subjects of human curiosity since well before they ever became subjects of modern science. In ancient Greece (broadly, 800 BC to 600 AD), naturalists Aristotle and Theophrastus performed and recorded exploratory dissections of dolphins, while historians Herodotus and Plutarch recounted human interactions with them. Plutarch noted that dolphins were particularly welcome sights to Greek mariners; dolphins were said to guard, rescue and lead lost sailors from danger, and in the absence of danger seemed playful and willing to interact (Plutarch, 1956).

Ostensibly because of encounters like these, the ancient Greek curiosity in dolphins was accompanied by admiration for them, and by an attribution of dignity normally reserved for humans – it was sacrilege to kill a dolphin, if not a human slave (Plutarch, 1956). Attesting to their special status, dolphins are depicted on ancient Greek frescoes and coinage, and are prominently featured in mythology. In the so-called "dolphin rider" myths, dolphins are given the honored duty of assisting the human transition between life and death (Beaulieu, 2008).

The reasons for the ancient Greeks' curiosity and admiration for dolphins are not defini-

tively established among historians, however oft-cited reasons include dolphins' perceived friendliness towards humans and, relatedly, their marked human-like social intelligence (Catton, 1995). That ancient Greek mythology places dolphins first among the sea god Poseidon's servants and as his messengers is a provocative hint of the special, communicative bond envisioned between the species (Oppian, 1928). And the perception of this bond generalizes beyond the historic Greeks. The Aborigines of Australia and the Maori of New Zealand, for instance, both characterized dolphins in mythology as capable of providing wise spiritual counsel to humans; the Maori called dolphins *humans of the sea* (Reiss, 2012).

More recent and detailed accounts of dolphins have served to maintain and/or enhance the perception of a kinship between dolphins and humans in the modern day United States rather than to diminish it. The ancient Greek mariners' stories of rescues by dolphins have been corroborated. Around the turn of the 20th century, a Risso's dolphin dubbed Pelorus Jack found fame for escorting ships in the vicinity of the perilous French Pass, a channel between Wellington and Nelson in New Zealand. After being spared by the crew of the schooner *Brindle*, he led them among the rocks and currents of the pass, and soon became such a predictable sight to ships that they would stop and wait for him. Reportedly no shipwrecks occurred in the more than twenty years he was active (Robbins, 1987). Stories such as this would be amplified for general consumption in media productions such as the 1960's television show *Flipper*, featuring a highly intelligent dolphin in communion with humans, and arguably even effect legislation by boosting public support for the Marine Mammal Protection Act of 1972, which marks marine mammals' receipt of more blanket protections than their terrestrial cousins.

The ongoing public fascination with dolphins for their perceived humanity has no doubt motivated and influenced science's treatment of them in the modern day. One of the most prominent dolphin biologists of the 20th century, John Lilly, did not hide an anthropomorphization of dolphins. Funded by NASA, he and colleagues famously attempted to teach the dolphin Peter to speak English in a flooded house (Lilly, 1965; Riley, 2014). Lilly would also

publish a collection of his own books and papers pointedly titled, *Lilly on Dolphins – Humans of the Sea*. While Lilly would later be considered extreme in his views on dolphin intelligence and its equivalence to human's, his outlook would leave a mark on dolphin cognitive science, increasing both public and scientific interest in the field. And this thesis undoubtedly inherits some of John Lilly's optimism for dolphins' capacity for communication, as it is based on the view that natural dolphin vocal communication may be far more complex, if not necessarily humanlike, than is known at present.

1.2 Bottlenose Dolphin Intelligence and the Potential for Language

There are neuroanatomical, behavioral, and cognitive studies support the perspective that dolphins – particularly for the Atlantic bottlenose dolphin, which is the "dolphin" of this thesis unless otherwise indicated – are intelligent and have the potential for complex communication. With regards to general intelligence, various neuroanatomical measures used to predict cognitive abilities of mammals (with variable support and reliability), such as absolute brain mass, brain-to-body mass ratio, and the encephalization quotient (the ratio of brain mass to predicted brain mass, given the mammal's overall size) consistently predict the Atlantic bottlenose dolphin and other odontocetes to be in the vicinity of primates in intelligence, sometimes above nonhuman primates (Marino et al., 2006). Another indicator of intelligence accepted among animal psychologists, the capacity for *mirror self-recognition* – the ability of an individual to recognize oneself in a mirror – places the bottlenose dolphin in an exclusive group with great apes, the Eurasian magpie and some Asiatic elephants (Gallup Jr, 1970; Plotnik et al., 2006; Reiss and Marino, 2001). Other evidence of dolphins' general intelligence includes reports of their culturally-transmitted tool use (Krutzen et al., 2005) and foraging tactics (Sargeant et al., 2005), their capacity for metacognition (Smith, 2010), and their possession of both symbolic declarative knowledge and procedural knowledge

(Herman, 2010).

As to dolphins' social intelligence, evidence is again extensive. Bottlenose dolphins live in a complex social landscape. They live in gender-mixed groups that usually comprise about 15 individuals but which can range in size from 15 to 100 individuals (Shirihai and Jarrett, 2006). For males in particular, groups are not organized based on genetic relatedness, and are not fixed in composition; a single male dolphin may associate with many groups throughout his life, and a single group may undergo composition changes on a daily or even hourly basis. This characterizes what has been termed the dolphin *fission-fusion* society (Connor et al., 2000). Groups can merge to become first-order, second-order, and third-order *alliances* based on complex male relationships – relationships believed to be nontrivial feats of "social intelligence," not based on a simple equivalence rule – and male-female relationships, comprising potentially 400 or more individuals (Connor et al., 2001; Krutzen et al., 2003). Incidentally but provocatively, the so-called *Dunbar number*, which puts a theoretical cap on the number of humans composing a stable social group based on fitting group size to the average brain size for various primate species (which, in turn, is also proposed to correlate with general intelligence), is speculated to lie between 150 and 250 (Dunbar, 1992). Dolphins will engage in cooperative activities such as play and pack hunting, and exhibit non-conceptive sex behavior as observed in primates (Furuichi et al., 2014; McCowan et al., 2000; Sargeant et al., 2005).

Given that general intelligence co-occurs with social intelligence in bottlenose dolphins, it is tempting to speculate that sophisticated communication exists, facilitating higher-level sharing of information among conspecifics. However, it cannot be presumed that such communication would be entirely vocal in nature. Bottlenose dolphins have also been shown to communicate tactiley (Dudzinski et al., 2012) and visually (Herman and Kastelein, 1990), the latter being consistent with studies showing dolphins possess broad-spectrum vision that is equipped with adaptations for low-light and in-air object detection (Griebel and Peichl, 2003; Griebel and Schmid, 2002). And while dolphin visual acuity is thought to be limited,

poorer for instance than the visual acuity of pinnipeds, studies suggest that it is enhanced by dolphins' echolocation – a modality that will be discussed more extensively later (Pack and Herman, 1995).

While we cannot disqualify tactile and visual signals as the potential bases of dolphin communication, dolphins also possess formidable acoustic machinery that would serve ocean-based communication well, particularly where touch and vision fails, which is at a distance: for vision, at distances farther than a few meters, and less as one descends into the light-barren depths.

Regarding the bottlenose dolphin vocal apparatus, dolphins possess at least two physically separable mechanisms for producing "vocal" (more properly termed nasal) sounds, one which predominantly produces tonal sounds (or *whistles*) in a 3 to 20 kHz frequency range and one which produces broadband impulses (*clicks* and *bursts*) with frequencies spanning between a few hundred Hertz to over a hundred kiloHertz (Au and Simmons, 2007; Oswald et al., 2003). Together these mechanisms can be employed, sometimes simultaneously, to produce highly variable sound trains that researchers argue are theoretically capable of supporting a language as versatile as a human one (Ridgway et al., 2015). McCowan asked whether Zipf's Law, a result from mathematical statistics that observes that a word from a language corpus possesses a frequency that is inversely proportional to its rank in the frequency table, applies to dolphin whistles (McCowan et al., 2002). A language corpus' adherence to Zipf's law has been suggested to indicate that it optimizes communicate capacity by not being too repetitive or too diverse (Zipf, 1949). McCowan found that a corpus of dolphin whistles adhered to Zipf's law as well as samples from many human langauges (McCowan et al., 1998a). While this result must be interpreted carefully because we do not know for certain what features of whistles are salient to dolphins, and because simulations suggesting adherence to Zipf's law is a necessary but not sufficient requirement of language, the result remains provocative (Suzuki et al., 2005).

Dolphin vocalizations hint at complex communication not only on the basis of their

attributes as individual, static units: these units also vary and are produced as biphonal signals that may be indicative of complex communication. Certain bottlenose whistles, referred to as *signature whistles*, have been shown to function as unique dolphin self-identifiers – they will be discussed more extensively later. These signature whistles can be mimicked by conspecifics, seemingly for the purpose of attracting the attention of individuals (Tyack, 1986). Bottlenose dolphins can also mimic artificial, whistle-like sounds (Reiss and McCowan, 1993; Richards et al., 1984). Dolphins can be trained to associate these same artificial sounds with objects and have been reported to spontaneously combine the sounds to exhibit behavioral concordance in use (Reiss and McCowan, 1993; Richards et al., 1984). Moreover, dolphins possess large natural repertoires of sounds, for which ontogeny has been reported and studied (McCowan and Reiss, 1995b).

I have briefly reviewed why scientists studying bottlenose dolphin vocalizations, particularly the whistles, believe that current research gives us only a small glimpse of richer vocal communication: dolphins are intelligent both with regards to problem solving and social coordination, and they command acoustic machinery capable of generating large vocal variety.

Before the aims of this thesis are presented in full, the current state of social bottlenose dolphin vocalization research will be presented.

1.3 The Signature Whistle Hypothesis and the State of Dolphin Vocal Communication Research

The first captive Atlantic bottlenose dolphins became available for study in the United States in the 1940s. Quickly whistles were distinguished from among their other vocalizations, including "snapping" and "barking" sounds, and were vaguely associated with social expression (McBride and Herb, 1948). By classifying whistles based on their time-frequency contours as visualized in sonograms, it was noticed that whistles within and among dolphins are numerous and varied, and that they might be interpreted as human-like words with individual meaning

(Dreher, 1961). Melba and David Caldwell were the first researchers to robustly correlate whistle contour with some external feature (Janik and Sayigh, 2013). The Caldwells recorded many wild Atlantic bottlenose dolphin vocalizations in isolation, compiling whistle-contour repertoires for individuals. They noticed that each dolphin seemed to possess a distinct whistle type, a result that was consistent across many populations (Caldwell and Caldwell, 1965; Caldwell et al., 1990; Sayigh et al., 2007).

As a result of these studies, the Caldwells proposed the *signature whistle hypothesis*, which many researchers in the field of Atlantic bottlenose dolphin communication currently accept. Including the contributions of newer research, the signature whistle hypothesis might be stated as follows: every bottlenose dolphin possesses an individually distinctive whistle that it tends to use more than any other whistle, particularly when it is isolated from conspecifics, and which broadcasts the owner's identity to conspecifics (Caldwell et al., 1990; Janik and Sayigh, 2013). The extent to which a dolphin uses its own signature whistle more than other whistles is dramatic: the signature whistle accounts for over 90% of whistles produced by a dolphin held in isolation, and 38-70% of whistles produced by a freely swimming, social dolphin (Caldwell et al., 1990; Janik and Sayigh, 2013; Janik and Slater, 1998; Sayigh et al., 2007).

The Caldwells' pioneering research and the introduction of the signature whistle hypothesis strongly impacted the ensuing decades of bottlenose dolphin communication research. By isolating dolphins and determining their signature whistles, researchers could rigorously study how an individual's whistle changed under different natural and experimental conditions. In some cases, researchers could even study how signature whistles changed in social groups, where the SigID heuristic was employed to identify non-mimicked signature whistles or else crude sound localization techniques that took statistical advantage of signature whistles' prevalence in recordings were employed to identify signature whistle vocalizers (Janik and Sayigh, 2013; Janik et al., 2013). By contrast, other studies recorded and focused on the overall vocal repertoires and also reported the use of predominant stereotypic whistles by

Figure 1.1: Melba and David Caldwell's Signature Whistles A figure from one of Melba and David Caldwell's first studies (Vocalization of Naive Captive Dolphins in Small Groups, 1968) outlining the signature whistle hypothesis. (A)-(E) display the predominant whistle of five different dolphins. (F) is a whistle from the vocalizer of (D), interpreted as a rendition of (D) interrupted by the simultaneous call of a conspecific. In (G) are "chirps overlaid by pulsed emissions." In (H) are "burst-pulsed barks."

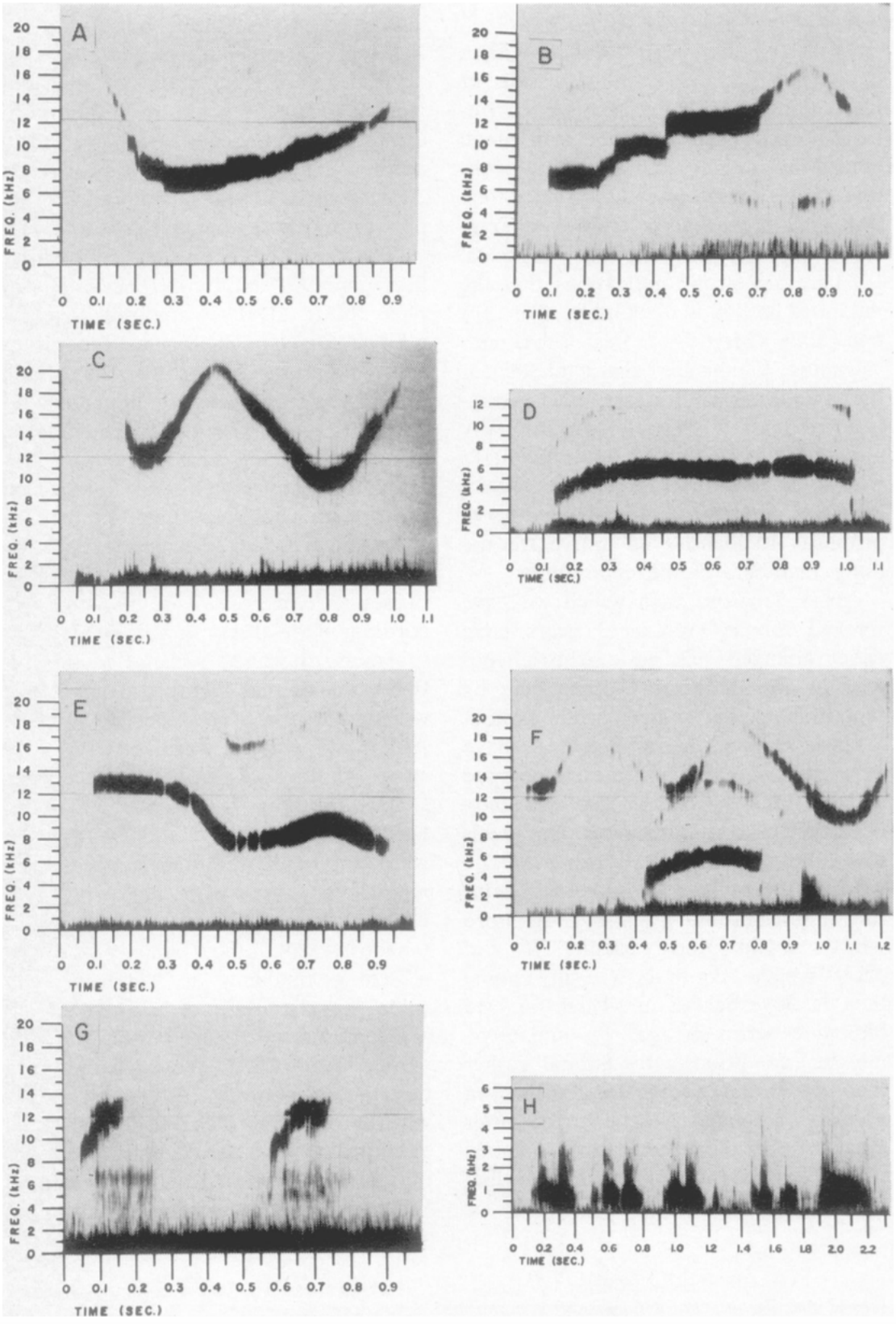


Figure 1.1

individuals when in temporary isolation and during social interactions and noted these calls shared the general characteristic of being rise type calls or looped rises (McCowan, 1995; McCowan and Reiss, 1995a,b, 2001; McCowan et al., 1998b). These studies suggested that, based on contour uniqueness, aquarium dolphins can possess individual whistles repertoires containing on-the-order-of-tens of whistles, and, with only a few being unique to individuals (based on bubble-stream identification, to be discussed), indicated the existence of shared repertoires. Thus, researchers who did not identify signature whistles were largely limited to studies of the whistle repertoires of groups.

Despite many research groups' preoccupation with signature whistles, the head of one such group, Janik, cautions that signature whistles are not the culmination of communication research in dolphins, stating that they are simply accessible tools owing to their stability and dominance (Janik and Sayigh, 2013). He concedes that little is known, for instance, about *non-signature whistles* (which include, unsurprisingly, any whistle that is not a signature whistle). These whistles are scarce, more varied across an individual dolphin's repertoire, and may be shared more extensively than signature whistles (Janik and Sayigh, 2013).

The primary reason non-signature whistles have been neglected is undoubtedly because researchers have lacked a reliable solution to the problem of sound attribution. Due to non-signature whistles' variability, crude methods of sound attribution that suffice for attributing signature whistles among freely swimming dolphins do not suffice for non-signature whistles.

Developing a better system of sound attribution to allow for the study of all whistles is the purpose of this thesis. I believe that with a more reliable system of sound attribution, a rigorous study of both signature and non-signature whistles might be performed. The field stands not only to gain an understanding of the individual vocal repertories, functionality, and development of non-signature whistles, but also a handle on the question of whether complex vocal exchanges occur among dolphins.

1.4 Whistle Acoustic Properties, Brief Review of Click Vocalizations

As was briefly mentioned earlier, *whistles* are tonal sounds with the majority of their power concentrated between 3 to 20 kHz (Herzing, 2000; Oswald et al., 2003). By *tonal*, we mean that whistles are narrowband, their energy spread over perhaps 200 Hz at any given time. A whistle can last for a fraction of a second to as long as four seconds, its center frequency varying continuously, creating the appearance of a worm in time-frequency space (see Figure 1.1). A whistle is said to be *looped* if it comprises identical elements spaced by less than 0.5 seconds (Esch et al., 2009). Moreover, whistles often consist not only of their lowest-frequency, typically highest-amplitude time-frequency component (called the *fundamental*), but of a stack of near-identical *harmonics* of frequencies that are integer multiples of the fundamental. Whistles containing as many as ten harmonics have been reported (Branstetter et al., 2012).

It is well known that the acoustic *clicks* used by dolphins and other odontocetes are highly directional, which aids in echolocation (Au et al., 1986). As clicks and whistles both originate in the nasal passages, some researchers believe that the same anatomical features that contribute to clicks' directionality – including a reflective concave skull, reflective cranial air sacks, and a density graded melon – may contribute meaningfully or incidentally to the directionality of whistles (Branstetter et al., 2012). While whistle directionality is not well studied, for a few species of dolphin (including *T. truncatus*) whistles have been found to be partially front-directional, with directionality increasing with frequency (and in turn harmonic level); the difference between "front" and "side" whistle intensity can be about 5 dB in 160 dB for a whistle's fundamental (Branstetter et al., 2012; Lammers and Au, 2003).

Clicks are broadband pulses: one click typically spans more than 2 kHz and lasts fewer than 300 microseconds (Tyack and Clark, 2000). A bundle of clicks spaced at intervals of ˜3 or more milliseconds with large, 60-70 kHz bandwidths belongs to a functional sub-class, the *click train*, used for echolocation (Au, 2000; Johnson, 1967); when echolocating, a dolphin

generates a click, analyzes its reflection to recognize objects in the vicinity, and repeats, usually for the purpose of foraging (Au, 2000). Alternatively, a bundle of clicks spaced at intervals of ˜0.5 or fewer milliseconds with small, ˜3 kHz bandwidths probably has social functionality (Herzing, 2000). Lumped together, bursts and click trains have been associated with emotions such as fear (Caldwell and Caldwell, 1968; Herzing, 2000), consortship (Connor and Smolker, 1996), and alarm (Caldwell and Caldwell, 1968; Herzing, 2000).

As an aside, note that whistles and clicks are potentially produced by two or more, likely lateral-cranially segregated vocal apparatuses, as evidence by common bottlenose and other species of dolphins ability to produce biphonations (Kaplan et al., 2017; Papale et al., 2015).

1.5 Sound Attribution

Sound attribution refers to matching a sound heard or received on some audio device with its source. Assuming the potential sources to be identical in all characteristics except location, sound attribution typically involves *sound source localization*. Sound source localization ascribes an area in space to a sound or *signal* based on predictable changes that alter it based on physical phenomena during its travel from source to sensor location(s) (Neunuebel et al., 2015). Sound source localization usually involves multiple spatially separated sensors, each receiving the signal of interest with changes that are unique to the particular source/sensor pair's physical relationship; such changes might include amplitude change, frequency composition change, and particularly phase change and time delay. Assuming the positions of the sensors are known, a set of source/sensor changes can be used to specify sound source location. There are many methods for accomplishing this that cannot all be reviewed in this thesis, but the mathematical details for a relevant subset will be reviewed in upcoming chapters. Following sound source localization, sound attribution itself will often take visual data into account to match the spatial output area from localization with an individual source; in other fields, some clever ways exist for doing this that leverage the known visual positions of potential

speakers to compensate for poor sound source localization, a point which the work in this thesis did not reach (Neunuebel et al., 2015; Warren et al., 2018). Sound source localization and attribution for cetacean calls from unrestrained animals has been attempted before wth varying degrees of success. Prior to five to ten years ago, I feel these methods fall into roughly five categories, based on the capabilities of the underlying sensor arrangements.

In the first category are methods that often involve large ships owned by organizations like NOAA or the US Navy, which pull arrays of three to dozens of hydrophones. These methods localize sound sources in the open ocean based on the relative time delays of received signals across sensors. Due to the large distances between the array and potential sources as compared to the distances among individuals within a social group, as well as a lack of visual data, these methods are largely limited to semi-blindly attributing sound to widely distant cetacean groups and/or solitary individuals (Barlow and Taylor, 2005; Watkins and Schevill, 1972; Yack et al., 2013).

In the second category are methods involving large fixed hydrophone arrays, for instance the US Navy's SOSUS (sound surveillance system), which was deployed in 1949 in both Atlantic and Pacific Oceans to monitor Soviet submarines (Ano; Whitman, 2005). As above, these methods localize sound sources in the open ocean base on the relative time delays of received signals across sensors. These methods benefit from widely-distributed sensors (at scales of hundreds and thousands of miles), and enjoy better coverage and localization performance at distance than the towed array methods, however suffer from the same large distances between sensors and sources as compared with within-social-group distances, as well as a lack of visualization, so are also limited to blindly attributing sound to widely distant cetacean groups and/or solitary individuals (Cummings and Thompson, 1994; Janik, 2000; Stafford et al., 1994; Watkins et al., 2000).

In the third category are methods employing portable hydrophone arrays in aquaria or in the field, often paired with cameras or high-quality observation. As above, these methods typically (but not always) localize sound sources based on the time delays of received signals

across sensors, but with the added benefits of the animals' close proximities to the sensors (the arrays are opportunistically deployed) and good visual data, these methods can attribute sound to non-solitary individuals. Such methods have been employed heavily in studying dolphin echolocation, and to a lesser extent dolphin whistles (Dudzinski et al., 1995; Tyack, 1991; Watkins and Schevill, 1972, 1974; Wisniewska et al., 2015). However, these methods have still not been employed with great success for studying non-signature whistles or vocal exchanges. Limitations include their short time of deployment in aquaria or the field and their tendency to lose accuracy significantly at just a few meters of source-sensor separation, leading to difficulties acquiring large, longitudinal data sets for dolphins distributed across a large space (Dudzinski et al., 1995; Tyack, 1991; Watkins and Schevill, 1972).

In the fourth category are methods, similar to the above, employing distributed sensor arrays (typically separations in excess of 1m) in aquaria or the field. Unlike the methods based on dense sensor arrays, these methods can be theoretically suited to performing comprehensive sound attribution for a large (10+ meter) enclosure. However, such methods have not yet attributed whistles among more than two dolphins/dolphin groups for a permanent population in a reverberant environment (i.e., a dolphinarium), for a variety of reasons including lack of sound localization precision and lack of visualization (Freitag and Tyack, 1993; Janik and Thompson, 2000; López-Rivas and Bazúa-Durán, 2010). Only one such method, using temporary hardware installed around a relatively non-reverberant lagoon serving as an interim dolphin holding pen, has successfully performed whistle attribution for a group of more than two dolphins, achieving 40% attribution over a few hundred whistles (Thomas et al., 2002). The sound localization system that is the subject of this thesis employs a method belonging to this fourth category, and it will be considered in more detail in upcoming chapters.

The fifth category consists of a single method of sound attribution that does not involve sound source localization; it relies on the observation that bottlenose dolphins release bubble streams concurrent with whistles (Herzing, 1996; McCowan and Reiss, 1995a). While sometimes this method can achieve sound attribution for individuals in a large enclosure with

a single hydrophone and suitable visualization, major weaknesses include the observation that bubble-streams are not always produced when dolphins whistle and therefore are not sufficient to indicate whistling, and moreover that the association seems biased towards particular whistle types (Fripp, 2005).

More recently, new methods of sound localization/attribution have been developed in the form of hydrophone-containing tags that are attached to the dolphins for which sound attribution is desired (Akamatsu et al., 2000; Tyack, 1985; Watwood et al., 2005). Once deployed, these tags solve the sound attribution problem for the equipped dolphins, based on either simple sound source localization from onboard recording devices or the timing of visual indicators emitted from onboard lights. While powerful, sound attribution methods based on these devices do have weaknesses. They include the frequent disruption of monitoring given the tendency of tags to fall off after a few days, that aquarium dolphins must be trained to wear these tags and that in general their effect on the wearers' behavior remains unclear, and that in the wild they are not usually accompanied by visualization – not an impediment to the attribution itself but to tracking group membership in the wild and finding vocalization behavioral correlates (van der Hoop et al., 2014).

The proposed method falls into the fourth category, and relies on deploying a distributed sensor array with accompanying cameras in the Dolphin Discovery pool of the National Aquarium in Baltimore, Maryland, a contained habitat with an invariable group of seven dolphins. As the equipment is permanent, we are uniquely situated to obtain data over months if not years; even with an imperfect solution for sound attribution, we are poised to acquire enough data, consisting of enough dolphin interactions, to mount a substantial study of dolphin (non)exchange and individual vocal repertoires.

1.6 Proposal

I endeavored to design a system that would produce data sufficient for the attribution of whistles to individual dolphins at Dolphin Discovery of the National Aquarium in Baltimore, Maryland. Successful sound attribution as we conceived it required data that contained (1) the locations, in real coordinates, of all dolphins or "potential vocalizers" (i.e., from camera recordings), and (2) the locations, in real coordinates, of the origin of all vocalizations (i.e., from hydrophone recordings). Requiring both camera and hydrophone recordings, whistle attribution would be achieved by matching the two data sets. At the time of writing, a fully autonomous attribution system has not been completed, however the fundamental hardware components are in place: several networked cameras, several networked hydrophones, and infrastructure for consolidating the data on a single computer. In addition, software for continuously recording synchronous audio and video has been designed, and various measurements and calibrations necessary for achieving whistle attribution have been made. Moreover, a strong proof of concept for a machine-learning-based approach for whistle attribution has been developed. All of these topics will be discussed in subsequent chapters.

Chapter 2

Dolphin Surveillance System: Hydrophone Panels

2.1 Pool Overview

The dolphin pool of the Dolphin Discovery exhibit at the National Aquarium, Baltimore is approximately a cylindrical tank; it is approximately 20 feet deep as measured by plumb line, and 108-112 feet in diameter as measured by laser rangefinder (Bosch 225 ft. Laser Measure). The cylindrical pool is partitioned into four sub-pools Figure 2.1. The largest of the sub-pools, the so-called *exhibition pool* or *EP*, is an approximate half-cylinder. The walls and floors are constructed of thick structural concrete coated with an industrial epoxy except in several large sections, particularly on the curved wall (which faces the public, which takes the perspective of Figure 2.1). These sections are composed of 5.25"-thick acrylic coated with a scratch/environment-proofing film to facilitate viewing. Two of these large sections occupy the upper third of the pool. Sharing the EP's flat concrete wall are two support pools that are nearly quarter-cylinders, denoted *SP 1* and *SP 2*. The walls of these pools are constructed entirely of epoxy-treated concrete except for one small acrylic viewing window each, looking into a small central observation room called the *pit*, into which the EP also has

a viewing window. Lastly, a cylindrical section with radius somewhat less than the total pool radius is cut symmetrically from SP1 and SP2; this is the *med pool*. The med pool has a depth substantially less than the 20 feet indicated earlier. Around the perimeter of the EP and SP's are a number of concrete ledges or *slide-outs* that serve as dolphin slide areas and trainer platforms.

Between each pair of sub-pools is a portal approximately 6' high and 4.25' wide that can be blocked by a removable *gate*, a 1"-thick board punctuated by approximately 1"-diameter pores. Because the gates are porous, the four sub-pools are always linked by a common water supply, and are therefore acoustically linked, except under the rare circumstances that the med pool portals are blocked by solid steel gates for the purpose of dropping the med pool water line below the level of the other pools.

The purpose of the portal-gate system is to separate the aquarium's seven dolphins (two male, five female) for husbandry, an aspect of pool operation that the research group has little control over. Potential reasons for separating dolphins include the prevention of unwanted sexual arousal or intercourse, the prevention of aggressive displays, the prevention of the spread of disease or stress, or in general the prevention of any interaction that the trainers feel is not optimal to the animals' collective wellbeing. The distribution of the dolphins among the pools can change hourly or daily depending on the short-term health/behavioral assessments as well as the long-term goals of Dolphin Discovery's medical and training staffs. Long-term goals might include the slow mixing of the male and female dolphins after a long period of separation, or the introduction of foreign hydrophone panels into the water.

Any attempt at tracking dolphin behavior must cope with the disruptions of natural behavior resulting from these separations as well as various other acts of husbandry. These include scheduled changes to the artificial lighting; scheduled interactions with guests (which do not strictly qualify as husbandry); scheduled feedings and medical procedures; scheduled *training sessions*, in which the trainers introduce and reinforce trained behaviors – some of these trained behaviors inspired by the behavior of wild dolphins, some not – in front of an

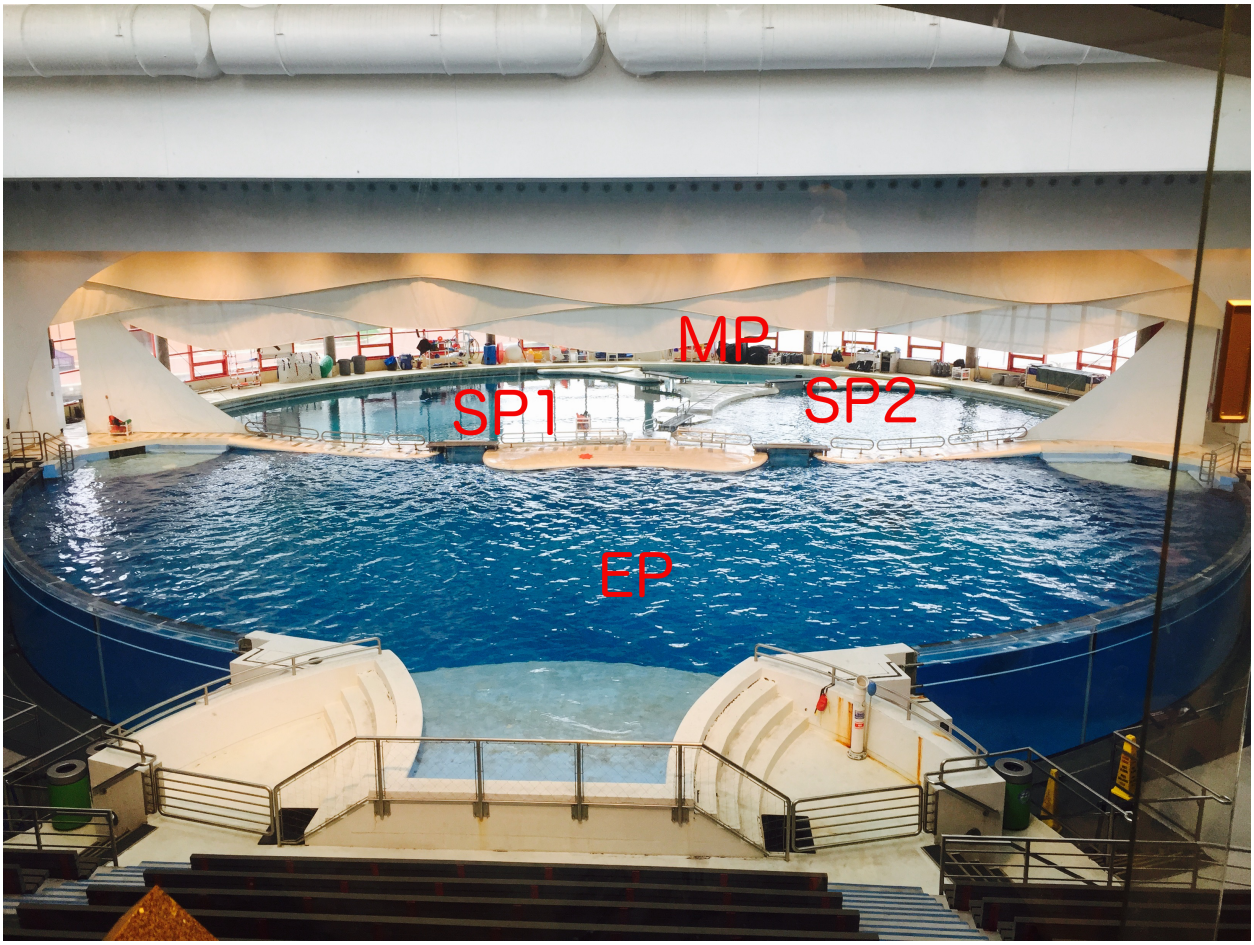


Figure 2.1: Overhead View of Dolphin Discovery Pool, National Aquarium, Baltimore The largest sub-pool, closest to the camera, is the EP. The two next-largest sub-pools, from left to right, are SP1 and SP2. The smallest, farthest sub-pool, barely visible, is the med pool. Note the portals and gates between each pair of sub-pools.

audience; scheduled *enrichment sessions*, in which the trainers expose the dolphins to novel sensory stimuli, typically in the form of colorful visual and auditory toys displayed on the acrylic window; and scheduled access to pool toys – non-realistic items such as pool noodles, rings. In general, any attempt at tracking the behavior of the Dolphin Discovery animals must accept that the dolphins' behavioral patterns and environment are tightly controlled by the trainers and, as a consequence, tend to be neither naturalistic nor uninterrupted.

The proposed surveillance system was originally intended to provide continuous audio and visual tracking of all Dolphin Discovery's seven dolphins. Such a system would require multiple hydrophones (four or more) placed in each of the four sub-pools, all hard-networked; additionally it would require multiple cameras achieving three-dimensional coverage of all sub-pools, all hard-networked as well. Owing to time and cost limitations, and due to a lack of overhead access for placing cameras around the rear pools and a general lack of accessibility for placing hydrophones (discussed in the next section), we limited our surveillance to the EP, employing a smaller system. Together with the husbandry considerations just discussed, this decision reduced our theoretical ability to track uninterrupted dolphin interactions, in that continuous tracking would necessarily end for any dolphin moving from the EP to one of the SP's.

Another drawback of the EP-limited hardware setup is the potential for whistle misattribution resulting from hearing but not seeing dolphins residing in the SP's. A successful software tracking system for the EP dolphins must not misattribute whistles from the SP's to dolphins visualized in the EP. To wit, this might be accomplished by excluding whistles using an amplitude thresholding based on comparing separate sets of whistles taken from inside and outside the EP, or by using a more sophisticated subroutine – potentially also based on a training set – drawing from the sound attribution/localization software itself. In any case, at present the system would not permit such distinctions.

In the next section, I will review the design considerations and basic design of the hydrophone component of the system.

2.2 Hydrophone Placement and General Design Considerations

While naively my task was to place hydrophone arrays around the EP in such a way as to optimize the performance of a sound localization algorithm, constrained only by pool geometry, in practice hydrophone array placement and design was constrained and dictated primarily by aquarium construction, aesthetic, and husbandry rules and considerations, and thriftiness.

As outlined in the previous section, the EP, the sub-pool in which we sought to achieve sound localization/attribution, is approximately a half-cylindrical tank measuring about 20' deep from the water line and about 54-56' in radius.

Hydrophone placement solutions that required exclusion from consideration included all those requiring drainage of the pool, drilling into or around the pool, and installing visible, irremovable fixtures in or around the pool. Draining the pool would require an excessive amount of aquarium labor (namely, spraying pool water into the harbor and refilling), drilling into the dolphin tank would to be too disruptive to its acoustically sensitive residents (and underwater drilling would likely require pool draining to prevent dust from contaminating the water), and most irremovable attachments that might be installed without draining or drilling could still be deemed a threat to animal husbandry or pool aesthetics. Note that these operations would not only be unacceptable for installing whatever assembly housed the hydrophones themselves, but for installing protections for the hydrophones' cables, which would necessarily run a distance between the hydrophones and the water surface.

The above considerations implied that our placement solution would involve a rigid assembly, or array, of hydrophones, secured to a preexisting fixture outside the pool. This meant that we could not place hydrophones on the flat wall of the EP. The flat wall of the EP is hidden by a concrete overhang that serves as a path for the training staff for walking, manipulating the gates in the EP/SP portals, and making presentations, so attaching a rigid

assembly to a preexisting fixture outside the pool in this area would likely pose a tripping risk to the trainers. Moreover, installation of hydrophones on this wall would likely create related safety problems in routing hydrophone cables to the perimeter of the pool, where audio interfaces could safely be installed.

The placement zone for the hydrophones was therefore limited to the curved section of the EP wall, including the central concrete slide-out area and and the two long stretches of acrylic wall. The oblong concrete sections abutting the central concrete slide-out (which itself was not a placement option for the same reason as the flat wall slide-outs) were determined to be unsuitable due to noise from pool outflow gratings, and so we would place our rigid hydrophone arrays along the two stretches of acrylic wall.

Proceeding further in the design would require deciding a method of sound source localization, which would dictate the minimum number of hydrophone arrays and hydrophones per panel. The methods available to us properly belong to the field of *passive acoustic monitoring* (Zimmer, 2011). Using standard hydrophones for detecting signals of unknown source frequency composition and intensity, the most valuable information available to us for the purpose of sound source localization was believed to be the differential transit times of the whistles to the hydrophones; thus, the methods available to us more specifically belong to the field of *time delay estimation* (or *TDE*) (Zimmer, 2011). TDE methods of sound source localization are not universally classified, however I roughly break them down into the classes of *beamforming*, *time-of-arrival*, and *time-difference-of-arrival* methods.

Beamforming is a well established and deep field of study and a full review is beyond the scope of this thesis, however one of the simpler versions, the *delay-and-sum* method, will be presented here for illustrative purposes. In short, beamforming involves an array of sensors, typically ten to hundreds, placed at close distances from one another – less than half of the source wavelength – in a regular pattern (lines, rectangular grids, and concentric circles are common choices) chosen based on problem properties (Van Veen and Buckley, 1998). In the delay-and-sum method, the signals from the various sensors are all simply summed

together, and a final output signal is produced (Greensted, 2012; Van Veen and Buckley, 1998). Because the impinging source sound wave (which we normally assume to be planar, i.e., in a far-field approximation) reaches the different sensors at different times depending on the array geometry and on the source position, summation maxima and minima resulting from interference will occur for different array geometries and source positions. For a fixed array geometry, the maxima and minima depend entirely on the source position, and an array geometry can be engineered to favor signal maximization for an arbitrary source position.

This becomes useful for sound source localization with the addition of *beamsteering*. With beamsteering, the array geometry is physically or digitally altered to change the favored source position in a predictable way. In delay-and-sum, the input signals are digitally delayed (Greensted, 2012). Figure 2.2 depicts delay-and-sum beamforming and beamsteering.

Sound localization can theoretically be achieved with beamsteering by recording a sound of interest from all sensors and performing a post-hoc beamsteering across a suspect space searching for maxima. However, notable drawbacks of this approach include the requirement of a large number of sensors confined to one location, the requirement of post-hoc data analysis (for the tracking of sparse source signals), the requirement of regular precise hydrophone calibrations, and known difficulties locating sources in highly reverberant environments due to the presence of erroneous maxima due to multipath interference (Di Claudio and Parisi, 2003). The last drawback is most significant, as will be detailed in subsequent chapters. While a sophisticated method of sound localization involving two beamforming arrays to reduce erroneous maxima might have been implemented, we thought this method to be cost-prohibitive. Moreover, beamforming is a particularly dense and sometimes esoteric field of study and, lacking any available expertise or opportunity for rigorous trial-and-error, the risk of investing in a suboptimal array that would not be as effective as a suboptimal setup designed to implement a time-of-arrival or time-difference-of-arrival method seemed high. Ultimately, we decided against a beamforming approach to sound localization.

