Do killer whales use pulsed calls as a communication system? Differentiating calls based on characteristic frequencies and harmonics

Dave Cade
Beam Reach Marine Science and Sustainability School
dave102@beamreach.org

Abstract-

A method of determining the distinguishability of killer whale calls is described using fundamental frequencies of harmonic structures, frequencies of maximum energy and characteristic frequencies that have properties similar to the formants of terrestrial mammals. A method for determining how spread out a dataset is within a confined space is also tested. No conclusions were drawn about the likelihood of killer whale calls as communication devices, but more analysis is needed. The potential importance of the second syllable to killer whale calls is discussed.

Introduction-

Much work has been done in recording, categorizing, describing and analyzing the sounds made by killer whales (e.g. Grebner et al. 2009). One of the best studied populations is the resident killer whale population of the NE Pacific, specifically those whales that summer along the Washington and British Columbia coasts. Killer whales (as well as all other members of delphinidae) rely heavily on vocalizations for foraging, navigating and communication (Wieland et al. 2010). Their vocalizations can be categorized into three types: echolocation clicks, tonal whistles and pulsed calls (Ford 1989) that are potentially made with different anatomical structures. It is well established that clicks and pulsed calls result from the movement of air through a structure unique to odontocetes called the *museau de singe* by whalers but more recently called the phonic lips. Traits like click repetition rate are determined by the periodic opening and closing of the lips. Whistles are produced similarly, but it is likely that they are regulated as well by nasal plugs and their nodes in conjunction with the ligaments of the blowhole (Berta et al. 2006).

The Southern Resident population can be divided into sub-groups (clans, pods and matrilines) based on geographic range, behavior (such as beach rubbing) and association, but also by call types that are characteristic of each group (Ford 1987). While whistles have proven to be more difficult to categorize (although Riesch et al. did have interesting results in 2005), non-echolocation pulsed calls that occur in discrete, repetitive patterns have been thoroughly categorized in both Northern and Southern resident populations. The calls of the Southern Resident population (the proposed focal population of this study) have been categorized into 26 discrete call types with 9 subtypes (Ford 1987). While slight changes in calls have been noted over time (Wieland et al. 2010), the classification of the call types has been virtually unchanged in the literature since the original catalogue in 1987.

The purpose of these calls has not been conclusively determined. Attempts have been made to glean information about specific behaviors from calls made by orcas with mixed results (Weiss et al. 2007, Grebner et al. 2009, Parijs et al. 2004). Ford (1991) and others postulate a communicative purpose at the very least relating to the use of calls to maintain contact among pod members. It is a basic assumption in many papers that vocalizations are used in communication and thus research is focused on what it is that they are communicating (e.g. Thomsen et al. 2002). Indeed, it is common when analyzing sounds that animals make for researchers to assume that any loud sound is a signal and thus to infer a communicative purpose (Bradbury and Vehrencamp 1998). Thus far, however, clear, unequivocal evidence that calls are used in killer whale communication remains elusive. The vast majority of scholars agree that animal communication “involves the provision of information by a sender to a receiver, and the subsequent use of that information by the receiver in deciding how to respond” (ibid, p.2). This definition, however, is unsatisfactory on its own because it begs the question of what is meant by information. Bradbury and Vehrencamp (1998) go on to discuss “true communication” as a signal sent that must benefit both the sender and the receiver in some fashion. It is not the purpose of this study to derive what the benefit of these pulsed calls may be to the sender and receiver, but to show that they have evolved in a way so as to maximize the chance that the message of the sender is received as intended. Put another way, the purpose of this study is to present initial evidence that killer whale calls are maximally distinguishable from each other. These findings imply that the call system has evolved over time in a way that maximizes its effectiveness as part of an acoustic communication system.

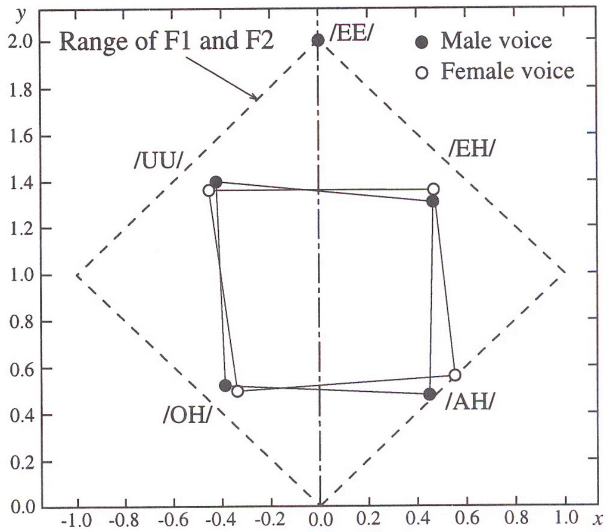
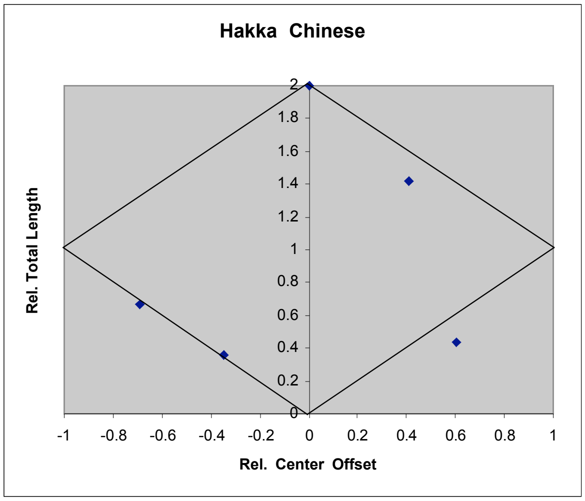
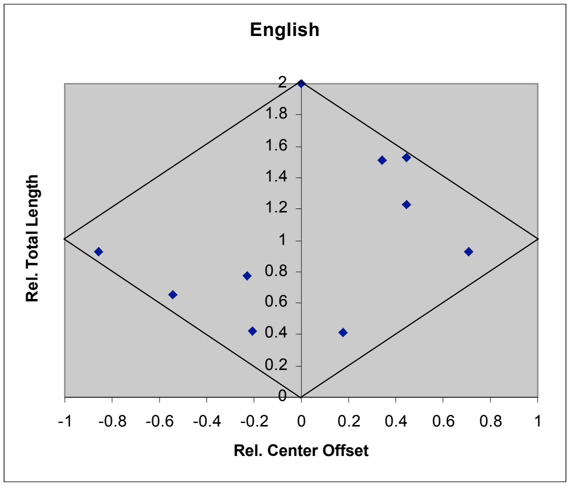


