



The fallacy of 'signature whistles' in bottlenose dolphins: a comparative perspective of 'signature information' in animal vocalizations

BRENDA McCOWAN* & DIANA REISS†

*School of Veterinary Medicine, University of California at Davis

†Osborn Laboratories for Marine Sciences, New York Aquarium

(Received 23 October 2000; initial acceptance 2 January 2001;
final acceptance 14 April 2001; MS. number: A8915)

In sharp contrast with descriptions of contact calls in all other species, the contact or cohesion calls used by bottlenose dolphins, *Tursiops truncatus*, in contexts of social isolation have been historically described as individually distinctive and categorically different whistle types, termed 'signature whistles'. These whistle types have been proposed to function as labels or names of conspecifics. Other studies have reported an absence of signature whistles and have demonstrated that dolphins, like other species, produce a predominant shared whistle type that probably contains individual variability in the acoustic parameters of this shared whistle type. To further understand the discrepancies between different studies on dolphin whistle communication and the vast differences reported between the isolation calls of dolphins and other species, we conducted a study replicating the approach and methodologies used in the studies that originally and subsequently characterized signature whistles. In contrast to these studies, we present clear evidence that, in contexts of isolation, dolphins use a predominant and shared whistle type rather than individually distinctive signature whistles. This general class of shared whistles was the predominant call of 10 of the 12 individuals, the same shared whistle type previously reported as predominant for individuals within both socially interactive and separation contexts. Results on the further classification of this predominant shared whistle type indicated that 14 subtle variations within this one whistle type could be partially attributed to individual identity. Thus, individual variability in the production of a shared contact call, as reported for other taxa, probably accounts for individual recognition in dolphins. No evidence was found for categorically different signature whistle types in isolated bottlenose dolphins using both qualitative and quantitative techniques.

© 2001 The Association for the Study of Animal Behaviour

Individual recognition through contact calls in a diversity of taxa has been well documented in the literature across several years. The function of a shared contact call within species or social groups has been described as maintaining individual recognition between parents and offspring, between mates, as well as between neighbours and strangers. Identity and recognition is based on individual variability in the acoustic parameters of shared contact calls. Yet in striking contrast to all other species, contact

or cohesion calls in bottlenose dolphins, *Tursiops truncatus*, have been historically reported by some researchers as stereotypic individually distinctive 'signature whistles'. Signature whistles have been defined as the predominant whistle type, accounting for 70–95% of the whistles produced by each individual (Tyack 1986; Caldwell et al. 1990). That the whistle type be predominant is absolutely essential to the hypothesis because otherwise it would be impossible to know which one of the many whistles each dolphin produces in contexts of isolation (and more social contexts) is its individualized signature whistle. Proponents of the 'signature whistle hypothesis' report that individual recognition is based upon the predominant production of an obviously categorically different whistle 'contour', the temporal modulation of the fundamental frequency of whistles. They claim that this predominant whistle type of each dolphin is used to label

Correspondence: Dr B. McCowan Comparative Behavioral Biology Laboratory, Department of Population Health and Reproduction, Veterinary Medicine Teaching and Research Center, School of Veterinary Medicine, University of California at Davis Tulare, CA 93274, U.S.A. (email: bmcowan@vmtrc.ucdavis.edu). D. Reiss is at the Osborn Laboratories for Marine Sciences, New York Aquarium, Wildlife Conservation Society, Brooklyn, NY 11224, U.S.A.

or name themselves and one another (Tyack 1986; Janik & Slater 1998; Janik 2000; reviewed in Caldwell et al. 1990). Thus the 'signature whistle hypothesis' predicts that each dolphin produces, and is individually identified by, a categorically different stereotypic whistle contour type, and not by individual acoustic variations of a shared contact whistle or call type like in other species.