The next class of algorithms, time-of-arrival (*TOA*), work from an explicit knowledge of the

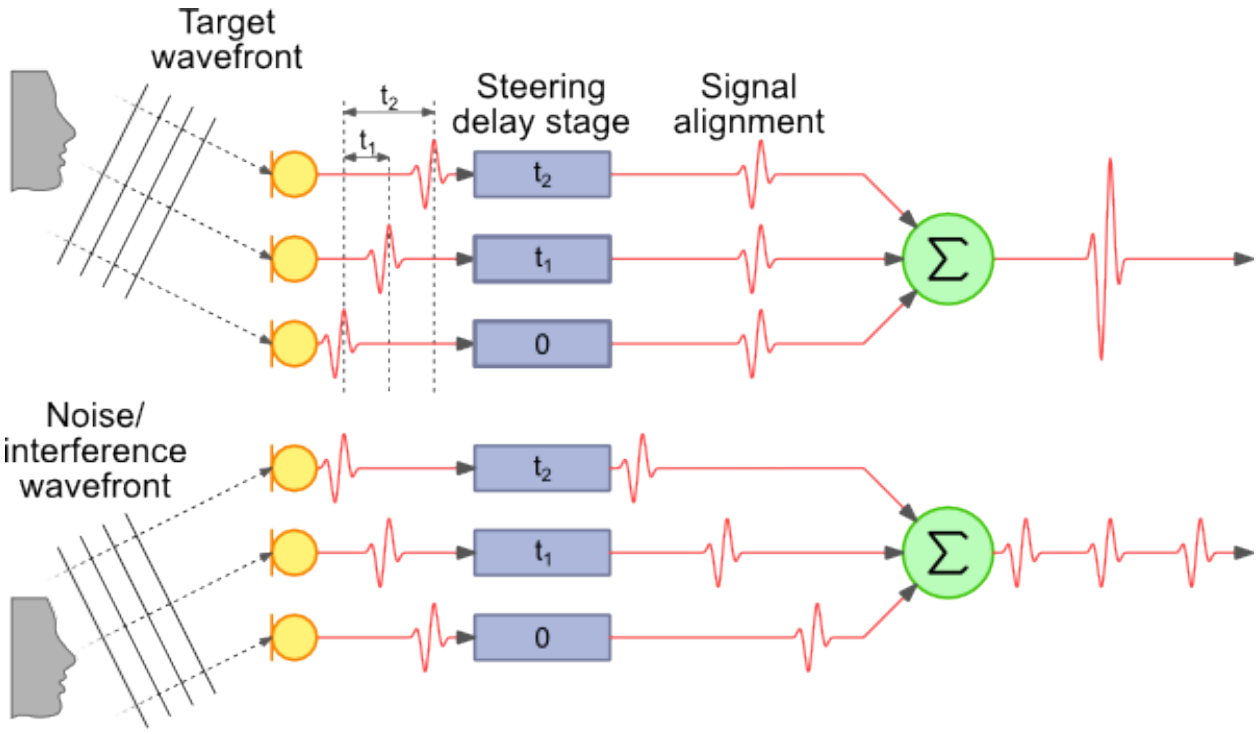


Figure 2.2: Depiction of Beamforming and Beamsteering A three-sensor beamforming array with a digital steering stage impinged upon by plane waves from two locations is depicted. The plane wave from neither location generates a maximum under a simple sum, but with digital steering, the plane wave from the first location generates a maximum under summation. From (Greensted, 2012).

time it takes the signal to reach each sensor in the array (Li et al., 2016). For a given speed of sound, through simple multiplication the times of arrival can be converted into direction-blind estimates of distance, one for each sensor, and so the source localization problem is reduced to one of finding the ideal intersection of spheres. Unfortunately, knowledge of arrival times implies knowledge of the time at which the signal is generated, which in our case we do not possess. Therefore, time-of-arrival algorithms can be trivially disregarded from consideration for our problem.

Thus, mainly by process of elimination, we sought to design a sound localization system suited to time-difference-of-arrival (or *TDOA*) algorithms. I will discuss the topic of the

TDOA approach now insofar as it informs our hydrophone array geometry – refer to Section 6.2 for mathematical details of the chosen approach. In general, TDOA algorithms are based on the time it takes a signal to reach one sensor relative to another sensor, across every sensor pair. While in the conventional TOA method every hydrophone localizes the source to the surface of a sphere, in the conventional TDOA method every hydrophone pair localizes the source to a hyperboloid. In one version of the TDOA method, the time delays are used to determine *direction of arrival* (or *DOA*), forming lines or cones from which the source point is localized. This is a version we ultimately did not consider owing to an inability to reliably determine time delays from closely adjacent pairs of hydrophones in our final hydrophone array.

I will briefly review the mathematics of the standard TDOA method and describe how it motivates the decision of an array geometry, following the abstract treatment for 2 dimensions of Isaacs et al. (2009). The problem will be revisited more practically in Section 6.2.

Given a sound source point $\boldsymbol{s} \in \mathbb{R}^2$ and M hydrophones $\boldsymbol{h} := [\boldsymbol{h_1}, \boldsymbol{h_2}, ..., \boldsymbol{h_M}] \in \mathbb{R}^{2\times M}$, we consider a sound emanating from $\boldsymbol{s}$ at some time t_o with speed v. We seek to solve for $\boldsymbol{s}$ starting from the time of arrivals to the hydrophones, $\hat{t}_i$. The noisy time measurement of hydrophone $\boldsymbol{h_i}$, assumed to suffer from a zero-mean, σ^2-variance Gaussian error (an idealization) is:

$$\hat{t}_i = t_\text{o} + \frac{d_i(\boldsymbol{s}, \boldsymbol{h})}{v} + \epsilon_i, \tag{2.1}$$

where,

$$d_i(\boldsymbol{s}, \boldsymbol{h}) := ||\boldsymbol{s} - \boldsymbol{h_i}||, \quad i = 1, ..., M \tag{2.2}$$

is the source-hydrophone distance. We know from our earlier discussion of the TOA method of sound localization that the onset time, t_o, is unknown, and so we eliminate it by subtracting time-of-arrivals, in the defining operation of the TDOA set of methods:

$$\hat{t}_{ij} := \hat{t}_i - \hat{t}_j = \frac{d_i(\boldsymbol{s}, \boldsymbol{h})}{v} + \epsilon_{ij}, \tag{2.3}$$

where

$$d_{ij}(\boldsymbol{s}, \boldsymbol{h}) := d_i(\boldsymbol{s}, \boldsymbol{h}) - d_j(\boldsymbol{s}, \boldsymbol{h}) \tag{2.4}$$

is a difference of source-hydrophone distances, and ϵ_{ij} is a subtraction of two Gaussian, zero-mean, σ^2-variance random variables possessing zero mean and $2\sigma^2$ variance. The treatment of Isaacs et al. (2009) states that, without loss of relevant information, we can limit our attention to the time-of-arrivals taken with respect to the first, so-called *reference* hydrophone (we note that, in practice, placing such importance on the reliability of a single hydrophone over the others is inappropriate). Thus, we can summarize the variables of interest in $(M-1) \times 1$ vectors:

$$\hat{\boldsymbol{t}}(\boldsymbol{s}, \boldsymbol{h}) := [\hat{t}_{i1}]_{(i=2,\ldots,M)} \tag{2.5}$$

$$\boldsymbol{d}(\boldsymbol{s}, \boldsymbol{h}) := [d_{i1}]_{(i=2,\ldots,M)} \tag{2.6}$$

$$\boldsymbol{\epsilon} := [\epsilon_{i1}]_{(i=2,\ldots,M)} \tag{2.7}$$

The TDOA vector can be shown to possess mean and covariance:

$$\bar{\boldsymbol{t}}(\boldsymbol{s}, \boldsymbol{h}) := E[\hat{\boldsymbol{t}}] = \frac{\boldsymbol{d}(\boldsymbol{s}, \boldsymbol{h})}{v} \tag{2.8}$$

$$\boldsymbol{Q} := E[(\hat{\boldsymbol{t}} - \bar{\boldsymbol{t}})(\hat{\boldsymbol{t}} - \bar{\boldsymbol{t}})^T] = \sigma^2[\boldsymbol{I} + \boldsymbol{1}\boldsymbol{1}^T] \tag{2.9}$$

where $\boldsymbol{1} := [1 \quad 1 \quad \ldots \quad 1]^T$. Our estimate of the sound source point, $\boldsymbol{s}$, is in statistics called an *estimator* of an *estimate*, which in this case is the TDOA vector $\hat{\boldsymbol{t}}$ and its constituents. As such, it is subject to the *Crameŕ-Rao bound*, a lower bound on the variance of an estimator. An estimator that minimizes the lower bound of its variance is statistically optimal, and so we seek to minimize the Crameŕ-Rao bound. The Crameŕ-Rao bound states that, for an unbiased estimator (one whose expected value of its difference from its expected value is zero) $\hat{\theta}$ of unknown deterministic parameter and estimate θ, distributed according to a probability density function $f(x; \theta)$ based on observed measurements x, the following holds:

$$var(\hat{\theta}) \geq \frac{1}{I(\theta)} \tag{2.10}$$

where $I(\theta)$ is the *Fisher information*. Very briefly stated, the Fisher information tells us the amount of information about parameter θ that the measurements x contain. Minimizing the Crameŕ-Rao bound or maximizing the Fisher information is the basis of many researchers' attempts to optimize the TDOA-based source location estimator for various parameters, in particular sensor geometry. In two dimensions, the Fisher information for the TDOA-based source location estimator has been derived by Chan and Ho (1994), assuming our noise vector ϵ is independent of $\boldsymbol{s}$:

$$\boldsymbol{I}(\boldsymbol{s}, \boldsymbol{h}) = \frac{\boldsymbol{G}^T \boldsymbol{Q}^{-1} \boldsymbol{G}}{v^2} \tag{2.11}$$

where $\boldsymbol{Q}$ is given by (2.9) and

$$\boldsymbol{G} := \begin{bmatrix} \boldsymbol{g}_2^T - \boldsymbol{g}_1^T \\ \vdots \\ \boldsymbol{g}_M^T - \boldsymbol{g}_1^T \end{bmatrix}_{(M-1 \times 2)} \tag{2.12}$$

with

$$\boldsymbol{g}_i := \frac{\boldsymbol{s} - \boldsymbol{h}_i}{||\boldsymbol{s} - \boldsymbol{h}_i||} \tag{2.13}$$

Using this result, Isaacs et al. (2009) look for sensor geometries that minimize the Crameŕ-Rao bound for a source point located somewhere inside a circle with radius r_1, chosen from a uniform distribution. The M sensors are limited to lie inside an annulus with $r_1 \leq r \leq r_2$. They find that an optimal configuration is for the M sensors is to remain maximally distant from the source circle, on the circle with radius r_2, with angles from the x-axis given by

$$\phi_j - \phi_i = \frac{2\pi}{M} \quad j = i + 1 \tag{2.14}$$

for each (i, j) pair. In words, the sensors are distributed in a *splay* configuration, with equal spacing along the perimeter of the circle r_2 between each sensor. This is the fundamental result that inspires the placement of our hydrophones. While it is not directly applicable, pertaining to two dimensions with source locations limited to a circle rather than half-circle, work done for 3-dimensional geometries for a known source point (a weaker problem setup)

also suggest that spacing sensors in a symmetric fashion around a source point such that the source-sensor direction vectors are well distributed often leads to an optimal solution (Meng et al., 2012; Yang, 2007). While it remains an open project to rigorously derive the optimal sensor geometry for our geometry, we felt that tentatively proceeding with a splay configuration was reasonable, especially given that hydrophone arrays were to be moveable.

We therefore endeavored to space as many hydrophones as possible equally along the EP's round acrylic wall. We hoped to place the sensors at at least two distinct depths, following same line of reasoning of maximizing the spread of source-sensor direction vectors, but at the time of writing this has not yet been achieved.

After a long process of habituating dolphins to the presence of each new hydrophone array, discussed in the next section, as well as concerns that the hydrophones would be unsightly and/or distracting to the public, the aquarium allowed us to install four hydrophone arrays – more arrays might have been possible were we able to present a strong argument that four was insufficient. Primarily for redundancy, but also to potentially accommodate a future partial beamforming approach to sound source localization to bolster the TDOA approach, we placed four hydrophones in each array. See Figure 2.3. The specifics of the array design will be discussed in a forthcoming section.

2.3 Hydrophone Panel Installation, Dolphin Habituation

We eventually designed and built two distinct generations of hydrophone arrays, denoted Mark 1 and Mark 2. One of these sets is semi-permanently installed in the EP, representing the culmination of several dozen planned and unplanned installations and removals spread over two years, from July 2015 to June 2017. The many installations and removals of the hydrophone arrays was required by the aquarium's medical staff, reasoning that they would reduce the dolphins' peak stress levels during introduction. The hydrophone arrays could be

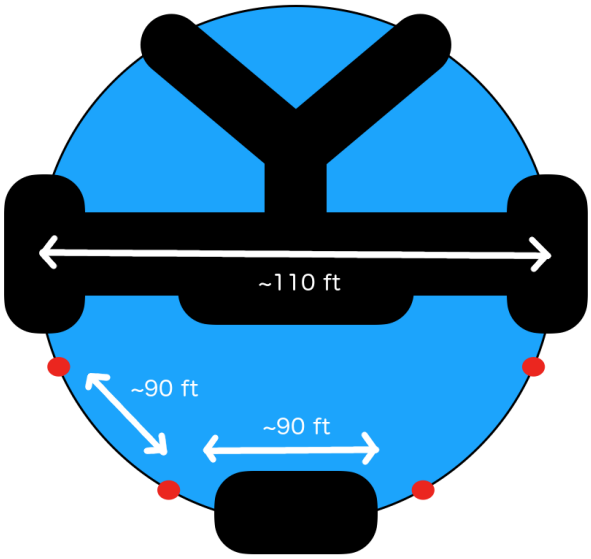


Figure 2.3: Proposed Array Placement Cartoon depiction of pool as seen from above. Red dots indicate approximate array placement.

potential stressors as they were to be new additions to what is otherwise a mostly barren pool and had to be resistant to dolphin tampering.

Even though the National Aquarium dolphins did not have their cortisol levels checked regularly during the array introduction, the staff did look for evidence of stress through behavioral indicators. General stress indicators included returning fish, aggression, and unresponsiveness during training sessions, and array-directed indicators included tail-fluking, ramming, and avoidance. The dolphins' reactions to the early array installations convinced the staff that stress-reducing protocols that allowed the dolphins to habituate to the individual hydrophone arrays must be created, and these changed over time. The reasoning of the staff was that the onset of chronic stress could be avoided by exposing the dolphins to the hydrophone arrays individually for slowly increasing durations. Thus, the strategy involved installing one array in the pool for a few hours for several days, positively reinforcing the

dolphins with fish in its vicinity, "permanently" installing the array, then introducing the next array similarly. Below is good representation of the protocol that we attempted to follow:

With a minimum of four days between steps for evaluation:

1. Start with 2 hydrophones in 3-4 hours per day
2. 2 hydrophones in 6 hours per day
3. 2 hydrophones in overnight and watch 4 nights
4. 3rd hydrophone in 3-4 hours per day
5. 3rd hydrophone in 6 hours per day
6. 3 hydrophones in overnight and watch 4 nights
7. 4th hydrophone in 3-4 hours per day
8. 4th hydrophone in 6 hours per day
9. 4 hydrophones in overnight and watch 4 nights
10. All 4 hydrophones in permanently

A process somewhat like this was repeated and disrupted multiple times before the Mk. II hydrophone panels were permanently installed, in June 2017. During the final habituation period which occurred over three months our research team and the NA training staff conducted systematic behavioral observations. Daily observations were conducted of the dolphins' behavior toward or in the proximity of the hydrophones by trainers from 7am -6 pm daily and our research team monitored behavior from 6pm to 7am. The final protocol stated that the trainers and/or researchers responsible for evaluating the dolphins' reactions to the

hydrophone panels were to score for the following behaviors: (1) head-butts to the panels, (2) tail-flukes directed towards the panels, and (3) general avoidance (left to the observer's discretion to define). Daytime observations were conducted by National Aquarium trainers and nighttime observations by myself and members of the Magnasco-Reiss research team; initially all observations were conducted in person (from the audience area, from the sound booth at the top of the audience area, and/or the pool "pit"), later nighttime observations were conducted from NYC using remotely accessed real-time surveillance video feed. During the final installations of the first two panels, only a few solitary, sporadic instances of any of the three behavioral indicators of stress were noted, in the first three of eight days of observation. The main behavior observed was genital rubbing on the arrays.

2.4 The Critical Element: the Hydrophone

The critical element of our hydrophone array is the single hydrophone. A hydrophone is essentially a microphone manufactured with an appropriate acoustic impedance – a measure of the amount of sound pressure generated by a given *acoustic flow*, which must be matched to a medium's properties to maximize power transfer – to receive sound underwater. It employs a *piezo-electric transducer* to convert the mechanical energy of the incoming sound into an analog current. Depending primarily on the properties of the piezo-electric transducer, the sound energy will be more or less effectively converted into electrical energy depending on the sound's frequency and its angle of approach to the transducer. Respecting financial and engineering limitations, it was important that we choose a hydrophone well suited to the frequencies of interest, minimally the 3-20 kHz range of the bottlenose dolphin whistle, and source locations of interest, lying between -90 and +90 degrees based on an x-axis projecting from the"face" of a forward-facing hydrophone.

For cost and acoustic properties, we chose the SQ-26-08 hydrophone from Cetacean Research Technology, whose frequency response plots are shown in Figure 2.4 and Figure 2.5.

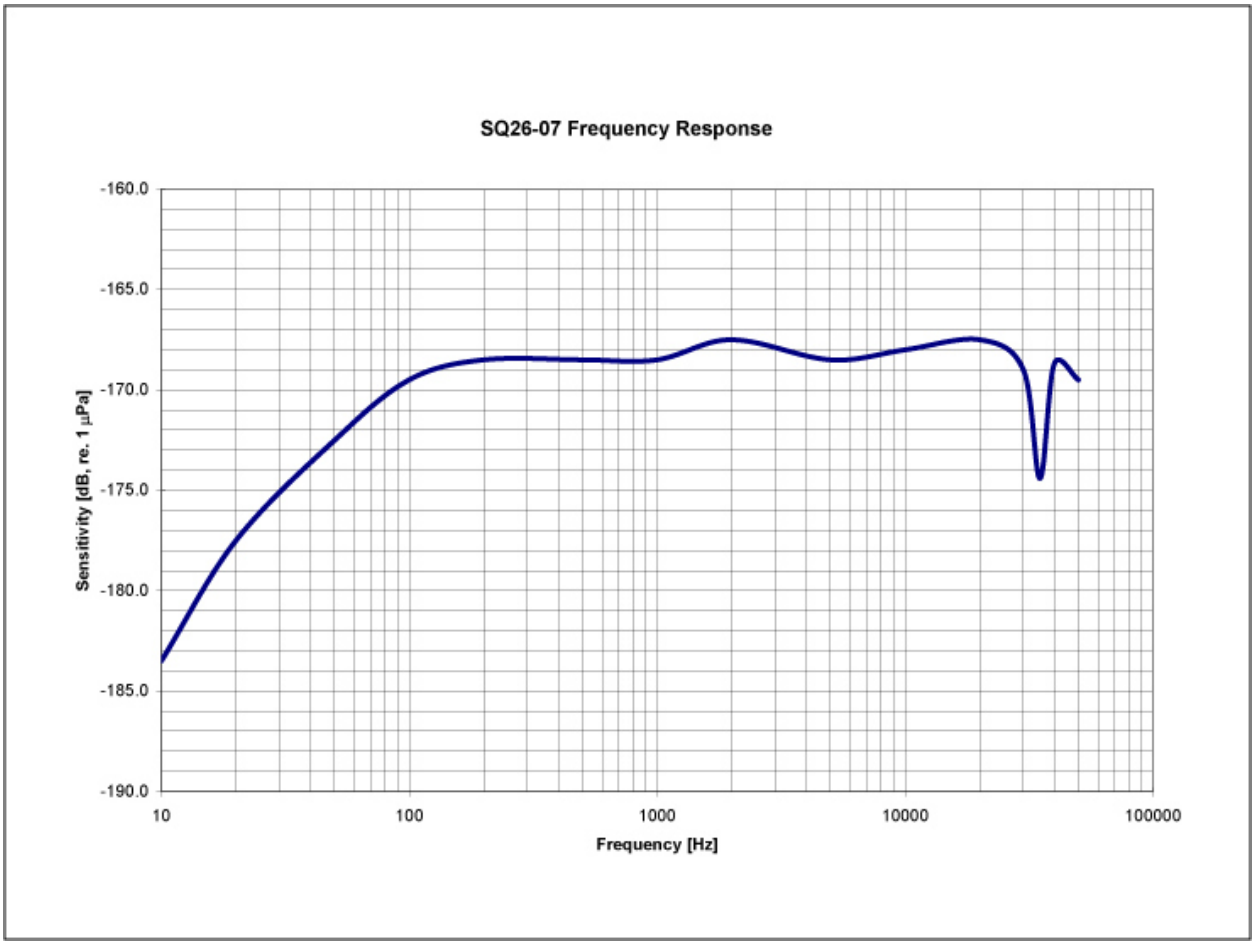


Figure 2.4: Frequency Response of SQ-26-07 Hydrophone Pressure sensitivity of SQ-26-07 hydrophone at frequencies between 10 Hz and 50 kHz at hydrophone face, measured by manufacturer. Data was communicated to be representative of the the SQ-26-08 hydrophone. Plot courtesy of Joe Olsen from Cetacean Research Technology.

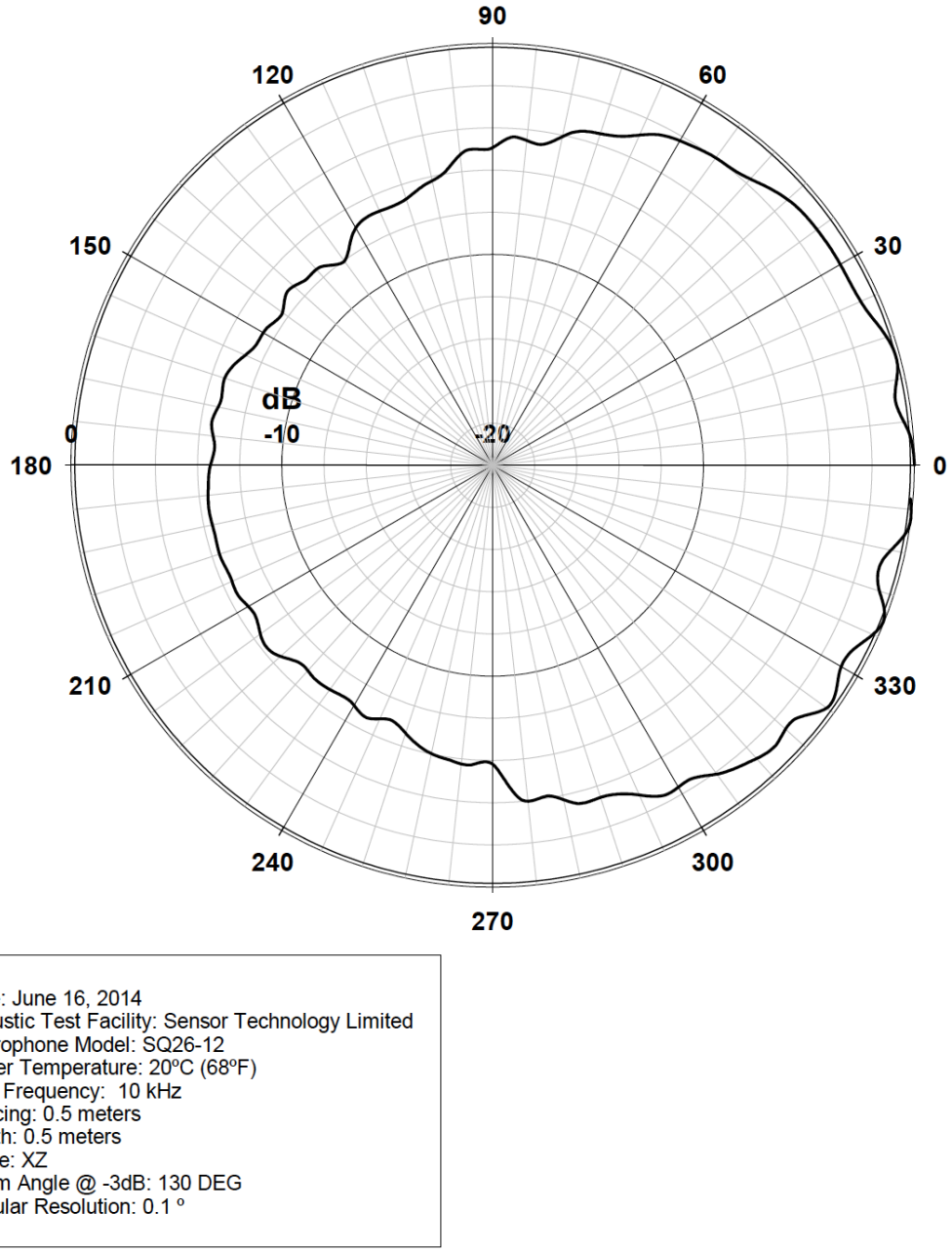


Figure 2.5: Vertical Sensitivity of SQ-26-12 Hydrophone at 10 kHz Pressure sensitivity of SQ-26-12 hydrophone at 10 kHz at different angles – with the x-axis pointing along the hydrophone body towards the face, y-axis pointing away from the body horizontally, and z-axis pointing away from the body vertically, angles are in the XZ plane with 0 degrees at the hydrophone face – measured by Sensor Technology Limited. Data was communicated to be representative of the SQ-26-08 hydrophone. Plot courtesy of Joe Olsen from Cetacean Technology.

As seen in Figure 2.4, we see that there is no more than a 1 dB fluctuation in sensitivity in our frequency range of interest. This is considered to be a "flat" frequency response, meaning that most users would approximate the relative power at different frequencies of a signal to be more a function of the sound source and propagation than hydrophone properties. Of course, the validity of this approximation depends on the application. During design we tentatively assumed this approximation to be valid for our purposes: if we eventually had reason to believe the frequency response was insufficiently flat, we could potentially, artificially flatten it further by adjusting the gain of our signals in the Fourier domain with the inverse of Figure 2.4.

As seen in Figure 2.5, we see a more significant 4 dB fluctuation in sensitivity between 0 and +/-90 degrees from the hydrophone face for a 10 kHz sound. Even so, in general the cylindrical transducer contained in sensors like the SQ-28-08 are said to produce an effectively "omnidirectional" response. Because the manufacturer suggested the hydrophone was sufficiently omnidirectional for our purposes, and because the more omnidirectional, spherical-transducer-containing hydrophones were both financially prohibitive and more difficult to protect from dolphin tampering, we settled for the SQ-28-08.

In both Mk. I and Mk. II versions of our hydrophone arrays, we embedded the hydrophones in small acrylic and later PTFE (Teflon) near-parabolic dishes (shape constrained by mill cutter availability) – see Figure 2.6 – both for protection from dolphin tampering and to potentially increase hydrophone sensitivity at greater angles from the face. We did not possess a sufficiently non-reverberant testing environment to rigorously produce a new plot like Figure 2.5 for the embedded hydrophone, however we we did qualitatively ensure that the dishes did not block signals (ratios of total of spectral power from glancing and direct signals were consistently above 0.5).

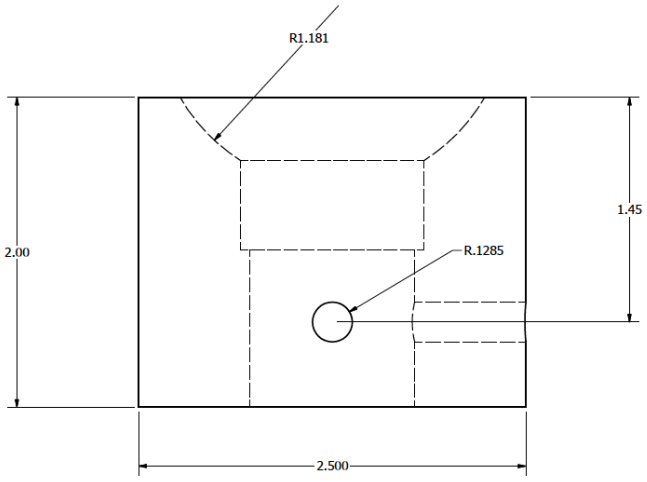


Figure 2.6: Hydrophone Insert Selection from a mechanical drawing for the Mk. II Hydrophone Array's hydrophone insert. Note the near-parabolic opening. The insert is itself held in place in the hydrophone panel and holds the hydrophone in place via two titanium set screws threaded through the panel and pressing into a rubber (EPDM)-stuffed clearance hole.

2.5 Hydrophone Array Mk. I

Recall that we determined that all arrays would be placed along the curved (radius of curvature 54-56'), 5.25"-thick acrylic sections of the EP wall, which is topped by a 2.25"-high lip extending 4" into the pool and 2.75" out of the pool. It was quickly decided that the only safe, versatile rigid attachment point for a hydrophone array was this lip, thus the hydrophone array was modeled as a "hanger" that rigidly attaches to this lip. The overall structure of the hydrophone array is depicted in Figure 2.7.

I will discuss the major pieces in turn together with the relevant design considerations. Aside from the consideration of corrosion resistance, given that dolphins live in 35 ppt salt-water, the major consideration for the Mk I. Hydrophone Array was its robustness to dolphin attack. Prior to the design process, the research team was informed by numerous aquarium officials of the aquariums' dolphins curiosity and "Vandal-like" qualities. One employee, Mark Turner, relayed two stories to this effect, one detailing the dolphins' destruction of a permanent, PVC hydrophone holder, and another detailing how the dolphins repeatedly rammed a small, dolphin-controlled submarine into a wall until destroyed.

Two C-channel brackets together with the plate lying atop the pool lip constitute the main attachment mechanism of the array. Large-gauge socket cap screws go through the brackets from the top and screw into the plate beneath; importantly, these screws go through slots in the brackets, allowing the brackets to slide along the plate and squeeze the pool lip before the screws are tightened. Given the pool radius (˜54') and the bracket length (20") the wall curvature has a negligible effect on the squeeze – the bracket length and pool curvature differ by approximately 1 part in 5000 – particularly given that the brackets are often installed atop adhesive protectors that can compress. To further ensure the tightness of the bracket assembly, three large-gauge, winged screws (manipulated by a diver) enter the bottom of the brackets and bite the bottom of the wall's lip.

To curb corrosion, the top plate was machined out of marine-grade 316 stainless steel, a difficult task requiring our primary machinist, Vadim Sherman, to use hard carbide tooling.

Heavily concerned with structural strength and the brackets' inability to splay over time, but unable to find appropriate marine-grade steel for them, the brackets were machined from wide carbon steel channels. As carbon steel is not corrosion resistant, I painted them with polyurethane.

Extending down from the bottom of the pool-side bracket are two rectangular support rods, for connecting the hydrophone panel with the bracket assembly; between the bracket and the support rods is a rectangular adaptor piece with counterbored holes to hide the heads of the screws entering the rods as well as those entering the bracket. The bracket assembly was designed so that the rods extend from its center of gravity, to avoid torque on the pool lip. Moreover, I was concerned that the subtle curvature of the pool together with the rectangular walls presented a danger to the pool's integrity were the dolphins to ram the array: potentially the force would be concentrated on the pool wall along the outer edges of the rods. Consequently, I machined several counterbored holes along the lengths of the support rods to house nylon cushions. Also with regards to safety, the support rods were designed to be six feet, which, as cantilevers, I calculated would only allow for an inch or two of splay from the wall; for fear of the assembly losing too much of its rigidity (a danger to its structural integrity, to the dolphins, and to audio quality) this was the longest I felt comfortable designing them to be, though for the quality of sound localization I would have liked to have made them longer.

Originally, these rods were machined out of 316 stainless steel for strength and corrosion resistance, but due to concerns about the effects of the assembly's weight (217 pounds out-of-water and 187 pounds in-water) on the ease of installation and removal they were again machined out of fiber-reinforced fiberglass (FRP), reducing the total weight by approximately 30 pounds.

Between the two support rods is a hollow 6061 structural aluminum pole for safely conveying four hydrophone cables from the hydrophone panel to the surface. To prevent the pole from applying torque to its attachment point on the panel should it be yanked by a

dolphin, it is held in place by an acrylic cross-piece bolted to the support rods. Note that 6061 aluminum tends to corrode in saltwater and so the pipe was fitted with threaded holes to accept an aluminum sacrificial anode, which must be replaced every few months.

Finally, at the bottom of the array is the sensor-containing hydrophone panel – seen in Figure 2.9 – the piece that the rest of the hydrophone assembly is designed to hold rigidly against the EP wall. The main function of the panel itself is to, without impairing audio quality, prevent its four hydrophones from being accessed by dolphins. Therefore, the panel was machined by Vadim out of a single piece of acrylic, and encases much of the body of the hydrophones, leaving the transducers mostly exposed. The panel's backside possesses channels for conveying the cables from the four hydrophones to the entrance of the cable pole. The panel is attached to the support rods with large-gauge socket cap screws, and to the cable rod with two smaller screws.

Even after substitution of fiberglass-reinforced plastic for steel in the support rods, the Mk. 1 Hydrophone Array is quite heavy: 187 pounds out-of-water. To reduce its effective weight, I developed a fast and safe process for assembling parts of the array on the pool lip itself, which nevertheless involved two people outside the water and one scuba diver relying heavily on a *buoyancy control device (BCD)*. The Mk. I's weight is one of several problems that would be addressed during development of the Mk. II.

2.6 Hydrophone Array Mk. II

We underwent the earlier-described process for habituating the National Aquarium's dolphins to the presence of the Mk. 1 Hydrophone Arrays. The arrays proved durable, withstanding not only dolphins and corrosion (though not entirely – explained below) but repeated installations, and after several months two of the four planned arrays had been "permanently" installed. However, one afternoon I received a phone call from the aquarium's head trainer: after investigating an array for some time, one curious female dolphin, Spirit,

had learned how to wedge her rostrum between the bottom panel and the pool wall. One of the male dolphins seemed to be picking up the technique. The Mk. I Arrays needed to be removed immediately to prevent potential injury to the dolphins.

Some time before this phone call I noticed the dolphins were interested in and successful at making use of the calculated one or two inches of splay to pry the bottom panel from the wall, and so I recorded their efforts while pondering solutions. Analysis of the footage indicated that slightly loosened screws had likely made the narrow adaptor piece (which connects the pool-side C-bracket and the support rods) susceptible to rotation, causing more splay and less restoring torque between panel and wall. Perhaps the Mk. I Hydrophone Array could be saved with the design of a wider adaptor and the use of a stronger threadlock, adhesive that fills space between screws and their holes and reduces slippage. However, I became concerned that the original cantilever predictions, which predict support rods of *any* material would weaken and splay with distance from the attachment site, did not adequately account for dolphin strength and the effect of transient water pressure gradients that could increase splay. Without reducing the length of the support rods, which was not a preferred solution for sound localization, a total overhaul of the hydrophone array was necessary.

Other problems with the Mk. I Hydrophone Array had surfaced that also motivated a redesign. First, the polyurethane coating on the carbon-steel C-brackets for preventing corrosion was beginning to chip; gradually corrosion of the carbon steel became visible, to the distress of the aquarium's exhibit staff. The aquarium's exhibit staff also voiced concern regarding the long-term potential for corrosion of even marine-grade stainless steel, as well as the array's general color scheme and its window-obstructing 20" width. I also hoped to reinforce certain pieces spread the four hydrophones over at least two different heights to improve the quality of sound localization. I will speak about the changes in turn. The whole assembly is shown in Figure 2.9.

The two C-channel brackets together with the plate lying atop the pool lip remain the same in principle and general design as in the Mk. I, apart from being reduced in width

(the whole assembly is reduced in width by 70%). A significant change is the material. To reduce the weight of the assembly and eliminate the need for a polyurethane coating, the marine stainless steel and carbon steel were replaced with structural 6061 aluminum; with the bottom panel fixed (to be discussed), calculations suggested that aluminum brackets made from solid 0.5"-thick aluminum would not splay. Moreover, to protect the aluminum from corrosion and avoid the need for sacrificial anodes, I sent the pieces (and all the aluminum pieces mentioned subsequently) for *Type III (hardcoat) anodization*. This process involves the aluminum piece being placed in a sulfuric acid bath to which a current is applied. The surface of the aluminum exposed to the bath is oxidized, creating an outer layer of aluminum oxide (rust). The 0.002" aluminum oxide layer is harder than the underlying aluminum (and is even harder than tool steel), does not chip, is electrically insulated, and protects the inner aluminum from corrosion.

The adaptor joining the pool-side bracket and the support rods grew in width as planned, to prevent rotation of the array around it. Two large holes were added to accommodate thumb-screws that enter the pool-side bracket and bite the pool lip, and one added to pass hydrophone cables. The adaptor was also modified to accommodate two new pieces, dubbed buttresses, which flank the support rods and fix them with horizontal bolts, and are secured to the adaptor above and the acrylic cross-piece below; apart from reinforcing the support rods' attachment to the adaptor, the buttresses are wide and also help to prevent rotation of of the assembly below the front bracket. The adaptor and buttresses are anodized aluminum.

The support rods are essentially unchanged apart from being one foot longer and attached to the buttresses. Instead of nylon cushions, which were degrading in the Mk. I, the support rods have plastic acetal foots atop rubber washers hidden in the counterbored holes. Also of note are four small holes that were added to accept a so called mini-panel, a reduced version of the bottom panel meant to allow two of the array's four hydrophones to be moved approximately four feet up. Due to financial constraints these could not be immediately built, impacting sound localization along the Z axis.

The cable pole is anodized and longer, and has additional attachments to the acrylic cross-piece and the bottom panel as a result of two approximately half-cylindrical, anodized aluminum collars; one can be seen in Figure 2.10. It has also been modified with a window to accommodate the mini-panel in the future.

The face of the bottom panel, seen in Figure 2.10, has been rotated by 90 degrees consistent with the array's width reduction. Its hydrophone inserts are constructed of chemically-resistant, low-friction PTFE (Teflon). Two acetal plastic arms are attached to the left and right hand side of the panel and hold the solution to the cantilever problem, marine suction cleats (Darby 6.5in Suction Cleat).

As their name implies, these cleats are designed to be used as underwater attachment points for boats. One cleat is rated to resist a force perpendicular to the cup surfaces of at least 100 pounds. Normally a cleat is manually engaged via two levers that increase the cup curvature and help to create suction against a surface; I significantly reduced the levers to "dolphin-proof" them, drilling holes into the stubs that could be manipulated by a custom tool machined by Vadim. Later acetal plastic boxes were fabricated that cover the cleats and their holding arms for additional security. The holding arms were designed such that the cleats could move freely by ˜0.5" in any given direction to allow the cups to adjust to the pool wall.

During installation, the engagement of the cleats and attachment of the protective boxes are performed by a diver; they must also be re-tightened by a diver approximately once every two weeks. Because of this complication and because the suction cups and boxes were deemed aesthetically displeasing by aquarium officials, a test was performed to ensure that the cleats are necessary to adequately reduce the splay of the bottom panel from the pool wall despite the strengthening of the upper array. It was proven that that they are.

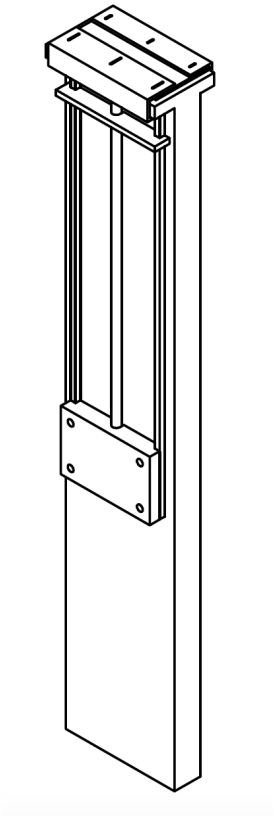
Due to imperfections of the suction cleats caused during manufacture, shipping, or usage, several replacements must be kept on hand for when one loses suction. Even among newly-arrived suction cleats, approximately 75% proved incapable of at least two weeks of suction

in air, and 25% in water during fish tank testing (success in air seemed to predict success in water). Only those maintaining suction for at least two weeks in water were among the first batch of suction cleats deployed.

Note that all formerly marine-grade stainless steel fasteners were replaced with acetal plastic fasteners or titanium fasteners, for fear of stainless steel's imperfect corrosion resistance in saltwater. As time went on many acetal fasteners were replaced with titanium fasteners for improved mechanical strength.

Four Mk. II Hydrophone Arrays have been permanently installed in the EP of Dolphin Discovery since June 2017. Apart from the need for occasional reattachment of the suction cleats and their boxes, for occasional algae cleaning, and for one on-the-spot mechanical modification, they have remained surprisingly robust to their environment for almost a year.

(a)



(b)

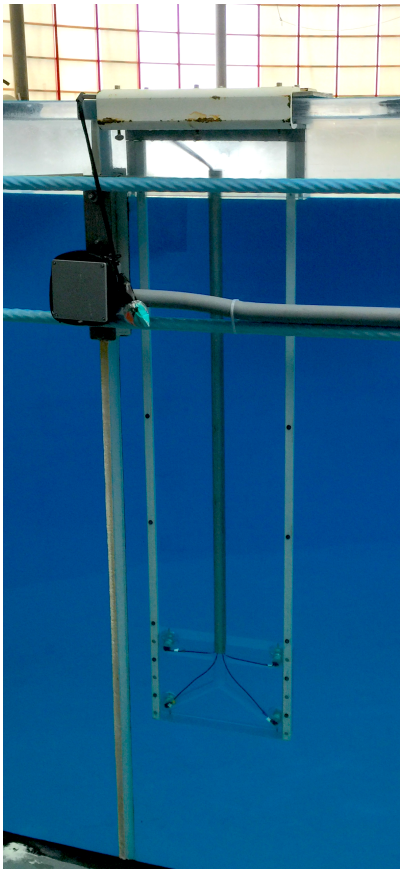


Figure 2.7: Mk. I Hydrophone Array

(a) Selection from a mechanical drawing. Array is depicted attached to a small section of acrylic wall, which visibly has an approximate "T" cross-section. Notable pieces include two C-channel brackets hugging both the pool lip and a rectangular plate atop the lip, a central cylindrical pole containing cabling between two square support rods and secured by a thin rectangular piece, and the bottom panel with four holes for hydrophones.

(b) A photo of an installed array.