Figure 1- The symmetrical distribution of the five Japanese vowel sounds. The sound /ee/ was shown to have the largest separation of formant values (F1 and F2). The offset from the center of the /ee/ sound and the distance between the F1 and F2 relative to /ee/’s distance were measured for the other four vowels (see discussion in methods section). A logarithmic scale was used since human hearing is based on the octave. The x-axis is offset and the y-axis is relative distance. The dashed line represents the theoretical possible boundaries of the formants. Note that the five sounds are well separated within the boundaries of the diamond. (Transnational College of LEX 2008)

Many current policies for killer whale management have been driven by the evident need of killer whales to communicate with each other vocally in an increasingly noisy environment (for a discussion of anthropogenic noise, see Holt 2008). Although anthropogenic factors are relatively recent additions to ambient noise on an evolutionary scale, the need for killer whales to be understood by other members of their pod has always been present and has likely pushed the development of call sequences into patterns that are readily differentiated. This need parallels the communicative needs of all species, including humans. In 1995, a group of Japanese students at the Transnational College of LEX published what they meant to be a readable account of Fourier analysis. Through their research into the property of waves, they developed a method using analysis of formant frequencies (see discussion of formants below) to describe how the Japanese language has naturally acquired a certain symmetry (figure 1) with regards to the five Japanese vowel sounds. That is, they suggest that:

> [o]ver the millennia, human beings have unconsciously isolated a few sounds that are the best suited for most clearly vocalizing language. The result is [a] symmetrical arrangement [whereby] each vowel has been selected to maximize this symmetry of sounds. (Transnational College of LEX 2002, p. 161)





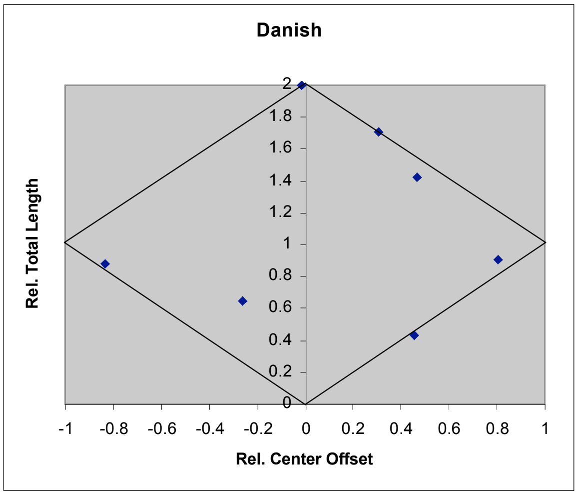


Figure 2- The relative spacing of the formant structure of the vowels of three different languages for adult males. Each point represents a distinct vowel sound. Note the wide spacing of the points that are (with one significant exception) still within the confines of the boundary. Data for Hakka Chinese from Cheung 2007, other chart data from Lindau 1978.

My own analysis (see methods section for description of process) suggests that other human languages also demonstrate this symmetry of vowel sounds (figure 2). If, as suggested above, this symmetry is a result of the unconscious morphing of sounds over time (as language developed) until there were discrete, replicable units (vowels) that were separate enough so as to be readily distinguishable, then it is also likely that vocal communication systems utilized by other animals would have evolved over time to be as aurally distinguishable as possible, and a similar symmetry would be present. If such a property could be demonstrated in killer whale calls, it would strongly suggest that their vocalizations developed and evolved as a communication system. A parallel structure to human vowels can be thought of as, like vowels, sounds that are distinguishable to the ear in a noisy environment. That is, on a windy day or on a bad (analog) telephone connection, vowel sounds are still distinguishable even when the overall meaning of the sound is garbled. As an example picture how easily something shouted over crowd noise can be heard but not understand. For instance, if a new couple is out at dinner in a crowded restaurant, and one person says quietly, “I love sushi,” it could foreseeably be heard as “I love you” because of the similarity of vowel sounds, causing confusion on all fronts. Analogously, it is probable that killer whale call lexicon has a similar structure of sounds that provide character to the overall call and that are distinguishable from each other even in noise or over long distances. This study discusses evidence that the calls of Southern Resident killer whales (SRKWs) demonstrate a pattern similar to that which has been found in the communication systems of humans.

Contributing to the idea that killer whale lexicon has had a chance to evolve over time in a way analogous to language evolution are studies pointing to its changes in the short term (Wieland et al. 2010) and pointing to the development of vocabulary by mimicry of other whales.

Vocal mimicry, the building block for developing language over successive generations, has been shown to exist in few non-human mammals, but does indeed exist in cetaceans (Fitch 2000). There is also evidence that young killer whales learn to make specific calls from more mature killer whales instead of genetically inheriting the ability (National Marine Fisheries Services 2008, Ford 1991). The debate over killer whale "culture" is robust (Rendell and Whitehead 2001), yet there exists significant evidence that certain killer whale behaviors like specific vocalizations or intentional beaching as a technique to hunt seals are not only learned behaviors (in the case of the former) but intentionally taught (in the case of the latter) (Ibid).

In human languages vowels are determined by the structure of their formant frequencies. That is, formants (specifically the first two formants) are regions that, if removed, change the overall character of the sound of the vowel. More formally, a formant is an identifiable region of high energy in the frequency spectrum of a vocalized sound and is assumed to result from the resonance of the vocal tract (O'Shaughnessy 2008). They can be thought of as bandpass filters that allow certain regions of frequencies to pass through. While the vocal cords determine the pitch of a speaker's voice, formants determine the sound. In whispered speech, while no pitch is audible, formant frequencies can still be detected (Fitch 2000).

It is well established that the perception of the sound of a stationary vowel is determined by the first two formants (Kollmeier et al. 2008) and this fact drove the analyses shown in figures 1 and 2. There is precedent for using formant frequencies in other mammals. In red deer, changes in formant frequencies have been used to successfully classify individual animals based on their common roar. The sounds produced by red deer are modulated because of resonances in the supralaryngeal vocal tract. Additionally, variation in the movement of the larynx, mandibles tongue and lips causes individual differences in the formant structure of the calls (Reby et al.