Both historical and current studies reporting signature whistles have primarily, but not always, recorded dolphins in contexts of voluntary social separation or temporary forced social isolation from conspecifics (Tyack 1986; Sayigh et al. 1990; Janik & Slater 1998; reviewed in Caldwell et al. 1990). Signature whistle researchers have conducted many of their studies in these isolation contexts to determine which one of the many whistle types produced by each individual dolphin is its actual signature whistle. Most recently, Janik & Slater (1998) described context-specific use of signature whistles and reported that they are produced only in contexts of voluntary social separation (or temporary forced social isolation) from group members. These data, however, do not correspond well to reports on signature whistle use from previous signature whistle studies in which dolphins were recorded under isolation or socially interactive contexts (Tyack 1986; reviewed in Caldwell et al. 1990). In these studies, signature whistles were found under both isolation and more socially interactive contexts.

In direct contrast to studies that report highly stereotypical individualized signature whistles, we and others have shown that dolphins from similar and different social groups produce a predominant whistle type that is shared, not an individualized whistle type, in both socially interactive and separation contexts (Dreher & Evans 1964; Dreher 1966; McCowan & Reiss 1995a, b, 1997). Many of these and other studies have also shown that dolphins produce a wide variety of other whistle types across different contexts both within and across social groups (Dreher & Evans 1964; Dreher 1966; Burdin et al. 1975; Kaznadzei et al. 1976; Ding et al. 1995; McCowan & Reiss 1995a, b, 1997). We also found that this predominant shared whistle type shows acoustic variability in more subtle acoustic parameters (McCowan et al. 1998). Results also revealed that social familiarity, social group membership and capture site location have a strong influence on spectral features of whistle acoustic structure. Therefore, this variability may be one mechanism contributing to 'regional' dialects in dolphins. Such acoustic variability within whistle types could also contribute to individual recognition.

Finding such disparity in reports of the predominant whistle types used by dolphins is puzzling. Current signature whistle proponents have claimed that the discrepancy between these two perspectives is simply due to differences in recording contexts and analytical methods (McCowan 1995; Janik & Slater 1998; Janik 1999). Thus, to clarify the nature and use of predominant whistles by isolated dolphins, we conducted a study replicating as closely as possible the original studies conducted by Caldwell & Caldwell (1965; reviewed in Caldwell et al. 1990) with the same approach and methods used in

historical and current studies reporting signature whistles (Sayigh et al. 1990; Janik & Slater 1998; reviewed in Caldwell et al. 1990). In contrast to the signature whistle hypothesis, we predicted that dolphins under different contexts of isolation from conspecifics would produce the same shared 'contact' whistle as reported previously by our research group, a rise whistle type, previously termed whistle type 2 (McCowan 1995; McCowan & Reiss 1995a, b, 1997; McCowan et al. 1998). We also predicted that this shared predominant whistle type would show individual variability or 'signature information' in its production by different individuals, as has been found in the contact calls of other species.

To test these predictions, we recorded the whistle vocalizations of 12 individual dolphins from three different captive social groups in contexts of voluntary and forced social separation and during temporary forced social isolation at poolside (before transport to another location at the same facility). We used these contexts of recording to ensure that we recorded dolphin subjects under the same conditions as those conducted in studies investigating the signature whistle hypothesis.

Finally, because past signature whistle studies have exclusively used qualitative techniques (e.g. human observers) to categorize dolphin whistle contours into types, we conducted an additional experiment to evaluate the comparative efficacy of qualitative and quantitative methods in classifying dolphin whistle contours (McCowan 1995; Janik 1999). Janik (1999) reported that the human observer technique was superior to known quantitative methods (such as McCowan 1995) for reliably categorizing dolphin whistles. To address this issue, we employed the human observer method to classify whistles for direct comparison to the signature whistle studies. But we also employed the contour similarity method (McCowan 1995) to categorize quantitatively a subset of these same whistles, allowing us to compare the results from each method and evaluate their respective efficacy in categorizing dolphin whistles (McCowan 1995; Janik 1999).