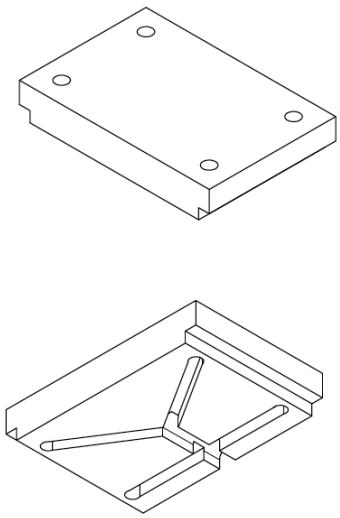
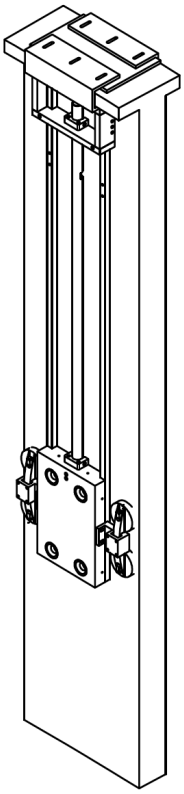


Figure 2.9: Mk. I Hydrophone Array Panel Selection from a mechanical drawing for Mk. 1 Hydrophone Array Panel. Visible in both the top and bottom images are four holes at the panel's four corner for holding hydrophones. These holes have two diameters along their lengths, fitted to the two-diameter geometry of the hydrophones, so that the hydrophones are held at fixed distances from the panel face. Note that the hydrophone inserts seen in Figure 2.6 are not included in this diagram as a simpler version (not shown) was designed and implemented at short notice. Also visible in the bottom image are grooves for the support rods, for the cable pole, and for the hydrophone cables.

(a)



(b)

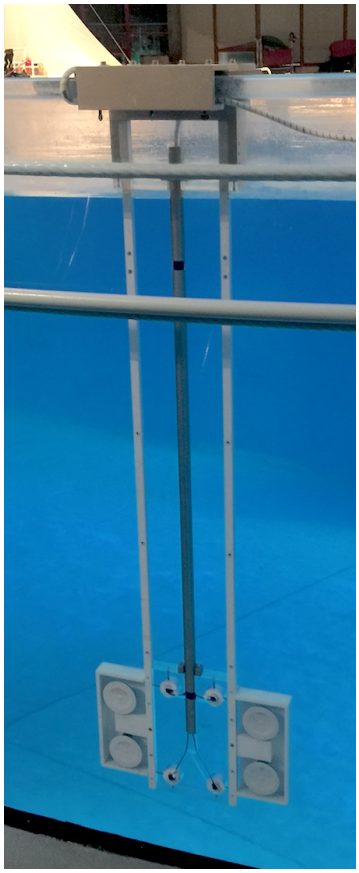


Figure 2.10: Mk. II Hydrophone Array

(a) Selection from a mechanical drawing. Array is depicted attached to a small section of acrylic wall, which visibly has an approximate "T" cross-section. Consult the main text for a full description of modifications. Protective boxes for suction cups are not shown.

(b) A photo of an installed array.

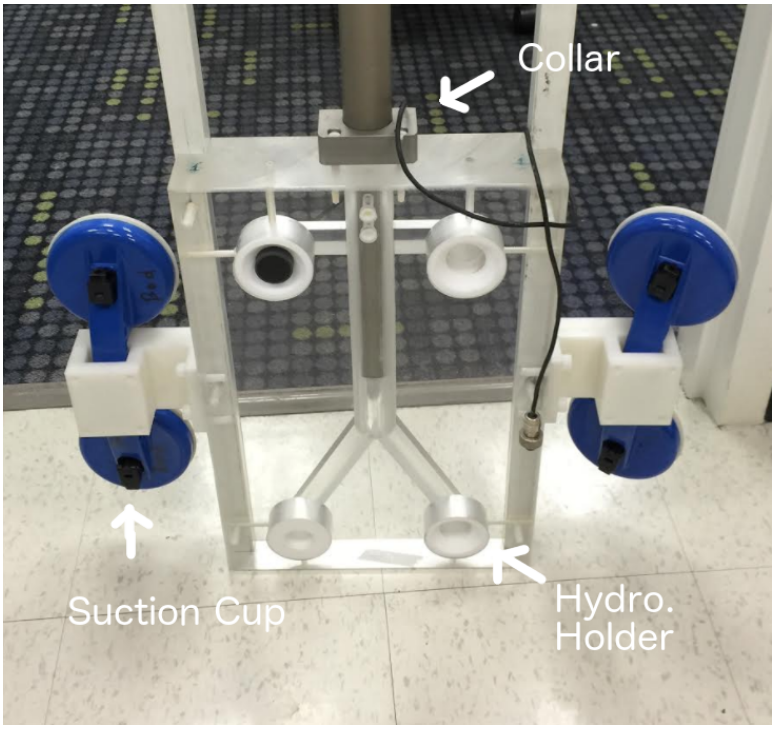


Figure 2.12: Mk. II Hydrophone Array Panel Photo of the Mk. II Hydrophone Array Panel, partially assembled. From the Mk. I the panel itself has been rotated 90 degrees, has had various holes added for fasteners, and includes fuller hydrophone inserts, from Figure 2.6, with near-parabolic cuts. Visible on the top of the panel is a collar securing the cable pole. Most notable is the addition of two acetal plastic arms holding two "suction cup cleats," described in the text.

Chapter 3

Dolphin Surveillance System: Cameras and System Infrastructure

3.1 Cameras

Recall that whistle attribution as we intend to achieve it at the National Aquarium consists of two tasks: sound source localization, accomplished by processing audio data received from the hydrophone arrays described last chapter, and visual localization of potential sound sources. The visual localization task involves not only locating where potential sources are at the times that sounds of interest are received, but identifying these potential sources uniquely, so that across time all sounds are properly assigned to their common sources.

The first-subtask, locating potential sources at particular times, starts with visually locating sound sources in photo or video feed based on *a priori* information about visual features that distinguish them from background. The result of this image processing task is locations for the potential sound sources in *camera coordinates*, which for a single camera consists of a set of 2D pixel points. To match these locations with the results of sound source localization for attribution, these camera-coordinate locations must be converted to locations in *world coordinates*, or real space. For a point represented in 2D pixel coordinates (here the

camera is assumed to be a lens-less *pinhole camera*) as $[u\ v\ 1]^T$ and in 3D world coordinates as $[x_w\ y_w\ z_w\ 1]^T$ the transformation can be expressed as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{3.1}$$

where K is a matrix of the camera's *intrinsic parameters* (including focal length, image sensor format, and principal point), and R and T are *extrinsic parameter* matrices describing the camera's position in the world – they denote rotation and translation matrices, respectively.

Obtaining the linear intrinsic parameters K is usually a simple task involving an in-lab calibration or asking the manufacturer. Obtaining extrinsic parameters R and T is more difficult, typically requiring an *in situ* calibration. Also, it is important to note that solving Equation 3.1 for a point's world coordinates with known camera coordinates produces the coordinates for a shaft of light; solving for a point's 3D world coordinates exactly requires at least two cameras and a more complex mathematical treatment, and thus requires finding the extrinsic parameters for two or more cameras.

At the aquarium, several factors complicated accomplishing the first-subtask of locating potential sources. For any camera placed so as to visualize the dolphins from above the pool, the air-water optical interface was troublesome: various algorithms I implemented for determining the camera coordinates of dolphins (excluded from this thesis) were disrupted by a "shimmering surface" or *sun glitter* effect (light reflecting off a dynamic water surface), and the mathematics for finding the mapping from camera coordinates to world coordinates were complicated by refraction. Also, the distance at which such a camera must be placed from the water and dolphins not only amplified both the linear and nonlinear effects of camera distortion but made the implementation of an ever-more-necessary camera calibration more difficult (see Figure 3.1). For any camera placed on the pool wall so as to not encounter

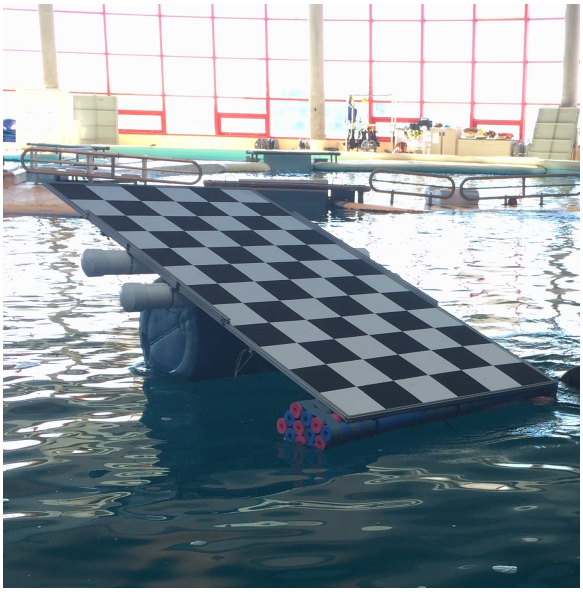


Figure 3.1: Camera Calibration Chessboard Standard, established practice for calibrating both a camera's intrinsic and extrinsic parameters involves recording a chessboard of known grid-size at many positions/orientations and optimizing the necessary coordinate transforms. Below is an 8'x4' floating chessboard made for this purpose.

an air-water optical interface (a setup which for some time was not feasible), the primary complicating factors were the need for multiple cameras to cover the space, and again the logistics of performing camera calibrations.

The second sub-task, uniquely identifying the potential sources, can be trivial if the different sources can be easily distinguished in any arbitrary photo or video frame. However, for dolphins it is at best possible to distinguish among sources in certain frames. In this case, the standard solution for extrapolating this information across frames is *object tracking*. Object tracking involves following objects over sequential video frames as they continuously move in space. At the aquarium, the unique difficulties of object tracking primarily consist of those that affected potential source localization, if to a lesser extent (i.e., it is easier to localize an object when its location a short time ago is known, which is the case in tracking).

With the overall task of visual localization having been introduced, I now discuss the

hardware installed towards achieving it. At the time of installation, installing underwater viewing cameras was not practical (and currently is to a limited extent). Considering this, it was quickly determined that the best views of the EP were obtained by cameras placed on a catwalk extending across the EP, measured to be 46 feet above the water surface. On this catwalk it was not only possible to place a single camera with a full overhead view of the EP, but several cameras in a line that collectively covered the full pool and achieved two-camera viewing overlap at all points, allowing for a *stereoscopic* approach to overcoming the refractive effects caused by the pool surface's air-water interface, and potentially to obtaining reliable 3D world coordinates for dolphins.

The camera type chosen, the AXIS P1435-LE, was suitable on the basis of its relatively wide field-of-view (˜95^o horizontal, ˜51^o vertical), low-light sensitivity (˜0.02 lux), resolution (1080p), frame-rate (theoretically 60 Hz), and for being a *powered-over-ethernet* (PoE) network camera (IP). As we required that the cameras be operated from positions tens of meters from their hub/power and share a common clock, this Cat-cable-mediated standard seemed suited to our needs.

The cameras were attached at the edge of the catwalk base Figure 3.2. They were directly attached to an optics-grade adjustable angle bracket that can in turn slide along an optical railing, which allowed for the camera view to be adjusted during installation.

Ultimately, a failed large-scale attempt at calibrating for the cameras' extrinsic parameters Figure 3.1 led me to abandon a precise stereoscopic approach involving all five cameras and opt for relying on the one, primary overhead camera. Towards performing a rough 2D mapping between camera coordinates and "pool coordinates," the aquarium staff helped me measure out a small grid of black lines visible in this camera on the bottom of the pool Figure 4.1a, and I measured the catwalk with respect to the same coordinate system. A reliable automated visualization localization algorithm is still being researched, as sun glitter foils standard approaches, including background subtraction/blob detection, adaptive correlation filters, and cascade objection detection.

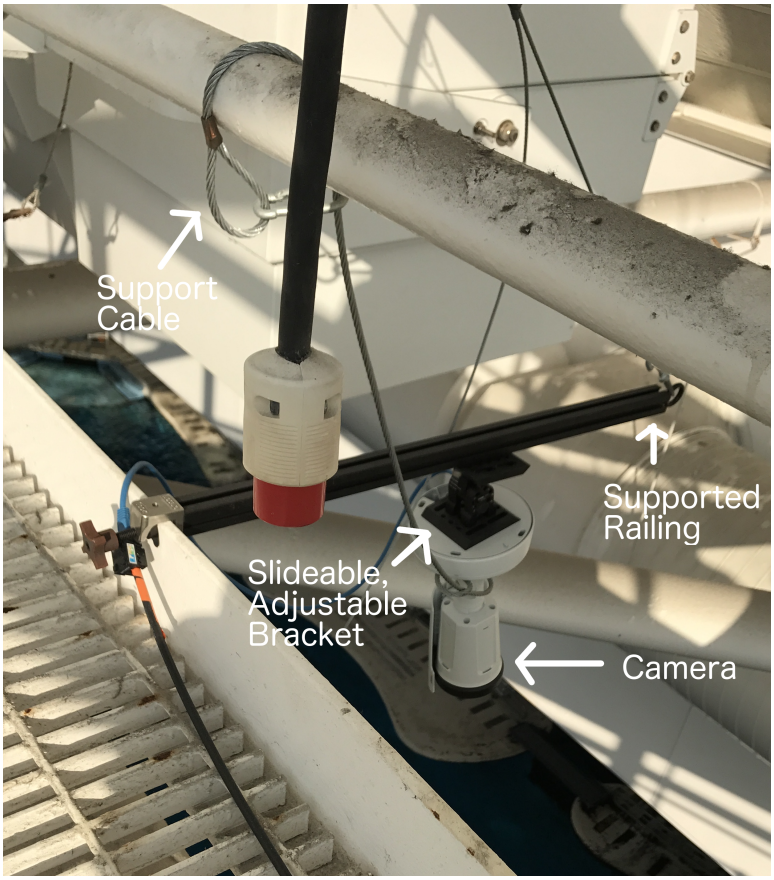


Figure 3.2: Mounted AXIS P1435-LE Camera Photo of a camera (AXIS P1435 LE) mounted on the catwalk spanning the EP. The camera is mounted to an adjustable angle bracket (Thorlabs Adjustable Angle Mounting Plate), which in turn is connected to a slideable piece on an optical railing (Thorlabs 25 mm Construction Rail) attached to the catwalk. The end of the optical camera is secured to the catwalk railing by a braided steel wire (McMaster Adjustable-Loop-to-Hook Wire Rope Lanyard, 1/8" Rope Diameter), and for safety the camera itself is attached to another braided steel wire.

3.2 Data Infrastructure/Lines

Each of the four installed hydrophone arrays, which hang over the lip of the curved, acrylic EP wall, contains four hydrophones. Outside each array, secured to the nearest metal seam connecting adjacent pieces of acrylic walling, the aquarium installed a 6"x6"x6" PVC box that was purportedly (but not actually) waterproof. A 2" hole was drilled into the bottom of each box, through which the four hydrophones' two-wire shielded cables enter, bound together upon exiting the array with polyethylene sheathing (McMaster Spiral Bundling Wrap, Polyethylene, 3/8" ID, 1/2" OD) and PTFE (Teflon) sheathing (McMaster Chemical-Resistant Expandable Sleeving for High Temperature, Heavy Duty PTFE, 1/2" ID); after any "permanent" array installation the hydrophone cables must be placed inside and the hole filled with silicone sealant (3M) Marine Grade Silicone Clear Sealant). Inside the PVC box, the four individual hydrophone cables terminate on a single 8-pin male connector, which in turn connects the corresponding female attached to a 0.5"-diameter "snake" cable: inside the snake are four shielded twisted wire pairs sharing an outer shield, carrying signals from the four hydrophones. From the PVC box, the snake cable travels through a 1.5"/1" PVC conduit along the EP perimeter towards the central concrete slide-out, eventually entering a waterproof, 1.5'x1.5'x0.5' stainless steel box. In total, two stainless steel boxes were installed, one on either side of the EP's central slide-out, each containing two snake cables carrying data from two arrays (eight hydrophones).

Also inside each stainless steel box is a MOTU 8M audio interface that digitizes eight analog hydrophone signals at 192 kHz and conveys them to the computer hub some distance away; the two audio interfaces are internally connected via a Cat5 cable that slaves one to the other (with regards to data as well as clock time), and a single optical cable from one MOTU sends data to the hub. A complicating factor is the need to provide a 5V *bias voltage* across the two wires of each hydrophone. This voltage is necessary in any condenser hydro/microphone for impedance conversion, helping to reduce signal loss. This voltage was provided by a 5 V DC power supply (powered by a 120 V outlet inside the

box, shared with the MOTU) powering an 8-channel voltage fan-out circuit commissioned from Sanjee Abeytunge. Technical details of Sanjee's board and my wiring are excluded (see the hydrophone manufacturer, Cetacean Research Technology, for standard hydrophone operation requirements).

The cameras, being *powered-over-ethernet* (PoE), are served by individual Cat5e cables running some distance along Dolphin Discovery's network of upper catwalks. The cables converge on a single Ethernet switch (Ubiquiti Networks EdgeSwitch 24 Port 250 Watt Managed PoE+ Gigabit Switch with SFP) connected to the hub.

At the computer hub (currently, a late 2013, 12-core Mac Pro running macOS Sierra), the hydrophone signals are received as standard audio input via the MOTU 8M audio interface; most audio settings were changed in proprietary MOTU mixing software. The cameras are accessible via IP addressing, and their settings were changed through firmware.

3.3 Recording Software

As the ultimate purpose of the hydrophone and camera sub-systems is to provide long-term audiovisual recordings of dolphin interactions, we wished to run software on the hub for recording time-synched audio and video continuously. No pre-made software was readily available for this purpose. I decided to develop software on the Matlab platform for the strength of its function packages (*toolboxes*), for its relative user-friendliness, and for consistency with other areas of the aquarium codebase.

The MOTU 8M audio interface and in turn the 16 hydrophone feeds can be straightforwardly accessed through Matlab's audio device reader and sequential 5-minute, 16-channel WAV-format files (this approaches the 2 GB WAV file limit) generated. Complicating matters slightly, two parallel programs (one taking odd and the other even counts of a recording series, as dictated by the system clock) must be run to accomplish truly continuous recording due to the extra time required by a single program to write the audio to disk.

Unfortunately, Matlab is not practically capable of independently accessing the feeds of multiple IP cameras and making parallel recordings, and so using an intermediary application has been necessary for recording video. The most suitable application found at present is the IP surveillance program *Xeoma*. It provides an interface for real-time viewing of all video feeds, and dumps these feeds into archives. Through Matlab, I access these archives and save video in 5 minute chunks time-synched to the 5-minute audio files.

A single master Matlab script activates the three above scripts on separate parallel workers (i.e., independent processing units) on the computer hub, and all data is recorded to a large, 48-Terabyte *RAID* (for *Redundant Array of Independent Disks*) device (Promise Technology 48TB Pegasus2 R8 Thunderbolt 2 RAID Storage Array). Various helper scripts have been written for small tasks such as discarding old data and consolidating daily recordings.

Chapter 4

Dolphin Surveillance System: System Calibration/Testing

4.1 The Impulse Response Function

With the system hardware installed, I sought to measure the acoustic *impulse response functions* (IRF) of the hydrophone sub-system. In this case, an "impulse response function" refers to the acoustic *response* received in one hydrophone (at a particular position in the pool) to an acoustic *impulse* – ideally, a *delta* function, a signal of infinite amplitude lasting for an infinitesimal time – generated at another particular position in the pool. Ideally, the distortion of the source signal manifesting in the received signal is a function of the physical environment through which the source signal travels (in actuality, it also includes the response properties of the sending and receiving devices).

I had two reasons for measuring the hydrophone system's IRF's. First, under certain circumstances, an IRF contains complete information about the *multipath* contributions of the received signal corresponding to *any* source signal, not just an impulse (but for constant source and receiver positions). As will be discussed more in the introductory section of the chapter on Time Delay Estimation (5.1), the multipath contributions of a received signal

result from the source signal reaching the sensor from many different paths (e.g., due to reflection), creating difficulties in time delay estimation for which, typically, only the direct path is important. Through a deconvolution process, a source/sensor pair's IRF can remove multipath effects from a received signal, leading to easier time delay estimation.

As an aside, note that the circumstances under which the IRF contains all the information about the multipath contributions of the received signal for *any* source signal include that the system is *linear* and that it is *time-invariant*. Linearity means that, if the responses of the system, $y_i[t]$, to the individual input signals $x_i[t]$ are known, then an input to the system consisting of the weighted addition of several $x_i[t]$ produces a predictable response consisting of the weighted addition of the corresponding $y_i[t]$:

$$\sum_i c_i x_i[t] \rightarrow \sum_i c_i y_i[t] \tag{4.1}$$

Time-invariance means that the time at which we apply some input signal $x_i[t]$ to the system has no effect on the response $y_i[t]$ other than shifting it in time. That is, for any time shift T, we have:

$$x[t-T] \rightarrow y[t-T] \tag{4.2}$$

While the response functions of the speaker-hydrophone-aquarium system are safely considered to be time-invariant, they are not linear, owing to a number of nonlinear contributions, in particular the amplitude and frequency dependent responses of the speaker and hydrophones, and electrical noise in the accompanying electronics. As the relative magnitude of these effects are unclear, it is possible that a *linear, time-invariant* assumption would still hold.

The second primary reason for measuring the hydrophone system's IRF's was for producing signals at known locations in the pool that were easy to extract time delay estimations from and therefore localize. This was important to ensuring that the various principles and large-scale geometrical measurements on which the hydrophone system is based are valid. Again, as will be reiterated in more detail in Section 5.1, time delay estimation for whistles is made difficult by multipath complications as well whistles' slow (and often irregular) onset

above the acoustic noise floor. By contrast, time delay estimation for delta-like impulses is relatively easy because of their abrupt onsets, which suits them to simple algorithms that are mostly insensitive to multipath effects. Indeed, it is for this reason that attempts at pool sound localization for echolocation signals have been successful.

4.2 Calibration Methodology

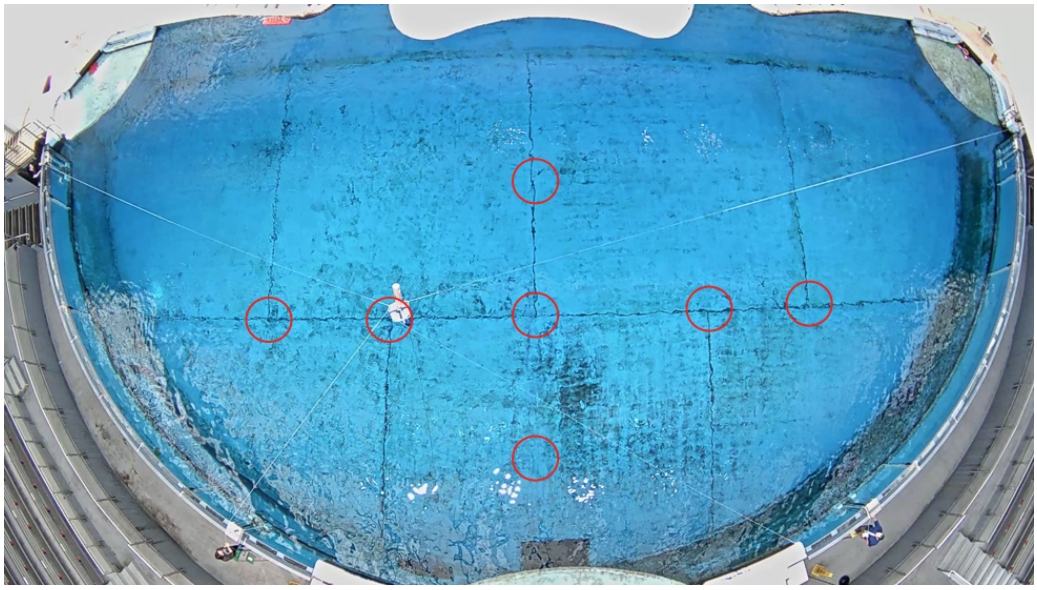
The overall strategy for obtaining IRF's was to play an appropriate signal on an underwater speaker, a Lubbell LL916H, suspended at known heights from a buoy at known distances with respect to the hydrophone arrays. As the heights of the hydrophone arrays as well as their projected positions with respect to an orthogonal, two-dimensional grid system on the bottom of the pool were already measured, the source points as well as all hydrophone arrays occupied known positions on a common coordinate system. Photos of aspects of the process are shown in Figure 4.1, and will be referred to throughout this section.

Before discussing the logistics, it is important to note that, though the goal of the process was to obtain *impulse* response functions, the signal I played from the speaker was not a delta-like impulse. Given the limitations of the speaker system and considerations of animal husbandry, it would not have been practical to play a sound approximating an infinite-amplitude, infinitesimal-duration delta function. However, with a quick derivation we can see that it is not strictly necessary to use an impulse to acquire the IRF.

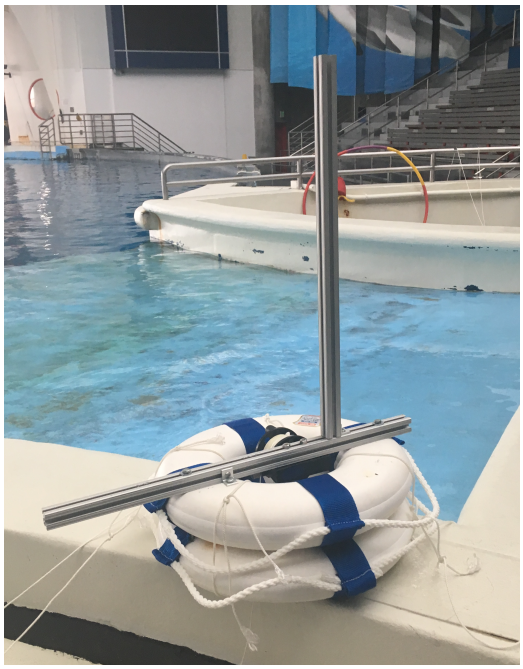
For a *linear, time-invariant system* (LTI), described last section, the received signal $y[t]$ for a known source signal $x[t]$ can be described as a convolution with the IRF, $h[t]$:

$$y[t] = \sum_{\tau=-\infty}^{\infty} x[\tau]h[t-\tau] \tag{4.3}$$

We can take the Fourier ($\mathscr{F}$) transform of both sides (noting $X[f] := \mathscr{F}\{x[t]\}$, $Y[f] := \mathscr{F}\{y[t]\}$, $H[f] := \mathscr{F}\{h[t]\}$), and, with the convolution simplified in Fourier space, rearrange the equation to pull out the Fourier-transformed IRF (called the *frequency response function*)



(a) Overhead view of a impulse response testing session. Four white ropes connected to the buoy are held by assistants at the four panels. Testing locations are circled in red; note that each designated location includes an upper and lower position



(b) A close-up of the buoy used for testing. The speaker is held by the winch in the center. Not shown is the corrugated aluminum cylinder around the vertical aluminum railing, the target of laser range-finding.

Figure 4.1: Impulse Response Testing

before taking the inverse Fourier transform of both sides to recover the IRF.

$$\mathscr{F}\{y[t]\} = \mathscr{F}\{\sum_{\tau=-\infty}^{\infty} h[t-\tau]x[\tau]\} \tag{4.4}$$

$$Y[f] = H[f]X[f] \tag{4.5}$$

$$\mathscr{F}^{-1}\{H[f]\} = \mathscr{F}^{-1}\{\frac{Y[f]}{X[f]}\} \tag{4.6}$$

$$h[t] = \mathscr{F}^{-1}\{\frac{\mathscr{F}\{y[t]\}}{\mathscr{F}\{x[t]\}}\} \tag{4.7}$$

The above suggests that we can obtain the impulse response given *any* pair of source and received signals. However, to ensure that the denominator is nowhere zero, and to avoid biasing for any frequency in particular, in practice it is best if the source signal's power is uniformly distributed across all frequencies (also a property of a true impulse). Such a signal can be obtained by inverse Fourier transforming a signal designed in complex frequency space that has unitary power at all frequencies, with random phase. Note that if the signal thus obtained is not played at the appropriate sampling rate in its entirety, its power spectrum will not be unitary, but rather random with powers falling on a Gamma distribution – this is consistent with the power reflecting the absolute value of Gaussian variable pairs in real and imaginary space. The duration of the signal should be longer than the longest expected multipath travel time (one second was used). Moreover, the signal can be repeated a number of times (360 was used) to account for various stochastic effects: the IRF is constructed from the median value for every time point.

The Lubbell LL916H speaker responsible for playing the calibration signal has a spliced, 125' cable that I connected to an amplifier-transformer-power-source circuit (see speaker manufacturer for standard operation), which received input from an analog output of one of the two MOTU 8M audio interfaces serving the hydrophone system. In this configuration, the 192 kHz calibration sound was played in synch with the 192 kHz audio recordings, and both were managed by the hub computer. Also on the hub computer, a video record of the calibration was kept by the software described last chapter. I personally managed the

hub computer during this time, and via an intercom system communicated to the assistants below.

Four assistants were positioned on ladders around the EP, one at each hydrophone array. The assistants each held a rope connected to a buoy Figure 4.1b from which the Lubbell speaker was suspended by a winch/spindle (maker/model unknown); the ropes were used to move the buoy (two JIM BUOY Catalina Series Economy Life Rings, 17" Diameter, bolted together) across the EP and stabilize it. Also on the buoy was a 6"-diameter, 2'-high corrugated aluminum cylinder that the assistants were instructed to target with a laser range-finder (indicated previously) whenever the buoy reached a designated position in the pool; these measurements were recorded and later used to find the speaker location with respect to the hydrophone arrays using a standard triangulation procedure.

In total, the buoy was moved to seven positions in the pool for two different speaker heights, seen in Figure 4.1a.

4.3 Whistle De-Reverb

IRF's for the fourteen testing locations, for each of 16 hydrophones, were obtained. An example is shown in Figure 4.2. Referring to Figure 4.1a, this IRF is for a source located at the upper-far-left point and a hydrophone located in a panel at the far right. The red line marks the first incidence of the impulse, and the green line marks my geometry-based prediction of the last of the primary reflections, about 12 milliseconds after the first incidence. We see that several peaks follow and gradually decay to the noise floor after about 110 milliseconds, indicating higher-order reflections, or *reverberation* (technically, anything after 50 milliseconds is *echo*). In general, we expect this picture to be unique for every pair of source and hydrophone locations, dependent on pool geometry and material properties. I have done preliminary work modeling these effects using a ray-tracing simulation approach, which is excluded.

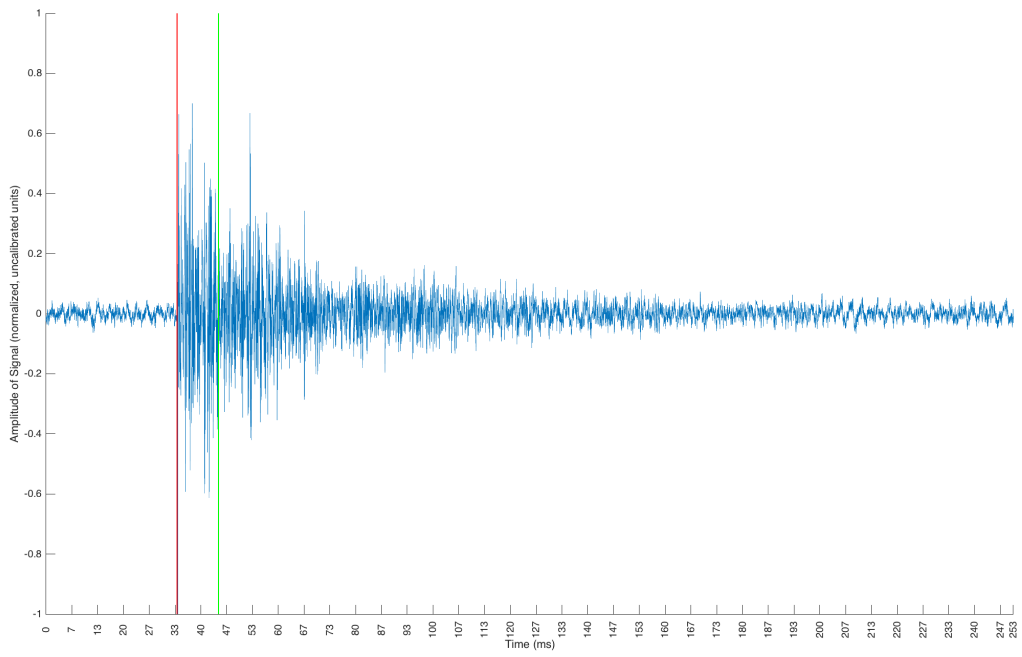


Figure 4.2: Example of Impulse Response Function (IRF) Example of an Impulse Response Function (IRF), obtained by playing a broad-frequency calibration signal at the upper-far-left calibration position and listening from a hydrophone in the far right hydrophone array, as seen in Figure 4.1a. The red line marks the receipt of the first incidence of the signal, the green line marks the estimated receipt of the last of the primary reflections.

Recall that the first reason for obtaining the IRF's was to remove multipath contributions – everything following the first incidence in Figure 4.2 – from received whistles. While in theory this can be performed with a simple deconvolution of the whistle signal by the IRF by inverting Equation 4.3, in practice it is better to use an approach that tries to minimize the effects of noise on the deconvolution. A standard method is the *Wiener deconvolution*. It will be stated without derivation. For a received signal $y[t]$, the estimate source signal $\hat{x}[t]$ can be written as the convolution $\hat{x}[t] = (g * y)[t]$. The Fourier transform of the Wiener kernel $g[t]$ is given by:

$$G[f] = \frac{H^*[f]S[f]}{|H[f]|^2 S[f] + N[f]} \tag{4.8}$$

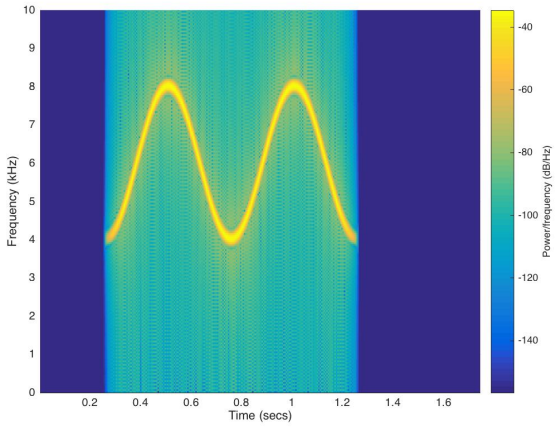
where $S[f]$ is the power spectral density of $x[t]$, $H[f]$ is the Fourier transform of the IRF, and $N[f]$ is the power spectral density of the estimated noise. A representation of the performance of the Wiener deconvolution method on a whistle-like sound played in the pool where the IRF signal was generated is given in Figure 4.3.

It should be readily apparent that the Wiener deconvolution operation did not recover the original signal. In fact, it further degraded the signal with noise in a number of frequency bands; outside these bands, traces of the original signal can be seen. The cause of this added noise is unlikely to be inaccurate values for the noise term $N[f]$ in Equation 4.8, as these were derived from real background noise and, moreover, are not present in the standard deconvolution, which produces similar results. More likely, the values for the estimated power spectral density of the original signal, $S[f]$, as well the IRF $H[f]$ are unreliable at certain frequency bands. One reason is that the frequency response of the Lubbell speaker and circuitry serving it are not accounted for whenever the input "source signal" is used in a calculation. A way to better account for this in the future would be to record the sound at the source point with a similar hydrophone as used in the hydrophone arrays, and to use this signal instead of the input signal as the source for deconvolution. However, were the speaker biasing particular frequency bands or in general adding unspecified distortion to source signals, we would still expect a non-optimal estimate for the IRF's. Perhaps with detailed knowledge of the speaker system response it would be possible to craft a compensating calibration signal.

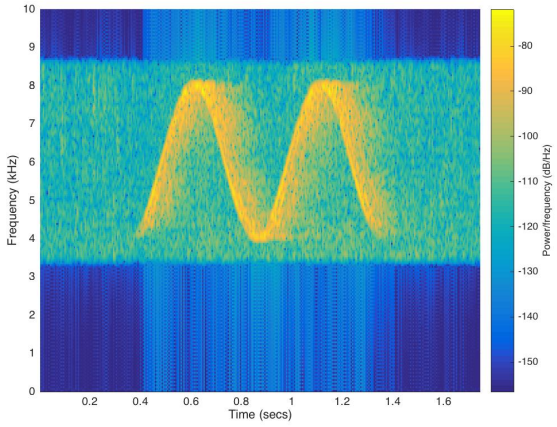
Note that similar results as Figure 4.3 were obtained across whistle/IRF pairs. Attempts were made to use filtering and an ad hoc approach of ignoring frequency bands expected to accumulate high noise in the deconvolution, and the standard deconvolution was also used; nothing tried significantly improved the quality of the recovered signal.

Figure 4.3: Wiener Deconvolution of a Whistle-like Signal with the IRF a: The original whistle-like tonal signal played at the source location (far right point in Figure 4.1a). Displayed in a standard 2048 Hamming-window spectrogram. **b**: The signal received at a hydrophone (from the far right array in Figure 4.1a). Note that a bandpass filter was applied between ˜3.25 and ˜8.75 kHz. **c**: The deconvolved signal.

(a)



(b)



(c)

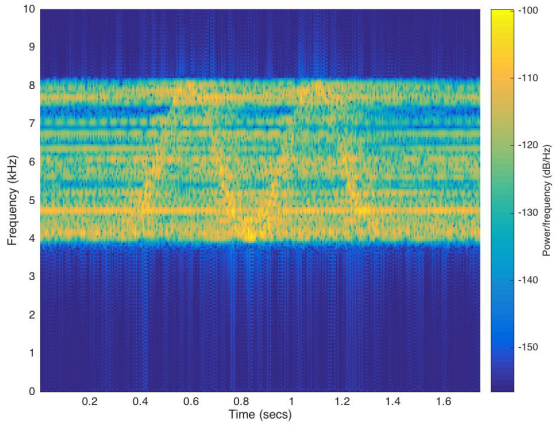


Figure 4.3

4.4 Impulse Response Function Localization

For the purposes of testing the aquariums system's capacity for signal localization, however, the IRF's proved helpful. First, I manually extracted the first incidence times for all testing locations in all hydrophones, from plots resembling Figure 4.2. While a peak-detecting algorithm might have accomplished this task, given the occasional appearance of false peaks and the relatively small size of the data set I determined it was safer and potentially more time efficient to proceed manually. Allowing for error, each individual time became a set of times, corresponding to the extracted time added to a set of Gaussian random variables (mean of zero, and standard deviation calculated to be a multiple of the intra-hydrophone-array time variability between upper and bottom rows – the differences between these channels were found to be mostly noise-based) added. By simple subtraction these times became arrival time differences (TDOA's) between pairs of hydrophones. The TDOA's were fed into a standard sound localization algorithm termed *spherical interpolation*, described in more detail in Section 6.2; each set TDOA's was mapped to a single point in space.

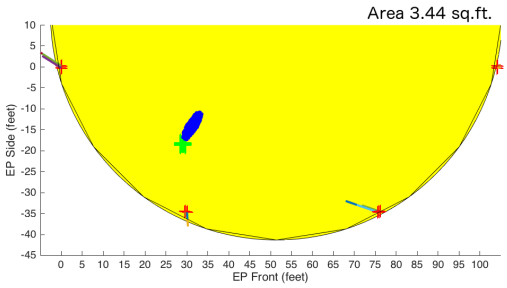
Figure 4.4 shows the results in fourteen plots, one for each IRF. While these plots are two-dimensional projections of three-dimensional results, significant information is not lost in the projections, as the hydrophone panels have effectively no localization precision along the Z-axis (substantiating my original plan to include a Z-displaced "mini-panel" in the Mk. II Hydrophone Array). In two dimensions the clouds of localized points can be approximated as ellipses and characterized by two orthogonal radiuses (a width and a length). Under this approximation, I have calculated the average areas of the clouds as well as the percentage of the EP they occupy, as well as the distance between the true calibration points from the nearest cloud points; this is done for all calibration locations as well as separately for midline and non-midline locations – as is visible from the plots, the former are localized more poorly, likely a consequence of the array and pool geometry that requires further examination. These data are in Table 4.1.

Table 4.1: Performance of Spherical Interpolation on IRF Signals

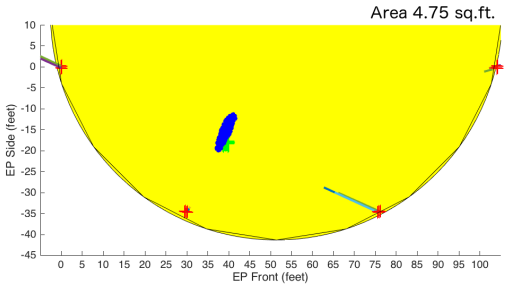
Statistic	Value
Mean Cloud Area (ft^2, % of pool)	10.84 +/- 9.63 (0.24 +/- 0.21)
Mean Cloud Area, non-midline (ft^2, % of pool)	6.310 +/- 2.51 (0.14 +/- 0.054)
Mean Cloud Area, midline (ft^2, % of pool)	16.88 +/- 12.48 (0.37 +/- 0.27)
Mean Distance of True Point from Cloud (ft)	2.63 +/- 3.76
Mean Distance of True Point from Cloud, non-midline (ft)	2.29 +/- 1.75
Mean Distance of True Point from Cloud, midline (ft)	3.08 +/- 5.86

Figure 4.4: IRF Localization Each plot represents a simplified overhead 2D projection of the EP. Indicated as a green cross is the true position of the IRF. Indicated by a red cross (not usually visible) is the position of the estimated position of the IRF, from estimated time of arrivals; each blue asterisk represents the estimated position of the IRF from estimated time of arrivals plus Gaussian random variables, described in the text. The four red cluster of crosses around the EP perimeter indicate hydrophone positions. The lines from them are proportional to the hydrophones' estimated time estimation error; when oriented towards the true point, they indicate that estimated times were too late; when towards, too early.