2006). Despite some claims that all mammalian calls have formant structures (Fitch 2000), a problem arises with killer whales (and all odontocetes) that use a completely different set of anatomical structures (as discussed above) to produce pulsed calls. Despite this, many methods developed for analysis of human speech and shown to be applicable to terrestrial mammals have also been used effectively to categorize many types of delphinid vocalizations. Cepstral analysis (which is also used in the identification of formants) has been used effectively to classify odontocetes by species (Roch et al. 2007). Because formant analysis depends on the resonances of structures of the vocal tract, however, it does not appear that killer whale calls can be classified by formant analysis alone. It has also been noted that the difficulties of using formant analysis for killer whale calls could be due to the high value of the fundamental frequency of the calls (Miller et al. 2007). This points to another problem with identifying the formants of killer whale calls by ear. Killer whales have a very different hearing sensitivity than do humans (Figure 3), thus determining the character of a killer whale call by human ear likely does not give the full picture of what the call sounds like to whales. Nonetheless, results for this type of analysis are included.

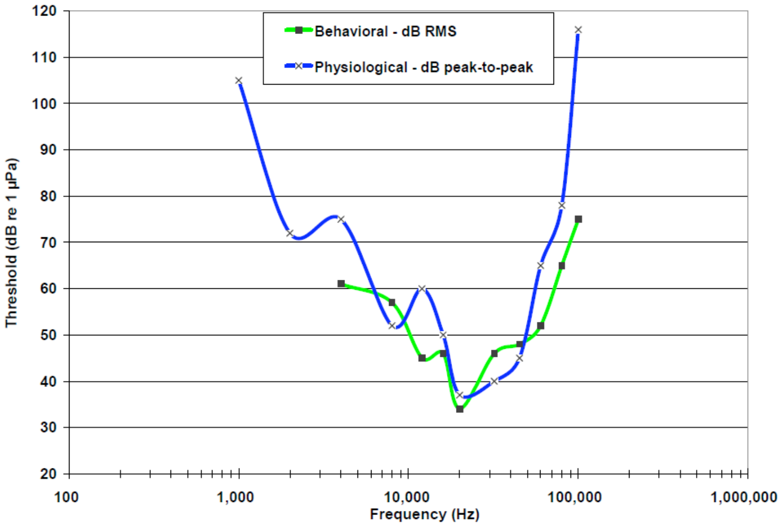


Figure 3- Hearing sensitivity of a captive killer whale (reprinted from Szymanski et al. 1999 as reported by Holt 2008). Compare this to humans, which have a peak frequency sensitivity around 4000 Hz (Sininger et al 1997).

Calls have also long been successfully classified by their harmonic structure (e.g. Ford 1987) and properties related to this structure are also analyzed in this study. All calls contain at least one frequency which, if removed from the spectrum, result in a fundamental change to the sound of the call.[1] Tracking this frequency (analogous to the first formant), any secondary characteristic frequencies (analogous to the second formant), the fundamental frequency (as calculated by the harmonic interval) and the frequency with the highest energy across different call types provided variables that are fundamental to the character of killer whale calls in a way analogous to the characterization of human speech (specifically vowel sounds) by formants (the variables in that case). Statistical methods are employed to determine if these calls indeed have a maximally differentiated structure.

My initial hypothesis was that each syllable of the calls commonly used by SRKWs will have a structure uniquely determined by two of the variables described above. Additionally, I predicted that each syllable would be highly distinguishable from each other in a way analogous to how the formant structure of vowels can be shown to be highly distinguishable. I predicted that this property would show that killer whale calls have undergone evolutionary changes similar to the changes undergone by human language, and thus demonstrate the likelihood of the importance of differentiating each syllable. This importance would indicate that each call has a designated, distinct communicative purpose.

Methods-

Southern Resident killer whale calls were recorded during the months of September and October 2010 in the San Juan Islands region of the Salish Sea. All recordings were done in Haro

---

[1] A prospect for future investigation is to determine whether the characteristic frequencies of each call match the pulse repetition rate. It has been calculated that the pulse repetition rate usually matches with one of the harmonic bands (Deecke et al. 1999)

Strait and Boundary Pass. We used a towed LabCore 40 four-hydrophone array which has a peak frequency response at 5 kHz and a single Cetacean Research Technology (CRT) C54 XRS/266 high-frequency hydrophone with a flat response curve from 1-30 kHz. The hydrophones were towed at a depth of 1.85 m by attaching each cable to a finned weight. The LabCore array was towed off the port stern and the CRT off the starboard stern. Only hydrophones 1, 3, and 4 were used in the array to account for the CRT channel in our 4-channel recording devices (a pair of 2-channel SoundDevices 702 recorders). Recordings were continuous digital files sampled at 192 kHz with a 16-bit depth rate. The files were then down-sampled to 48 kHz in processing and spliced into 1-minute, 4-channel wav files.

General speed of movement during recordings was 2.5 knots through the water. Our vessel was propelled by electric motors which minimize background noise. For most recordings, the array was towed and thus horizontally deployed behind the boat. At times, however, the boat was forced to stop or slow down to comply with the "Be Whale Wise" guidelines and Washington state laws. In such scenarios, the hydrophones lie vertical in the water, but this does not affect the quality of the recordings.

Calls were then typed based on Ford's 1987 call catalog and digital audio examples put together by Val Veirs for the Whale Museum. Many of these sample files were also examined in order to provide examples of known call type for comparison. Each call type was broken up in to syllables based on abrupt frequency changes. These designations are more specific than Ford's original determination of "parts" for Southern Resident calls since his parts sometimes include abrupt frequency shifts, peaks and dips in the spectrum. To be included in analysis, a syllable must have a flat (slope flatter than ± .2) section of length > .1 ms.

The CRT was generally used for analyzing the calls because it has a flatter overall representation of background noise across a larger frequency bandwidth (figure 4). The LabCore array was used only when the received level of the call on the CRT is not at a high enough level to note the differences in aural character. Note in the figure that the peak which registers the highest amplitude in the CRT spectrum is the 3rd peak, whereas in hydrophone 3 (H3) it is the 4th peak. Recordings made before September 20th were generally not analyzed on the CRT because of low gain settings at that time.