METHODS

Subjects

A total of 12 subjects were recorded from three different captive social groups for this study. Subjects included eight dolphins from Six Flags Marine World in Vallejo, California (four wild-born adult females: CHE, JAS, SAD, STO; two captive-born subadult males: LIB, NOR; and two captive-born juvenile males: AVA, BRI); two dolphins from an independent social group from the former Marine World Africa U.S.A. in Vallejo, California (one wild-born adult female: TER; and one captive-born juvenile male: DEL); and two dolphins from the New York Aquarium in Brooklyn, New York (two captive-born adult males: PRE, TAB). The two subadult and two juvenile male subjects were the offspring of three of the adult female subjects at Six Flags Marine World (LIB and BRI from CHE; NOR from STO; AVA from JAS). The juvenile male was unrelated to the adult female at Marine World

Africa U.S.A. and neither were physically, socially or acoustically exposed to the dolphins at Six Flags Marine World prior to the onset of this study. All adult females from both Six Flags Marine World and Marine World Africa U.S.A. were unrelated and acquired from various wild populations (McCowan et al. 1998). The two male subjects from the New York Aquarium were unrelated.

Vocal Recordings

We recorded subjects during periods of temporary isolation from conspecifics under three conditions: voluntary separation in the same pool ($N=115$ whistles), voluntary isolation in a separate pool ($N=298$ whistles), and during involuntary isolation (temporary capture) at poolside ($N=380$ whistles). These three contexts are the same as those reported in either the original or more recent studies reporting signature whistles (Tyack 1986; Janik & Slater 1998; Sayigh et al. 1990; reviewed in Caldwell et al. 1990). We recorded a total of 793 whistles from the subjects during these isolation contexts. We positively identified all whistles recorded to individuals based on the subject's use of concurrent bubble stream production (McCowan 1995; McCowan & Reiss 1995a, b) or in-air localization during vocalization (Caldwell et al. 1990; reviewed in Sayigh et al. 1990). Because the use of concurrent bubble stream production had the potential to bias the frequency of occurrence of the whistle types represented (since it frequently occurs with whistle production in younger animals but not always in adults), we conducted an analysis of the frequency of occurrence of whistle types produced with bubble streams to those produced in air. We compared the frequency of occurrence of each whistle type recorded with concurrent bubble stream emission (summing the occurrence of all whistles of a specific type across identified vocalizing subjects, $N=388$) to the frequency of occurrence of each whistle type recorded in air (again summing the occurrence of all whistles of a specific type across identified vocalizing subjects, $N=395$). Under both conditions, one whistle type, whistle type 2, was the most predominant, accounting for 71% of the identified whistles recorded with concurrent bubble stream emission and 75% of the identified whistles recorded in air from these subjects. This test validates the use of bubble stream emission as an appropriate method that sufficiently represents the whistle production rates of, at least, our subjects under contexts of isolation. These rates of production also correspond well to the whistle type production rates with and without concurrent bubble stream production in both separation and more socially interactive contexts (McCowan & Reiss 1995a, b). In fact, in one recent comparison, over 90 000 whistles collected continuously over a 1-month period showed similar rates of whistle type production across our subjects (B. McCowan & B. Garlick, unpublished data).

We used a Cetacean Research Technology hydrophone (flat frequency response up to 30 kHz) and a Labcore hydrophone (flat frequency response up to 30 kHz) to record the vocalizations during isolation and/or separation. We used an Audio-technica unidirectional

microphone (flat frequency response up to 25 kHz) and a Shure unidirectional microphone (flat frequency response up to 25 kHz) to record the vocalizations during temporary capture at poolside. All recordings were made with a SONY or Tascam DAT recorder at a sampling rate of 44.1 kHz.

Whistle Categorization

We categorized whistles using one qualitative method, human observers, and one quantitative method, the contour similarity technique (McCowan 1995) to classify whistles into types. The use of these two methods allowed us to compare the efficacy of qualitative and quantitative methods for categorizing dolphin whistles.