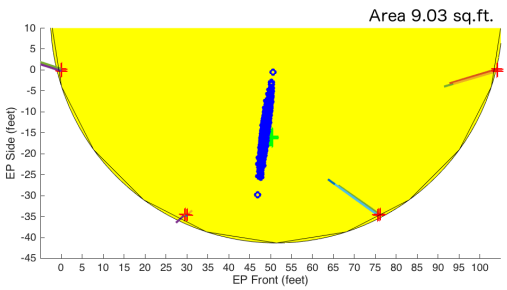
(a)



(b)



(c)



(d)

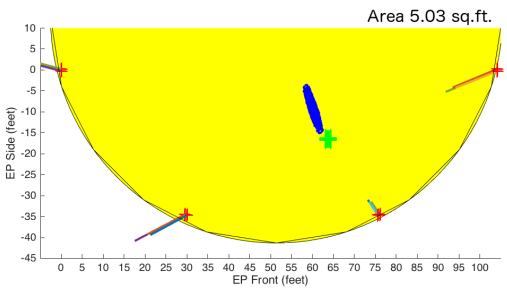
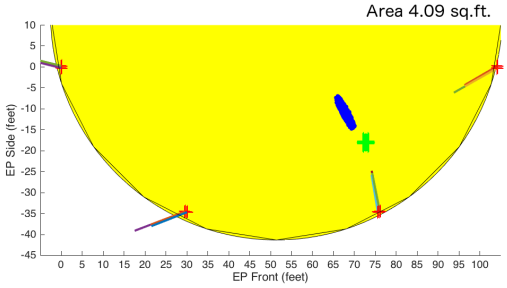
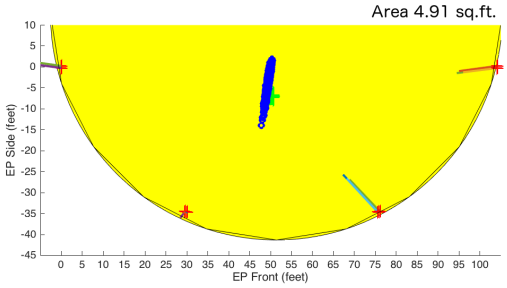


Figure 4.4

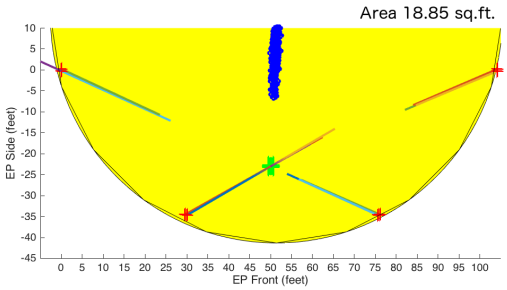
(e)



(f)



(g)



(h)

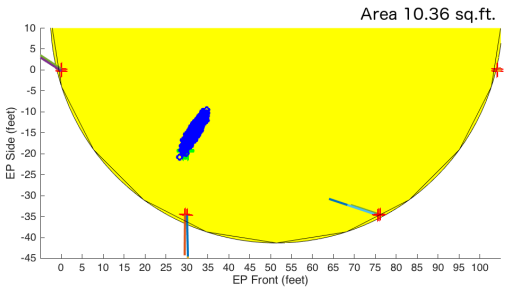
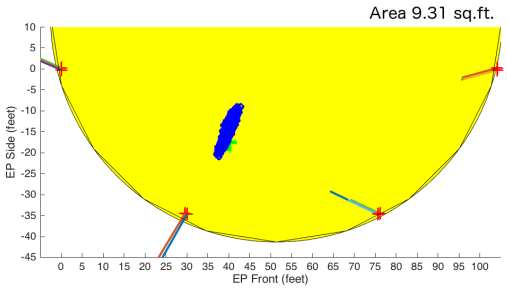
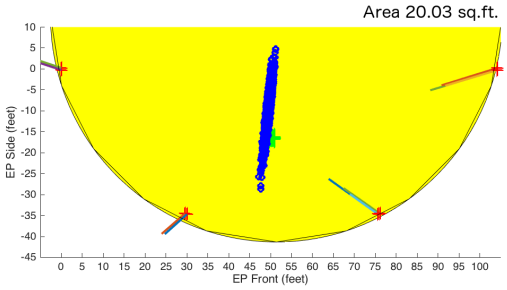


Figure 4.4

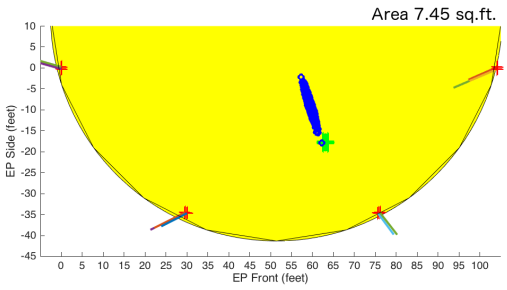
(i)



(j)



(k)



(l)

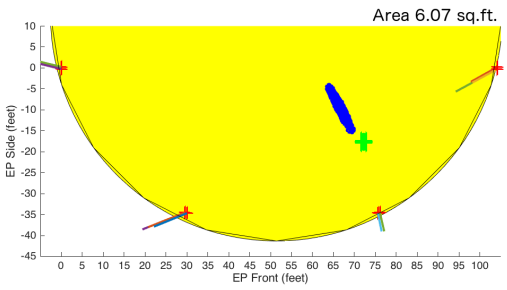


Figure 4.4

(m) **(n)**

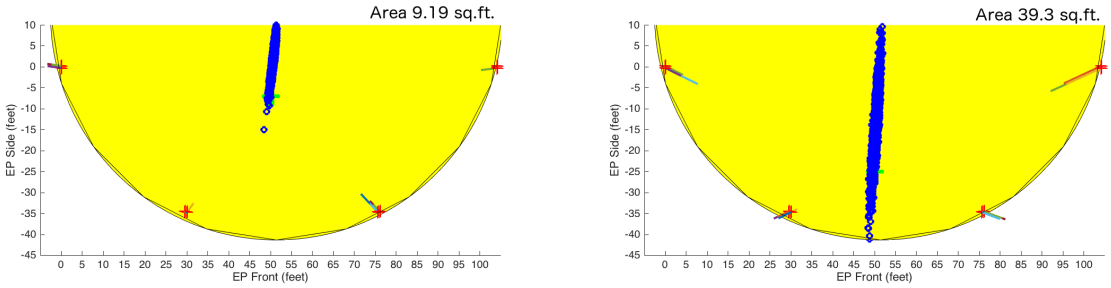

Area 9.19 sq.ft.
EP Side (feet)
EP Front (feet)
Area 39.3 sq.ft.
EP Side (feet)
EP Front (feet)

Figure 4.4

The data indicate that the cloud of localization points consistently occupies less than 1% of pool area (or, equivalently in this case, volume) – note that the plot markers are somewhat exaggerated in size for visibility – and that the true sound source is reliably within 5 feet of it. There is no appreciable overlap of clouds belonging to unique calibration locations in XY except at midline positions, where distinguishing among the calibration points is not realistically feasible. In practice, however, I expect the need to distinguish between two midline sources to be relatively rare; the dolphins in the EP tend to circulate around the pool perimeter. In general, these data seem to suggest that, were dolphins to vocalize signals resembling an impulse – their echolocation clicks might qualify – depending on their alignment it would be possible to distinguish them at a separation of two or three body-lengths. Depending on dolphin number and clustering, this might certainly be adequate to achieve successful sound attribution for most vocalizations. It is important to note that cloud size was manually chosen to minimize the ratio of distance-to-point/area, and that there is room for a more rigorous quantitative optimization. Moreover, with a more substantial set of IRF's it might be possible to develop a correction function that compensates for not only the localization clouds' spread but their offsets from the expected source points.

As an aside, it is obvious that the cloud of localized points is always oriented towards the pool center, which is a result of the spherical interpolation method in combination with our sensor geometry that deserves further investigation. If it were possible to collapse the distribution with modifications in sensor geometry or the algorithm itself, the system's capacity for sound localization might improve drastically.

As a quick check, the estimated time delays of the IRF's were also used to determine whether any of the 16 hydrophones consistently underperforms. For each calibration point, the ideal arrival time differences were calculated (requiring a knowledge of the speed of sound and the location of source and hydrophones), and deviations from the estimated arrival time differences calculated. The mean and standard deviations across all calibration points were calculated for every hydrophone and are plotted in Figure 4.5. No significant difference

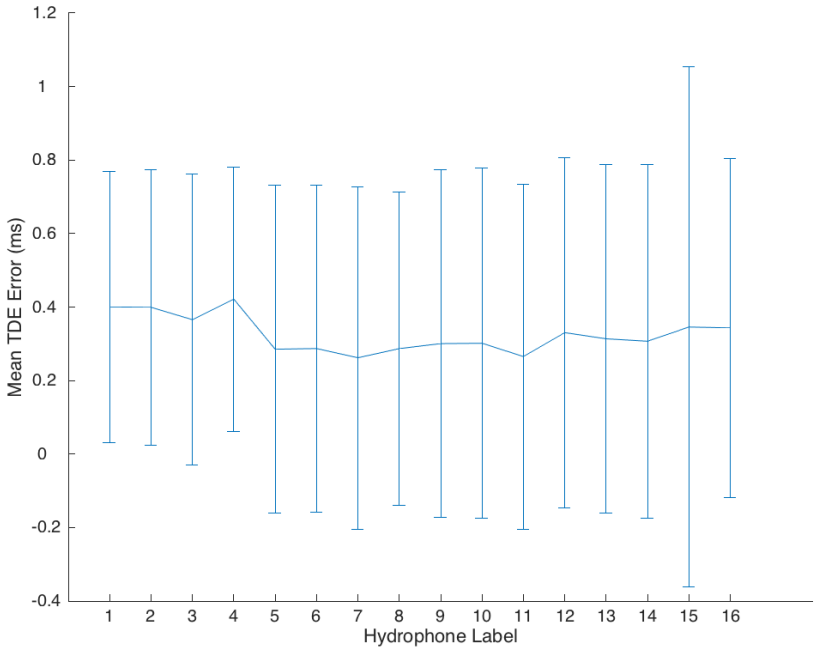


Figure 4.5: Mean Time Delay Estimation (TDE) Error of Hydrophones Plot of mean Time Delay Estimation (TDE) error of hydrophones, calculated from estimated and theoretical time delays for IRF's. Error bars indicate standard deviations.

among hydrophones is visible.

Overall, I take these results as an indication that the system and the various measurements it depends upon to be valid. Whether whistles can be localized as well as IRF's will investigated later in this thesis.

Chapter 5

Time Delay Estimation for Whistles

5.1 Introduction

The method of sound attribution proposed here consists of two parts: sound source localization and matching the sound origin with a potential speaker identified in video feed. Recall that sound source localization ascribes an area in space to a signal of interest based on predictable changes that alter it during its travel from source to sensor location(s). Matching the sound origin with a potential speaker involves a comparison of the sound source coordinates from sound source localization with video-derived coordinates of potential speakers to choose the most probable speaker. Sound source localization for whistles is the focus of this thesis; performing speaker matching efficiently would require robust visual tracking of dolphins across our cameras, which has not yet been developed.

Across Sections 1.5 and 2.2 I broadly reviewed different approaches to sound source localization for whistles and argued for a *splay* configuration of hydrophones around the EP that is suited to a *time-difference-of-arrival* (TDOA) approach to sound source localization, which is based on using the time-differences-of-arrival (or TDOA's) of a signal of interest between pairs of sensors to generate a set of hyperbolae that each contains a potential set of potential sound origin points. I will discuss the problem of reconciling the hyperbolae to

settle on a single source point in Section 6.2. What currently concerns us is a prerequisite to obtaining a source point from the set of hyperbolae: obtaining the differences in arrival times for whistles.

Obtaining good TDOA's is crucial for mounting a successful attempt at geometric sound source localization. The precision necessary is high: consider that it takes a sound wave approximately 0.5 ms to travel one meter in water, which represents a mere 0.1% of an average 0.5-second whistle's length. As a whistle can take as many as 5 ms to reach its maximum energy in a hydrophone, rising gradually above the noise floor, 0.5 ms can be easily lost depending on whistle source intensity, distance of the dolphin from the hydrophone, and directional effects on amplitude from either dolphin or hydrophone orientation. As high precision is required of time delay estimation for many applications, many strategies for time delay estimation do not simply compare the timing of a signal's perceived onset across sensors, but compare the timing of many of a signal's features across sensors: the most-used comparison, the so-called *cross-correlation*, compares entire waveforms between sensors. Unfortunately, even this method is stunted in the current problem due to the influence of *multipath propagation*.

Multipath propagation occurs when a signal travels from source to sensor along multiple independent paths. The most common reason for multipath propagation is the existence of reflective boundaries, such was the various concrete walls and the water-air interface of the EP. With such boundaries present, a signal not only travels from source to sensor directly, but travels from source to sensor along every reflective path that unites the two; if you imagine the sound source as emanating a sphere composed of billiard balls in all directions, you can imagine that some of the balls that initially travel away from a fixed sensor will be redirected towards the sensor after bouncing off an appropriate combination of walls. For a source signal $s(t)$, M hydrophones, N_j multipaths reaching hydrophone j, and path-dependent noise term

$e_j(t)$, we can idealize the multipath signal received in hydrophone j as (Spiesberger, 1998):

$$r_j(t) := \sum_{n=1}^{N_j} a_j[n]s(t - t_j[n]) + e_j(t) \tag{5.1}$$

where $a_j[n]$ is a scalar that modulates the signal amplitude for a given hydrophone and multipath. Even in this idealized, linear representation of the multipath problem, note that a single source signal generates a unique signal in every sensor, one that depends on the locations of the source, sensor, and overall acoustical properties of the environment. As a result, just as we cannot expect to properly estimate hydrophone arrival time differences by comparing perceived onset times, we cannot expect to properly estimate hydrophone time arrival differences by methods like the cross-correlation, which generally expect every sensor to have received an identical signal.

In the upcoming section I will review some strategies I implemented for estimating TDOA's. To assess their performance, I used a set of 60 artificial signals played from a Lubbell LL916 underwater speaker at 13 distinct, known (i.e., with respect to the hydrophone array) locations in the EP of the National Aquarium as depicted in Figure 5.1; these signals were played in the same session with the same apparatus and methods as the pseudo-white noise played to find impulse response functions. With the source locations known, the theoretical delay times can be calculated straightforwardly from the absolute time-of-arrivals, obtained by dividing each source-sensor distance by the speed of sound in water, and compared with observed time-delay-of-arrivals for any pair of hydrophones.

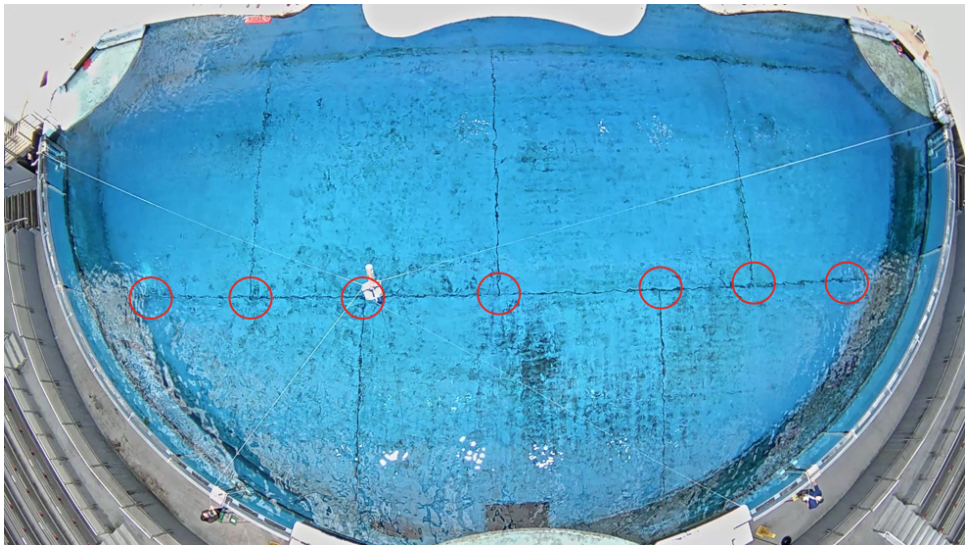


Figure 5.1: Approximate XY Locations of Tonal Sounds for Evaluation of Time Delay Estimation Methods Each red circle denotes two calibration locations, one higher (6 feet deep) and one lower (15 feet deep). Due to time constraints the lower, far left point was not reached.

The shape of the signal as well as the number of times it was played was heavily impacted by husbandry concerns and potential conflicts with another aquarium project (the tonal signal was not to be too "dolphin-like"). The waveform was obtained in Matlab by calculating the instantaneous phase by integration of the desired frequency function, and played through Matlab.

The problem of estimating arrival time differences is an active area of research in signals engineering, and has been for some decades. I was limited in the number of methods I could design and/or implement. Below I review the following methods: a custom-made waveform onset detector, a custom implementation of the cross-correlation maximizer as well as Matlab's internal "finddelay" implementation, a custom implementation of the Generalized Cross Correlation with Phase Transform (GCC-PHAT) maximizer (Matlab's implementation is excluded, having been found deeply ineffective), and a custom-made two-dimensional cross-correlation maximizer of spectrograms. One obvious way to extend the work of this thesis would be to implement more methods.

Every TDOA estimation method that follows was implemented on whistles that were subjected to *bandpass filtering*. Preliminary investigations suggested that bandpass filtering of whistles is essential for obtaining reasonable results with every included method. The bandpass filter, which takes takes an input signal and outputs a signal with all frequency components above some maximum and below some minimum removed, was employed computationally using Matlab on the digital whistle audio files. Note that a bandpass filter that performs the frequency thresholding task perfectly (e.g., that passes all frequency components in the passband without bias, that completely rejects all frequency components outside the passband, and does not suffer distortion at the edges) does not exist, and so different bandpass filter constructions are suited to different problems. For our problem, I emphasized that the filter possess a flat frequency response in the bandpass region. Additionally, I demanded that the filter possess the *zero-phase property*, which eliminates the possibility of *filter delay*, signal shifting in time resulting from filtering, and can be implemented by passing a signal through

a filter in both forward and backward directions.

Additionally, an arrival time difference requires two signals, and I had some choice in this decision. While sometimes the set of time delays for every possible pair of received signals is considered, it can be more manageable to use a single received signal as a universal reference, which I did for the purposes of algorithm comparison – after the best approach is identified I will use all pairs for completeness. Moreover, since I knew the exact waveform of the source signal (if not including the distortion caused by the speaker – this might be fixed in the future), for half of the evaluations I "cheated" and used this ideal signal as a universal reference. Using this ideal signal as a reference in the cross-correlation method of time delay estimation represents an additional named method of time delay estimation, called the *matched filtering* approach.

5.2 Signal Onset Method

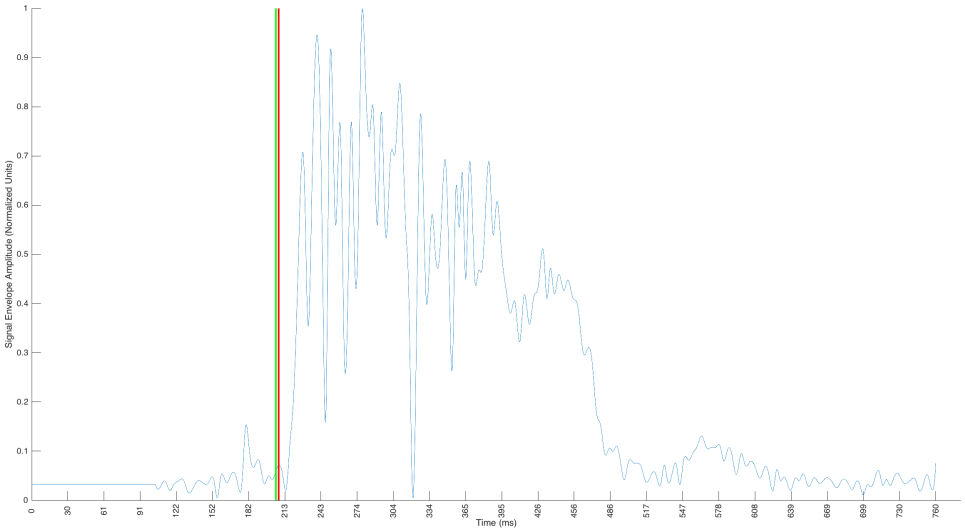
The first method I will explore is perhaps the most naive approach one can take to time delay estimation, involving detecting the signal's onset (or a comparable landmark) in each hydrophone based on the amplitude characteristics of the signal waveform. The advantage of this method is that it avoids the difficulties created by multipath contributions to the signal, as theoretically the first source signal copy in any received signal results from the direct path. Such methods, even simpler than the one proposed, have been employed successfully for the purpose of sound source localization.

Note that it was immediately obvious that this method would only be effective for a tightly band-passed signal, whose waveform contains a significantly reduced number of false onset peaks. The next section, pertaining to the cross-correlation method of time delay estimation, will more explicitly consider the advantage of signal band-passing. Moreover, it was found that this method was most effective operating on a signal's envelope, the signal "magnitude" excluding oscillations resulting from phase, rather than the signal itself. The

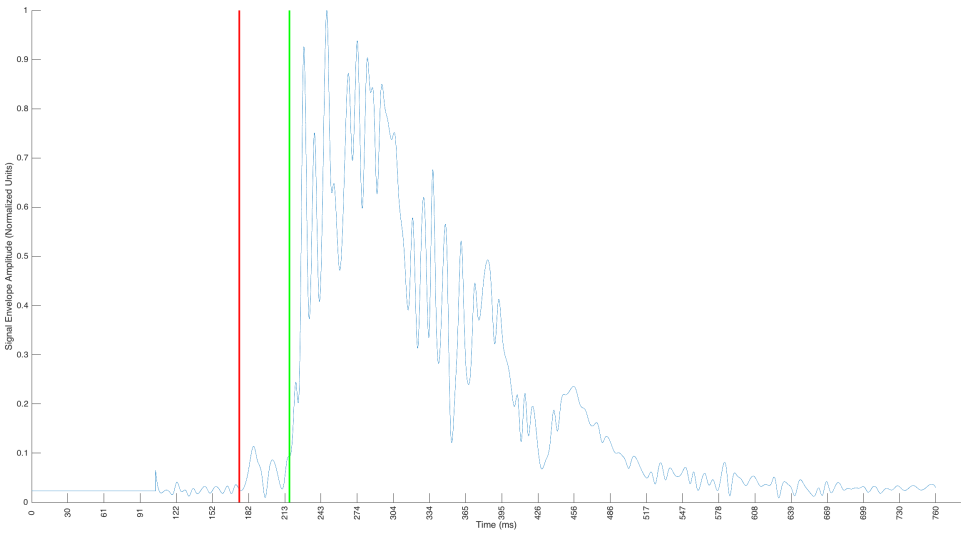
envelope can be obtained by taking the Hilbert transform of the signal, which obtains the so-called *analytic representation* of the signal from which the envelope can be extracted.

In order to better avoid the onset of false signals, the algorithm contains two distinct steps. First, it steps along the signal, comparing the statistics of large windows preceding and succeeding the current step until a peak of approximately the appropriate width and mean level is located. Second, with the current step still ahead of the peak, the algorithm continues stepping, comparing smaller windows until it finds the onset peak, characterized by its steep rise and high maximum.

On many signals that were not explicitly used for crafting it, the algorithm performs reasonably well. However, its success is not guaranteed. Examples of both a successful and unsuccessful detection of onset is shown in Figure 5.2.



(a) A successful detection. The detected onset is detected is marked by a red line. Green is an estimation of the true onset. Separation is 2.4 ms.



(b) An unsuccessful detection. Separation is 42.2 ms.

Figure 5.2: Signal Onset (Landmark) Detection

Table 5.1: Performance of Signal Onset Method in Estimation of Arrival Time Delays

Method	Reference Signal	Bandpass Range (kHz)	Mean of Abs. of Deviations (ms)	Mean of Deviations (ms)	$\surd$ (Mean Square of TMHE) (ms)
Signal Onset	N/A	1.5 - 3.25	10.8 ± 67.9	-1.00 ± 71.6	67.8 ± 53.9
Signal Onset	N/A	1.5 - 5.0	13.6 ± 81.5	-2.00 ± 86.2	77.4 ± 58.3

* TMHE := Truncated Mean Hydrophone Error

Note that we cannot say for certain where the true onset of the signal is (the yellow marker is a relative measure, based on the deviation of the TDOA's derived from the algorithm's found time delays from the geometrically-estimated TDOA's), whether it is the at the base of the ˜.05 peak or the ˜0.5+ peak, however the latter is more commonly seen and is therefore the target of detection. Nevertheless, distinguishing between the two in detection has proven tricky, as different signals contain different mergers of the two peaks (and other "subpeaks", as is evident to the right of the true peak in the lower example). While the current algorithm stands to be refined, not only to improve performance but computational speed, the signal onset variability we observe across many hydrophones and play locations for the *single* test sound suggests that adequately refining the algorithm to be effective on many whistle types might be unrealistic without a quantitative training set suited to machine learning.

In Table 5.1 is a presentation of the algorithm's performance across all whistles. Unfortunately, with an error consistently over 10 ms, equating to a distance of over 15 meters, the algorithm does not perform sufficiently well in its current state.

5.3 Cross-Correlation Method

Imagine that hydrophone i receives a discrete signal in time denoted $r_i[t]$, and that hydrophone j similarly receives a discrete signal in time denoted $r_j[t]$. Assuming the signals

are both real, the cross-correlation between $r_i[t]$ and $r_j[t]$ is defined as:

$$(r_i * r_j)[\tau] := \sum_{t=-\infty}^{\infty} r_i[t] r_j[t+\tau] \tag{5.2}$$

The cross-correlation method of arrival time difference estimation expects that the cross-correlation will reach a maximum when $\tau = t_{delay}$, where t_{delay} is the desired time. Mathematically this is stated as:

$$t_{delay} = \underset{\tau}{\operatorname{argmax}} (r_i * r_j)[\tau] \tag{5.3}$$

This assertion can be made plausible for a simple, deterministic case if we let $r_i[t] = s[t]$ and $r_j[t] = s[t - t_{delay}]$:

$$\sum_{t=-\infty}^{\infty} (r_i[t] \pm r_j[t+\tau])^2 \geq 0 \tag{5.4}$$

$$\sum_{t=-\infty}^{\infty} (r_i[t]^2 + r_j[t+\tau]^2 \pm 2 r_i[t] r_j[t+\tau]) \geq 0 \tag{5.5}$$

$$\sum_{t=-\infty}^{\infty} (r_i[t]^2 + r_j[t+\tau]^2) \geq \mp 2 \sum_{t=-\infty}^{\infty} r_i[t] r_j[t+\tau] \tag{5.6}$$

$$\sum_{t=-\infty}^{\infty} (r_i[t]^2 + r_j[t+\tau]^2) \geq 2 |(r_i * r_j)[\tau]| \tag{5.7}$$

$$\sum_{t=-\infty}^{\infty} (s[t]^2 + s[t+\tau-t_{delay}]^2) \geq 2 |(r_i * r_j)[\tau]| \tag{5.8}$$

$$\sum_{t=-\infty}^{\infty} s[t]^2 + \sum_{t=-\infty}^{\infty} s[t+\tau-t_{delay}]^2 \geq 2 |(r_i * r_j)[\tau]| \tag{5.9}$$

$$2 \sum_{t=-\infty}^{\infty} s[t]^2 \geq 2 |(r_i * r_j)[\tau]| \tag{5.10}$$

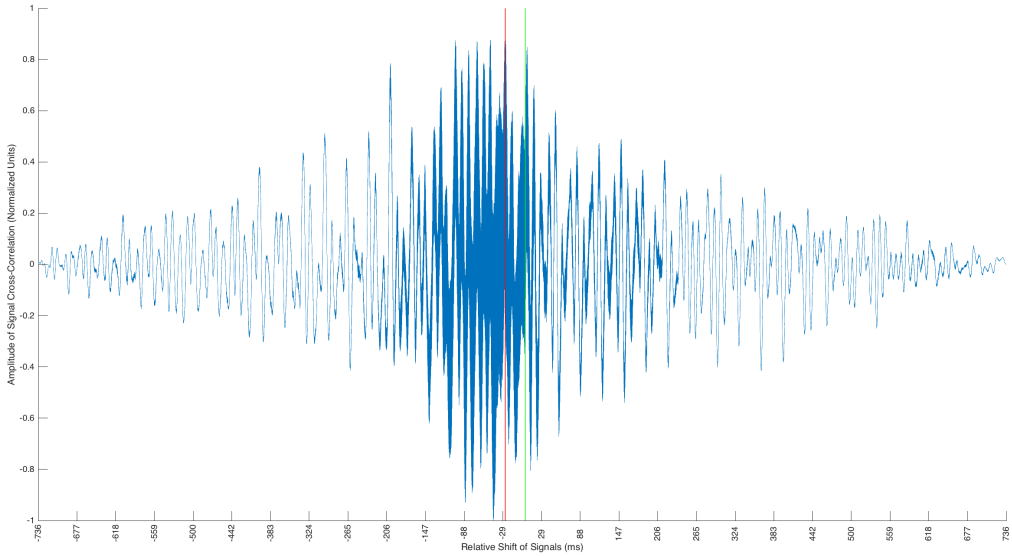
$$\sum_{t=-\infty}^{\infty} s[t] s[t + \overbrace{\tau - t_{delay}}^{\tau = t_{delay}}] \geq |(r_i * r_j)[\tau]| \tag{5.11}$$

$$(r_i * r_j)[\tau = t_{delay}] \geq |(r_i * r_j)[\tau]| \tag{5.12}$$

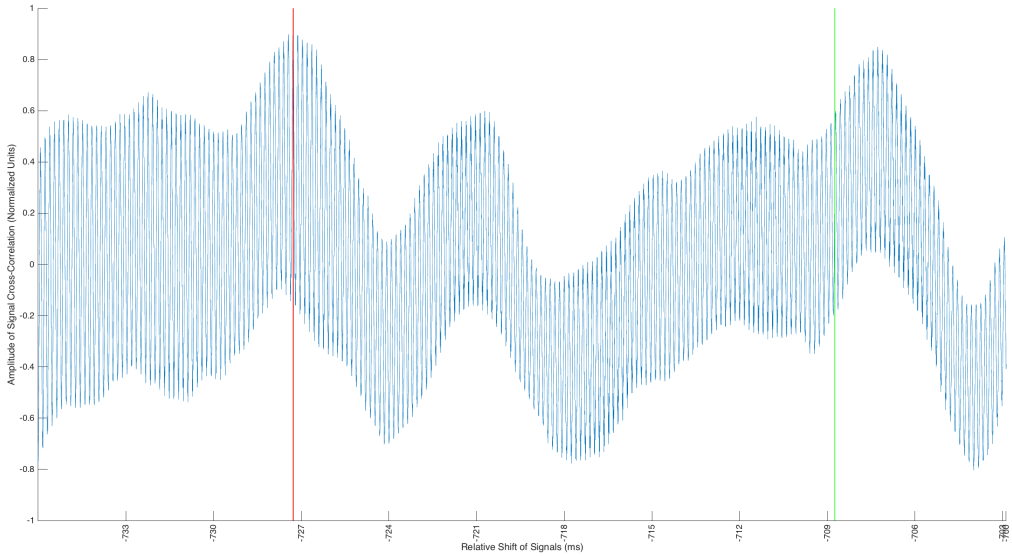
In general, of course, $r_i[t]$ and $r_j[t]$ cannot be assumed to contain exact, shifted copies of $s[t]$ in any problem. Variable signal attenuation and distortion across sensors as well as general linear and nonlinear noise contributions are both standard problem features. Despite the

simplifications of the above derivation, however, the method of maximizing cross-correlations to estimate arrival time differences has been found effective for many problems, and its use is standard in estimating arrival time differences for the purposes of sound source localization.

An additional reason we cannot expect $r_i[t]$ and $r_j[t]$ to contain exact, shifted copies of $s[t]$ in the present problem is what I alluded to with Equation 5.1: multipath contributions, causing the source signal to appear numerous times in the received signals, once for each reflection. To illustrate this, in Figure 5.3 I plot a particularly gruesome example of cross-correlation for unfiltered signals received from hydrophones belonging to two adjacent hydrophone arrays.



(a) Entire cross-correlation. The red line indicates the global maximum, green an estimation of the real shift.



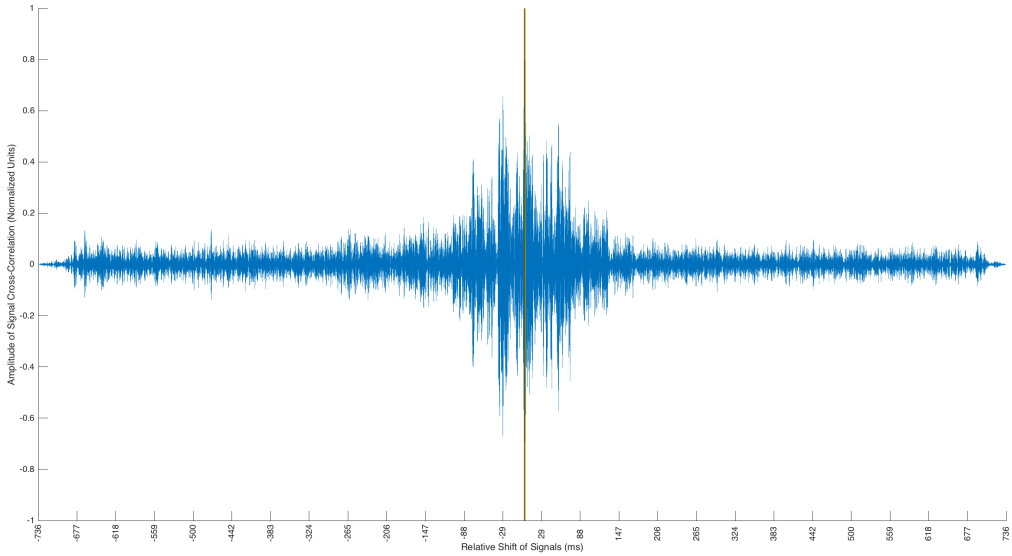
(b) Zoomed version of above.

Figure 5.3: Basic Cross-Correlation of Two Hydrophone Signals

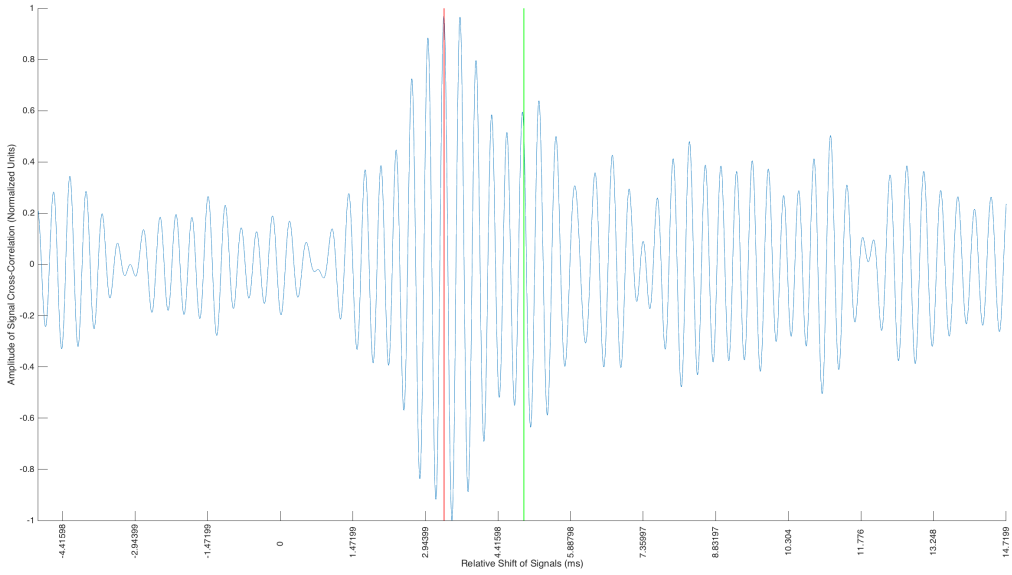
The upper plot shows the entire cross-correlation of the two signals; note that the zero relative shift position is at the center. The presence of many peaks should at once be obvious. Each peak indicates a relative shift where the two signals are temporarily in alignment; the greater the peak, the "better" the alignment. We expect the desired arrival time difference to be indicated by one of these peaks (in fact, based on our treatment above, the largest one), and other peaks to indicate both chance noise alignments as well as alignments of the many source signals from Equation 5.1; if each received signal contains N copies of the source signal, we can expect N^2 such peaks. The oscillatory peaks that occur on a very short time-scale, separated by about 0.5 ms, are a result of the phases of the ˜2-4 kHz source signal itself; these are more visible in the bottom plot.

In the bottom plot, the red line indicates the global maximum of the cross-correlation, and the yellow line indicates approximately where we expect to find this peak based on our knowledge of the source's position and thus the real signal delay between the receivers. We see that these lie on different peaks of approximately the same height, almost 30 milliseconds (corresponding to about 40 meters) away from each other. Fortunately, this error is an extreme case.

Adding a bandpass filter, introduced above, helps the situation significantly – we will take this for granted in the future. Another simple improvement we can make is to *circularize* the cross-correlation, which ensures that the two signals are always overlapping. This prevents the peaks from artificially shortening the farther we get from 0-delay, seen in Figure 5.3. New plots are Figure 5.4.



(a) Entire cross-correlation. The red line indicates the global maximum, yellow an estimation of the real shift.



(b) Zoomed version of above.

Figure 5.4: Basic Cross-Correlation of Two Hydrophone Signals

Now the cross-correlation peaks are clustered in a narrower range around the 0-shift mark; peaks resulting from alignment of noise outside the frequency range of interest have been eliminated, and it is more likely that the remaining peaks are a result of the source signal. Unfortunately, we see that the fundamental problem of many similar-sized peaks at significant separations persists. Nevertheless, in this particular case the error in time estimation has become a significantly improved 1.6 milliseconds (˜2.4 meters).

Another alternative to the basic cross-correlation procedure is found in Matlab's **finddelay** function. The main difference is that it gives higher weight to peaks closer to 0-shift. This was also implemented.

Performing the process outlined above (using either my or Matlab's cross-correlation approach), using either hydrophone one or the source signal as reference, for all known-position calibration signals, we can compare the estimated arrival time delays with the real arrival time delays. Note that when we cross-correlate with the source signal, we are implementing the *matched filter* approach to estimating arrival time difference. Briefly, a matched filter is a linear filter that optimizes the signal-to-noise of an unknown signal relative to a known template signal.

Referring to Table 5.2, we see that all cross-correlation methods estimate arrival time delay with no less error than 5 milliseconds, which corresponds to more than 7.5 meters in water.

Table 5.2: Performance of Cross-Correlation Methods in Estimation of Arrival Time Delays

Method	Reference Signal	Bandpass Range (kHz)	Mean of Abs. of Deviations (ms)	Mean of Deviations (ms)	$\surd$ (Mean Square of TMHE) (ms)
Cross-Correlation	Hydro. 1	1.5 - 3.25	5.90 ± 44.2	-0.670 ± 45.2	48.0 ± 64.0
Cross-Correlation	Hydro. 1	1.5 - 5.0	12.9 ± 22.8	-1.70 ± 27.0	28.5 ± 12.9
Cross-Correlation	Source Signal	1.5 - 3.25	6.5 ± 80.6	0.770 ± 41.2	38.7 ± 48.0
Cross-Correlation	Source Signal	1.5 - 5.0	11.1 ± 11.4	2.6 ± 16.3	23.0 ± 8.44
Matlab "FindDelay"	Hydro. 1	1.5 - 3.25	5.80 ± 48.4	-0.562 ± 49.4	58.3 ± 79.4
Matlab "FindDelay"	Hydro. 1	1.5 - 5.0	12.8 ± 22.8	-1.8 ± 27.0	28.4 ± 13.0
Matlab "FindDelay"	Source Signal	1.5 - 3.25	6.50 ± 40.0	0.797 ± 41.1	38.7 ± 48.0
Matlab "FindDelay"	Source Signal	1.5 - 5.0	11.1 ± 11.4	2.70 ± 16.3	23.1 ± 8.42

* TMHE := Truncated Mean Hydrophone Error

5.4 GCC-PHAT Method

The Generalized Cross Correlation (GCC) method was developed to be an improvement on the standard cross-correlation method for estimating time delays (Knapp and Carter, 1976). It involves including a weighting term in frequency space, or a *processor*, in the standard cross-correlation. The processor can be chosen in any number of ways to highlight desired properties of the input signals. As this weighting term is typically described in frequency space, to express the modification I start by writing the definition of the cross-correlation, Equation 5.15, for the signals as functions of frequency rather than time (where again the Fourier transform into frequency space is denoted by $\mathscr{F}$). Incidentally, implementing the standard cross-correlation in frequency space is one way of performing the *circular cross-correlation* mentioned last section.

$$(r_i * r_j)[\tau] := \mathscr{F}^{-1}\{\mathscr{F}\{r_i[t]\}^* \cdot \mathscr{F}\{r_j[t]\}\} \tag{5.13}$$

where * indicates the complex conjugate, and we remember $r_i(t)$ and $r_j(t)$ are signals received by hydrophones i and j.

We define the GCC for $r_i(t)$ and $r_j(t)$ as follows:

$$GCC_{r_i,r_j}[\tau] := \mathscr{F}^{-1}\{\Psi[f] \cdot \mathscr{F}\{r_i[t]\}^* \cdot \mathscr{F}\{r_j[t]\}\} \tag{5.14}$$

where $\Psi[f]$ is the processor or weighting function of interest.