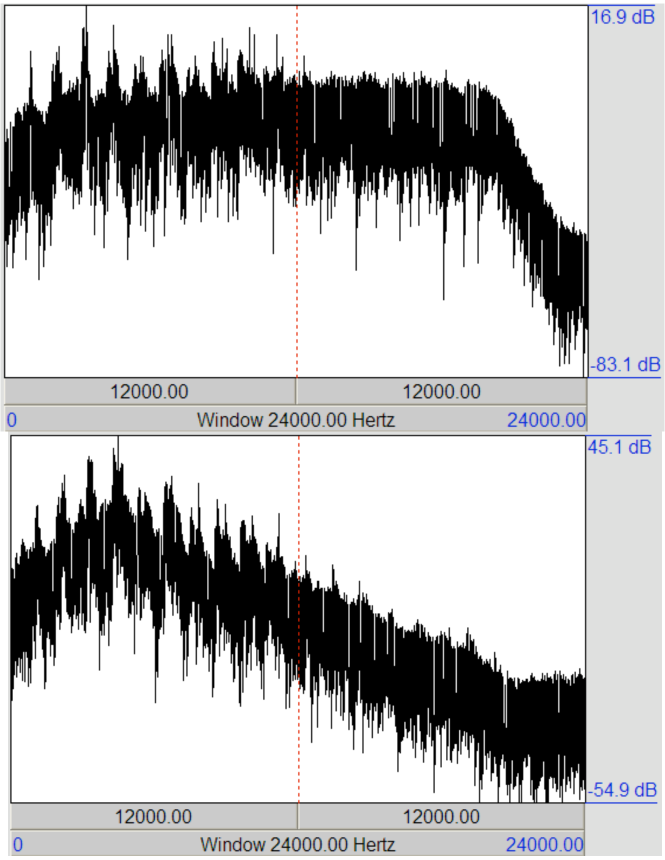


Figure 4- CRT spectrum of the first syllable of an S1 call (top) and from hydrophone 3 (bottom)

Individual syllables were analyzed (using the spectrum produced by PRAAT) for features that are *characteristic* of the sound, much like how the frequency values of the first two formants determine the sound of human language vowels.[2] The key piece of this analysis was looking for elements of a call that are *characteristic.* Some progress has been made in identifying frequency peaks that have this property. That is, a peak is characteristic if, when it is stopped along with a 1 kHz band, the sound quality changes. This change is usually marked by a change in timbre, frequency and volume. The frequency of the spectral peak around which the band was removed can be noted

---

[2] Another technique that shows promise and is used commonly to describe individual orca sounds as well as characterize groups of mammal sounds involves cepstral analysis. This was not attempted in this project but has potential for future application.

(using the peak calculator in Audacity or the formant calculator in PRAAT, which takes an average as the peak shifts). If the peak appears to be bimodal, the average frequency of the two highest peaks will be recorded. The lowest frequency is noted as F1, the next lowest F2 etc. In the call pictured in figure 4, there is only one characteristic peak. That is, there is only one peak which, if removed, changes the character of the sound. On both hydrophones, the characteristic peak is similar (3329 for H3, 3372 for the CRT), but on H3 it is not the peak with the highest amplitude. This discrepancy again suggests use of the CRT data as much as possible.

Because some calls (like the ones above) have proven to have only one characteristic peak, and yet that peak alone is not sufficient to dictate the timbre, frequency and volume of the call, there is another facet of the call that needs to be analyzed. Average harmonic interval (often called the side band interval (SBI) in the literature, see Ford 1987) over the visible spectral peaks and the frequency of the peak of maximum energy were also examined. Harmonic interval is a measure of the fundamental frequency. To determine this measure, the smallest and largest easily distinguishable frequencies of a given set of harmonic bands were measured. Then their difference will be divided by one less than the total number of bands to give an average interval. While these results should be consistent with previously published data about individual calls (e.g. Ford 1987) in instances where my identified syllables overlap with published call parts, there is evidence that killer whale calls change over time (Wieland et al 2010), so it was important to calculate this interval for the data I am using.

Together, these measures (characteristic frequencies, harmonic interval and frequency of maximum energy) were examined pair-wise to determine the degree of characterization of the call syllable. That is, if a whale were to change either element in the pair, the call would sound fundamentally different. Once these measures were procured for a variety of calls, analysis was performed in a way analogous to that performed on the formant structure of vowels. For a detailed description of the following methodology, see Transnational College of LEX 2002.

With human language, the general procedure is to have a number of speakers of the standard grammatical form of the language produce the same set of sounds into a recording device. The F1 and F2 values are extracted from a spectrogram, and then average results are reported for a variety of categories, usually male, female and children. In all languages examined in preparation for this study (Lindau 1978, Watrous 1991, Cheung 2007, Johnson and Martin 2001, Liu and Ng 2007, Hillenbrand 1995, Kiefte et al 2010, Fant et al. 1969, Peterson and Barney 1951) there is a vowel sound that has a smaller F1 and a larger F2 than all other vowels, leading to the conclusion that it operates in the largest possible space for F1 and F2. In languages with vowel nomenclature with which I was familiar, that sound was always akin to the long "ee" sound as in "beat" or "reed." If F1 and F2 are plotted for each vowel (figure 5) the result is akin to those reported by the Transnational College of LEX. Note that the x-axis in this scale is the natural logarithm of the frequency values (for reasons discussed in the figure 1 caption). For my analysis, I used the logarithmic unit cents to report frequency as this unit is commonly applied to musical instruments and has been used to classify killer whale call frequencies as well (Brown et al. 2006). The dashed center line in the figure is the center of the logged frequency of F1 and F2 of the longest vowel. It is interesting to note that only rarely in all languages examined does F1 move to the right of this line or F2 to the left. This center line is thus taken as an upper and lower bound, respectively, for F1 and F2. The center of each of the

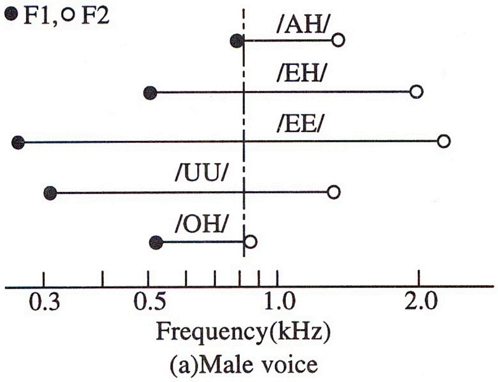


Figure 5
(Transnational College of Lex 2002)

other vowels is calculated and the positive or negative distance to the center of the long vowel is termed the "offset." A ratio of this offset to the longest possible offset (which occurs for a center halfway between the long vowel's center and end points) is calculated. This is the x-axis in figures 1 and 2. The y-axis is the length of each vowel relative to the longest vowel length (which is the longest possible vowel length). The diamond encasing the points in the figures is the theoretical boundaries of these values given the assumptions (described above) of the locations of F1 and F2 relative to the longest vowel.