Qualitative human observer method

Because past research on signature whistles has historically and primarily used human observers to visually sort spectrograms of whistles into types based upon whistle contour, we first used the human observer technique to categorize the whistles in this study. We categorized whistles in this manner to ensure that different methodologies could not account for any differences in results between this study and the studies reporting signature whistles (Janik & Slater 1998; Janik 1999). Spectrograms of all 793 whistles were generated with a sampling rate of 44.1 kHz, a resolution of 516 FFT and a Hamming windowing function using Signalyze software. Spectrograms of 783 whistles (10 whistles were of marginal signal-to-noise ratio and thus were removed from the data set) were categorized by the three human observers on the basis of whistle contour. The three observers (two naïve and one experienced in the analysis of dolphin whistles), blind to the identity of the whistler and context, were independently asked to visually sort all 783 spectrograms based upon whistle contour alone into as many types as they saw fit. They were given no instruction on how different the contour had to be to categorize a whistle into a different type. However observers were provided with templates (line representations) of whistle contours as found in McCowan & Reiss (1995a, b; e.g. 1, 2, 5, 6, 7, 130, 131, 165). We provided a total of 168 different whistle type templates and instructed observers to label whistles with reference to these whistle types if, and only if, the observers perceived a match to these previous categories. The remaining whistle types ($N=4$; i.e. 207, 208, 209, 301) were assigned novel numerical labels independently by each observer. We instructed the observers in this manner to ensure that we could make direct comparisons to our previous quantitative analyses (McCowan 1995; McCowan & Reiss 1995a, b). After independent categorization by each human observer, we determined the final categories for whistles by consensus. If two or more observers categorized a whistle into the same category, then we considered the whistle the same whistle type. This method is comparable to that used in the most recent signature whistle studies (Janik & Slater 1998; Janik 1999). We statistically compared observers' classifications using the Cohen's kappa test in StatXact

4.01 statistical software. Statistical agreement among the observers was significant (kappa coefficient range 0.54–0.74; $P < 0.00001$ for each test) for each comparison among the observers' classifications of the 783 whistles.

We subsequently asked human observers (two experienced and one naïve) to further categorize into subcategories the whistles from one shared whistle type, whistle type 2, which was the most predominant whistle type for all but two of the 12 subjects from the three different social groups ($N = 396$; 185 whistles of the 581 whistle type 2 whistles were removed from the data set by the observers while classifying the subtle contour variation due to marginal signal-to-noise ratio). We asked observers to classify these whistles into as many subtypes as they saw fit to determine whether subtle changes in contour or other acoustic parameters could be visually sorted into types that corresponded to individual identity. We provided no templates for this analysis. The same blind procedure as described above was used to determine the whistle subcategories, which were labelled as subcategories of whistle type 2 (2A1, 2C, 2J, etc.). We again statistically compared the observers' classifications. Statistical agreement among the observers was significant (kappa coefficient range 0.45–0.66; $P < 0.00001$ for each test) for each comparison among the observers' classifications of the 396 whistles.

McCowan's method: quantitative contour similarity

We conducted a quantitative analysis because in a recent study, Janik (1999) reported that his qualitative human observer method was superior to three different quantitative methods, including McCowan's method (1995), for reliably identifying signature whistles (but see both Results and Discussion on this issue). To compare our quantitative method to the human observer method, we categorized a subset of the whistles ($N = 125$) recorded from five different subjects of one social group (AVA, BRI, CHE, NOR, SAD) using a modified version of McCowan's method (1995), in which we extracted 60 frequency and time points (instead of 20 points) across the duration of each whistle contour. Whistles were measured using Cool Edit Pro signal software (with a resolution of 1024 FFT and a Hamming windowing function) and macros developed by the first author. We compared the frequency measurements from the entire data set using correlation and then subjected the correlation coefficients to principal component analysis as described in McCowan (1995). We then subjected the principal components to k -means cluster analysis to obtain final group membership of the whistle contours, which was determined by the maximum number of nonoverlapping clusters in the iterative use of k -means cluster analysis, as described in McCowan (1995). We statistically compared the categorization of this subset of whistles analysed using McCowan's method to the categorization of the same whistles by the human observers, with human observer results serving as the basis of comparison (as in Janik 1999), using Cohen's kappa test in StatXact 4.01 statistical software. In addition, we conducted cross-validation tests using discriminant analysis in S-Plus statistical software to evaluate the internal validity of the categories generated

from the quantitative comparison of whistles (McCowan 1995). The percentage of whistles correctly classified to their original categories ranged from 98 to 100% in this cross-validation test.