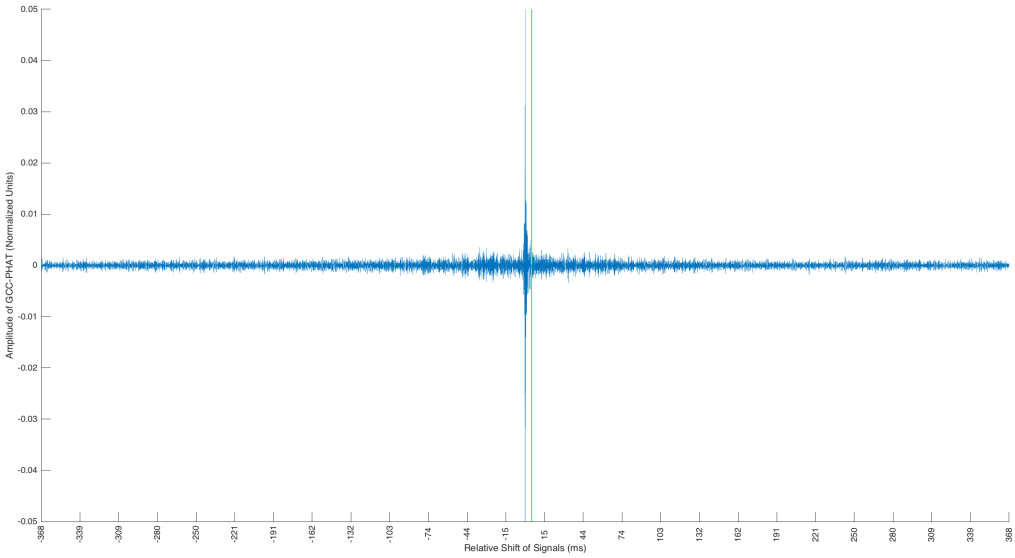
The Generalized Cross Correlation with Phase Transform (GCC-PHAT) method gives a particular form to the processor:

$$\Psi[f] := \frac{1}{|\mathscr{F}\{r_i[t]\}^* \cdot \mathscr{F}\{r_j[t]\}|} \tag{5.15}$$

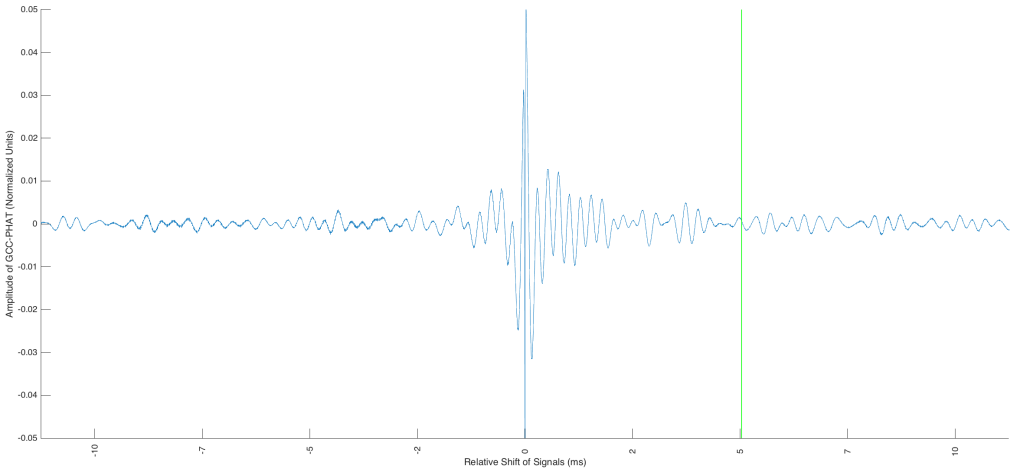
The effect of this processor is to equalize the gains of all cross-frequency bands in order to emphasize the phase information of the signals, which contains information about the signals' relative delay. In theory, the desired delay will appear in the GCC-PHAT series

as a sharp, narrow peak, and literature suggests that the first incident peak appears more robustly than the secondary peaks characteristic for signals from reverberant environments (Van Den Broeck et al., 2013).

For the same filtered hydrophone signals used in Figure 5.3 for an example of the standard cross-correlation method, an example of the GCC-PHAT method is plotted in Figure 5.5. Now there is only one prominent peak, rising to an amplitude of 1.0, and numerous small peaks magnitudes smaller, between 0.005 and 0.01. In this case, the presence of a single prominent peak is deceptive, however: the predicted signal delay is 7.9 milliseconds (˜12 meters) away, near one of the smaller peaks. In this case, GCC-PHAT performed worse than the standard cross-correlation method.



(a) Entire GCC-PHAT for same signals as Figure 5.3. The blue line, part of the trace, indicates the global maximum and predicted shfit; it reaches an amplitude of 1.0. The yellow line is an estimation of the real shift.



(b) Zoomed version of above.

Figure 5.5: GCC-PHAT of Two Filtered Hydrophone Signals

Table 5.3: Performance of GCC-PHAT Methods in Estimation of Arrival Time Delays

Method	Reference Signal	Bandpass Range (kHz)	Mean of Abs. of Deviations (ms)	Mean of Deviations (ms)	$\sqrt{}$ (Mean Square of TMHE) (ms)
GCC-PHAT	Hydro. 1	1.5 - 3.25	3.10 ± 2.20	-0.273 ± 3.90	6.48 ± 2.29
GCC-PHAT	Hydro. 1	1.5 - 5.0	3.10 ± 2.20	-0.273 ± 3.90	6.48 ± 2.29
GCC-PHAT	Source Signal	1.5 - 3.25	239 ± 354	-143 ± 422	577 ± 289
GCC-PHAT	Source Signal	1.5 - 5.0	227 ± 344	143 ± 405	568 ± 304

* TMHE := Truncated Mean Hydrophone Error

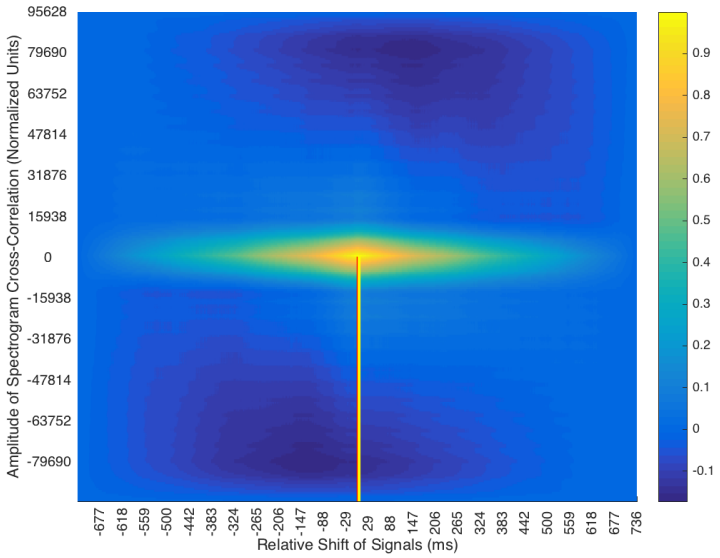
The results of the GCC-PHAT method for all signals are in Table 5.3. The GCC-PHAT method using the first hydrophone as reference is the best-performing method thus far, with approximately 3.1 millisecond error (˜4.6 meters). However, this will still prove insufficient for the purposes of geometric sound source localization.

5.5 Spectrographic Cross-Correlation Method

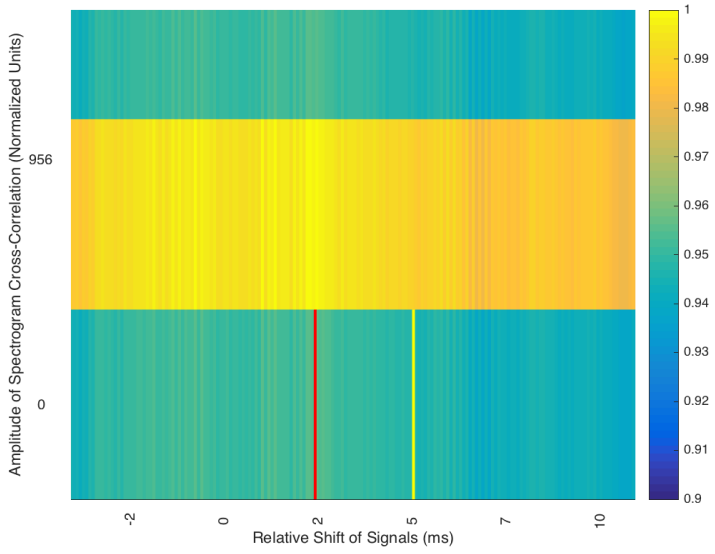
One method that follows from the standard cross-correlation peak-finding method for two signals is a two-dimensional cross-correlation peak finding-method performed on the signals' spectrograms. While not commonly used, this method has been used for time-delay estimation of bowhead whale calls (Mellinger and Clark, 2000). There are clear disadvantages to the method: phase information is lost in the construction of the spectrogram, additionally time is coarse-grained. Nevertheless, the method is simple to implement, and it is conceivable that the method's modest insensitivity to shifts in frequency would confer it some advantage.

The largest window deemed reasonable for estimating time delay was 64 samples, corresponding to 1/3 milliseconds or ˜0.5 meters. A window size of 32 samples was also evaluated.

A plot of the spectrographic cross-correlation for the same two signals used in the last two sections is shown Figure 5.6. The error is greater than 5 milliseconds or ˜7.5 meters. The performance of the method across all samples is tabulated in Table 5.4, and suggests that this method is no more reliable than those already considered.



(a) 2D cross-correlation of spectrograms belonging to the same signals used for Figure 5.3. The heat map is in normalized units of amplitude cross-correlation. The red line indicates the global maximum and predicted shift; the yellow line is an estimation of the real shift. The error is 2.86 milliseconds (˜4.3 meters)



(b) Zoomed version of above.

Figure 5.6: 2D Cross-Correlation of Two Signals' 32-Sample Spectrograms

Table 5.4: Performance of Spectrographic Cross-Correlation Methods in Estimation of Time Delays

Method	Reference Signal	Bandpass Range (kHz)	Mean of Abs. of Deviations (ms)	Mean of Deviations (ms)	$\surd$ (Mean Square of TMHE) (ms)
32-Sample Spec. Cross-Correlation	Hydro. 1	1.5 - 3.25	4.90 ± 28.7	-0.648 ± 29.4	17.2 ± 18.6
32-Sample Spec. Cross-Correlation	Hydro. 1	1.5 - 5.0	5.90 ± 6.10	-0.939 ± 8.8	13.2 ± 5.86
32-Sample Spec. Cross-Correlation	Source Signal	1.5 - 3.25	10.7 ± 65.5	-1.10 ± 68.4	64.0 ± 64.8
32-Sample Spec. Cross-Correlation	Source Signal	1.5 - 5.0	22.5 ± 45.3	8.70 ± 51.6	54.8 ± 37.4
64-Sample Spec. Cross-Correlation	Hydro. 1	1.5 - 3.25	5.10 ± 29.2	-0.383 ± 30.0	17.7 ± 18.8
64-Sample Spec. Cross-Correlation	Hydro. 1	1.5 - 5.0	5.80 ± 6.80	-1.00 ± 9.40	13.5 ± 6.78
64-Sample Spec. Cross-Correlation	Source Signal	1.5 - 3.25	12.5 ± 59.9	-1.50 ± 63.3	70.7 ± 72.8
64-Sample Spec. Cross-Correlation	Source Signal	1.5 - 5.0	29.7 ± 51.5	11.0 ± 60.7	73.5 ± 52.9

* TMHE := Truncated Mean Hydrophone Error

* TMSE := Truncated Mean Sample Error

Chapter 6

Sound Localization

6.1 Introduction

After obtaining estimations of the time-differences-of-arrival (TDOA's) of a whistle for the system's sixteen hydrophones, the subject of the previous chapter, sound source localization can be performed for the ultimate purpose of sound attribution. The standard method of obtaining an origin point with TDOA's will be presented in the next section.

However, though this method was successfully applied in Section 4.4 to localize the impulse response functions obtained from the pseudo-white noise calibration data to demonstrate the system's basic functionality, a rigorous analysis of the TDOA method's performance on tonal sounds and/or real whistles will not be included here: it was quickly realized that the errors associated with the estimated TDOA's for these sounds prevented me from obtaining even approximate whistle source locations, and plots analogous to the series in Figure 4.4 were found not to be meaningful.

Considering the high degree of failure of this first TDOA-based localization method, rooted in estimations of the TDOA's known to be inaccurate, I did not think it was optimally productive to implement additional methods of explicitly TDOA-based sound source localization. Instead, I decided to pursue a data-driven approach to sound source localization

by building a dataset of recordings of tonal sounds played in the aquarium EP at known locations, and by using machine prediction to distinguish among sounds played at the different locations. The same machine prediction scheme thus built might also be effective for dolphin whistles recorded from arbitrary locations in the pool. The details of this approach occupy the majority of this chapter.

6.2 Spherical Interpolation

Spherical interpolation is arguably the standard method for solving for a sound source position from a set of estimated TDOA's, beamforming methods aside. The proper derivation begins with explicitly introducing and subsequently minimizing the error in a geometric source range estimation based on of the estimated TDOA's. The optimization problem thus obtained is a difficult nonlinear, non-convex minimization problem in the unknown source position. While brute-force iterative techniques exist for solving it, they are in general computationally expensive, require a good estimation of the source position, and in general cannot guarantee convergence (Li et al., 2016). One such method was implemented for the current work with less success than the spherical interpolation method (Li et al., 2014). Seemingly more common approaches to the minimization problem, including the spherical interpolation method, are based on eliminating the problem's nonlinear dependence on the unknown source position (Li et al., 2016; Smith and Abel, 1987b; Spiesberger, 1999). The spherical interpolation method is representative of these methods and has been suggested to be the most robust to noise, and can be proven optimal (i.e., to represent a *maximum likelihood estimation*) given that the estimated error in the TDOA's is Gaussian (Li et al., 2016; Smith and Abel, 1987a,b). Below I present a simplified derivation of the spherical interpolation technique adapted from Zimmer (2011).

The problem is phrased as it was in Section 2.2, where optimal sensor placement was discussed. Given that a sound is generated at time t_0 from the source point $\boldsymbol{s} := (x_s, y_s, z_s) \in$

$\mathbb{R}^3$ and is detected by M hydrophones $\boldsymbol{h} := [\boldsymbol{h_1}, \boldsymbol{h_2}, ..., \boldsymbol{h_M} = (x_M, y_M, z_M)] \in \mathbb{R}^{3\times M}$, we are interested in recovering $\boldsymbol{s}$. While t_0 is unknown, we have knowledge of the *time-difference-of-arrivals* (TDOA's) between $\boldsymbol{h_i}$ and $\boldsymbol{h_j}$, denoted T_{ij}.

Again we consider the distance or *range* of the source from each hydrophone, $d_i := ||\boldsymbol{h_i} - \boldsymbol{s}||$, which can be written as the M equations:

$$d_i{}^2 = (x_i - x_s)^2 + (y_i - y_s)^2 + (z_i - z_s)^2 \tag{6.1}$$

Let $\boldsymbol{h_1}$ be defined as the *reference* hydrophone. Noting that range is a one-dimensional measure of radial distance from the source, we can write the range of all hydrophones in terms of d_1 and the range differences, $d_{1i} := d_i - d_1$:

$$d_i = d_1 + d_{1i} \tag{6.2}$$

The motivation for this is that, while the d_i's are unknown, the d_{1i}'s can be written in terms of the TDOA's:

$$d_{1i} = cT_{1i} \tag{6.3}$$

where c is speed of sound in water.

We can rewrite Equation 6.1 in terms of Equation 6.2 for the reference hydrophone and the other hydrophones:

$$d_1{}^2 = x_1^2 - 2x_1x_s + x_s^2 + y_1^2 - 2y_1y_s + y_s^2 + z_1^2 - 2z_1z_s + z_s^2 \tag{6.4}$$

$$d_1{}^2 + 2d_1d_{1i} + d_{1i}{}^2 = x_i^2 - 2x_ix_s + x_s^2 + y_i^2 - 2y_iy_s + y_s^2 + z_i^2 - 2z_iz_s + z_s^2 \tag{6.5}$$

where $i \forall [2, 3, ..., M]$.

Equation 6.5 constitutes a set of nonlinear, hyperbolic equations in the four unknowns, $d_1{}^2$, $x_s{}^2$, $y_s{}^2$, and $z_s{}^2$. There is more than one approach to solving them. The *spherical interpolation* approach is characterized by the next step, which eliminates quadratics in the unknowns. We subtract Equation 6.4 from Equation 6.5 to obtain:

$$2d_1d_{1i} + d_{1i}{}^2 = x_i^2 - x_1^2 - 2(x_i - x_1)x_s + y_i^2 - y_1^2 - 2(y_i - y_1)y_s + z_i^2 - z_1^2 - 2(z_i - z_1)z_s \tag{6.6}$$

where $i\forall[2,3,...,M]$.

We can write Equation 6.6 in the matrix form $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$, where

$$\boldsymbol{b} = \begin{bmatrix} (x_2^2 - x_1^2) + (y_2^2 - y_1^2) + (z_2^2 - z_1^2) - {d_{1,2}}^2 \\ (x_3^2 - x_1^2) + (y_3^2 - y_1^2) + (z_3^2 - z_1^2) - {d_{1,3}}^2 \\ \vdots \\ (x_M^2 - x_1^2) + (y_M^2 - y_1^2) + (z_M^2 - z_1^2) - {d_{1,M}}^2 \end{bmatrix} \tag{6.7}$$

$$\boldsymbol{A} = 2 \begin{bmatrix} d_{1,2} & (x_2 - x_1) & (y_2 - y_1) & (z_2 - z_1) \\ d_{1,3} & (x_3 - x_1) & (y_3 - y_1) & (z_3 - z_1) \\ \vdots & \vdots & \vdots & \vdots \\ d_{1,M} & (x_M - x_1) & (y_M - y_1) & (z_M - z_1) \end{bmatrix} \tag{6.8}$$

$$\boldsymbol{x} = \begin{bmatrix} d_1 \\ x_s \\ y_s \\ z_s \end{bmatrix} \tag{6.9}$$

For five hydrophones ($M = 5$) this equation can be solved exactly as simply $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, and for more hydrophones the standard least-mean-square solution applies:

$$\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} \tag{6.10}$$

Note that $\boldsymbol{A}$ must be invertible and its determinant never zero, meaning that the hydrophones should not share one coordinate component (e.g., be at the same depth), a condition that our own sensors narrowly avoid.

As an aside, note that the speed of sound in water, c, is is a function of physical-chemical properties of the water. Various equations for calculating the speed of sound in water have been proposed based on various data and constraints: some are based on precise data from water held in laboratory chambers, some on oceanographic data, some have been simplified for field computation, some are complex, and some are based on more dubious assumptions (e.g.,

regarding the conversion between pressure and depth, or for seawater, on suspect calculations of the speed of sound in low-saline water) than others. The physical-chemical variables common to them all are salinity, temperature, and pressure (Leroy et al., 2008).

To my knowledge, no equation has been developed specifically to calculate the speed of sound in saline aquaria. As it seems that the validity of the different equations vary considerably with water salinity, I chose to use an equation developed to calculate the speed of sound in oceans, which have salinities in a range of ˜30-40 ppt, consistent with the mean salinity of the the National Aquarium's EP, 31.50 ppt (Leroy et al., 2008). The one chosen, the Del Grosso equation, is one of the "complicated" variety of equations – it is a large polynomial of temperature, salinity and pressure and will not be stated here – and is reported by various sources to be the most accurate (Del Grosso, 1974; Spiesberger and Fristrup, 1990; Spiesberger and Metzger, 1991). For the EP mean salinity of 31.50 ppt and mean temperature of 26.04 °C (numbers provided by National Aquarium life support personnel), and a mid-pool pressure calculated to be 30.23 kPa (excluding atmospheric pressure), the speed of sound of 5030 ft/s has been calculated (Leroy and Parthiot, 1998).

Note that a 1.0 °C change in temperature corresponds to a roughly 2.5 ft/s change in the speed of sound and that a 30 kPa change in pressure (corresponding to moving to the surface or to the pool floor) corresponds to a 0.050 ft/s change in the speed of sound. Over one hundred feet (approximately the maximum hydrophone separation) a 2.5 ft/s change in the speed of sound can alter a TDOA by up to 10.0×10^{-3} milliseconds, which is significantly less than the currently observed error in TDOA estimation, over 5 milliseconds. However, there is room for further investigation of the amount of temperature fluctuation in the pool and its effect on the speed of sound.

It is also reasonable to ask whether dispersion – the phenomenon of different frequencies of sound traveling at different velocities – is significant in the aquarium. As the effects of dispersion are correlated with the distance a signal travels, along with the effects of amplitude attenuation and multipath propagation it could cause different hydrophones

to receive different versions of the same whistle, possibly affecting the performance of TDOA estimators. Moreover, along with amplitude attenuation and multipath propagation, dispersion could represent a source of information apart from TDOA's that could be used for range estimation. However, none of the authors of the speed of sound equations considered have taken dispersion into account, though they sometimes address it. Even over large distances (e.g., miles), dispersion has been shown to affect signals at around the same level as sensor error (Dushaw et al., 1993; Horton Sr., 1974). One study found that between 500 and 1500 Hertz, dispersion effects a difference in velocity on the order of one part in 100,000, which is comfortably insignificant (McCubbin Jr., 1954). Thus, for our purposes dispersion is not taken into account.

6.3 Prediction of Tonal Sound Locations: Building a Training Set

My proposal to perform sound source localization via machine learning classification involved two major steps: creating a data set on which to train the classifier, and building/evaluating the classifier itself. The purpose of the training set is to provide the classifier with an exhaustive supply of "examples" of the problem to be solved, which in this case is the association of 16 hydrophone signal samples of a dolphin whistle with its 3-dimensional point of origin in the EP. The ideal training set would consist of thousands (or more!) of different, real dolphin whistles – representative of the ones we intend to localize – each recorded from all points on a measured, rectangular, three-dimensional grid in the pool with inter-point spacing no greater than distance we might expect our hydrophones able to resolve (for a 192 kHz sampling rate, perhaps a hundredth of a meter). Of course, building such a training set would be impossible. My goal in constructing the actual training set was to come as close to this ideal as possible given the constraints.

First, I could not use real whistles from the dolphins themselves. An automated system for

extracting and matching real dolphin whistles with the dolphins' coordinates was underway but not yet adequate, and, even if it were, the dolphins' natural swimming and vocalization habits did not guarantee a readily accessible, geometrically distributed set of whistles like what was described above – months if not years of recordings might be necessary. Thus, I would need to deploy a speaker and play sounds using methodology previously described in Section 4.2.

Second, I was not allowed to play the resident dolphins' whistles back to them, or whistles too closely resembling real bottlenose dolphin whistles. This meant that I needed to play artificial tonal sounds that I felt somehow representative of the pools' bottlenose dolphins repertoires.

Third, and perhaps most significantly, I was severely constrained in the number of sounds I could play, limiting both the number (and thus the variety) of sounds played at a single location and the total number of locations at which sounds were played. The reasons for this included the aquarium's strict limit of 1.5-hours/day of experimentation time in the EP, and the time constraints of the 4-person research team that assisted me along with the NA staff. Obtaining the data was a complex, multi-person operation.

Based on the above, I settled to play 127 sounds at each of 14 locations inside the EP, which, with setup time, I calculated to fill the 3 hours of time allotted to us split between two consecutive days. Note that, while the original intention was to play sounds on a dense grid of points spanning the EP to effectively perform true 3-dimensional sound localization, the reduction of sampling points to a mere 14 locations meant that this would not be possible, and that even a successful classifier might only be able to distinguish among real dolphin whistles emitted at these particular places; the extent generalizability of the classifier to elsewhere in the pool became another unknown. Note that, under these circumstances, it might seem reasonable to re-phrase the problem in terms of regression rather classification, where 3D position was to be predicted rather than one of a non-ordinal set of sample points. However, given the complexity of reverberative effects and my inability to play real dolphin

Table 6.1: Parameter Set of Training Set Sinusoids

Parameter	Value Set
Duration (sec)	[0.3, 1]
Number of Cycles	[1, 2]
Center Frequency (Hz)	[6000, 10500]
Cycle Amplitude (Hz)	[2000, 5000]
Phase (rad)	$[-\frac{\pi}{2}, \frac{\pi}{2}]$
Power Onset/Decay Rate *	[0.1, 0.25]

* Values indicate fraction of signal length over which a sin^2 rise/falls occurs.

whistles, I decided to emphasize predictor generalizability in whistle space rather than real space in the training set, and not surprisingly it would become clear that 14 distinct spatial points were insufficient for regression. From this point forward the classification project necessarily became more a proof of concept than a practical attempt at achieving good sound localization.

The sounds used were sinusoids in time-frequency space, resembling whistles from the dolphin species *Stenella frontalis*, whose whistles were deemed permissible to play, as well as sounds from an "off-sinusoid" family. The latter family is characterized by expressions of the form $\frac{arcsin(m \cdot sin(2\pi f))}{arcsin(2\pi f)}$ that, for an appropriate choice of frequencies, take the appearance of stacked, rounded triangular traces in time-frequency space, adjusted by the parameter m. Apart from resembling *S. frontalis* whistles, another reason to play sinusoid sounds was that it was straightforward to systematically explore the typical parameter space that researchers use to characterize dolphin whistles. The parameter space used is described in Table 6.1.

The 14 chosen locations in the EP were the same for which impulse response functions were found earlier, on a 3 x 5 x 2 cross in the pool shown in Figure 4.1a. Sounds were played at a regular 4-second interval, set precisely to the computer system clock, for ease of post-hoc extraction. However, lag was created somewhere in the computer-audio-interface-speaker

pipeline. As preliminary tests suggested this might occur, all signals were prepended with a flat 3 kHz tag, whose onsets I tracked across all extracted sounds; these onsets were fit to a straight line whose coefficients I used to compensate for this drift.

As an aside, it is reasonable to ask whether actual whistle (or whistle-like) training sounds are necessary for building a successful classifier *for* whistles. More specifically, perhaps the training sounds could be samples from a "whistle eigenspace." Such a space might be constructed, for instance, by performing *principal component analysis* (PCA) on a set of whistles. However, if real whistles (in time space) were used for PCA, they would first need to be be made of equal length, using a transformation process (such as *dynamic time warping*) that could not guarantee the conservation of relevant signal features. If a set of artificial whistles constructed to be of identical length were used for PCA, it would remain unclear whether their set of eigenvectors, nonlinearly transformed by their transit through the speaker-water-hydrophone system, would be representative of the original whistles transformed by the same transit (so that we could still claim to have a whistle eigenspace). Performing a PCA of whistles in frequency space is another possibility that could be examined, but would certainly bring its own uncertainties. Ultimately, given the time constraints, the conservative method of using a training set composed of whistle-like sounds seemed best.

6.4 Classifying Tonal Sound Locations: Feature Selection

Before continuing, I will define classification and classifier training more clearly. Broadly, classification is the process of predicting some unknown, discrete, non-ordinal characteristic of a data sample (such as "season," "car manufacturer," or "is-yellow"), called the *label* or *class*, from an array of other, known characteristics of that same data sample – which in general may be ordinal or not, discrete or continuous – called the *features*. The classifier is the mathematical-statistical tool that performs the mapping from the feature array to the

class for a particular data sample. Many forms of classifier exist, having different strengths and weaknesses. Constructing the classifier, or training it, requires a large set of data samples for which both the features and classes are known; this set is called the *training set*. The success of a classifier can be evaluated by using it to predict the classes of a set of data samples for which both the features and labels are known, but which were not part of the training set; this is the *test set*.

In general, the two most significant decisions that must be made during classifier construction are the selection of the form of the classifier and the selection of the features that it relies on for prediction – these are both topics of immense interest within the machine learning field. The second decision is the concern of this section. A good set of initial features is computationally tractable and includes all the information that the classifier needs to perform prediction successfully, contained no more implicitly than the classifier is capable of using.

For the current problem, the feature set must be drawn from 16 hydrophone recordings of a whistle-like sound. Naively, one might suggest that the feature set simply include all values of the signal waveforms. However, this suggestion can be quickly dismissed by noting that this would require that the feature set contain a computationally impractical $192000 * 16 * 2$ values – downsampling the signal from 192 kHz is not an option given the small size of the expected time delays – and that across data samples the signals would likely need to be temporally aligned in some precise and meaningful way.

I decided on the following feature set. First, I included the TDOA's from the best-performing method from Chapter 5. While these TDOA's were not sufficiently accurate to be used for localization by spherical interpolation, they still potentially included information valuable to and not otherwise accessible to a classifier. Next, I included the normalized cross-correlation series between all pairs of hydrophone signals. Apart from including information that was potentially unavailable to the classifier otherwise, the advantages of using the cross-correlations over the raw signals was two-fold: first, the cross-correlation series are perfectly temporally aligned across all data samples, and second, while keeping the hydrophones'

192 kHz sampling rate I could make the series manageable in length by excluding all cross-correlation terms that corresponded to time-shifts between hydrophones that are larger than the largest possible signal delay. Lastly, I included the discrete Fourier transforms of all hydrophone signals, again truncated to exclude frequencies outside the range of interest. These series were again insensitive to time shifts among samples, and potentially reflected frequency-dependent dispersion and absorption information about the signals' paths.

The total feature set consisted of 1,331,887 one-dimensional numerical features, across 1,605 samples in the training set and 178 (a random 10% of the total) in the test set, which would only be used for classifier evaluation. At this size the feature set was already computationally cumbersome. In fact, preliminary investigations using basic decision trees (to be discussed), one of the only classifier types capable of handling the data at this stage, suggested that the discrete Fourier transforms of the signals did not affect classification performance, and I took the opportunity to discard them, reducing the feature set to 897,891 numerical features. The remaining feature set was still computationally troublesome but manageable.

6.5 Classifying Tonal Sound Locations: Random Forest Classification

At this point, I began training a *random forest* classifier, which is an ensemble of *decision tree* classifiers. They will be described in turn.

A decision tree is a powerful type of nonlinear, multi-class classifier that is constructed by applying simple construction rules to a training set and that produces classifications based on a transparent set of decisions; in many ways it is more suited to a rigorous understanding of data than more opaque classifiers that have recently enjoyed celebrity for their raw power, namely those based on neural networks. The first algorithm for Classification and Regression Trees (CART) was introduced by Leo Breiman et al. in 1984 (Breiman et al., 1984).

A generic, binary decision tree algorithm proceeds as follows. We start with the full training set at the top of the tree, the so called *root node*. We seek to split the samples in the training set between two lower *child nodes*. We perform the split by selecting a separation point in a single feature. For instance, if the TDOA between hydrophone i and j is the chosen feature, we might split all samples based on whether this value is above or below 1 millisecond – splits in ordered features such as a TDOA, or any of the features in my data, will always take the form of an inequality. The feature to split on and the location of the split are not chosen arbitrarily. We seek the break point in the feature that will maximize the difference in class distribution, or *inhomogeneity*, between samples in the two child nodes. There is not a single way to define the inhomogeneity, and the choice of this definition is one way in which decision tree algorithms can differ. The definition used in this thesis and that will be stated here is based on the *Gini Diversity index*. Let S denote a set of training samples where each sample $s \in S$ possesses a class $j \in 1, 2, ..., J$. If $p_i(S)$ denotes the proportion of samples in S belonging to class j (alternatively, the probability of a random sample possessing class j), the Gini Diversity of S is defined as:

$$G(S) := \sum_{j=1}^{J} p_i(S)(1 - p_i(S)) = 1 - \sum_{j=1}^{J} p_i(S)^2 \tag{6.11}$$

In general, $G(S)$ will be larger the less homogeneous the set S is. Let us split set S into sets $A(c)$ and $B(c)$ according to some splitting rule (recall: an inequality on a single feature) $c \in C$, where C is the set of all possible splits. We define the ideal splitting rule c_{best} as follows:

$$c_{best} := \underset{c \in C}{\operatorname{argmax}} \{G(S) - \frac{|A(c)|}{|S|} G(A(c)) - \frac{|B(c)|}{|S|} G(B(c))\} \tag{6.12}$$

where $|A|$, $|B|$, and $|S|$ are the number of samples in sets A, B, and S. In essence, the best split is that which produces two child sample sets A and B that jointly possess as small values for Gini Diversity as possible, and therefore as much inhomogeneity (as it is here defined) as possible.

After the best split is chosen at the parent node, producing two child nodes (connected by

branches to the parent node), the above split process is repeated at the two child nodes. This continues at each successive child until some termination condition is met, naively that all samples in a child node are of the same class. A node without children is called a *leaf*. Note that it is not necessarily optimal to continue the splitting process so that all leaves contain samples from a single class, as this often means that noise in the training set, which can not be expected to generalize to the test set, was fit. This creates a classifier said to have high *variance*. Various other termination conditions exist that prevent such overfitting.

A random forest classifier is a classifier constructed from many decision trees. It was developed to address weaknesses in the solitary decision tree classifier, particularly its tendency towards error resulting from high variance, noted above. The random forest reduces variance by diluting the erroneous, noise-based classifications of any one decision tree by polling many trees. The polling process can be as simple as taking a majority vote of the classifications of the individual classifiers (similarly, the random forest's classification error can be straightforwardly characterized based on the proportion of dissenting trees). Since a single training set usually lends itself to a single decision tree classifier, the many unique decision trees composing a random forest are constructed by generating many training sets from one using a technique called *Bagging* (for ***B****ootstrap* ***agg****regat****ing***). Bagging creates random child training sets by randomly sampling the original set (with replacement). The random forest classifier further diversifies its trees by demanding that every feature split of every decision tree be constrained to a random subset of features. Given n total features, a random subset of $\sqrt{n}$ is usually used at each split.

I trained a generic random forest on the training set of 897,891 features including TDOA's and truncated cross-correlations. In general, the more decision trees added to the random forest, the more accurate and less prone to overfitting it is. Over five days I constructed a 1,400-tree random forest from the training data (after 1,400 trees I encountered significant computational slow-down), where each tree was trained on a random subset of 75% of the total training set of 1,605 samples and at each branch one of $\sqrt{897,891}$ random features

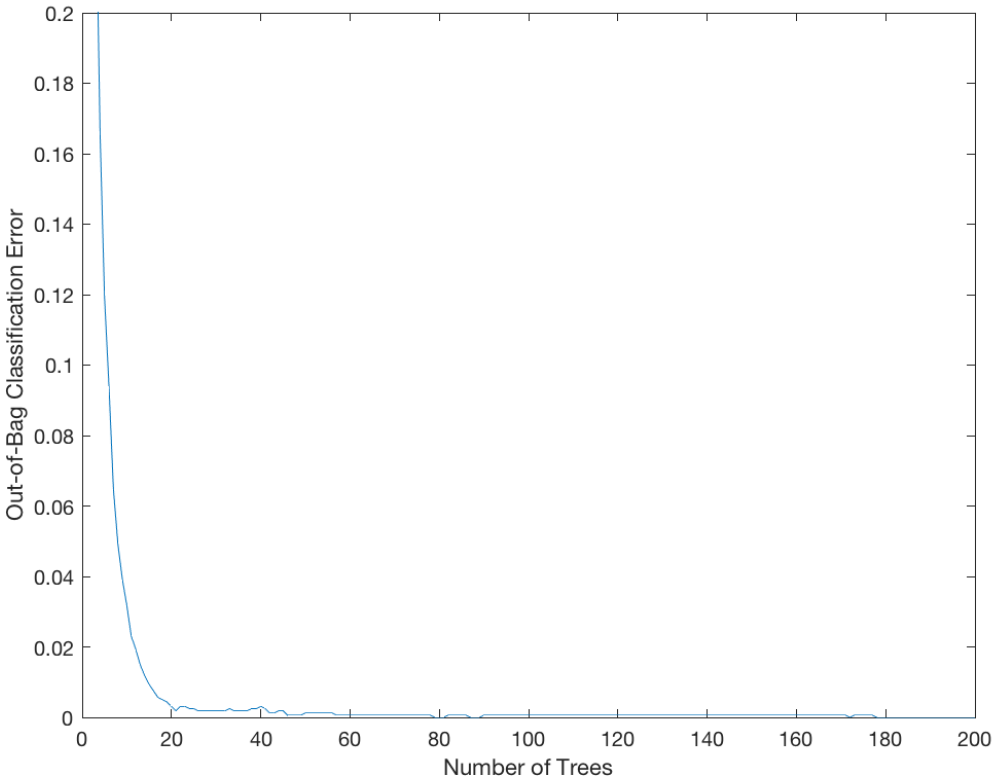


Figure 6.1: Full-Feature Random Forest Classification Growth Error The out-of-bag classification error of a random forest trained on the complete feature set (all TDOA and cross-correlation pairs) as decision trees are added. Note the real peak below 5 trees is at an error of 0.6.

was split on. Surprisingly, the random forest achieved 100% classification accuracy on the test set. Figure 6.1 shows the forest's improvement in classification as trees were added; it shows a steep increase in classification accuracy up to 20 trees, and perfect classification at approximately 180 trees. This is a significant result and constitutes proof that classification is a promising avenue for whistle localization, assuming the set of artificial whistle-like sounds suitably sample the space of real whistles, which remains to be seen.

Of course, one might ask how the classifier achieved this success. Because this random forest polls 1,400 individual decision tree classifiers to reach its decision, its decision process is

somewhat opaque, particularly given that even that the logic of a "decision tree," while transparent, does not necessarily lend itself to intuition. One way of gaining some understanding of the classifier is by assessing *feature importance*, or the relative influence of each feature on total classification success. The random forest is regarded for its strength at assessing feature importance, owing to the fact that the random bagging and feature sub-selection methods suit them for generating statistics from a single training and feature set. The measure of feature importance used here is termed *delta error*. To calculate the delta error, we first evaluate the change in class-prediction accuracy of a single decision tree based on the *out-of-bag samples* – the training samples that were not randomly selected during the bagging process for training that tree – when the values of a particular feature are permuted across all the samples. This is done for every feature in the tree to obtain an error value for each, and feature values thus obtained are averaged across all decision trees within the random forest. The larger a feature's delta error, the more important it is for successful classification.

In the random forest generated, 66,384 features possess importance greater than zero. The full plot can be seen in Figure 6.2. While it should be expected that not all of the features were sampled in the construction of the forest, and that a random subset of features have been inappropriately assigned zero importance, the forest's classification success combined with the expectation that the cross-correlation arrays carry redundant information suggest that these features are sufficient for consideration.

I used feature importance to ask whether the random forest prioritizes features representing particular inter-array measurements, separately for TDOA's and cross-correlations, and whether the random forest prioritizes features representing certain delays among the cross-correlations, potentially as a function of which pair of panels the cross-correlated signals belong to.

With regards to whether the random forest prioritizes features representing particular inter-array difference measurements, refer to Figure 6.3. One might expect that features corresponding to TDOA's and cross-correlations between hydrophones in different arrays

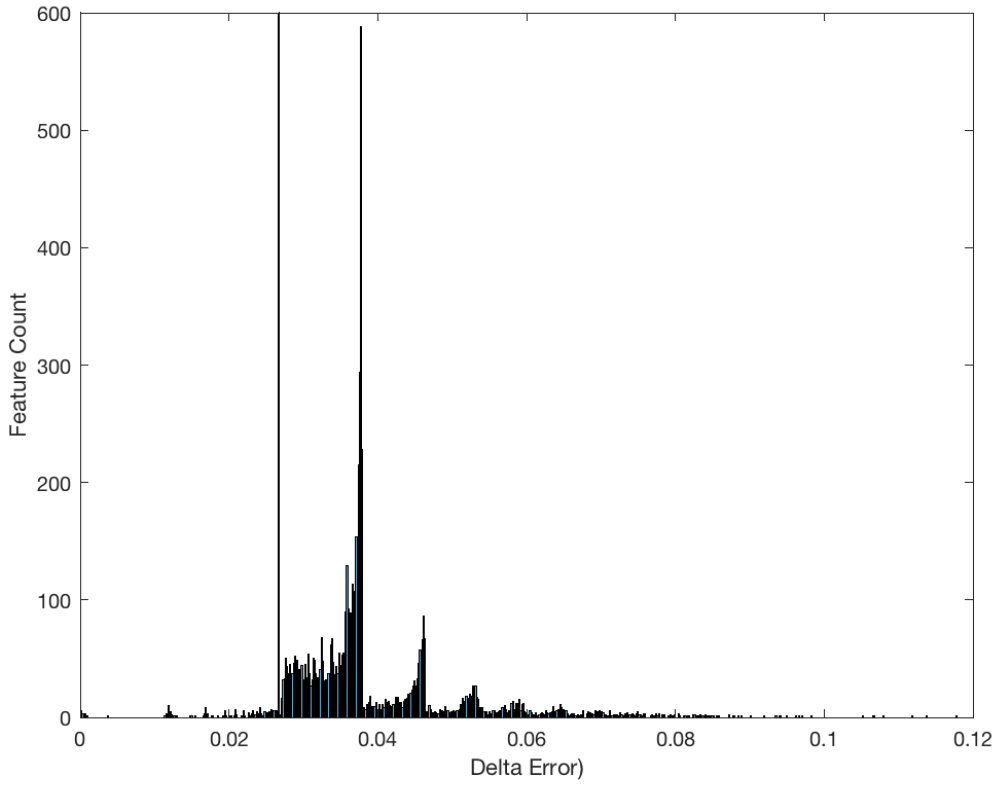


Figure 6.2: Histogram of Random-Forest-Generated Feature Importances 1,000-bin histogram of Random-Forest-generated feature importances, as measured by delta error, explained in the text. Note that there is a peak at 0 of 66,384 features, and a peak at 0.0267 of 59,145 features.

would have higher importance than features corresponding to TDOA's and cross-correlations between hydrophones in the same arrays, as the former reflect larger delay times that are less susceptible to destruction by noise. However, only with respect to the TDOA's do we find cross-array importances more emphasized, particularly between hydrophone arrays on opposite sides of the pool's XY vertical midline, which seems reasonable considering that their greater displacements (as compared with the vertical pairs) generally lead to greater TDOA's. Unexpectedly, TDOA's between hydrophones belonging to different arrays on the same side of the pool (1 and 2, 3 and 4) are even less emphasized than TDOA's between

hydrophones in the same arrays. For the cross-correlation features, no obvious pattern exists between hydrophone array location and feature importance. That hydrophone/panel pair importances are not intuitive might be because the fourteen testing locations depend significantly on different array comparisons. However, for both TDOA's and cross-correlation features, we would trivially expect symmetry in all four plots about the upward diagonal owing to the approximate symmetry of the arrays and the symmetry of testing locations across the XY vertical, however this is not observed. Together with the importance of array auto-correlations, the meaning of the feature importances remains unclear. It is possible they are partly artifacts of the random forest creation process' random feature selection.