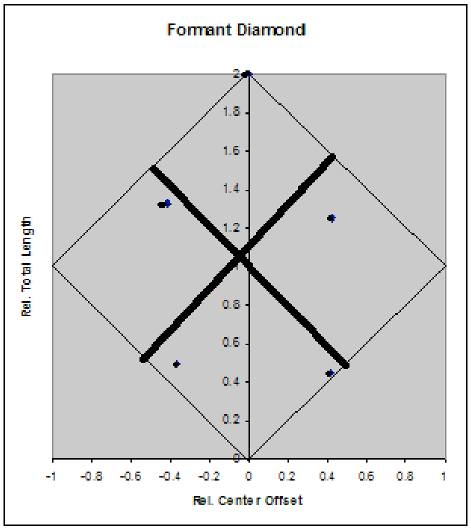


Figure 6- The male data from figure 1 is graphed above and the region is divided into 4 (the largest perfect square number less than 5). There is 1 point in each region except for the bottommost which has 2 points.

It is obvious in figure 1 that the five Japanese values are spread very far apart and thus very distinguishable in aural character. The results in figure 2 are a little less obvious, but still demonstrate large spacing. To give this qualitative impression of "spreadoutness" a quantitative measure, the formant diamond region was broken up into n equal sized squares (where n is the first perfect square number less than the number of points) as in figure 6. The number of points falling into each square (termed "bins") was calculated and then the mean and standard deviation of these values was determined. If a perfect distribution is found (one point in each bin), then the standard deviation would be 0. If the number of points is not a perfect square, however, then this perfect distribution is not possible and a higher standard deviation results For instance the standard deviation of figure 6 is 0.5, but compare this to the worst possible distribution (all points in one bin), which has a standard deviation of 2.5. The coefficient of variation (the standard deviation divided by the mean times 100) was the measure settled on to report the "spreadoutness" of the data since dividing by the mean accounts to some degree for the different number of bins used as determined by the number of data points. It should be noted that a perfect distribution (1 or 2 points in each bin) has coefficients of variation (cv) ranging from 0 to 35 and that the worst possible distributions (all points in one bin) can have cvs in excess of 800 (near 100*sqrt of the number of data points) when using 64 bins. It should be noted that this measurement has a slight bias towards empty bins. That is, if a large portion of the data is outside the diamond, it gives a better value (near 100) than if all the points are in 1 bin. This cv value of the spreadoutness of the data is referred to throughout as the spreadoutness factor (SONF). Different pair-wise arrangements of variables and different selections of syllables were evaluated in this manner and results recorded.

Results-

Overall, 282 syllables from 126 calls were analyzed. Of these, 93 syllables were data collected this fall, while the remainder were from previously collected and typed sample calls. These numbers were due entirely to the time consuming nature of typing calls, which proved to be one of the most challenging aspects of this analysis. Although much of the literature refers to well-established call types that are easily classifiable by humans (e.g. Brown 2006), I found numerous calls that fit no clear call type. While some calls were ubiquitous and easily identifiable (like S1 and S16), other types were not heard at all. Thus, for some calls the only analyzed samples are from archived files. I often made my best estimate during analysis, but then ended up not using the results after I compared the frequency values I got to the frequency values from sample files of known type, judging from the sample files that my estimate was inaccurate. Of the 93 syllables I typed and analyzed, 29 were discarded as unreliable types before average values were calculated.

Before using the spreadoutness factor described above on killer whale data, I tested it on vowel formant data collected from various published works. Results are summarized in table 1. Note that the Japanese data referred to throughout this paper (see figure 1) has the lowest SONF (40) and that the mean SONF for all languages was 79.29.

Data was then organized into call type and syllable number, and average values were taken of the calculated fundamental frequency (f_0), first characteristic peak of the spectrum (f_1), second characteristic peak of the spectrum (f_2) and frequency in the spectrum which had the most energy (f_{max}). It is worth noting that sometimes these values overlap. That is, the fundamental frequency at times had the highest energy, or the f_1 (the first region when, if removed, fundamentally changed the character of the sound) was coincident with the fundamental or with f_{max}. These data were recorded as is, but if the f_{max} value coincided with f_1 or f_0, then the peak with the second most energy was also recorded for purposes of pair-wise analysis.

Table 1- SONFs of various languages from published formant averages for comparison. N is the number of sounds analyzed; for human languages this represents the number of vowels recognized by the author. The extended SONF column represents the SONF calculated from an area extended 30% outside of the regular formant data. This measurement was calculated because some killer whale data do not conform to the predicted boundaries of the data. For the most part, this statistic gave slightly higher results (indicating less spreadoutness). These data match the graphs in figures 1 and 2.

Langauge	Author	n	SONF	Extended Sonf
Akan	Lindau	9	111.8	106.7
Danish	Lindau	7	68.01	28.57
English	Watrous	10	54.08	54.08
English	Peterson et al	10	54.08	54.08
English	Hillenbrand	12	75	75
English	Lindau	10	94.87	94.87
Hakka Chinese	Cheung	5	89.44	69.28
Japanese	TNCoL	5	40	40
Mandarain	Liu	7	86.07	86.07
Swedish	Fant et al	10	119.5	119.5
Average		9	79.29	72.815

Analysis was done on three pairs: f_1 vs f_2, f_0 vs f_{max}, and f_1 vs f_{max}. These pairs were chosen because in each case the first value is lower than the second value (or the second highest

Table 2- Results of SONF tests performed on collected averages of analyzed syllables. N>2 means that only syllables whose values were derived from an average of more than 2 syllables of the same type were used in that test. Length method shows whether the length of the longest syllable was used (length) or the min and max values of the pair (min/max).