In addition, to determine whether whistles could be attributed to individual subjects based upon subtle variations in the quantitative contour measurements of the predominant shared whistle type, we subjected the principal components of the correlation coefficients from the subset of whistles of the whistle type 2 category ($N = 86$) to discriminant analysis using individual subject as the grouping factor. We statistically compared results from this analysis to the results generated from the human observer method on the same subset of whistle type 2 whistles from the five subjects.

Finally, we conducted a final test to evaluate whether our quantitative technique could distinguish the calls among individuals in a functionally and acoustically similar call type of a phylogenetically unrelated species. We subjected the same analytical procedure described above for whistle type 2 whistles to 83 isolation peeps from four adult female squirrel monkeys, *Saimiri sciureus peruvinses*, of one social group from the California Regional Primate Research Center at the University of California at Davis, which were recorded during a 6-month experimental playback study on chuck call recognition (McCowan & Newman 2000). Squirrel monkeys are known to produce individually distinctive isolation peeps when separated from social group members both in captivity and in the wild. The isolation peeps therefore served as a direct analogue to the dolphin vocal data.

RESULTS

Frequency of Whistle Type Use by Isolated Dolphins

Table 1 presents the number of whistles produced by each individual for each whistle type as determined by the human observer method (see Fig. 1 for representative spectrograms of each whistle category). As indicated by this table, the most predominant whistle type was whistle type 2 (also known as the 'rise' whistle type, see Fig. 1), as observed in previous studies in which multiple dolphins from several different social groups were recorded during normal social interactions including both socially interactive and separation contexts (Dreher & Evans 1964; Dreher 1966; Ding et al. 1995; McCowan & Reiss 1995a, b, 1997). This predominant whistle type produced by isolated dolphins was the same whistle type that was reported by McCowan & Reiss (1995a, b, 1997) as predominant for all adult subjects and for infant subjects by the end of their first year in both socially interactive and separation contexts. We also conducted permutations of whistle type production by facility, context, captive versus wild-born, sex and age. With the exception of two individuals (i.e. AVA, BRI) in the analysis of different contexts, no differences in rates of whistle type production were found among any of these categories that were not attributable to individual differences (i.e. BRI, PRE) in whistle production (see Table 2). For example, AVA

Table 1. Frequency of occurrence of each whistle type ($N=783$ whistles) and each subcategory of the most predominant whistle type, whistle type 2 ($N=396$ whistles), for all subjects as determined by the human observer method (predominant whistle type for each subject is in bold)

Whistle type	Six Flags Marine World								Marine World Africa U.S.A.		NY Aquarium	
	AVA	BRI	CHE	JAS	LIB	NOR	SAD	STO	TER	DEL	PRE	TAB
1	3	2	5	19	16	3	13	3	1	21		2
2	280	19	19	19	117	64	13	3	1			25
5	6	1			14	5		1				
6		16										
7	8	71				1						
130						1						
131		1										
165	13	3	1	2	1		2	1			9	9
207						1						
208						1						
209						1						
301		2										
Total whistles	310	115	25	21	148	77	15	5	1	21	9	36
Type 2 subcategory												
2A1	167	2					1					
2A2	15	2										
2A3	10											
2B	1	1										
2C					13	8						
2E					36							
2F					3							
2J	1				8							
2R1	3		1	2	11	23	7	1	1	17		2
2R2	2	3	1	2	5	5						
2R3		1	12	1	1	7						
2R4			1			1						
2U1										1		2
2U2					13	2						
Total whistles	199	9	15	5	90	46	8	1	1	18		4