With regards to whether the random forest is prioritizing features representing certain time-delays among the cross-correlations, refer to Figure **??**. The importances of features corresponding to every cross-correlation delay time were averaged, across all hydrophone pairs (plots not shown) and across groups of hydrophone pairs corresponding to comparisons between different hydrophone arrays. For the latter groupings, the weighted means and standard deviations of importances across all delay times were computed for each array pair (delay times weighted by mean importance). Across all hydrophone pairs no concentration of feature importance among time delays was discerned, and among groups of hydrophone pairs belonging to different array pairs no concentration of feature importances among time delays – or differences in these concentrations – was discerned. Speaking to the latter: every weighted mean time delay importance is between approximately -0.5 and 0 milliseconds with standard deviations on the order of 10 milliseconds, almost a third of the length of the entire cross-correlation. Nevertheless, it is not necessarily unexpected that feature importances are not concentrated, as for even a single array pair the expected peak indicating the first incidence arrival might vary by as much as 20 milliseconds among the testing locations (indeed, the cross-correlation window was chosen specifically to cover this range). It is possible that a deeper analysis looking for clustering of feature importances around the expected peaks in the cross-correlations (corresponding to expected TDOA's) for every panel pair for every test

Figure 6.3: Cross-Hydrophone and Cross-Array Classification Importances Classification importance values across hydrophone and hydrophone array pairs are formed by summing the importances (delta error) of features across hydrophone pairs, and subsequently averaging hydrophone pair sums across array pairs. **a**: Cross-hydrophone importances for cross-correlation features. Hydrophones belonging to common panels (1-4, 5-8, 9-12, 13-16) are grouped by red boxes. **b**: Cross-array importances for cross-correlation features. Facing the pool from the audience area, panels increase in number from from left to right. **c**: Cross-hydrophone importances for TDOA features. d: Cross-array importances for TDOA features.

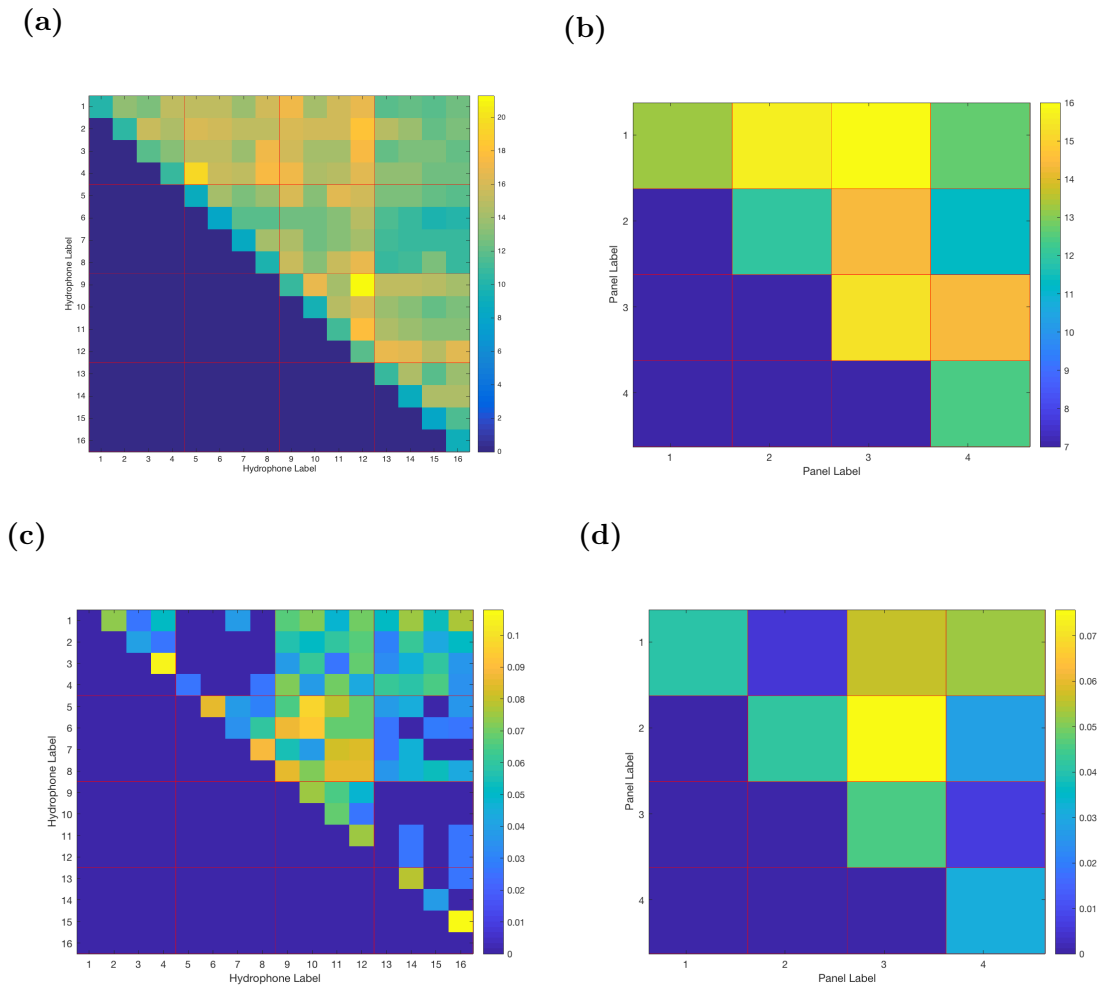


Figure 6.3

site would be successful.

Similar plots were created using the first principal components obtained via a *principal components analysis* of the training set in place of delta error. Rather than representing a feature's direct importance to classification, a high-magnitude feature in a principal component vector reflects that feature's tendency to co-vary with other high-magnitude features; different sets of co-varying features (or eigenvectors) account for different amounts of variation in the overall data. One might expect features representing cross-panel differences to account for more overall variance in the feature sets. However, this was not found. The plots are numerous and are excluded here.

While it is clear that a large random forest classifier trained on the full feature set consisting of all hydrophone cross-correlations and estimated TDOA's can obtain high classification accuracy, the lack of intuition behind the feature importances compelled me to find an effective classifier on a minimal feature set; this minimal feature set might lend itself to intuition more than the full feature set examined above. Rather than manually selecting an arbitrary number of features based on the importances already obtained, I found a complete minimal feature set by training a classifier on the full feature set that naturally performs a high degree of feature selection. A stronger classifier could then be trained on the features selected.

Towards this end, I trained a sparse decision tree classifier on all features of the full training set. The resulting decision tree achieved 96.63% classification accuracy on the test set using only 22 of the original 897,871 features. I then trained a random forest on the same features, achieving 99.69% out-of-bag classification accuracy and 98.88% classification accuracy on the test set. The delta error-based feature importances for the 22 features were then mapped back to hydrophone and hydrophone array pairs as was done previously; these plots are shown in Figure 6.4.

With respect to the cross-correlation features of the minimal classifier, Figure 6.4 shows that, rather being concentrated to any one hydrophone or hydrophone array, the 13 selected

Figure 6.4: Minimal Cross-Hydrophone and Cross-Array Classification Importances Classification importance values for minimal set of 22 features across hydrophone and hydrophone array pairs are formed by summing the importances (delta error) of features across hydrophone pairs, and subsequently averaging hydrophone pair sums across array pairs. **a**: Cross-hydrophone importances for cross-correlation features. Hydrophones belonging to common panels (1-4, 5-8, 9-12, 13-16) are grouped by red boxes. **b**: Cross-array importances for cross-correlation features. Facing the pool from the panels increase in number from left to right. **c**: Cross-hydrophone importances for TDOA features. d: Cross-array importances for TDOA features.

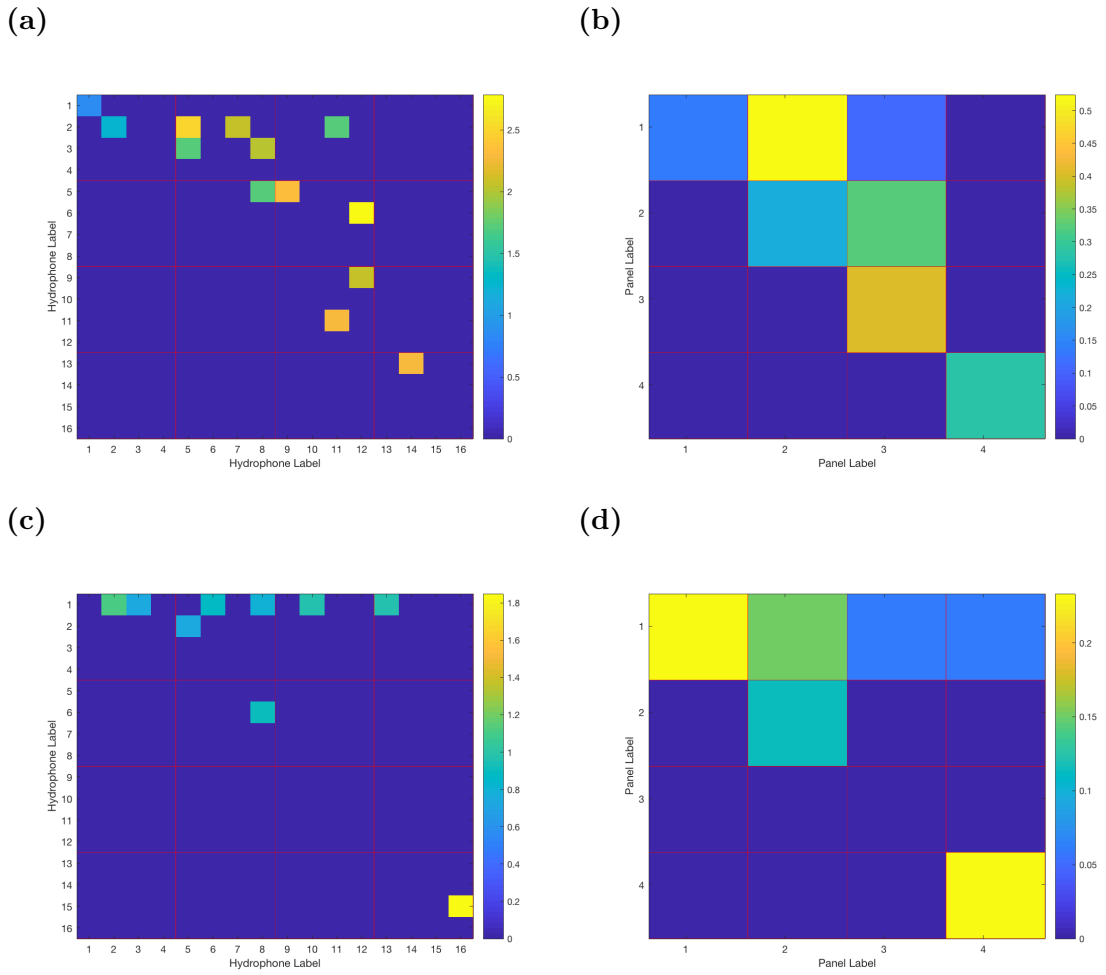

(a)
(b)
Hydrophone Label
Hydrophone Label
Panel Label
Panel Label
(c)
(d)
Hydrophone Label
Hydrophone Label
Panel Label
Panel Label

Figure 6.4

features are spread across hydrophone and hydrophone array pairs, consistent with the geometric expectation that the classifier's implicit localization would require information from a diverse and distant set of hydrophones. Curiously, only one feature involves hydrophones in the fourth (furthest right) hydrophone array, and it involves an intra-array comparison. As was the case for the full feature set, the general importance of intra-array features and the lack of symmetry across the plots' upward diagonal despite array and testing point symmetry (across the pool XY vertical) remain unexpected, but in this case the latter might be explained by the decision tree's tendency towards feature spareness. The 9 TDOA features of the minimal classifier are distributed among hydrophone and hydrophone array pairs much as the cross-correlation features are. In fact, that a TDOA between a hydrophone in the first array and every other array is used is strongly indicative that the classifier is performing implicit localization. For both the cross-correlation and TDOA features, no pattern in the specific importance values given to each feature is readily apparent.

All 13 selected cross-correlation features were averaged across their respective time-delay positions, however no concentration was apparent; as for the full feature set, a comparison to theoretically expected peaks for every array pair for every testing location might be required to uncover a deeper intuition.

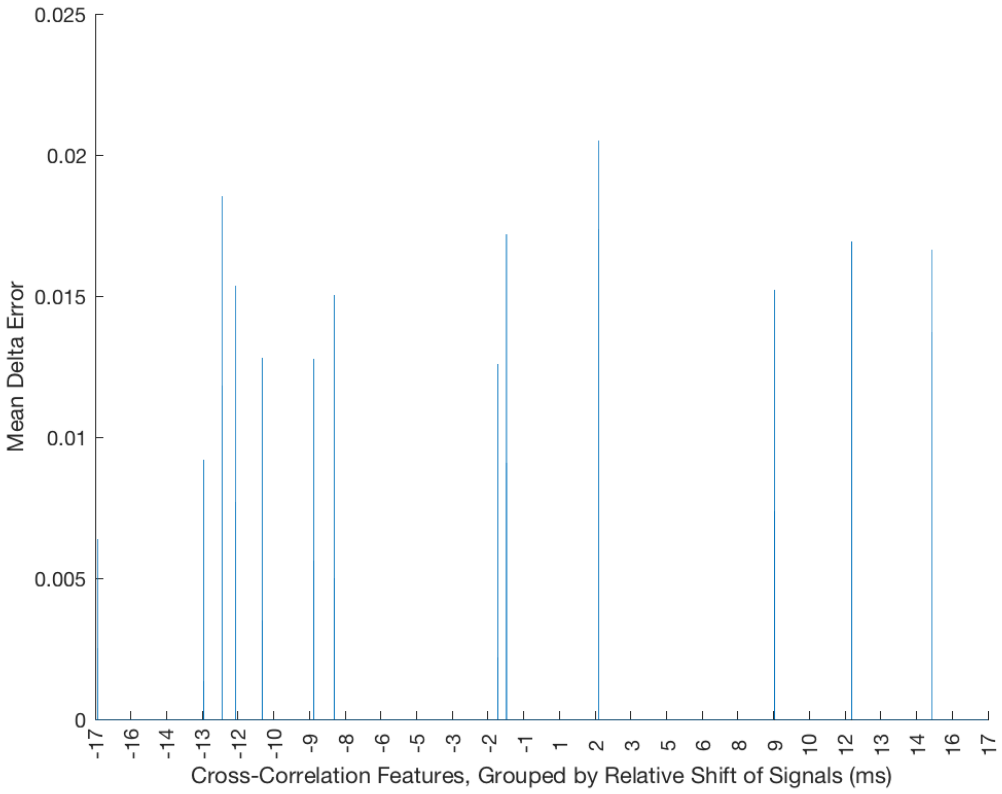


Figure 6.5: Minimal Cross-Correlation Delay Classification Importance Classification importance values across cross-correlation delays are formed by averaging the importances (delta error) across corresponding features, for all hydrophone pairs.

Chapter 7

Discussion, Future Directions

Presented in this thesis are the first step towards the completion an audiovisual system for performing whistle attribution and general behavioral tracking for the seven dolphins residing at Dolphin Discovery at the National Aquarium, Baltimore. While truly continuous, full-time sound attribution would require audiovisual tracking of all dolphins across the facility's three primary sub-pools, the proposed system is aimed at continuous sound attribution for subsets of the seven dolphins during the periods they reside in the facility's primary pool, the *exhibit pool* or EP. The two audiovisual tasks required for performing whistle attribution include sound source localization and visual tracking of dolphins, both on a shared coordinate system. This thesis presents installed hardware, software and general methodology that approaches achievement of sound source localization, including evidence of this sub-system's ability to localize idealized sound sources and a proof of concept for using a machine classifier for performing sound source localization for whistle-like sounds, as well as installed hardware and early software that requires extension for the achievement of visual tracking of dolphins. The completion of the sound attribution system in its entirety and, with it, the generation of whistle-attributed behavioral data for the aquarium's seven dolphins over months if not years, would represent a unique accomplishment in the field of cetacean communication and new access to questions about the distinctions between individual vocal repertoires and the

existence of vocal exchange.

The core of the proposed sound source localization subsystem are four arrays of four hydrophones distributed in an approximate "splay" configuration around the EP. After a two-year process involving over one hundred planned and unplanned installations and removals, several weeks of 24-hour monitoring periods, and two major and multiple minor redesigns, four duplicates of the newest iteration of the array have been "permanently" installed for almost one year; they have been proven suitably robust to corrosion, dolphins, and the discerning eye of the National Aquarium exhibit staff. The four hydrophone arrays constitute a theoretically minimum system for achieving two-dimensional sound localization using a conventional time-delay-of-arrival (TDOA) treatment, shown in Section 4.4 by their ability to localize idealized *impulse response functions* (IRF's) to elliptical regions occupying less than 1% of pool area/volume offset from the measured source points by less than 5 feet.

Nevertheless, while theoretically sufficient for sound localization, the system of hydrophone arrays does possess room for improvement that might be addressed to reduce the observed practical difficulties of localizing whistle-like sounds. Possible extensions of the hardware include completion of my planned "mini-panels" for vertically spreading the four hydrophones in the existing arrays, which would allow the system to perform a degree of vertical localization, and in general the installation of more hydrophone arrays placed consistent with the principle of maximizing the spread of the source-sensor direction vectors, discussed in Section 2.2. While a solution for placing hydrophones deeper beneath the water surface than the length of a hydrophone array (˜6') would herald a substantial increase in sound localization capacity, any re-design of the arrays would likely trigger higher stress levels in the National Aquarium's resident dolphins and require a lengthy desensitization period Section 2.2. Regardless of the arrangement of hydrophones, the inter-hydrophone distance measurements necessary for sound localization might be improved by using a more sophisticated system of "pings" generated from the hydrophones themselves. The three scenarios for which I would expect the system hardware to substantially improve in its capacity for sound localization without

pushing against the constraints imposed by the National Aquarium include the discovery of adequate contact microphones for recording whistles from the outside of the EP acrylic and their deployment, the development of a beamforming algorithm Section 2.2 requiring two to four localized hydrophone arrays (likely consisting of more than four hydrophones each) and the arrays' deployment, and the purchase and use of wearable hydrophone-recorder systems Section 1.5.

Regarding the computational side of sound source localization, the present body of work suggests that a conventional, geometric approach to localizing whistle-like sounds is inadequate given the available hardware and IRF data. Specifically, the process of estimating time-differences-of-arrival (TDOA's) for geometric sound source localization has not been performed successfully. The best-performing TDOA estimation method implemented, involving a comparison of bandpass-filtered hydrophone signals using a General Cross-Correlation Phase Transform (GCC-PHAT), produces TDOA's with greater than ˜5 millisecond (˜7.5 meter) error, and produces no meaningful spatial localization using a standard TDOA-based localization algorithm, spherical interpolation Section 5.4. The failure of conventional approaches to TDOA estimation is due to both whistles' slow and variable onset above the acoustic noise floor, which disturbs a conventional peak-finding approach to extracting time delays based on the first incidences of received signals, and the overlap of many source signals in the received signal as a result of multipath travel, which disturbs conventional cross-correlational approaches to extracting time delays based on whole received signals. Improving the system's capacity for sound (and specifically whistle) source localization at the computational stage would likely require coping with multipath effects either explicitly or implicitly. Explicit treatments have been proposed in the signal processing literature, in which the multipath problem features prominently, which were not fully implemented and surveyed in favor of an implicit machine learning treatment. Promising explicit treatments include those based on acoustic tomography, specialized correlations, and multidimensional matched filtering (Bell and Ewart, 1986; Spiesberger, 1998). Another explicit treatment might include using IRF's

to remove multipath, a topic discussed in Section 4.4.

A promising alternative to the explicit treatment of the multipath problem (as well as other sources of complex noise) with conventional signal processing techniques, which tend to be specific and can require a high level of expertise to choose among, is a machine learning-based approach to sound localization, relying on training data consisting of representative sounds played at known locations in the pool. Despite the difficulties associated with obtaining these training data, this approach has been shown to be effective for distinguishing artificial whistle-like sounds among fourteen testing locations using machine classification. In particular, the classifier can distinguish these sounds along the center XY vertical line of the EP, for which even IRF's could not be distinguished using spherical interpolation. Nevertheless, the generalizability of the current training set is uncertain, both in regards to how representative of real bottlenose dolphin whistles the training sounds are, and how close a source whistle – real or artificial – must be to a training location to produce a reasonable classification. To be more specific, the classifier's probability of classifying an off-testing-point sound to the nearest testing point or to another testing point as a function of its ranges to the various testing points must be expected to be complicated and to require significant probing before the current classifier might be used for practical signal localization. While a regression-based approach to the prediction problem would be expected to provide more spatial generality, current data have proven insufficient for this approach, and in general we expect the regression problem harder to solve well. Before addressing questions of the classifier's generalizability directly, it would seem reasonable to build a classifier or regressor on more training data covering more points in both whistle and real space. With enough sounds played at enough locations, the question of a classifier's generalizability might not need to be addressed. Lacking real dolphin whistles from known positions in the pool, the data currently available can be used to address the classification's approach generalizability in whistle space (and also real space) by selectively excluding subsets of whistles from training; a thorough analysis of this sort could not be prepared for this thesis. However, as noted above, securing these data is a complex,

multi-person operation requiring a five-person team from our research staff plus the help of aquarium staff. Therefore, increasing the training set size is not expected to be logistically easy. Perhaps a more efficient set of training sounds, for instance reflecting a sampling of a whistle eigenspace, might simplify the process. In any case, at this time attempts to reverse-engineer the successful random forest classifier to inspire a non-data-driven solution to the whistle localization problem have not been successful.

Aside from finalization of a method of whistle source localization, completion of the full proposed system for whistle attribution would require the creation of dolphin-video-tracking software paired with a reliable methodology for determining dolphin identity in one or more frames for every unbroken video follow of a dolphin. Preliminary work using the feed of the central overhead camera to achieve approximate 2-dimensional tracking suggests that conventional object tracking methods, including background subtraction and adaptive correlation filters, are significantly disrupted by sun glitter, which results in changing regions of light and dark across the surface of the pool over the dolphins. Ongoing work based on a large coarse-graining and dynamic thresholding of pool pixels might achieve success. Otherwise, a brute-force approach involving a pre-trained convolutional network should be considered. Apart from the lack of video tracking software, a factor that might preclude the achievement of whistle attribution is high levels of network latency and software latency associated with the current system of twelve IP surveillance camera systems, which are managed by the same computer system responsible for managing all sixteen, 192 kHz hydrophone recordings. Barring a complete system renovation, this problem might be solved by removing all unnecessary cameras and managing the remainder with simpler software. However, this might be at odds with other lab goals.

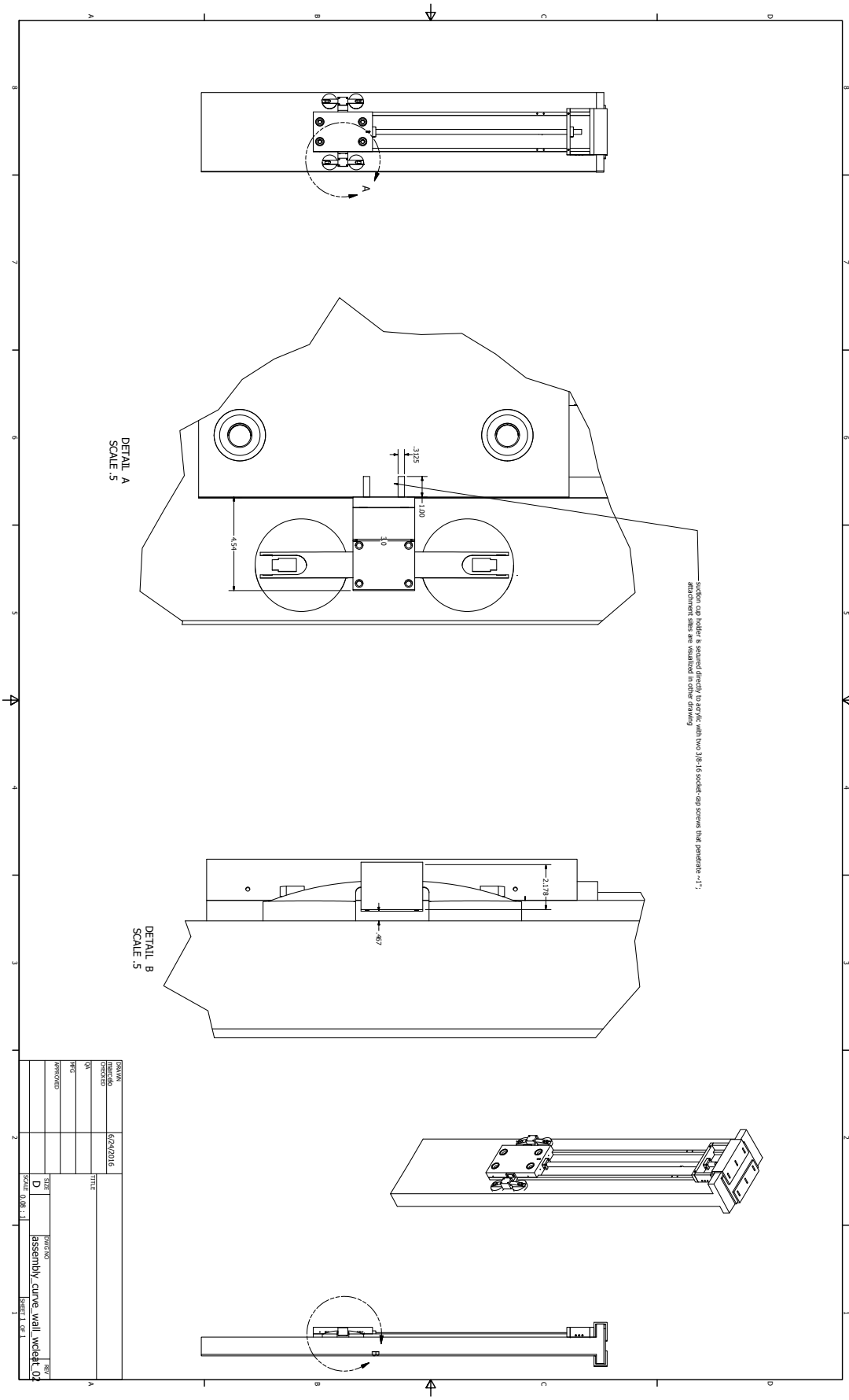
The proposed system for whistle attribution, if complete, would represent the first such system poised to collect data from a fixed group of dolphins on a time-scale of months or years. Even among similar systems, only one, consisting of eight hydrophones placed inside an irregularly-shaped, ocean-connected, relatively non-reverberant lagoon has met any degree

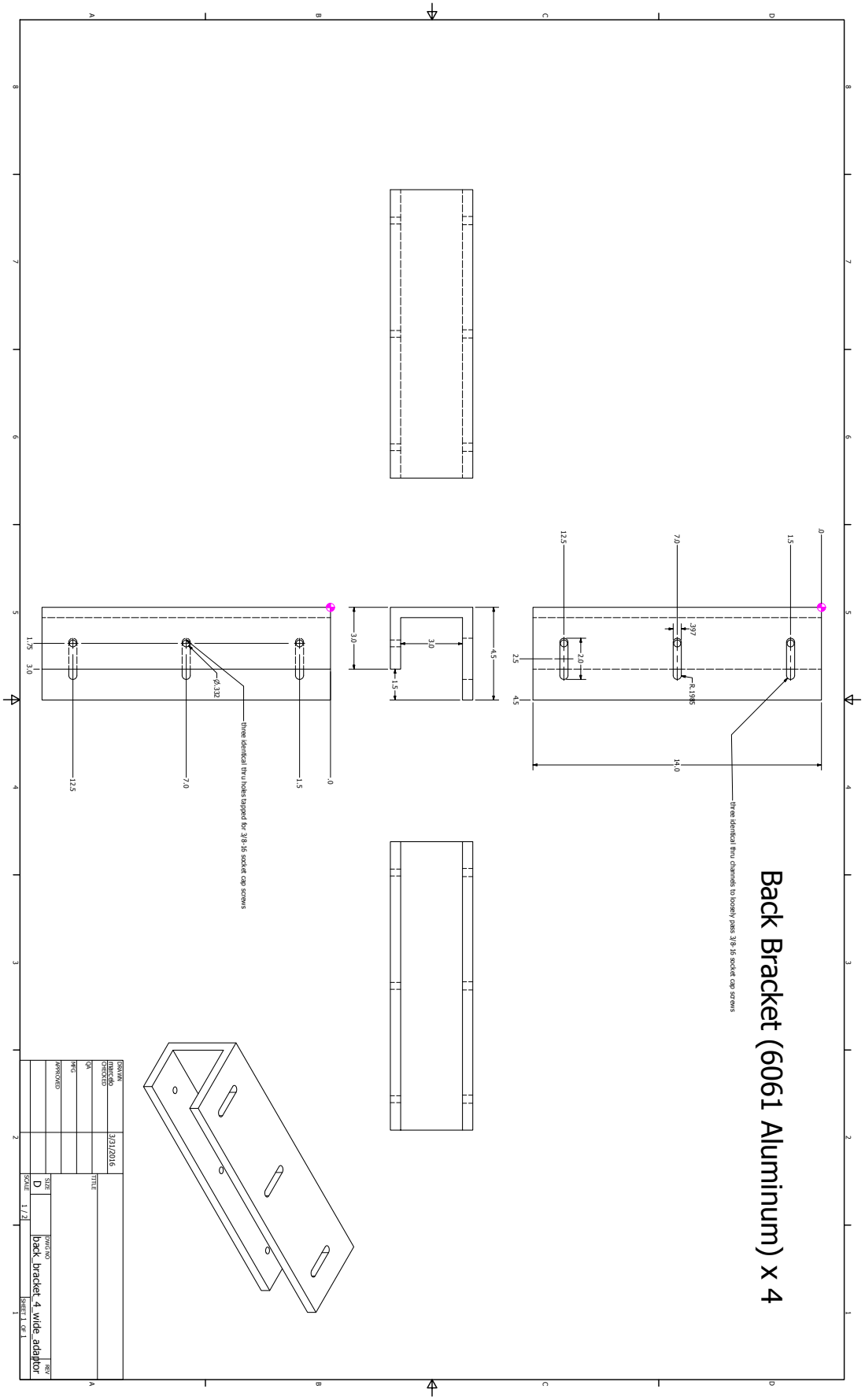
of success performing whistle attribution, and was temporary (Thomas et al., 2002). Other similar systems have failed to consistently attribute whistles across individuals primarily due to an inability to localize whistles as a result of multipath (Freitag and Tyack, 1993; Janik and Thompson, 2000; López-Rivas and Bazúa-Durán, 2010). Together with the use of wearable hydrophones, which possess the advantage of unambiguous sound attribution but the disadvantages of temporality and lack of visualization, the technology proposed here can help researchers to explore the currently unstudied area of combined dolphin signature and non-signature whistle exchange. Research in this area is a prerequisite to an understanding of the level of information exchange mediated by dolphin vocal communication.

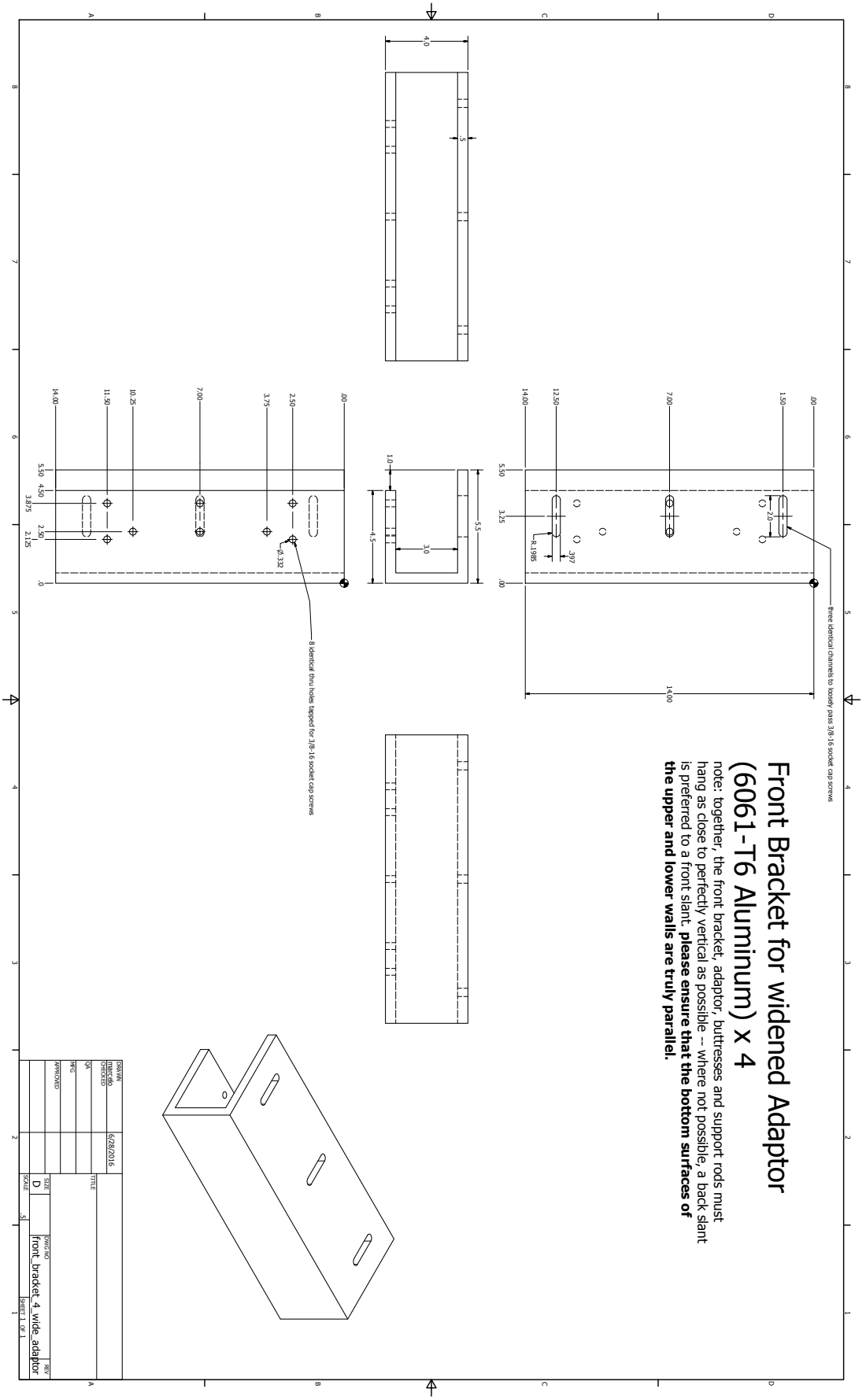
Chapter 8

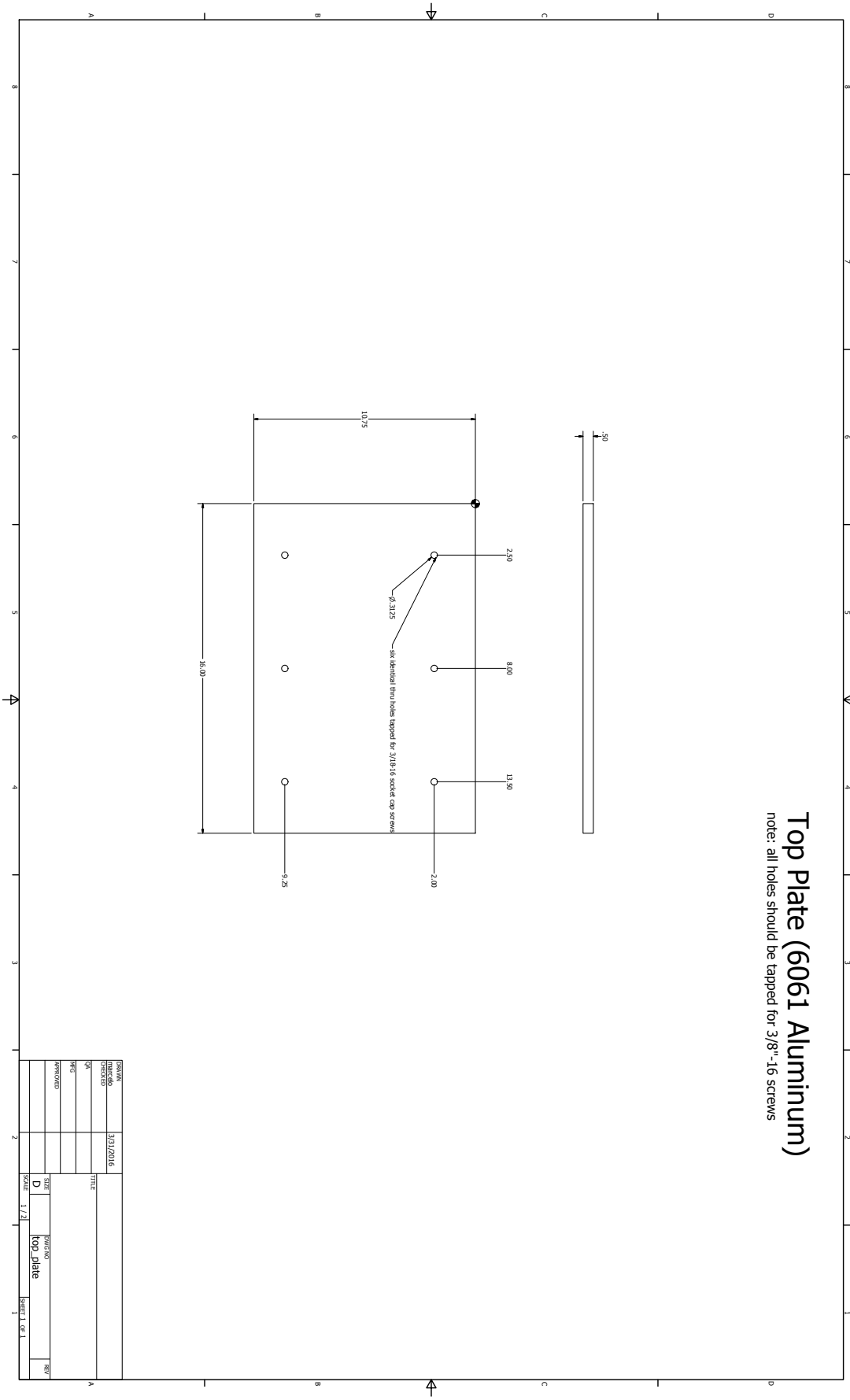
Appendix

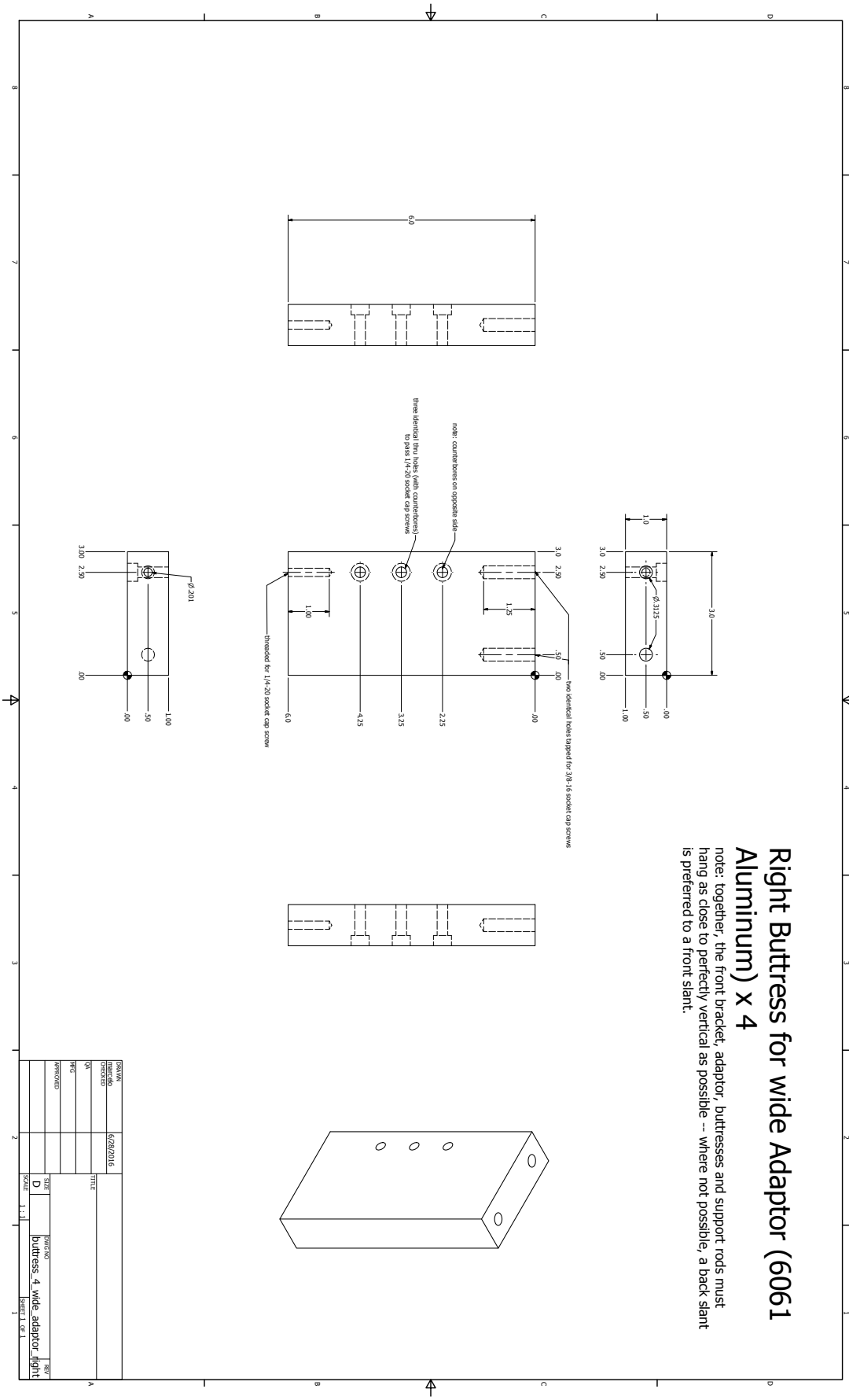
8.1 Mk. II Array Schematics

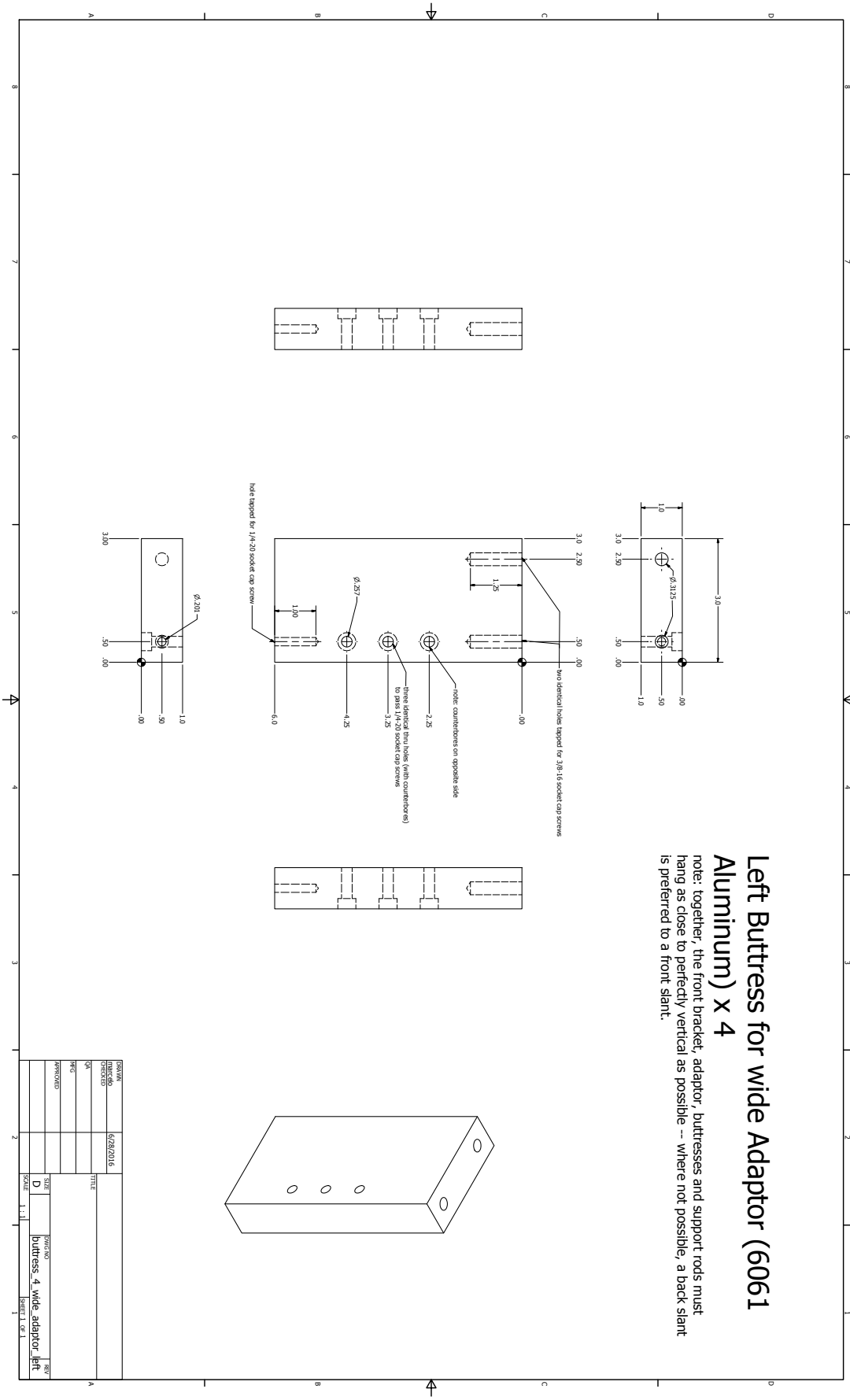

A
DETAIL A
SCALE .5
3.125
1.00
4.54
.10
suction cup holder is secured directly to acrylic with two 3/8-16 socket-cap screws that penetrate ~1";
attachment pats are visualized in other drawing
2.178
.467
DETAIL B
SCALE .5
B
DO NOT SCALE DRAWING
MATERIAL
FINISH
CHECKED
QA
MFG
APPROVED
DATE
6/24/2016
SIZE
D
SCALE 0.08 : 1
TITLE
DWG NO
assembly_curve_wall_wcleat_02
REV
SHEET 1 OF 1