			SONF			**Extended SONF**		
Description of Test	**#**	**Length Method**	**f_1 vs f_2**	**f_0 vs f_{max}**	**f_1 vs f_{max}**	**f_1 vs f_2**	**f_0 vs f_{max}**	**f_1 vs f_{max}**
All syllables	66	Length	85.63	101.67	86.12	105.93	113.30	96.20
All syllables	66	Min/Max	95.42	102.90	95.56	104.92	107.81	88.78
All syllables, n>2, σ/μ of f0 < 0.5	26	Length	73.23	91.56	82.71	96.87	116.32	98.37
All syllables, n>2, σ/μ of f0 < 0.5	26	Min/Max	107.07	144.28	98.68	107.57	153.24	101.91
All syllables, n>3	22	Length	77.99	82.92	67.70	92.47	93.42	61.57
All syllables, n>3	22	Min/Max	89.57	81.31	70.26	94.36	97.57	73.33
All syllables, n>3, σ/μ of f0 < 0.5	19	Length	73.53	88.82	66.80	97.09	117.47	65.90
All syllables, n>3, σ/μ of f0 < 0.5	19	Min/Max	94.31	91.77	75.12	91.77	93.36	88.15
All syllables, σ/μ of f0 < 0.3	52	Length	88.81	102.80	85.23	114.93	108.09	88.19
All syllables, σ/μ of f0 < 0.3	52	Min/Max	116.61	111.06	89.36	130.93	128.88	105.17
All syllables, σ/μ of f0&fmax < 0.3	28	Length	102.32	119.18	96.44	90.32	121.00	89.43
All syllables, σ/μ of f0&fmax < 0.3	28	Min/Max	113.88	133.43	107.58	143.04	187.29	152.05
All syllables, σ/μ of f1 < 0.3	36	Length	86.19	95.62	92.58	108.23	112.12	121.89
All syllables, σ/μ of f1 < 0.3	36	Min/Max	98.56	112.12	98.56	114.64	125.36	118.32
All syllables, σ/μ of f1 < 0.5	56	Length	88.79	90.57	84.72	115.36	91.82	81.43
All syllables, σ/μ of f1 < 0.5	56	Min/Max	94.75	112.15	79.30	110.82	118.67	104.36
All syllables, σ/μ of f1&f2 < 0.3	29	Length	89.52	97.15	90.60	71.70	101.51	107.44
All syllables, σ/μ of f1&f2 < 0.3	29	Min/Max	109.27	101.51	93.05	116.72	142.22	113.11
All syllables, σ/μ of f1&f2 < 0.5	42	Length	78.91	81.53	85.70	110.00	104.81	119.20
All syllables, σ/μ of f1&f2 < 0.5	42	Min/Max	90.48	95.01	97.19	103.47	123.22	116.21
All syllables, σ/μ of fmax < 0.3	34	Length	93.39	111.02	88.08	81.99	109.23	104.72
All syllables, σ/μ of fmax < 0.3	34	Min/Max	101.00	131.91	98.37	124.89	156.67	130.19
syllable 1	25	Length	104.08	124.16	108.01	97.89	120.76	106.07
syllable 1	25	Min/Max	88.98	159.43	100.00	117.26	125.83	97.89
syllable 2	23	Length	68.25	68.54	68.54	96.62	82.68	78.42
syllable 2	23	Min/Max	77.65	78.93	65.23	78.68	70.00	70.00
syllable 2, σ/μ of f0 < 0.3	18	Length	66.90	71.66	74.37	97.37	77.83	81.95
syllable 2, σ/μ of f0 < 0.3	18	Min/Max	83.54	94.63	86.64	90.36	77.83	86.64
syllable 2, σ/μ of f0&fmax < 0.3	8	Length	91.29	86.60	86.60	104.08	86.60	104.08
syllable 2, σ/μ of f0&fmax < 0.3	8	Min/Max	91.29	119.02	50.00	91.29	100.00	50.00
syllable 2, σ/μ of fmax < 0.3	11	Length	93.98	92.49	96.42	82.95	92.49	82.95
syllable 2, σ/μ of fmax < 0.3	11	Min/Max	85.70	105.63	97.38	85.70	120.43	97.38
syllable 3	13	Length	75.22	88.63	83.60	75.66	82.80	102.88
syllable 3	13	Min/Max	89.38	76.97	68.70	83.60	68.26	90.85
syllable 3 and 4	15	Length	74.50	83.96	83.96	73.48	85.73	103.20
syllable 3 and 4	15	Min/Max	92.47	76.49	82.16	78.42	71.41	88.32
syllable 3, σ/μ of f1 < 0.5	10	Length	76.49	105.53	103.38	81.47	121.40	124.60
syllable 3, σ/μ of f1 < 0.5	10	Min/Max	67.92	103.38	119.53	67.92	97.21	124.60
mean	28		88.865	100.43	86.9534211	98.1782	108.016842	97.7828947

peak was used in the case of f_{max}) and because they represent a cross section of analytical techniques: f_1 and f_2 were determined by human ears and f_0 and f_{max} by visible features of the spectrum. Analysis was done on a variety of samples from the data at large and results are presented in full in tables 2 and 3. In order to determine the position within a formant diamond, the length of the distance between the pair (in cents) must be determined and compared to the length of the longest distance. This longest distance was calculated in two ways: by using the longest distance and also by finding the distance between the minimum f_1 and maximum f_2 (in the case of f_1 vs f_2 analysis) even if these minimum and maximum were not from the same syllable.

Because no clear-cut pattern of spreadoutness was observed, different restrictions were tested including limiting analysis to syllables whose values were close to each other (low standard deviation). When clear-cut patterns were still opaque, the original data was combed through again, and syllables whose values were far away from the mean (suggesting a problem with data collection) were relabeled as outliers and excluded from the data. Results from these analyses are in table 3.

Table 3- Results when syllables whose analyzed frequency values were significantly different than the average were removed from the calculations.

			SONF			Extended SONF		
Description of Test	**#**	**Length Method**	**f_1 vs f_2**	**f_0 vs f_{max}**	**f_1 vs f_{max}**	**f_1 vs f_2**	**f_0 vs f_{max}**	**f_1 vs f_{max}**
All syllables	66	Length	101.21	110.33	95.22	122.09	136.42	127.91
All syllables	66	Min/Max	124.96	127.44	118.96	149.50	129.08	123.16
syllable 1	25	Length	97.90	109.92	104.08	109.90	148.60	120.76
syllable 1	25	Min/Max	117.26	151.38	141.42	145.77	145.77	147.20
syllable 2	23	Length	77.90	94.08	74.19	78.68	104.06	66.15
syllable 2	23	Min/Max	86.02	93.87	65.84	112.80	103.86	79.19
syllable 3	13	Length	96.19	102.56	89.37	76.10	68.26	68.26
syllable 3	13	Min/Max	76.10	76.97	76.10	84.00	84.40	84.40
syllable 1, σ/μ of f0 < 0.3	21	Length	87.29	118.55	110.52	84.47	121.17	114.60
syllable 1, σ/μ of f0 < 0.3	21	Min/Max	111.39	126.06	120.57	147.13	124.90	139.02
syllable 2, σ/μ of f1 < 0.3	18	Length	74.37	95.32	84.33	76.98	91.08	81.95
syllable 2, σ/μ of f1 < 0.3	18	Min/Max	76.98	96.01	77.83	101.35	111.26	84.33
syllable 3, σ/μ of f1 < 0.31	10	Length	91.24	109.20	91.24	88.74	67.92	93.07
syllable 3, σ/μ of f1 < 0.31	10	Min/Max	76.49	86.17	81.47	76.49	81.47	81.47
mean	25		92.5214	106.99	95.0814286	103.857	108.446429	100.819286