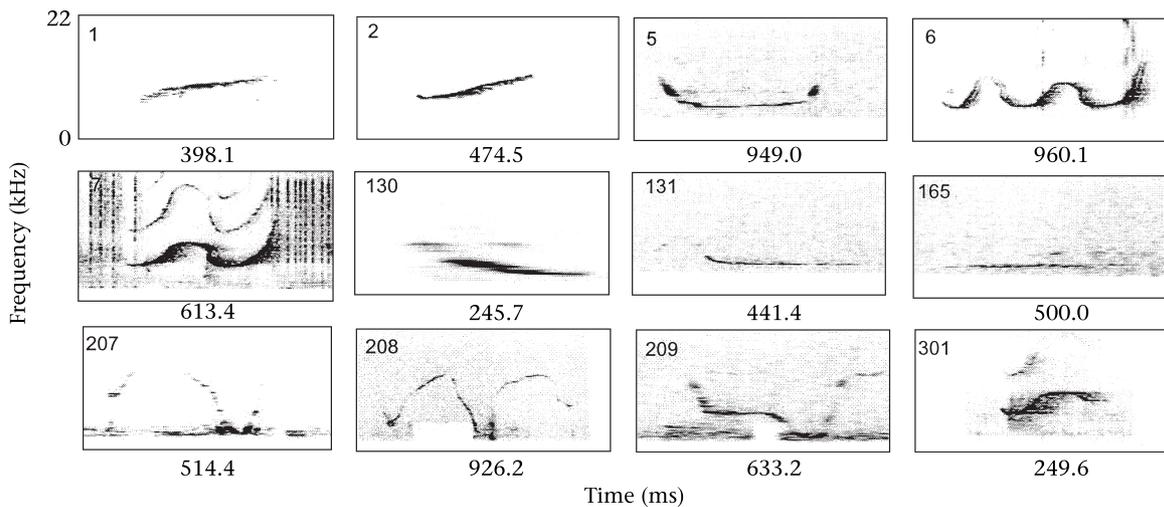


Figure 1. Representative spectrograms of each whistle category as determined by the human observer method. Spectrograms were generated using a 44.1 kHz sampling rate and 512-point FFT with Hamming windowing). Numbers correspond to the whistle types in Table 1. Whistle type 2 was the predominant whistle type for 10 of 12 subjects. (Also see McCowan & Reiss 1995a, b for extracted line representations of corresponding whistle types.)

Table 2. Frequency of occurrence of each whistle type ($N=783$ whistles) by context of recording for all subjects as determined by the human observer method (predominant whistle type for each subject is in bold)

Whistle type	Isolation in separate pool				Separation in same pool			Temporary capture at poolside										
	AVA	BRI	DEL	SAD	JAS	AVA	BRI	JAS	AVA	BRI	CHE	LIB	NOR	PRE	SAD	STO	TAB	TER
1	1	2				1			1		5	16	3					
2	227	5	21	2	1	36	8	18	17	6	19	117	64		11	3	25	1
5	6									1		14	5			1		
6		5					11											
7	7	20				1	51											
130													1					
131										1								
165					2				13	3	1	1		9	2	1	9	
207														1				
208														1				
209														1				
301							2											
Total whistles	241	32	21	2	3	38	72	18	31	11	25	148	77	9	13	5	36	1

predominantly produced whistle type 2 in the contexts of isolation in a separate pool and separation in the same pool, but whistle type 165 in the temporary capture context.