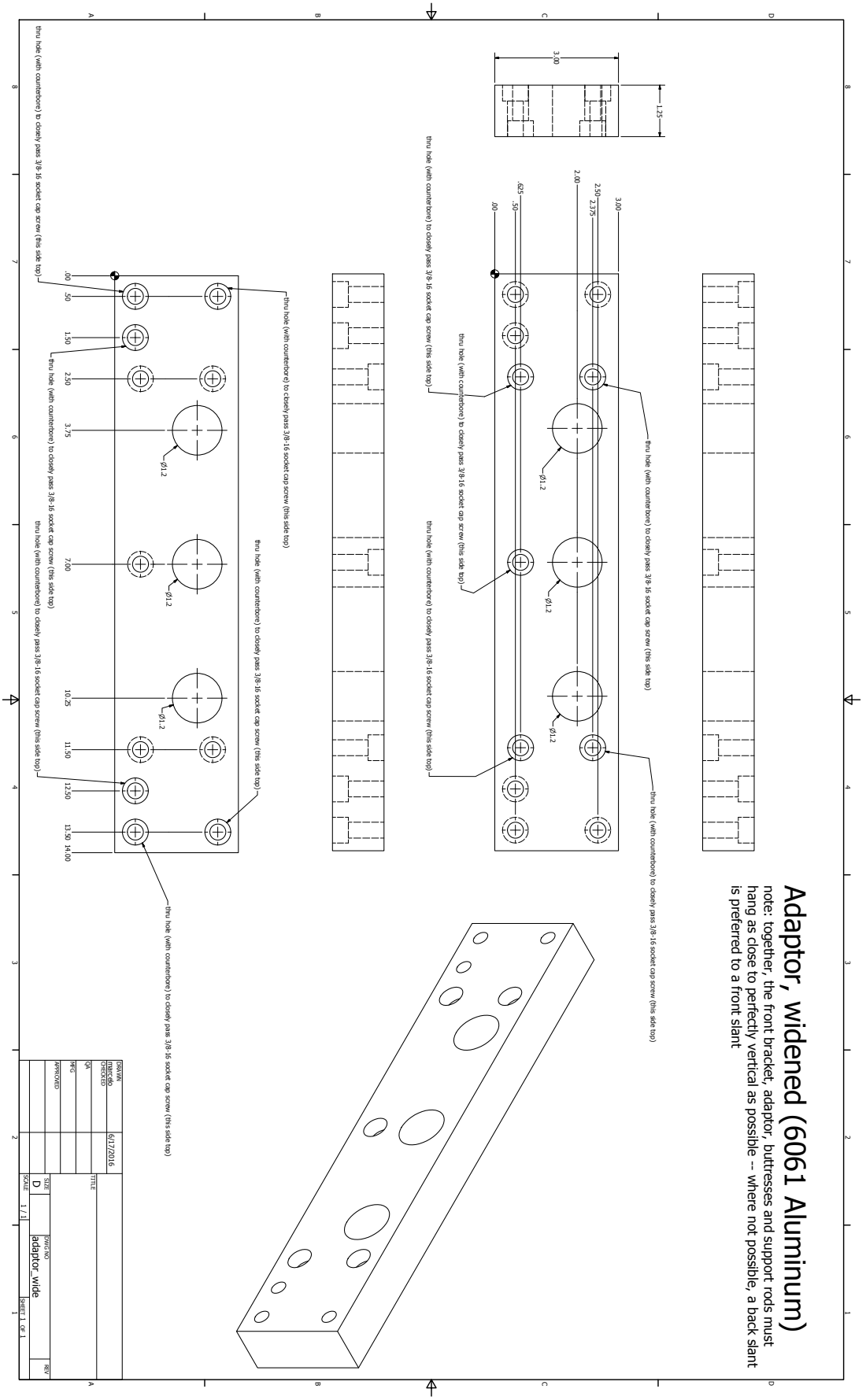

Back Bracket (6061 Aluminum) x 4
three identical thru channels to loosely pass 3/8-16 socket cap screws
three identical thru holes tapped for 3/8-16 socket cap screws
14.0
12.5
7.0
1.5
0
2.5
4.5
3.0
2.0
1.75
Ø.332
Ø.397
Ø.1985
DRAWN
CHECKED
QA
MFG
APPROVED
3/31/2016
TITLE
SIZE
D
SCALE
1 / 2
DWG NO
back_bracket_4_wide_adaptor
REV
SHEET 1 OF 1

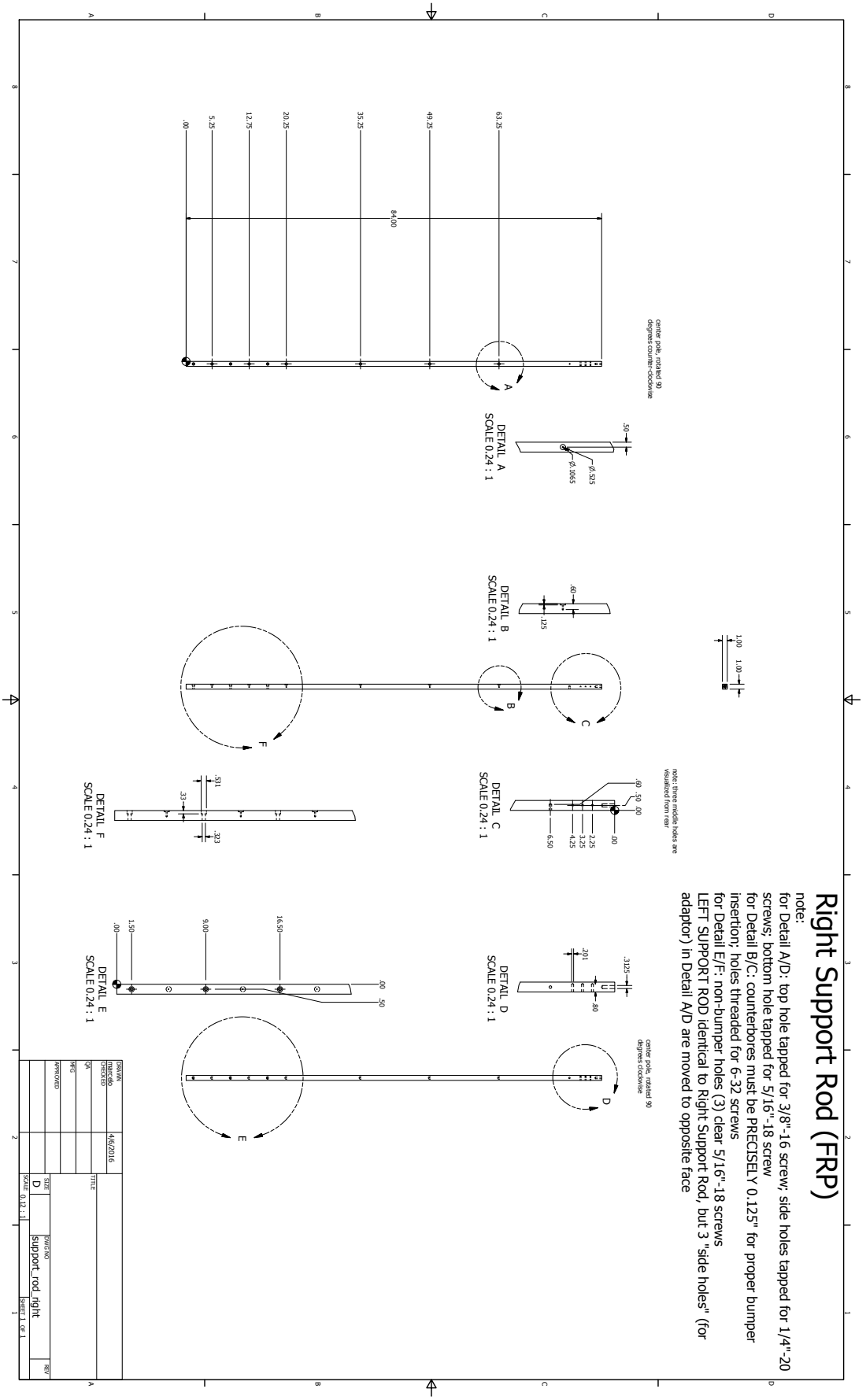

Front Bracket for widened Adaptor
(6061-T6 Aluminum) x 4
note: together, the front bracket, adaptor, buttresses and support rods must hang as close to perfectly vertical as possible -- where not possible, a back slant is preferred to a front slant. **please ensure that the bottom surfaces of the upper and lower walls are truly parallel.**
three identical channels to loosely pass 3/8-16 socket cap screws
8 identical thru holes tapped for 3/8-16 socket cap screws
front_bracket_4_wide_adaptor
6/28/2016
SIZE
D
SCALE
.5
SHEET 1 OF 1

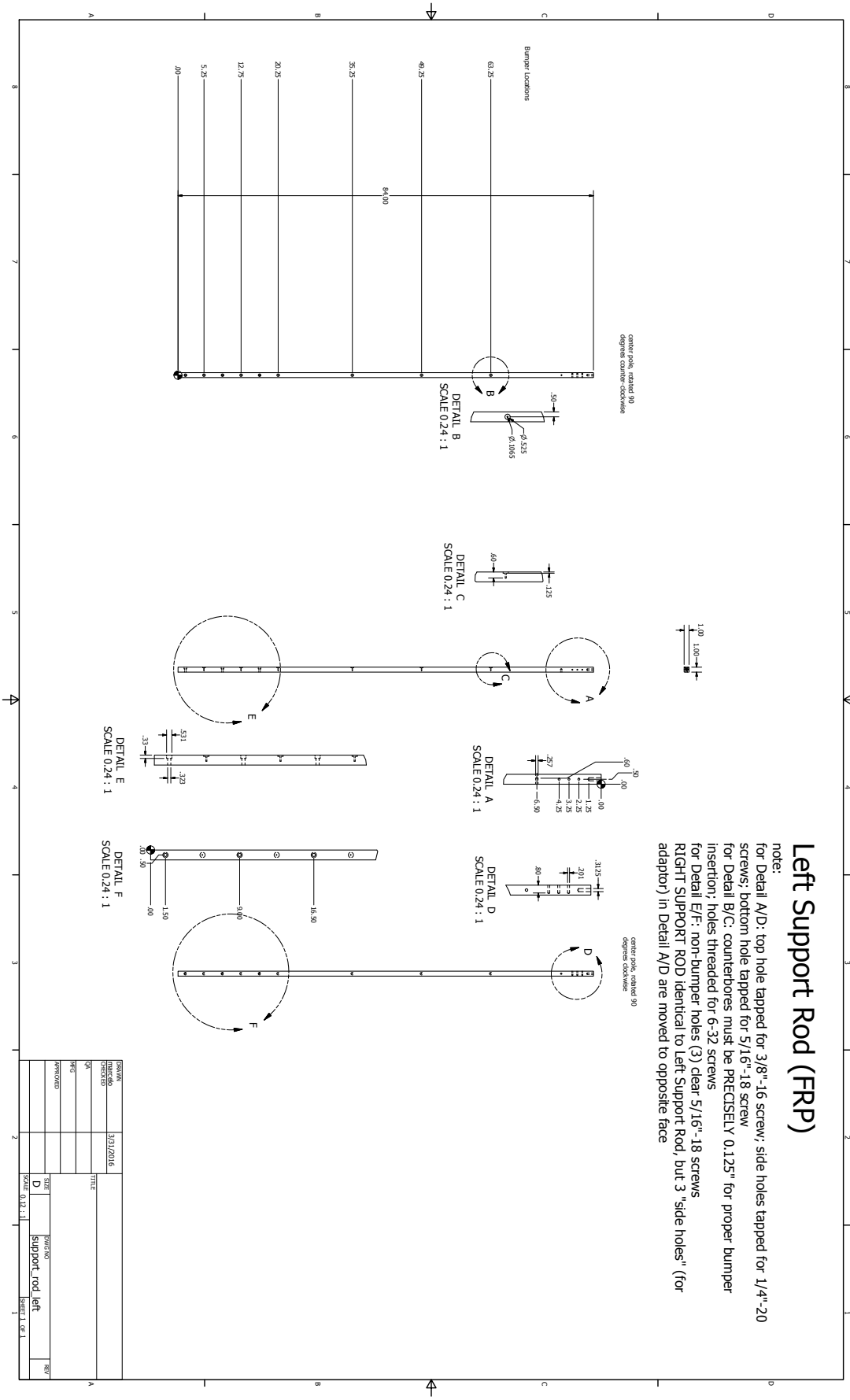

Top Plate (6061 Aluminum)
note: all holes should be tapped for 3/8"-16 screws
.50
10.75
16.00
2.50
8.00
13.50
9.25
2.00
Ø.3125
six identical thru holes tapped for 3/18-16 socket cap screws
DRAWN
CHECKED
QA
MFG
APPROVED
3/31/2016
TITLE
SIZE
D
SCALE
1 / 2
DWG NO
top_plate
REV
SHEET 1 OF 1

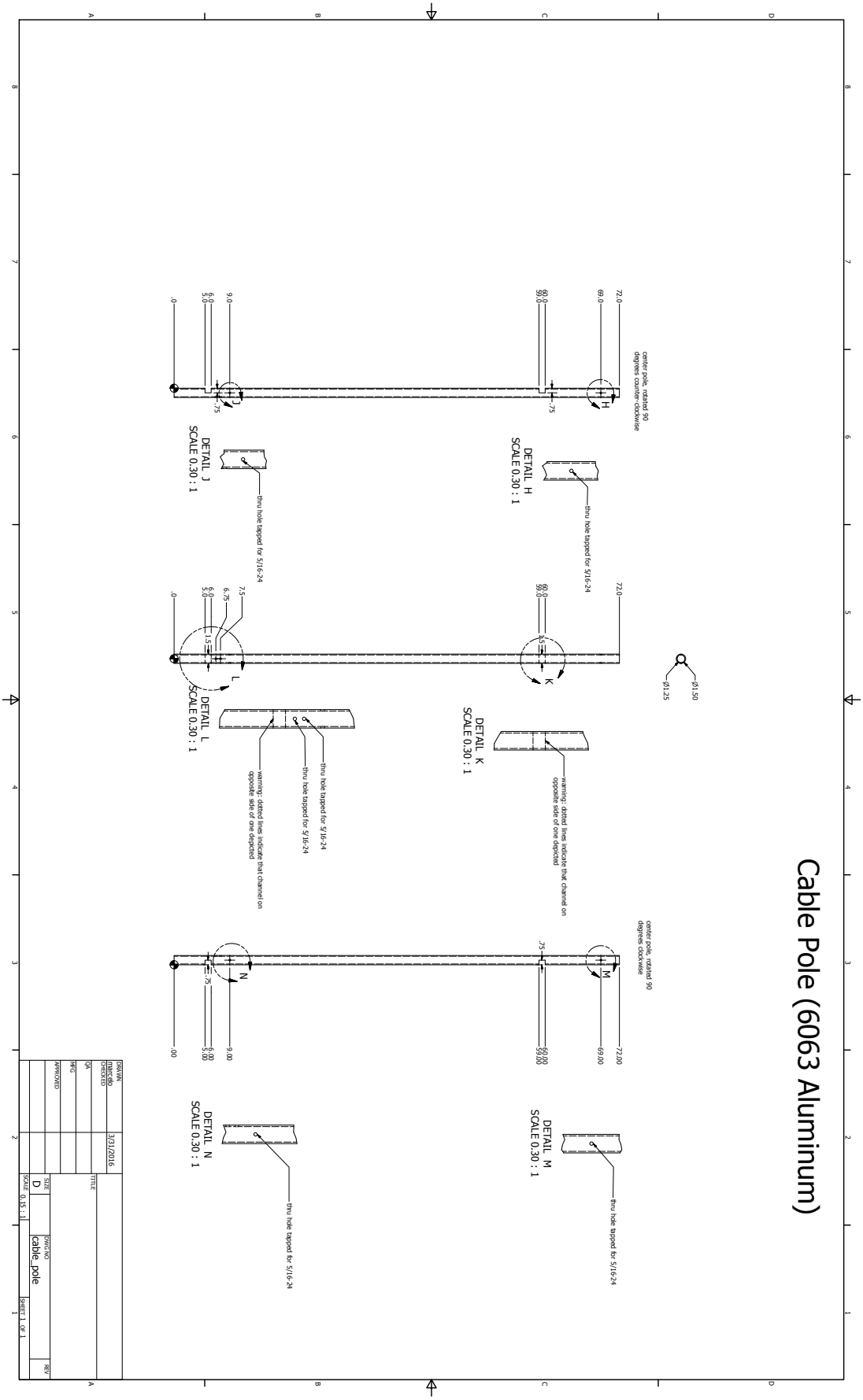

Right Buttress for wide Adaptor (6061 Aluminum) x 4
note: together, the front bracket, adaptor, buttresses and support rods must hang as close to perfectly vertical as possible -- where not possible, a back slant is preferred to a front slant.
note: counterbores on opposite side
three identical thru holes (with counterbores) to pass 1/4-20 socket cap screws
two identical holes tapped for 3/8-16 socket cap screws
threaded for 1/4-20 socket cap screw
Ø.3125
Ø.201
3.0
2.30
1.0
.50
.00
1.00
1.25
2.25
3.25
4.25
6.0
3.00
2.50
DRAWN
CHECKED
QA
MFG
APPROVED
6/28/2016
TITLE
SIZE
D
SCALE
1:1
DWG NO
buttress_4_wide_adaptor_right
REV
SHEET 1 OF 1

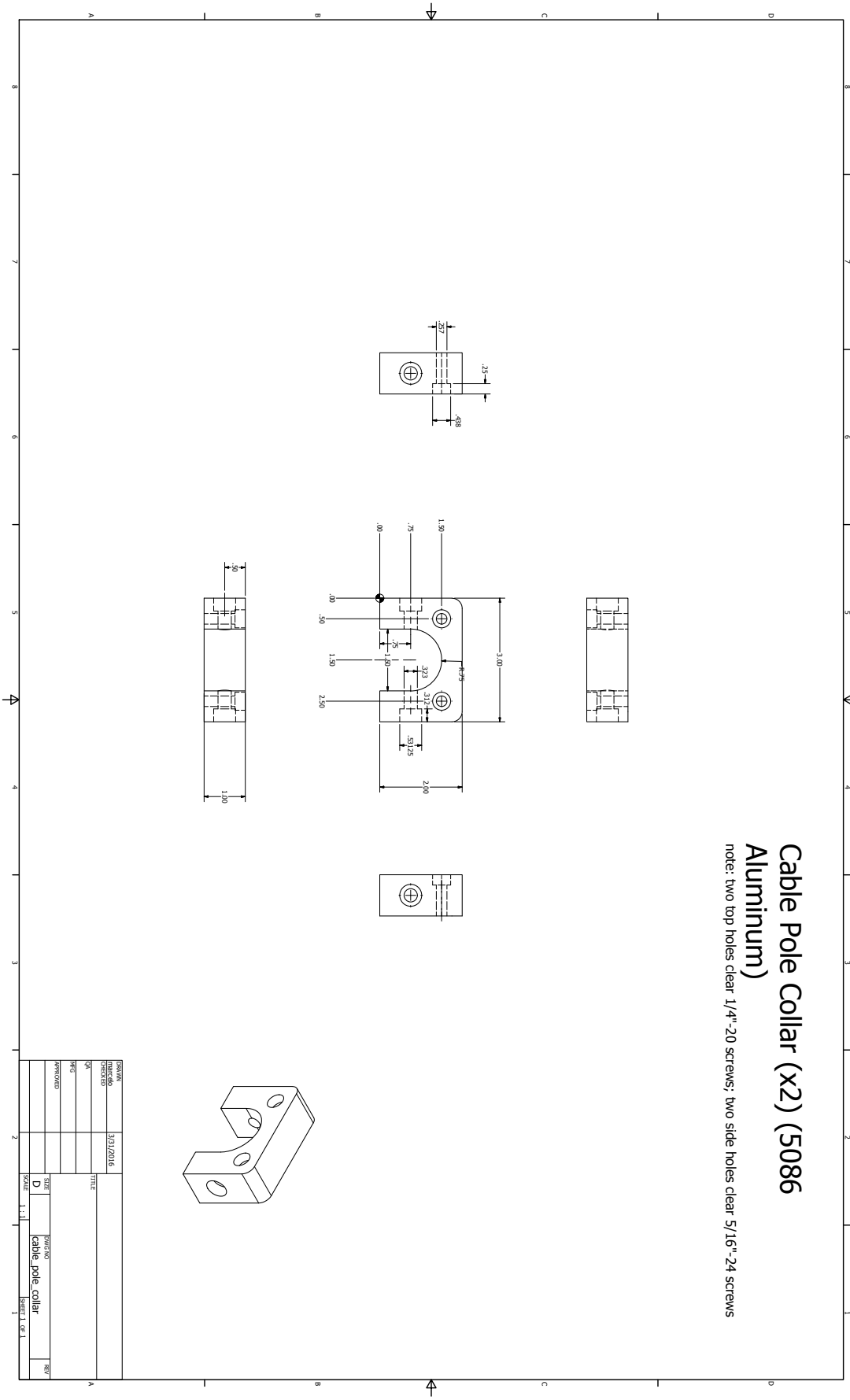

Left Buttress for wide Adaptor (6061 Aluminum) x 4
note: together, the front bracket, adaptor, buttresses and support rods must hang as close to perfectly vertical as possible -- where not possible, a back slant is preferred to a front slant.
Ø.3125
3.0
2.50
1.0
.50
.00
two identical holes tapped for 3/8-16 socket cap screws
note: counterbores on opposite side
2.25
1.25
three identical thru holes (with counterbores) to pass 1/4-20 socket cap screws
3.25
4.25
Ø.257
1.00
hole tapped for 1/4-20 socket cap screw
6.0
Ø.201
3.00
1.0
.50
.00
DRAWN
mwc@0
CHECKED
QA
MFG
APPROVED
6/28/2016
TITLE
SIZE
D
DWG NO
buttress_4_wide_adaptor_left
REV
SCALE
1 : 1
SHEET 1 OF 1

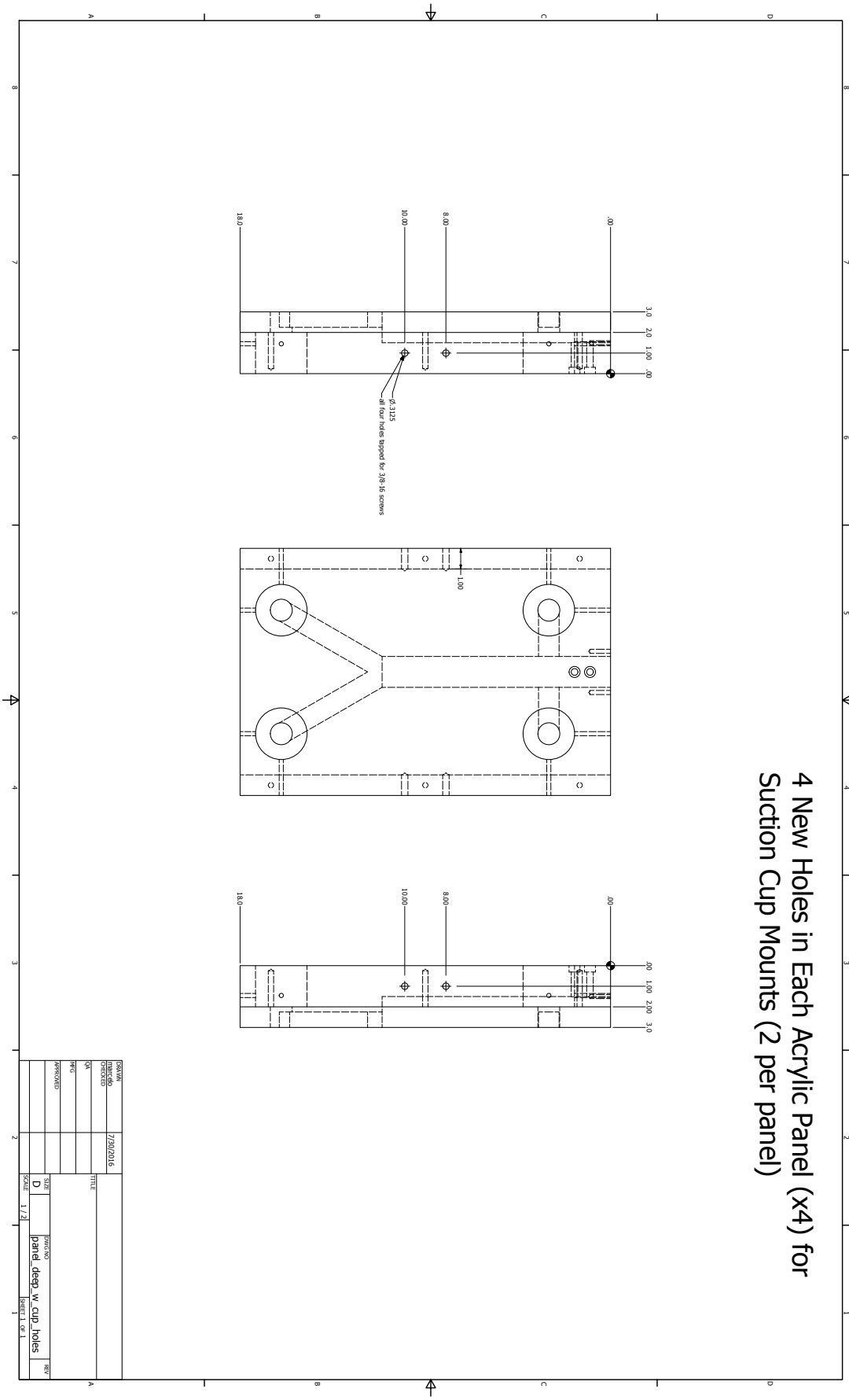

Adaptor, widened (6061 Aluminum)
note: together, the front bracket, adaptor, buttresses and support rods must hang as close to perfectly vertical as possible -- where not possible, a back slant is preferred to a front slant
thru hole (with counterbore) to closely pass 3/8-16 socket cap screw (this side top)
Ø1.2
3.00
1.25
2.50
2.375
2.00
.625
.50
.00
.50
1.50
2.50
3.75
7.00
10.25
11.50
12.50
13.50
14.00
DRAWN
mwc@b
CHECKED
QA
MFG
APPROVED
6/17/2016
TITLE
SIZE
D
DWG NO
adaptor_wide
REV
SCALE
1 / 1
SHEET 1 OF 1

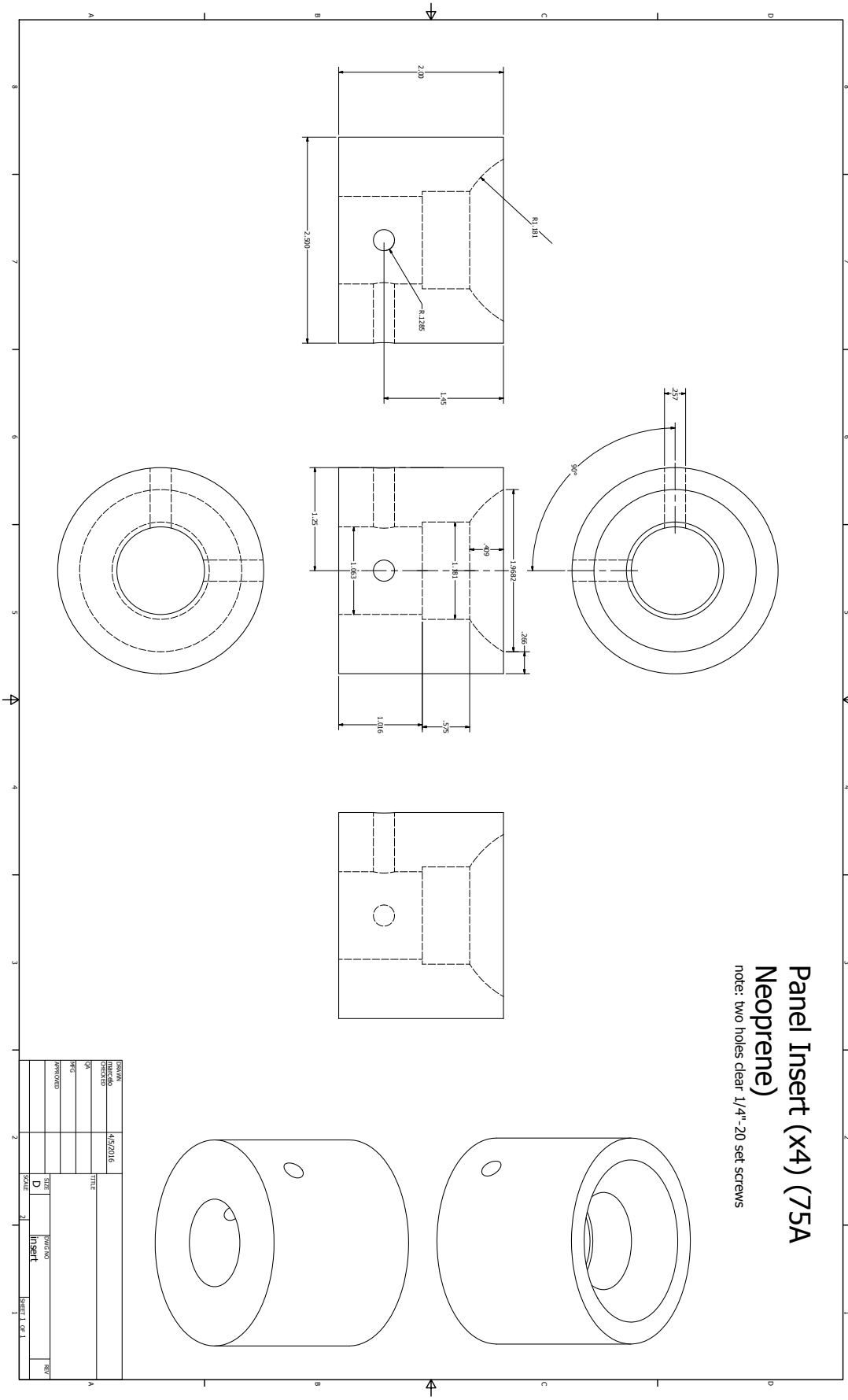

Right Support Rod (FRP)
note:
for Detail A/D: top hole tapped for 3/8"-16 screw; side holes tapped for 1/4"-20 screws; bottom hole tapped for 5/16"-18 screw
for Detail B/C: counterbores must be PRECISELY 0.125" for proper bumper insertion; holes threaded for 6-32 screws
for Detail E/F: non-bumper holes (3) clear 5/16"-18 screws
LEFT SUPPORT ROD identical to Right Support Rod, but 3 "side holes" (for adaptor) in Detail A/D are moved to opposite face
DETAIL A SCALE 0.24 : 1
DETAIL B SCALE 0.24 : 1
DETAIL C SCALE 0.24 : 1
DETAIL D SCALE 0.24 : 1
DETAIL E SCALE 0.24 : 1
DETAIL F SCALE 0.24 : 1
center pole, rotated 90 degrees clockwise
note: three middle holes are visualized from rear
84.00
63.25
49.25
35.25
20.25
12.75
5.25
.00
DRAWN
CHECKED
QA
MFG
APPROVED
4/6/2016
TITLE
SIZE
D
DWG NO
support_rod_right
REV
SCALE 0.12 : 1
SHEET 1 OF 1

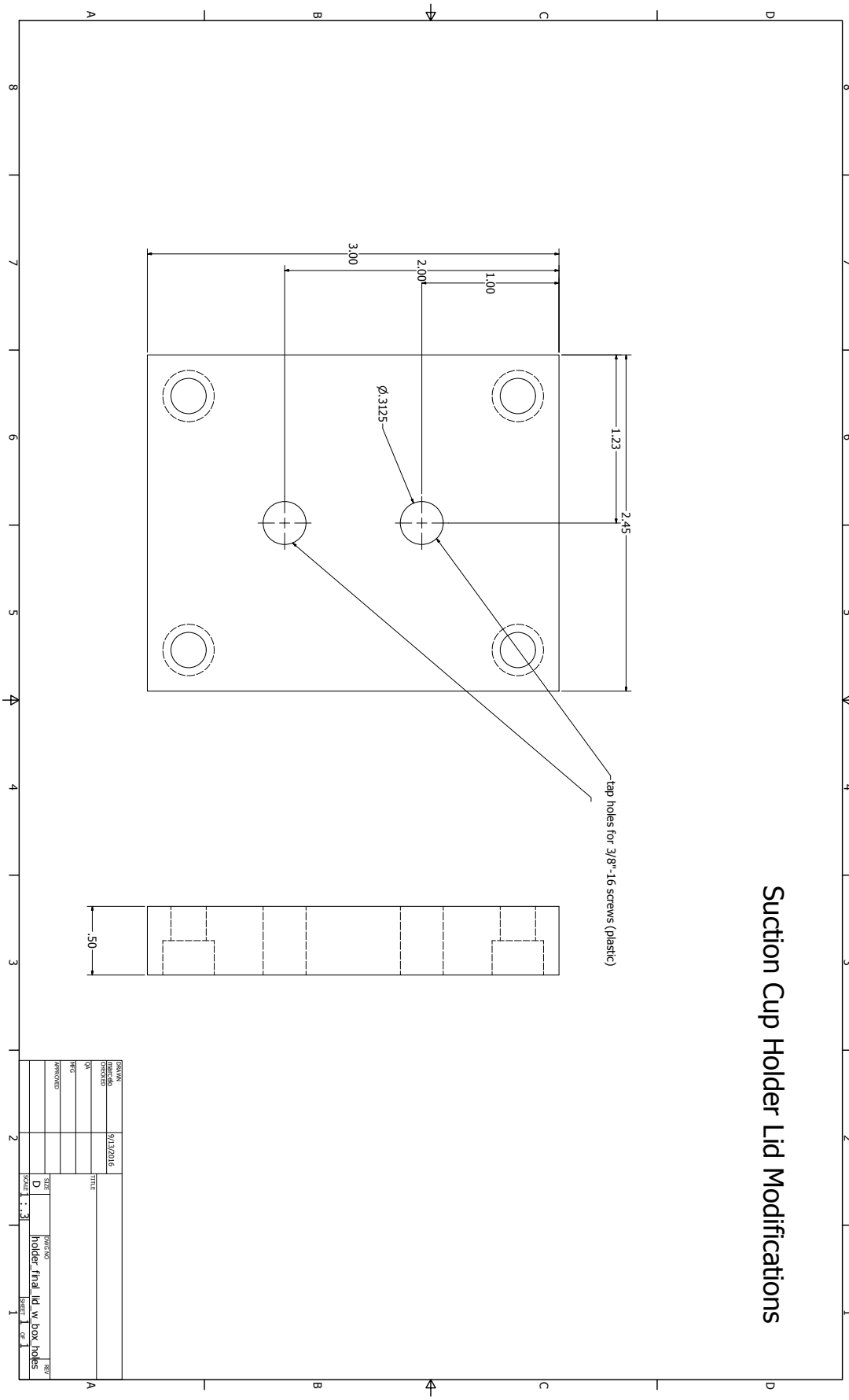

Left Support Rod (FRP)
note:
for Detail A/D: top hole tapped for 3/8"-16 screw; side holes tapped for 1/4"-20 screws; bottom hole tapped for 5/16"-18 screw
for Detail B/C: counterbores must be PRECISELY 0.125" for proper bumper insertion; holes threaded for 6-32 screws
for Detail E/F: non-bumper holes (3) clear 5/16"-18 screws
RIGHT SUPPORT ROD identical to Left Support Rod, but 3 "side holes" (for adaptor) in Detail A/D are moved to opposite face
Bumper Locations
84.00
center pole, rotated 90 degrees counter-clockwise
DETAIL B
SCALE 0.24 : 1
DETAIL C
SCALE 0.24 : 1
DETAIL A
SCALE 0.24 : 1
DETAIL E
SCALE 0.24 : 1
DETAIL D
SCALE 0.24 : 1
DETAIL F
SCALE 0.24 : 1
center pole, rotated 90 degrees clockwise
DRAWN
CHECKED
QA
MFG
APPROVED
3/31/2016
TITLE
SIZE
D
SCALE 0.12 : 1
DWG NO
support_rod_left
REV
SHEET 1 OF 1