It should be noted that even the lowest value observed, (f_1 vs f_{max}, syllable 2, σ/μ of f0&fmax < 0.3, table 2), was 25% higher than the SONF value of 40 for the Japanese language data. Figure 7 shows this data as an example of the best SONF value. When the total mean SONF of all the tests (μ = 93.7) is compared to the SONFs of the human languages (μ = 79.3) it is apparent that the killer whale SONFs are significantly higher since with F(1,122)= 3.92, p = .03

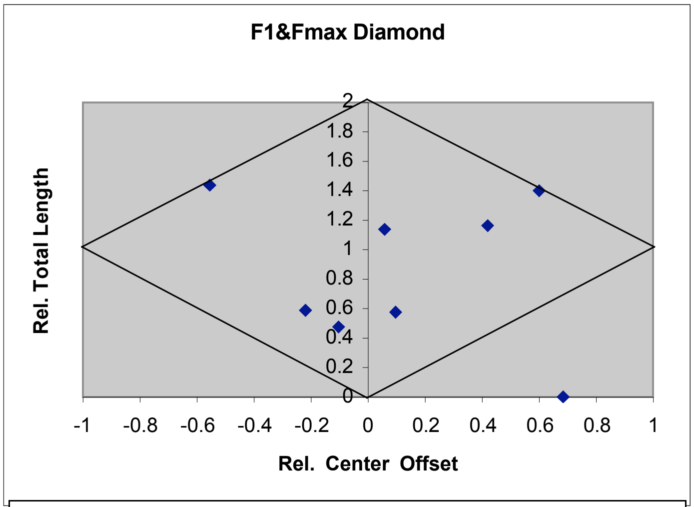


Figure 7- The data set with the lowest SONF value calculated (50). The σ/μ calculation for each of these syllables was low across all four frequency values calculated. Out of the 32 σ/μ values for these 8 syllables, one was 0.41, one 0.71 and the rest were ≤ 0.07.

Other significant values of note include the observation that in both table 2 (F(1,112) = 3.9, p = 0.0001) and table 3 (F(1,26) = 4.2, p = 0.043) f_0 vs f_{max} had a much higher SONF value on average than f_1 vs f_2, while f_1 vs f_2 was not significantly different (table 2: F(1,74) = 3.97, p = 0.53, table 3: F(1,26) = 4.2, p = 0.73) than f_1 vs f_{max}.

Also of note, syllable 2 (μ = 85.16) had significantly lower SONFs on average (table 2: F(1,86)=3.95, p= 0.03) than tests without syllable 2 included (μ = 94.26). Comparing the second syllable to the human language SONFs gives a p value of 0.31 (F(1,48)= 4.04) suggesting that the hypothesis that syllable 2 has comparable values to human language cannot be rejected.

Lastly, several of the syllables for which I separated out outliers before doing the analysis for table 3 had outlier sets that, while significantly different than the values that were kept, had patterns of their own. For instance, out of 15 S1 1st syllables, 5 were removed as outliers for the table 4 analysis based on their differences in f_1 (μ = 2656 for the S1s and μ = 1072 for the outliers) and f_{max} (μ = 3749 vs μ = 2147). For f_{max}, F(1,13) = 4.7, p = 0.003 and for f_1, F(1,13)= 4.7, p = 0.003. However, the standard deviation of the outliers for f_1 and f_{max} was only 53 and 104, respectively, implying that these outlier calls have well-clustered frequencies within themselves.

Discussion-

While any attempt to describe a previously unexamined aspect of a system with biological origins is ambitious in a 10-week time frame, this study was particularly hindered by a lack of time for proper analysis. Given the 26 discrete call types identified by Ford (1987) and the amount of data needed to properly find averages and identify outliers, more dedicated researchers and funding are needed to complete this work and adequately map characteristic frequencies of SRKW calls. Because of the difficulties inherent in identifying calls (as described above), the previous data used lent itself to some discrepancies since many of the calls were collected on machines with clearly different specifications (for instance, some spectrums stopped at 8000 Hz with no recorded frequencies above that point). Given more time, pulling more calls from data that was collected consistently will be key to more fully understanding the potential patterns in the calls. So, despite a dataset that is representative but perhaps not all-inclusive of the characteristics of killer whale call, some interesting facts can be safely drawn out.

Firstly, the methodology used to conduct this analysis seems valid. The SONF measurement is an indicative measure of how spread out is a set of data points. Although one might expect all human languages examined to have SONFs as low as that for the Japanese vowels, it is important to note that all studies except for the Japanese were done with their own purposes in mind. Many only looked at one type of speaker (for instance, Lindau (1978) was interested only in speakers of “standard” English and he divided his subjects into male and female voices, and Liu et al. were interested in the differences between esophageal and laryngeal speakers) and there is even disagreement about how many vowels there are in a given language (Hillenbrand et al. 1995 and Lindau 1978). Further, more broad-scale applications should be attempted to determine what exactly a “good” SONF value would be and its potential as a statistical tool. Equally important, it is interesting to note that characteristics like f_{max} and f_l, which are new to the literature, have some consistency across a given call type and syllable (as measured by standard deviation). This was an important finding because it was difficult to be remain consistent as an analyzer given different conditions. When the background noise was higher, for instance, it was difficult to tell if earlier, low-frequency, peaks changed the character of the sound I was hearing because of the background noise component or because it was a character of a call. In some such situations, the reported f_l values may have been low. Additionally, although the f_{max} value was the easiest to recognize on a spectrum, perhaps an even more descriptive value would be the frequency peak which has the highest energy relative to the threshold of killer whale hearing. Such a dB_{orca} calibration could prove invaluable in the quest to understand the importance of call variability to pods of SRKWs.

It is important to note as well that no broad pattern of killer whale calls having evolved to maximize auditory differentiability can be shown from this data, and thus the null hypothesis cannot be rejected in this case. Some somewhat surprising results can be inferred, however, including the relevance of the spreadoutness of syllable 2. One possible interpretation of this observation is that killer whales use the second syllable of a call as a vowel equivalent. That is, if call type is indeed some sort of communication system, it could be that the second syllable is what they cue in on to derive the bulk of the message. Anecdotal evidence from this study period confirms that killer whales often repeat calls. It would be informative to look at how these repeated calls change over the course of a call set, and specifically how the second syllable changes.

One would expect, given the variability in human hearing from orca hearing, that frequency observations not based on hearing might give better results. This was not the case however, as the f_0 vs f_{max} comparisons had significantly higher SONFs. This lends credence to the idea that formant equivalents may exist in odontocete communication despite the different anatomical structures involved in sound production and filtering.