In contrast to past 'signature whistle' studies (Tyack 1986; Janik & Slater 1998; see Caldwell et al. 1990 for review), no evidence for individually distinctive signature whistle contours was found in these bottlenose dolphins. Ten of 12 individuals produced one shared whistle type as their most predominant whistle type during contexts of isolation. For the two other individuals, two other predominant whistle types (i.e. 7, 165) were found but these whistle types could not be considered signature whistles because both whistle types were shared among several different individuals within and across independent captive social groups (see Tables 1, 2; McCowan & Reiss 1995a, b, 1997). The fact that whistle type 165 occurred more often in contexts of temporary capture at poolside (see Table 2) may also suggest that this whistle type is one variant (less modulated than that) of whistle type 2 produced under more restrictive isolation conditions. Nevertheless, neither whistle type 7 nor whistle type 165 can be considered individually distinctive 'signature whistle' contours as defined by the signature whistle hypothesis (Caldwell et al. 1990; Janik & Slater 1998).

Frequency of Use of Whistle Type 2 Subcategories by Isolated Dolphins

Table 1 presents the number of whistles produced by each individual subject for each subcategory of whistle type 2, as determined by the human observer method (see Fig. 2 for representative spectrograms of each whistle type 2 subcategory). Subtle variations in contour generally corresponded to individual distinctiveness in this shared call type. This pattern is also found in other mammalian species (e.g. individuality in the call category of isolation peeps in squirrel monkeys: Symmes et al. 1979; also see

below). Difficulties in categorizing subtle variations of whistle type 2, however, were apparent when using the human observer method. While individual distinctiveness in the whistles for some animals was found, not all individuals could be reliably distinguished using this qualitative method (Table 1). Thus if 'signature information' (as opposed to a 'signature whistle', which requires that whistles are obviously categorically different in contour) exists in the subtle variations of the whistle type 2 contour, human observers could not reliably discriminate all individuals by the visual sorting of spectrograms. A quantitative analysis of these whistles is predicted to be more capable of distinguishing such subtle individual variation (McCowan 1995; McCowan & Reiss 1995a; McCowan et al. 1998), as in the analysis of squirrel monkey isolation peeps (Symmes et al. 1979; also see below). Therefore we conducted our quantitative technique on a subset of whistle type 2 whistles from five individuals to determine whether this prediction was indeed true (see Quantifying 'Signature Information' in a Shared Call Type below).

Comparison of Qualitative versus Quantitative Methods

Table 3 presents the frequency of occurrence of each whistle type generated from our quantitative method in comparison to the human observer method for a total of 125 whistles from five individuals representing eight different whistle types as determined by the human observer method. As seen in this table, McCowan's method and the human observer method agreed in the categorization of this subset of whistle types, showing significant statistical agreement ($\kappa=0.82$, $P<0.0001$). This result contrasts with Janik's (1999) evaluation of these two methods. The discrepancy, however, is not due to methodological differences but rather to differences in the methods used to compare each technique. When we