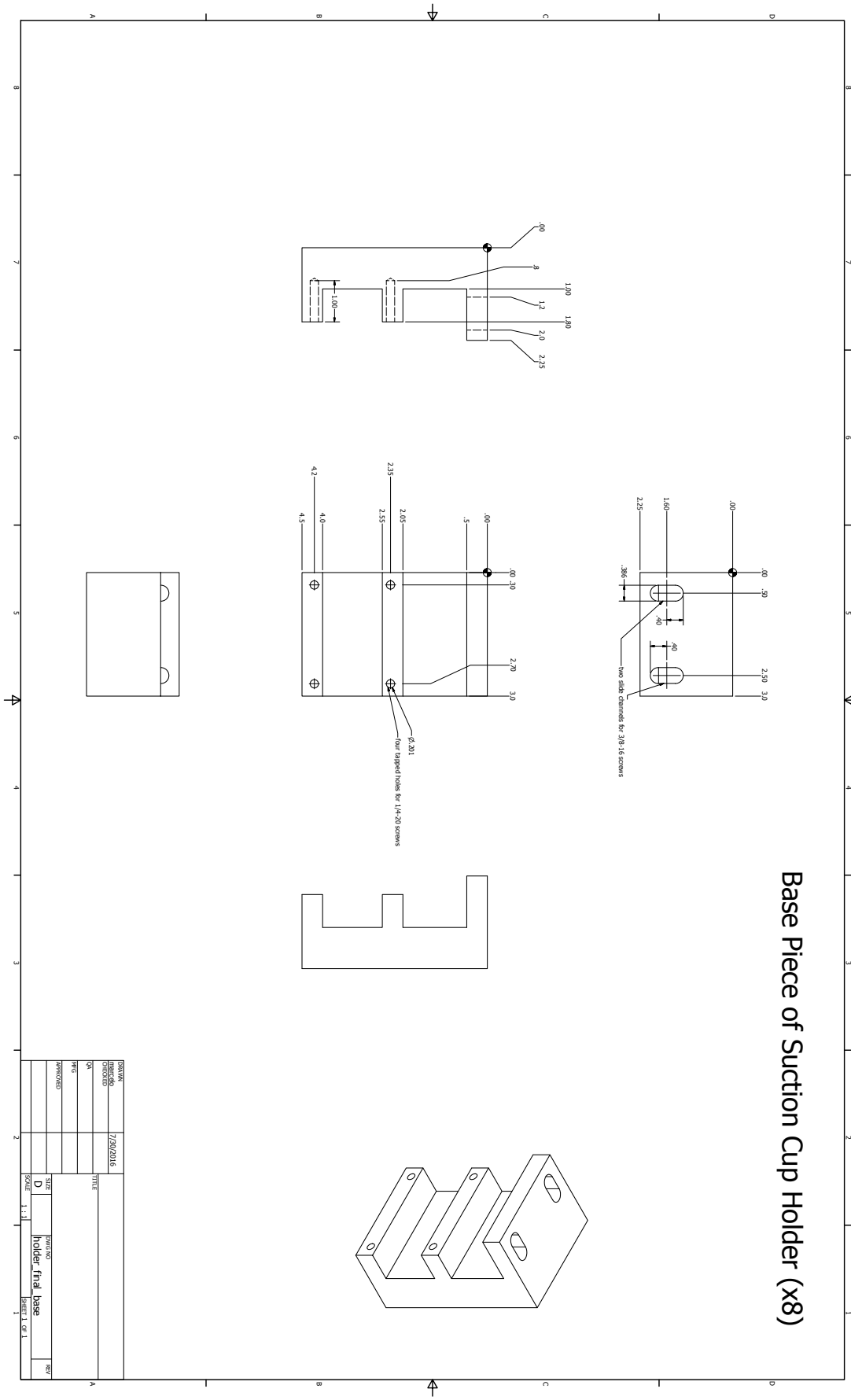

Cable Pole (6063 Aluminum)
center pole, rotated 90 degrees counter clockwise
center pole, rotated 90 degrees clockwise
DETAIL H SCALE 0.30 : 1
DETAIL J SCALE 0.30 : 1
DETAIL K SCALE 0.30 : 1
DETAIL L SCALE 0.30 : 1
DETAIL M SCALE 0.30 : 1
DETAIL N SCALE 0.30 : 1
thru hole tapped for 5/16-24
warning: dotted lines indicate that channel on opposite side of one depicted
Ø1.50
Ø1.25
72.0
69.0
60.0
59.0
9.0
6.0
5.0
.0
7.5
6.75
1.5
.75
72.00
69.00
60.00
59.00
9.00
6.00
5.00
.00
DRAWN
CHECKED
QA
MFG
APPROVED
3/31/2016
TITLE
SIZE
D
DWG NO
cable_pole
REV
SCALE 0.15 : 1
SHEET 1 OF 1

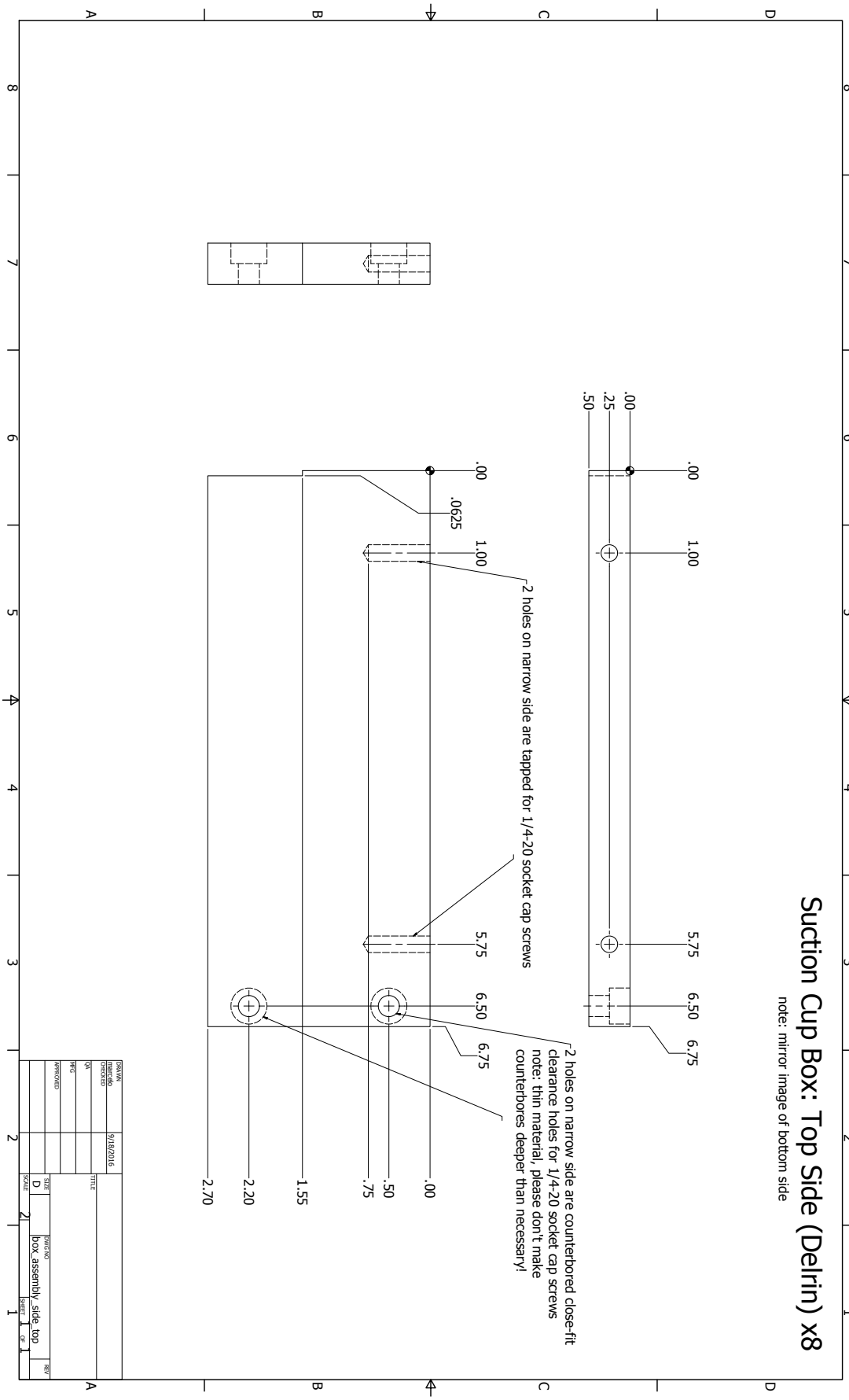

Cable Pole Collar (x2) (5086 Aluminum)
note: two top holes clear 1/4"-20 screws; two side holes clear 5/16"-24 screws
DRAWN
CHECKED
QA
MFG
APPROVED
3/31/2016
TITLE
SIZE
D
SCALE
1 : 1
DWG NO
cable_pole_collar
REV
SHEET 1 OF 1

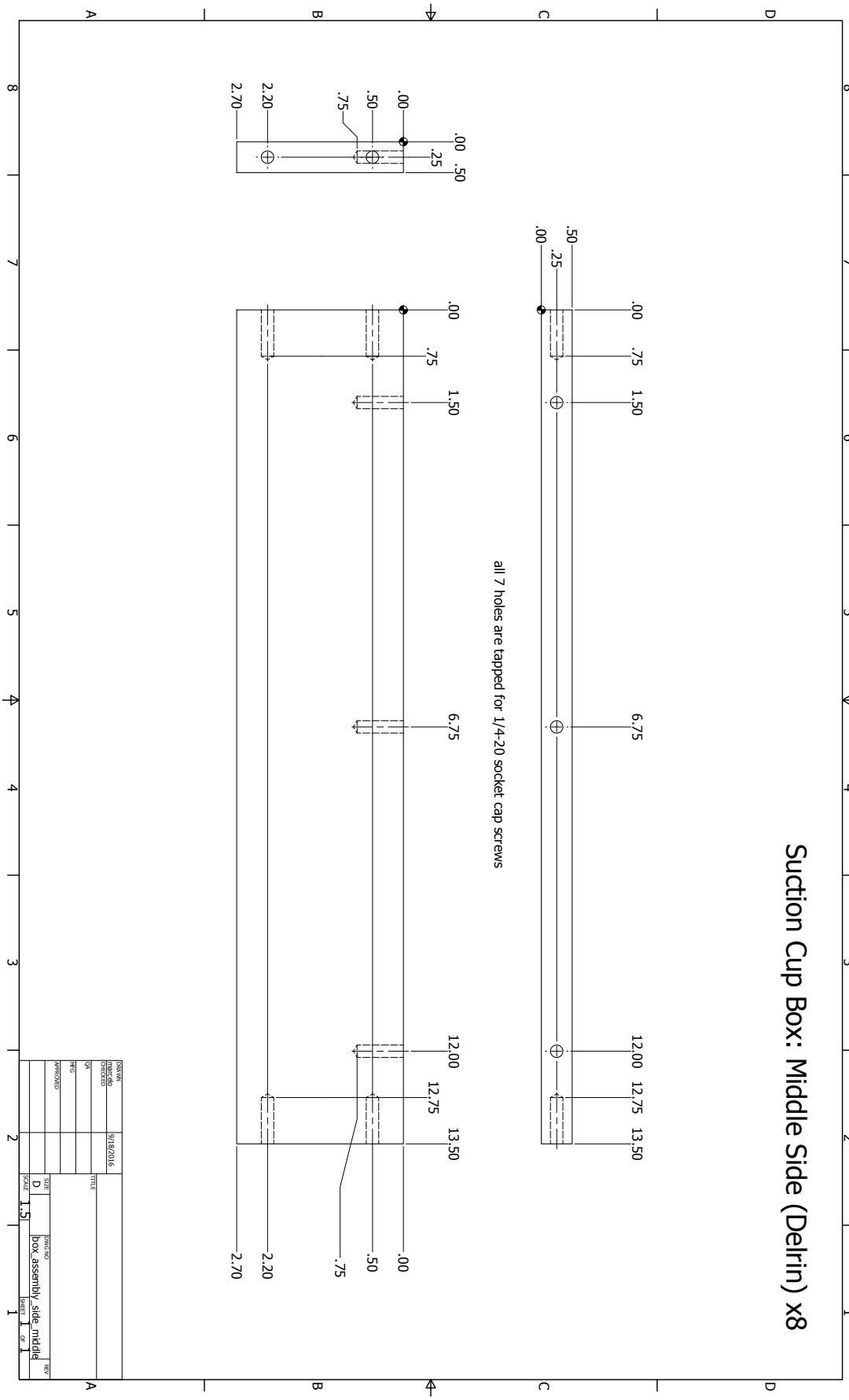

4 New Holes in Each Acrylic Panel (x4) for
Suction Cup Mounts (2 per panel)
18.0
10.00
8.00
.00
3.0
2.0
1.00
.00
Ø.3125
all four holes tapped for 3/8-16 sizes
1.00
18.0
10.00
8.00
.00
.00
1.00
2.00
3.0
DRAWN
mwc@ro
CHECKED
QA
MFG
APPROVED
7/30/2016
TITLE
SIZE
D
SCALE
1 / 2
DWG NO
panel_deep_w_cup_holes
REV
SHEET 1 OF 1

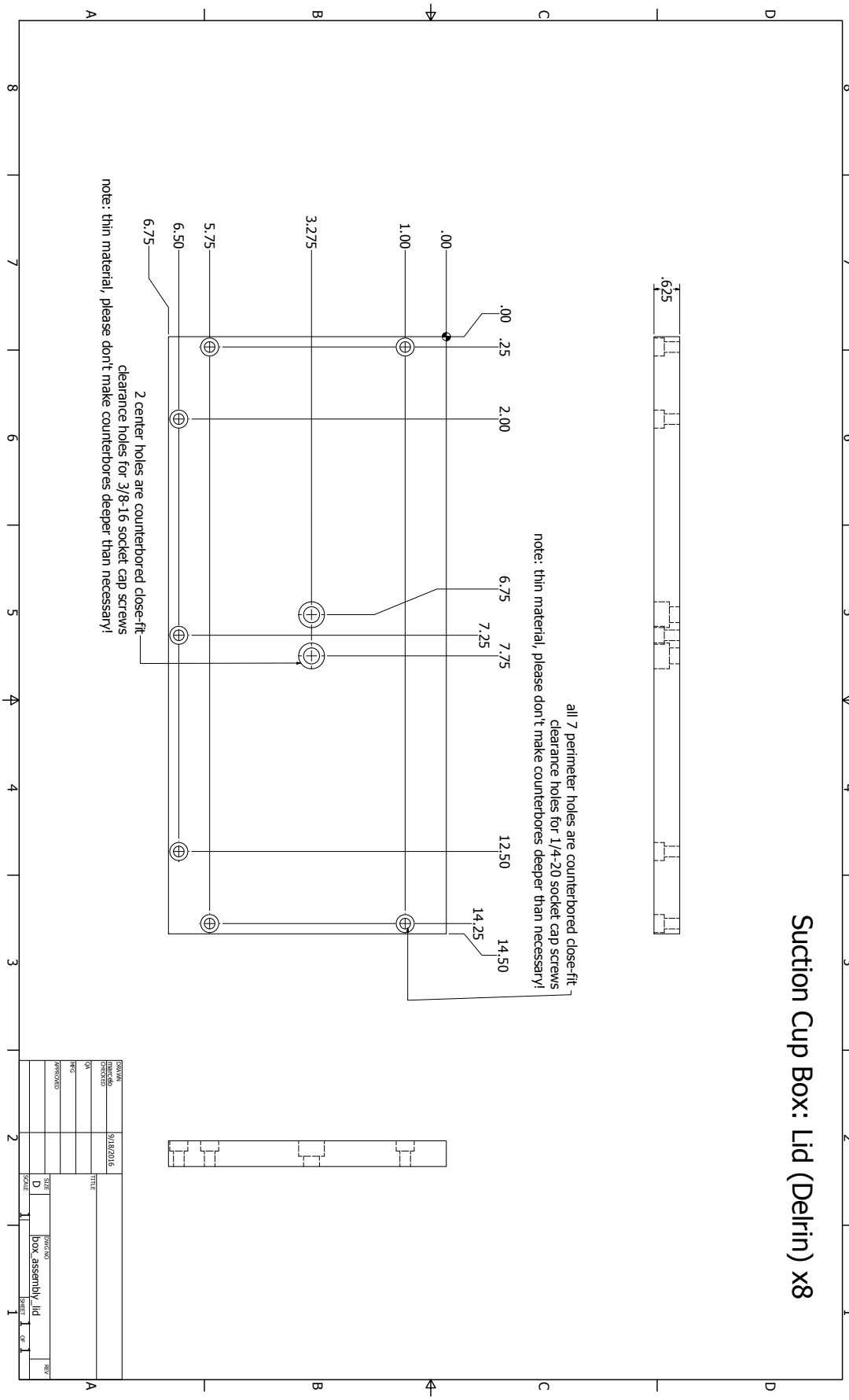

Panel Insert (x4) (75A Neoprene)
note: two holes clear 1/4"-20 set screws
2.00
2.500
R1.181
R.1285
1.45
.257
90°
1.25
.409
1.063
1.181
1.9682
.266
1.016
.575
4/5/2016
Insert
SIZE
D
SCALE
SHEET 1 OF 1
REV


Suction Cup Holder Lid Modifications
3.00
2.00
1.00
Ø.3125
1.23
2.45
tap holes for 3/8"-16 screws (plastic)
.50
DRAWN
CHECKED
QA
MFG
APPROVED
9/13/2016
TITLE
SIZE
D
SCALE 1 : 3
DWG NO
holder_final_lid_w_box_holes
REV
SHEET 1 OF 1


Base Piece of Suction Cup Holder (x8)
.00
.8
1.00
1.2
1.80
2.0
2.25
1.00
.00
1.60
2.25
.386
.50
.40
.40
2.50
3.0
two slide channels for 3/8-16 screws
4.2
4.0
4.5
2.35
2.05
2.55
.5
.00
.30
2.70
3.0
Ø.201
four tapped holes for 1/4-20 screws
DRAWN
CHECKED
QA
MFG
APPROVED
7/30/2016
TITLE
SIZE
D
SCALE
1:3
DWG NO
holder_final_base
REV
SHEET 1 OF 1


Suction Cup Box: Top Side (Delrin) x8
note: mirror image of bottom side
.00
.25
.50
1.00
5.75
6.50
6.75
2 holes on narrow side are counterbored close-fit clearance holes for 1/4-20 socket cap screws
note: thin material, please don't make counterbores deeper than necessary!
.00
.0625
1.00
2 holes on narrow side are tapped for 1/4-20 socket cap screws
5.75
6.50
6.75
.00
.50
.75
1.55
2.20
2.70
DRAWN
CHECKED
QA
MFG
APPROVED
9/18/2016
TITLE
SIZE
D
SCALE
2:1
DWG NO
box_assembly_side_top
SHEET 1 OF 1
REV


Suction Cup Box: Middle Side (Delrin) x8
all 7 holes are tapped for 1/4-20 socket cap screws
.00
.25
.50
.00
.75
1.50
6.75
12.00
12.75
13.50
.00
.50
.75
2.20
2.70
DRAWN
CHECKED
QA
MFG
APPROVED
9/18/2016
TITLE
SIZE
D
SCALE
1.5
DWG NO
box_assembly_side_middle
REV
SHEET 1 OF 1


Suction Cup Box: Lid (Delrin) x8
.625
all 7 perimeter holes are counterbored close-fit
clearance holes for 1/4-20 socket cap screws
note: thin material, please don't make counterbores deeper than necessary!
.00
1.00
3.275
5.75
6.50
6.75
.00
.25
2.00
6.75
7.25
7.75
12.50
14.25
14.50
2 center holes are counterbored close-fit
clearance holes for 3/8-16 socket cap screws
note: thin material, please don't make counterbores deeper than necessary!
DRAWN
CHECKED
QA
MFG
APPROVED
9/18/2016
TITLE
SIZE
D
SCALE
1:1
DWG NO
box_assembly_lid
SHEET 1 OF 1
REV

MASTER 4-PANEL SCREW LIST

DELRIN in white/light color if possible

SCREWS BY COMPONENT
6 (x4) 1.5″-long 3/8-16 winged SS thumb screws (see previous orders — HOME DEPOT) — bracket tightening —> REVISED USE SOCKET CAP
6 (x4) 1″-long 3/8-16 fully-threaded SS cap-head screws WITH washers — top plate /brackets
6 (x6) 1.5″-long, fully-threaded 3/8-16 DELRIN socket cap screws — bracket tighteners
6 (x6) 1″-long, fully-threaded 3/8-16 DELRIN socket cap screws — top plate/ brackets
6 (x6) PLASTIC washers for 3/18-16 screws — top plate/brackets
5 (x6) 1.25″-long, fully-threaded 3/8-16 DELRIN socket cap screws — adaptor/ front bracket
2 (x6) 2″-long, fully-threaded 3/8-16 DELRIN socket cap screws — adaptor/ support rods
4 (x6) 2″-long, fully-threaded 3/8-16 DELRIN socket cap screws — adaptor/ buttresses
6 (x6) 1.5″-long, partially-threaded 1/4-20 DELRIN socket cap screws — buttresses/support rods
2 (x6) 1.75″-long, fully-threaded 3/8-16 DELRIN socket cap screws — buttresses/ cross bar
2 (x6) 2.25″-long, partially-threaded 5/16-18 DELRIN socket cap screws — cross bar/ support rods

4 (x6) 1.5″-long, fully-threaded 1/4-20 DELRIN socket cap screws — cable pole collars/acrylic
4 (x6) 0.75″-long, fully-threaded 5/16-28 DELRIN socket cap screws — cable pole collars/pole
2 (x6) 1.5″-long, fully-threaded 5/16-28 DELRIN socket cap screws — pole/panel

6 (x6) 2″-long, fully-threaded 5/16-18 titanium socket cap screws — panel/ support rods

8 (x4) 2.5″-long 1/4-20 set screws in PLASTIC — INSERTS/panel

12 (x6) 0.5″-long 10-32 wood screws in titanium — bumpers/support rods

Bibliography

SOund SUrveillance System (SOSUS). URL `https://www.pmel.noaa.gov/acoustics/sosus.html`.

Tomonari Akamatsu, Ding Wang, Kexiong Wang, and Yasuhiko Naito. A method for individual identification of echolocation signals in free-ranging finless porpoises carrying data loggers. *The Journal of the Acoustical Society of America*, 108(3):1353–5, 2000.

W W L Au. Echolocation in Dolphins. In *Hearing by Whales and Dolphins*. Springer, New York City, New York, 2000.

Whitlow W L Au and James A Simmons. Echolocation in dolphins and bats. *Physics Today*, 60(9):40–45, September 2007.

Whitlow W L Au, Patrick W B Moore, and Deborah Pawloski. Echolocation transmitting beam of the Atlantic bottlenose dolphin. *The Journal of the Acoustical Society of America*, 80:688–691, 1986.

Jay Barlow and Barbara L Taylor. Estimates of Sperm Whale Abundance in the Northeastern Temperate Pacific from a Combined Acoustic and Visual Survey. *Marine Mammal Science*, 21(3):429–445, 2005.

Marie-Claire Anne Beaulieu. *The Sea as a Two-Way Passage between Life and Death in Greek Mythology*. PhD thesis, The University of Texas at Austin, April 2008.

Bradley M Bell and Terry E Ewart. Separating Multipaths by Global Optimization of a Multidimensional Matched Filter. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, ASSP-34(5):1029–1036, October 1986.

Brian K Branstetter, Patrick W Moore, James J Finneran, Megan N Tormey, and Hitomi Aihara. Directional properties of bottlenose dolphin (Tursiops truncatus) clicks, burst-pulse, and whistle sounds. *The Journal of the Acoustical Society of America*, 131(2): 1613–1621, February 2012.

Leo Breiman, Jerome Friedman, Charles J Stone, and R A Olshen. *Classification and Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1984.

Melba C Caldwell and David K Caldwell. Individualized Whistle Contours in Bottlenosed Dolphins (*Tursiops truncatus*). *Nature*, 207(1):434–435, July 1965.

Melba C Caldwell and David K Caldwell. Vocalization of Naive Captive Dolphins in Small Groups. *Science*, 159(3819):1121–1123, March 1968.

Melba C Caldwell, David K Caldwell, and Peter L Tyack. Review of the signature-whistle-hypothesis for the Atlantic bottlenose dolphin. In S Leatherwood and R R Reeves, editors, *The Bottlenose Dolphin*, pages 199–234. San Diego, 1990.

Chris Catton. *Dolphins*. St. Martin's Press, October 1995.

Y T Chan and K C Ho. A Simple and Efficient Estimator for Hyperbolic Location. *IEEE Transactions on Signal Processing*, 42(8):1905–1915, 1994.

R C Connor, M R Heithaus, and L M Barre. Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'. *Proceedings of the Royal Society B: Biological Sciences*, 268(1464):263–267, February 2001.

Richard C Connor and Rachel A Smolker. 'Pop' Goes the Dolphin: A Vocalization Male Bottlenose Dolphins Produce during Consortships. *Behaviour*, 133(9/10):643–662, August 1996.

Richard C Connor, Randall S Wells, Janet Mann, and Andrew J Read. The bottlenose dolphin: Social relationships in a fission-fusion society. In *Cetacean societies Field studies of whales and dolphins.*, pages 1–37. 2000.

William C Cummings and Paul O Thompson. Characteristics and seasons of blue and finback whale sounds along the U.S. west coast as recorded at SOSUS stations. *The Journal of the Acoustical Society of America*, 95(5):2853–2853, May 1994.

V A Del Grosso. New equation for the speed of sound in natural waters (with comparisons to other equations). *The Journal of the Acoustical Society of America*, 56(4):1084–1091, October 1974.

E D Di Claudio and R Parisi. Robust ML wideband beamforming in reverberant fields. *IEEE Transactions on Signal Processing*, 51(2):338–349, February 2003.

John J Dreher. Linguistic Considerations of Porpoise Sounds. *The Journal of the Acoustical Society of America*, 33(12):1799–1800, December 1961.

K M Dudzinski, C W Clark, and B Wursig. A mobile video/acoustic system for simultaneous underwater recording of dolphin interactions. *Aquatic Mammals*, 21(3):187–193, 1995.

Kathleen M Dudzinski, Justin Gregg, Kelly Melillo-Sweeting, Briana Seay, Alexis Levengood, and Stan A Kuczaj II. Tactile Contact Exchanges Between Dolphins: Self-rubbing versus Inter-Individual Contact in Three Species from Three Geographies. *International Journal of Comparative Psychology*, pages 1–24, January 2012.

R I M Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 20:469–493, 1992.

Brian D Dushaw, Peter F Worcester, Bruce D Cornuelle, and Bruce M Howe. On equations for the speed of sound in seawater. *The Journal of the Acoustical Society of America*, 93 (1):255–275, January 1993.

H Carter Esch, Laela S Sayigh, and Randall S Wells. Quantifying parameters of bottlenose dolphin signature whistles. *Marine Mammal Science*, 25(4):976–986, October 2009.

Lee E Freitag and Peter L Tyack. Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks. *The Journal of the Acoustical Society of America*, 93(4):2197–2205, April 1993.

Deborah Redish Fripp. Bubblestream Whistles Are Not Representative of a Bottlenose Dolphin's Vocal Repertoire. *Marine Mammal Science*, 21(1):29–44, January 2005.

Takeshi Furuichi, Richard Connor, and Chie Hashimoto. Non-conceptive Sexual Interactions in Monkeys, Apes and Dolphins. In J Yamagiwa and L Karczmarski, editors, *Primates and Cetaceans*. Springer, Tokyo, 2014.

Gordon Gallup Jr. Chimpanzees: Self-Recognition. *Science*, 167:86–87, January 1970.

Andrew Greensted. Delay Sum Beamforming, 2012. URL `http://www.labbookpages.co.uk/audio/beamforming/delaySum.html`.

Ulrike Griebel and Leo Peichl. Colour vision in aquatic mammals—facts and open questions. *Aquatic Mammals*, 29(1):18–30, January 2003.

Ulrike Griebel and Axel Schmid. Spectral sensitivity and Color Vision in the Bottlenose Dolphin (Tursiops Truncatus). *Marine and Freshwater Behaviour and Physiology*, 35(3): 129–137, January 2002.

L M Herman and R Kastelein. Cognitive performance of dolphins in visually-guided tasks. In J Thomas, editor, *Sensory Abilities of Cetaceans*. New York, 1990.

Louis M Herman. What Laboratory Research has Told Us about Dolphin Cognition. *International Journal of Comparative Psychology*, 23:310–330, August 2010.

Denise L Herzing. Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*. *Aquatic Mammals*, 22(2):61–79, 1996.

Denise L Herzing. Acoustics and Social Behavior of Wild Dolphins: Implications for a Sound Society. In Whitlow W L Au, Arthur N Popper, and Richard R Fay, editors, *Hearing by Whales and Dolphins*, pages 225–272. New York, 2000.

C W Horton Sr. Dispersion relationships in sediments and sea water. *The Journal of the Acoustical Society of America*, 55(3):547–549, March 1974.

Jason T Isaacs, Daniel J Klein, and Joao P Hespanha. Optimal Sensor Placement for Time Difference of Arrival Localization. In *Joint th IEEE Conference on Decision and Control and th Chinese Control Conference*, pages 7878–7844, October 2009.

V M Janik and M Thompson. A Two-Dimensional Acoustic Localization System for Marine Mammals. *Marine Mammal Science*, 16(2):437–447, April 2000.

Vincent M Janik. Source levels and the estimated active space of bottlenose dolphin (Tursiops truncatus) whistles in the Moray Firth, Scotland. *Journal of Comparative Physiology A*, 186(7-8):673–680, June 2000.

Vincent M Janik and Laela S Sayigh. Communication in bottlenose dolphins: 50 years of signature whistle research. *Journal of Comparative Physiology A*, 199(6):479–489, May 2013.

Vincent M Janik and Peter J B Slater. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behavior*, 56:829–838, September 1998.

Vincent M Janik, Stephanie L King, Laela S Sayigh, and Randall S Wells. Identifying signature whistles from recordings of groups of unrestrained bottlenose dolphins (*Tursiops truncatus*). *Marine Mammal Science*, 29(1):109–122, January 2013.

C S Johnson. Discussion. In R G Busnel, editor, *Animal Sonar Systems Biology and Bionics*, pages 384–398. Jouy-en-Josas, France, 1967.

J Daisy Kaplan, Kelly Melillo-Sweeting, and Diana Reiss. Biphonal calls in Atlantic spotted dolphins (Stenella frontalis): bitonal and burst-pulse whistles. *Bioacoustics*, 27(2):145–164, April 2017.

Charles H Knapp and Clifford Carter. The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, 24(4):320–327, August 1976.

M Krutzen, W B Sherwin, R C Connor, L M Barre, T Van de Casteele, J Mann, and R Brooks. Contrasting relatedness patterns in bottlenose dolphins (Tursiops sp.) with different alliance strategies. *Proceedings of the Royal Society B: Biological Sciences*, 270 (1514):497–502, March 2003.

Michael Krutzen, Janet Mann, Michael R Heithaus, Richard C Connor, Lars Bejder, and William B Sherwin. Cultural transmission of tool use in bottlenose dolphins. *Proceedings of the National Academy of Sciences*, 102(25):8939–8943, June 2005.

Marc O Lammers and Whitlow W L Au. Directionality in the whistles of Hawaiian Spinner Dolphins (*Stenella longirostris*): a signal feature to cue direction of movement? *Marine Mammal Science*, 19(2):249–264, April 2003.

Claude C Leroy and F Parthiot. Depth-pressure relationships in the oceans and seas. *The Journal of the Acoustical Society of America*, 103:1346–1352, March 1998.

Claude C Leroy, Stephen P Robinson, and Mike J Goldsmith. A new equation for the accurate calculation of sound speed in all oceans. *The Journal of the Acoustical Society of America*, 124(5):2774–2782, November 2008.

Xinya Li, Z Daniel Deng, Yannan Sun, Jayson J Martinez, Tao Fu, Geoffrey A McMichael, and Thomas J Carlson. A 3D approximate maximum likelihood solver for localization of fish implanted with acoustic transmitters. *Scientific Reports*, 4(1):668–9, November 2014.

Xinya Li, Zhiqun Daniel Deng, Lynn T Rauchenstein, and Thomas J Carlson. Source-localization algorithms and applications using time of arrival and time difference of arrival measurements. *Review of Scientific Instruments*, 87(4):041502–13, April 2016.

John C Lilly. Vocal Mimicry in Tursiops: Ability to Match Numbers and Durations of Human Vocal Bursts. *Science*, 147(3655):300–301, January 1965.

R M López-Rivas and C Bazúa-Durán. Who is whistling? Localizing and identifying phonating dolphins in captivity. *Applied Acoustics*, 71(11):1057–1062, November 2010.

Lori Marino, Daniel Sol, Kristen Toren, and Louis Lefebvre. Does diving limit brain size in Cetaceans? *Marine Mammal Science*, 22(2):413–425, April 2006.

Arthur F McBride and D O Herb. Behavior of the Captive Bottle-nose Dolphin, *Tursiops truncatus*. *Journal of Comparative and Physiological Psychology*, 41(2):111–123, 1948.

Brenda McCowan. A New Quantitative Technique for Categorizing Whistles Using Simulated Signals and Whistles from Captive Bottlenose Dolphins (Delphinidae, *Tursiops truncatus*). *Ethology*, 100:177–193, 1995.

Brenda McCowan and Diana Reiss. Quantitative Comparison of Whistle Repertoires from Captive Adult Bottlenose Dolphins (Delphinidae, *Tursiops truncatus*): a Re-evaluation of the Signature Whistle Hypothesis. *Ethology*, 100:194–209, 1995a.

Brenda McCowan and Diana Reiss. Whistle Contour Development in Captive-Born Infant Dolphins (*Tursiops truncatus*): Role of Learning. *Journal of Comparative Psychology*, 109 (3):242–260, 1995b.

Brenda McCowan and Diana Reiss. The fallacy of 'signature whistles' in bottlenose dolphins: a comparative perspective of 'signature information' in animal vocalizations. *Animal Behaviour*, 62(6):1151–1162, December 2001.

Brenda McCowan, Sean F Hanser, and Laurance R Doyle. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, 57(2):409–419, 1998a.

Brenda McCowan, Diana Reiss, and C Gubbins. Social familiarity influences whistle acoustic structure in adult female bottlenose dolphins (*Tursiops truncatus*). *Aquatic Mammals*, 24 (1):27–40, 1998b.

Brenda McCowan, Lori Marino, Erik Vance, Leah Walke, and Diana Reiss. Bubble Ring Play of Bottlenose Dolphins (*Tursiops truncatus*): Implications for Cognition. *Journal of Comparative Psychology*, 114(1):98–106, 2000.

Brenda McCowan, Laurance R Doyle, and Sean F Hanser. Using Information Theory to Assess the Diversity, Complexity, and Development of Communicative Repertoires. *Journal of Comparative Psychology*, 116(2):166–172, 2002.

T King McCubbin Jr. The Dispersion of the Velocity of Sound in Water between 500 and 1500 Kilocycles. *The Journal of the Acoustical Society of America*, 26(2):247–249, March 1954.

David K Mellinger and Christopher W Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. *The Journal of the Acoustical Society of America*, 107 (6):1–12, May 2000.

Wei Meng, Lihua Xie, and Wendong Xiao. Optimal Sensor Pairing for TDOA based Source Localization and Tracking in Sensor Networks. In *th International Conference on Information Fusion FUSION,*, pages 1897–1902, June 2012.

Joshua P Neunuebel, Adam L Taylor, Ben J Arthur, and SE Roian Egnor. Female mice ultrasonically interact with males during courtship displays. *eLife*, 4:e06203–24, May 2015.

Oppian. Halieutica. In *Oppian, Colluthus, Tryphiodorus*, pages 1–1. 1928.

Julie N Oswald, Jay Barlow, and Thomas F Norris. Acoustic Identification of Nine Delphinid Species in the Eastern Tropical Pacific Ocean. *Marine Mammal Science*, 19(1):20–37, January 2003.

Adam A Pack and Louis M Herman. Sensory integration in the bottlenosed dolphin: Immediate recognition of complex shapes across the senses of echolocation and vision. *The Journal of the Acoustical Society of America*, 98(2):722–733, August 1995.

Elena Papale, Gaspare Buffa, Francesco Filiciotto, Vincenzo Maccarrone, Salvatore Mazzola, Maria Ceraulo, Cristina Giacoma, and Giuseppa Buscaino. Biphonic calls as signature whistles in a free-ranging bottlenose dolphin. *Bioacoustics*, 24(3):223–231, May 2015.

Joshua M Plotnik, Frans B M de Waal, and Diana Reiss. Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, 103(45):17053–17057, November 2006.

Plutarch. De sollertia animalium. In *Moralia, Vol. XII*, pages 1–1. 1956.

Diana Reiss. *The Dolphin in the Mirror*. Exploring Dolphin Minds and Saving Dolphin Lives. Mariner Books, New York, September 2012.

Diana Reiss and Lori Marino. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences*, 98(10):5937–5942, May 2001.

Diana Reiss and Brenda McCowan. Spontaneous Vocal Mimicry and Production by Bottlenose Dolphins (*Tursiops truncatus*): Evidence for Vocal Learning. *Journal of Comparative Psychology*, 107(3):301–312, 1993.

Douglas G Richards, James P Wolz, and Louis M Herman. Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, Tursiops truncatus.. *Journal of Comparative Psychology*, 98(1):10–28, 1984.

S Ridgway, D Samuelson Dibble, K Van Alstyne, and D Price. On doing two things at once: dolphin brain and nose coordinate sonar clicks, buzzes and emotional squeals with social sounds during fish capture. *Journal of Experimental Biology*, 218(24):3987–3995, 2015.

Christopher Riley. The dolphin who loved me: the Nasa funded project that went wrong. *The Guardian*, pages 1–1, February 2014.

John Robbins. *Diet for a New America*. How Your Food Choices Affect Your Health, Happiness, and the Future Life on Earth. Stillpoint Publishing, Walpole, NH, 1987.

B L Sargeant, J Mann, P Berggren, and M Krutzen. Specialization and development of beach hunting, a rare foraging behavior, by wild bottlenose dolphins (Tursiop sp.). *Canadian Journal of Zoology*, 83(11):1400–1410, November 2005.

Laela S Sayigh, H Carter Esch, Randall S Wells, and Vincent M Janik. Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*. *Animal Behaviour*, 74(6):1631–1642, December 2007.

H Shirihai and B Jarrett. *Whales, Dolphins, and Other Marine Mammals of the World.* Princeton University Press, Princeton and Oxford, 2006.

J David Smith. Inaugurating the Study of Animal Metacognition. *International Journal of Comparative Psychology*, 23(3):401–413, 2010.

Julius O Smith and Jonathan S Abel. Closed-Form Least-Squares Source Location Estimation from Range-Difference Measurements. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, ASSP-35(12):1661–1669, December 1987a.

Julius O Smith and Jonathan S Abel. The Spherical Interpolation Method of Source Localization. *IEEE Journal of Oceanic Engineering*, OE-12(1):246–252, January 1987b.

John L Spiesberger. Linking auto- and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipath. *The Journal of the Acoustical Society of America*, 104(1):300–312, July 1998.

John L Spiesberger. Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration. *The Journal of the Acoustical Society of America*, 106(2): 837–846, August 1999.

John L Spiesberger and Kurt M Fristrup. Passive Localization of Calling Animals and Sensing of their Acoustic Environment Using Acoustic Tomography. *The American Naturalist*, 135 (1):107–153, January 1990.

John L Spiesberger and Kurt Metzger. A new algorithm for sound speed in seawater. *The Journal of the Acoustical Society of America*, 89(6):2677–2688, June 1991.

Kathleen M Stafford, Christopher G Fox, and Bruce R Mate. Acoustic detection and location of blue whales (*Balaenoptera musculus*) from SOSUS data by matched filtering. *The Journal of the Acoustical Society of America*, 96(5):3250–3251, November 1994.

Ryuji Suzuki, John R Buck, and Peter L Tyack. The use of Zipf's law in animal communication analysis. *Animal Behaviour*, 69(1):F9–F17, January 2005.

Rebecca E Thomas, Kurt M Fristrup, and Peter L Tyack. Linking the sounds of dolphins to their locations and behavior using video and multichannel acoustic recordings. *The Journal of the Acoustical Society of America*, 112(4):1692–1701, October 2002.

P L Tyack and C W Clark. Acoustic Communication in Whales and Dolphins. In *Hearing by Whales and Dolphins*. Springer, New York City, New York, 2000.

Peter L Tyack. An optical telemetry device to identify which dolphin produces a sound. *The Journal of the Acoustical Society of America*, 78(5):1892–1895, November 1985.

Peter L Tyack. Whistle Repertoires of Two Bottlenosed Dolphins, *Tursiops truncatus*: Mimicry of Signature Whistles? *Behavioral Ecology and Sociobiology*, 18(4):251–257, 1986.

Peter L Tyack. Use of a Telemetry Device to Identify which Dolphin Produces a Sound. In Karen Pryor and Kenneth S Norris, editors, *Dolphin Societies*. Berkeley, 1991.

Bert Van Den Broeck, Alexander Bertrand, Peter Karsmakers, Bart Vanrumste, Hugo Van Hamme, and Marc Moonen. Time-domain generalized cross correlation phase transform sound source localization for small microphone arrays. In *Education and Research Conference EDERC, th European DSP*, September 2013.

J M van der Hoop, A Fahlman, T Hurst, J Rocho-Levine, K A Shorter, V Petrov, and M J Moore. Bottlenose dolphins modify behavior to reduce metabolic effect of tag attachment. *Journal of Experimental Biology*, 217(23):4229–4236, December 2014.

Barry D Van Veen and Kevin M Buckley. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Magazine*, pages 1–21, April 1998.

Megan R Warren, Daniel T Sangiamo, and Joshua P Neunuebel. High channel count microphone array accurately and precisely localizes ultrasonic signals from freely-moving mice. *Journal of Neuroscience Methods*, 297:44–60, March 2018.

William A Watkins and William E Schevill. Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. *Deep Sea Research and Oceanographic Abstracts*, 19(10):691–706, October 1972.

William A Watkins and William E Schevill. Listening to Hawaiian Spinner Porpoises, *Stenella* Cf. *Longirostris*, with a Three-Dimensional Hydrophone Array. *Journal of Mammalogy*, 55 (2):319–328, May 1974.

William A Watkins, Joseph E George, Mary Ann Daher, Kristina Mullin, Darel L Martin, Scott H Haga, and Nancy A DiMarzio. Whale Call Data for the North Pacific November 1995 through July 1999 Occurrence of Calling Whales and Source Locations from SOSUS and Other Acoustic Systems. Technical Report WHOI-00-02, February 2000.

Stephanie L Watwood, Edward C G Owen, Peter L Tyack, and Randall S Wells. Signature whistle use by temporarily restrained and free-swimming bottlenose dolphins, Tursiops truncatus. *Animal Behaviour*, 69(6):1373–1386, June 2005.

Edward C Whitman. SOSUS. *Undersea Warfare*, 7(1), January 2005.

Danuta M Wisniewska, John M Ratcliffe, Kristian Beedholm, Christian B Christensen, Mark Johnson, Jens Koblitz, Magnus Wahlberg, and Peter T Madsen. Range-dependent flexibility in the acoustic field of view of echolocating porpoises(Phocoena phocoena). *eLife*, pages 1–16, March 2015.

Tina M Yack, Jay Barlow, John Calambokidis, Brandon Southall, and Shannon Coates. Passive acoustic monitoring using a towed hydrophone array results in identification of a previouslyunknown beaked whale habitat. *The Journal of the Acoustical Society of America*, 134(3):1–7, September 2013.

Bin Yang. Different Sensor Placement Stategies for TDOA Based Localization. In *IEEE International Conference on Acoustics, Speech and Signal Processing, .*, pages 1093–1096, April 2007.

Walter M X Zimmer. *Passive Acoustic Monitoring of Cetaceans*. Cambridge University Press, Cambridge, 2011.

George Kingsley Zipf. *Human Behavior and the Principle of Least Effort.* Addison-Wesley Press, Cambridge, Mass, 1949.