A final possible application of this analysis in the future could be its use in reclassifying killer whale call types. Even if two calls sound the same to the human ear and look comparable on a spectrogram, characteristics like f_{max} were shown in this study to divide calls like S1 into two categories. It is imperative to examine these calls more closely for signs that a killer whale might hear two S1 calls very differently.

Although this study did not demonstrate conclusively that killer whale call types likely carry a communicative purpose, such a conclusion can not yet be ruled out, and further research is needed.

Literature Cited-

Berta, A., J. Sumich and K. Kovacs. 2006. Marine Mammals: Evolutionary Biology. 2nd ed. Elsevier Inc., Burlington, MA.

Bradbury, J and S. Vehrencamp. 1998. Principles of Animal Communication. Sinauer Associates, Inc., Sunderland, MA.

Brown, J., A. Hodgins-Davis and P. Miller. 2006. Classification of vocalizations of killer whales using dynamic time warping. Journal of the Acoustical Society of America Express Letters 119(3).

Cheung, Y. M. 2007. An acoustical analysis of the vowels, diphthongs and triphthongs in Hakka Chinese. 16th International Congress of Phonetic Sciences Publication. Saarbrücken, Germany. www.icphs2007.de: 841-844

Deecke, V., J. Ford and P. Spong. 1999. Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects. Journal of the Acoustical Society of America 105(4):2499-2507.

Fant, G., G. Henningsson and U. Stålhammar. 1969. Formant frequency of Swedish vowels. Department of Speech, Music and Hearing: Quarterly Progress and Status Report 10(4):26-31.

Fitch, W. T. 2000. The evolution of speech: a comparative review. Trends in Cognitive Sciences 4(7).

Ford, J. 1987. A catalogue of underwater calls produced by killer whales *[Orcinus orca]* in British Columbia. Canadian Data Report of Fisheries and Aquatic Sciences 633.

Ford, J. 1989. Acoustic behavior of resident killer whales *(Orcinus orca)* off Vancouver Island, British Columbia. Canadian Journal of Zoology 67:727-745.

Ford, J. 1991. Vocal traditions among resident killer whales *(Orcinus orca)* in coastal waters of British Columbia. Canadian Journal of Zoology 69:1451-1483.

Grebner, D., D. Bradley, D. Capone, S. Parks, J. Miksis-Olds and J. Ford. 2009. Killer whale discrete pulsed call variation. Journal of the Acoustical Society of America 126(4):2272.

Hillenbrand J., L. Getty, M. Clark and K. Wheeler. 1995. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America 97(5):3099-3111.

Holt, M. 2008. Sound exposure and southern resident killer whales *(Orcinus orca):* A review of current knowledge and data gaps. U.S. Dept. of Commerce, NOAA Tech. Memo. NMFS-NWFSC-89, 59 p.

Johnson, K. and J. Martin. 2001. Acoustic vowel reduction in creek: effect of distintive length and position in the word. Phonetica 58(1-2):81-102.

Kiefte, M., T. Enright and L. Marshall. 2010. The role of formant amplitude in the perception of /i/ and /u/. Journal of the Acoustical Society of America 127(4):2611-2621.

Kollmeier, B., T. Brand and B. Meyer. 2008. Perception of speech and sound. The Springer Handbook of Speech Processing. Benesty, J., M. Sandhi and Y. Huan, eds. Springer-Verlag, Berlin, Germany.

Lindau, M. 1978. Vowel features. Language 54(3):541-563.

Liu, H. and M. Ng. 2009. Formant characteristics of vowels produced by Mandarin esophageal speakers. Journal of Voice 23(2):255-260.

Miller, P., F. Samarra and A. Perthuison. 2007. Journal of the Acoustical Society of America 121(6):3932-3937.

National Marine Fisheries Service. 2008. Recovery Plan for Southern Resident Killer Whales (*Orcinus orca*). National Marine Fisheries Service, Northwest Region, Seattle, Washington.

O'Shaughnessy, D. 2008. Formant estimation and tracking. The Springer Handbook of Speech Processing. Benesty, J, M. Sandhi and Y. Huan, eds. Springer-Verlag, Berlin, Germany.

Parijs, S., T. Leyssen and T. Similä. 2004. Sounds produced by Norwegian killer whales (Orcinus orca) during capture. Journal of the Acoustical Society of America 116(1):556-560.

Peterson, G. and H. Barney. 1952. Control methods used in a study of vowels. Journal of the Acoustical Society of America 24(2):175-184.

Reby, D., R. André-Obrecht, A. Galinier, J. Farinas and B. Cargnelutti. 2006. Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags. Journal of the Acoustical Society of America. 120(6):4080-4089.

Rendell, L. and H. Whitehead. 2001. Culture in whales and dolphins. Behavioral and Brain Sciences 24:309-382.

Riesch, R., J. Ford and F. Thomsen. 2005. Stability and group specificity of stereotyped whistles in resident killer whales, Orcinus orca, off British Columbia. Animal Behavior 71(1):79-91.

Roch, M., M. Soldevilla, J. Bortenshae, E. Henderson and J. Hildebrand. 2007. Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. Journal of the Acoustical Society of America 121(3):1737-1748.

Sininger, Y., C. Abdala and B. Cone-Wesson. 1997. Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. Hearing Research 104:27-38.

Szymanski, M., D. Bain, K. Kiehl, S. Pennington, S. Wong, and K. Henry. 1999. Killer whale (Orcinus orca) hearing: Auditory brainstem response and behavioral audiograms. Journal of the Acoustical Society of America 106: 1134-1141.

Thomsen, F., D. Franck and J. Ford. 2002. On the communicative significance of whistles in wild killer whales (Orcinus orca). Die Naturwissenschaften 89(9) 404-407.

Transnational College of LEX. 2002. Who is Fourier? A Mathematical Adventure. A. Gleason, trans. Language Research Foundation, Boston, MA.

Watrous, R. 1991. Current status of Peterson-Barney vowel formant data. Journal of the Acoustical Society of America 89(5): 2459-2460.

Weiss B., H. Symonds, P. Spong and F. Ladich. 2007. Intra and Inter-group vocal behavior in resident killer whales, Orcinus orca. Journal of the Acoustical Society of America 122(6):3710-3716.

Wieland, M., A. Jones and S. Renn. 2010. Changing durations of southern resident killer whale *(Orcinus orca)* discrete calls between two periods spanning 28 years. Marine Mammal Science 26(1):195-201.