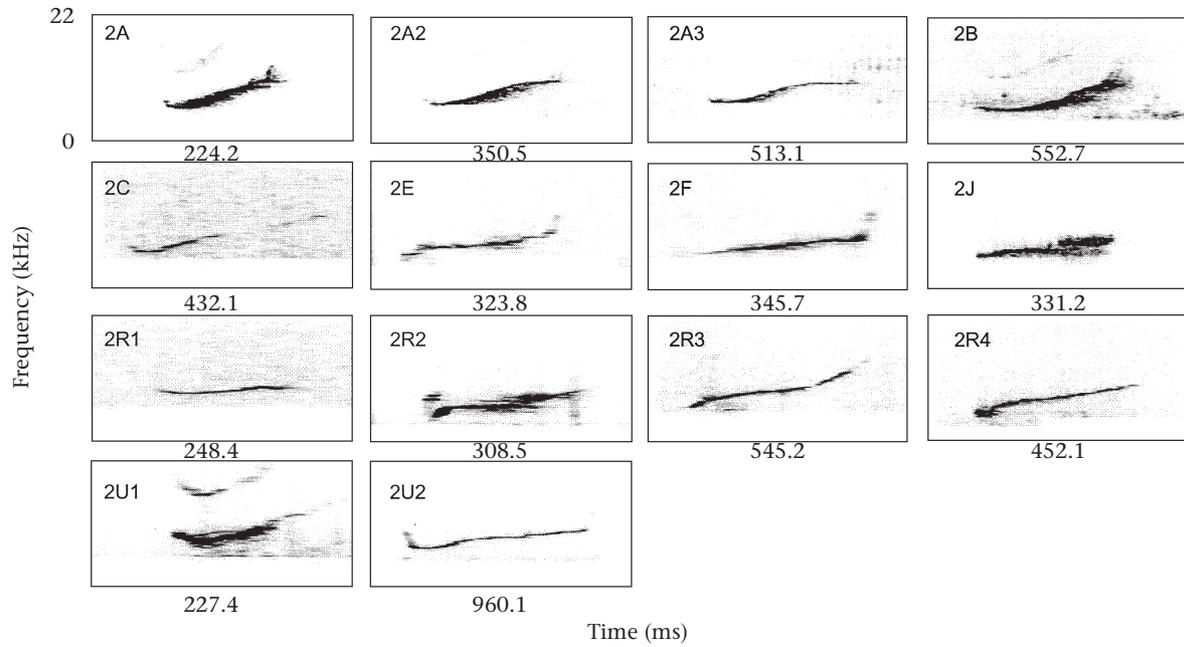


Figure 2. Representative spectrograms of each whistle type 2 subcategory as determined by the human observer method (spectrograms were generated using a 44.1 kHz sampling rate and 512-point FFT with Hamming windowing). Numbers correspond to the whistle subtypes in Table 1.

Table 3. Frequency of use of whistle types by five individual subjects for a subset of 125 whistles as determined by McCowan’s (1995) method and the human observer method (predominant whistle type for each subject is in bold)

Method	Whistle type	Individual subject					Total
		AVA	BRI	CHE	NOR	SAD	
McCowan’s	1						
	2	23	3	20	43	5	94
	5				1		1
	6		4				4
	7		26				26
	165						
	207						
Human observer	1			3	1		4
	2	23	2	16	40	5	86
	5				1		1
	6		4				4
	7		27				27
	165			1			1
	207				1		1
209				1		1	

For comparison: Janik (1999) classified a subset of 104 bottlenose dolphin whistles recorded in an isolation context into six categories (five categories of stereotypical signature whistle contours and one category of nonstereotypical whistle contours, $N=42$ and 62 , respectively) using one qualitative method (the human observer method, Figure 3 in Janik 1999) and three quantitative methods (McCowan’s 1995 method, cross-correlation method and frequency difference method; Figures 5, 6 and 7 in Janik 1999, respectively). In contrast to Janik’s (1999) results, however, our reanalysis of these data revealed significant agreement between the human observer method and McCowan’s quantitative method. All other pairwise comparisons between the four methods in Janik (1999) failed to show statistical agreement for the reasons discussed in McCowan (1995). In our reanalysis of the cross-correlation and frequency difference techniques, we used whistle classification results from the second level in the cluster tree, which represented the best possible ‘fit’ to the five signature whistle categories defined by Janik (1999).

statistically analysed the data provided in Janik (1999), the agreement between Janik’s human observer method and McCowan’s method was statistically significant

(kappa=0.78, $P<0.00001$), very similar to the agreement between the two methods as determined in the present study (see footnote in Table 3 for details).

Table 4. Discriminant analysis results on the percentage of type 2 whistles ($N=86$) correctly classified to five subjects from the same social group using McCowan's method (predominant number of classified whistles for each subject is in bold)

	Number of whistles categorized to each subject					Total	% Correct	χ^2
	AVA	BRI	CHE	NOR	SAD			
AVA	18	0	0	2	3	23	78.3	<0.00001
BRI	0	2	0	0	0	2	100.0	<0.00001
CHE	0	0	12	3	1	16	75.0	<0.00001
NOR	7	0	9	16	8	40	40.0	0.003
SAD	1	0	0	0	4	5	80.0	<0.00001

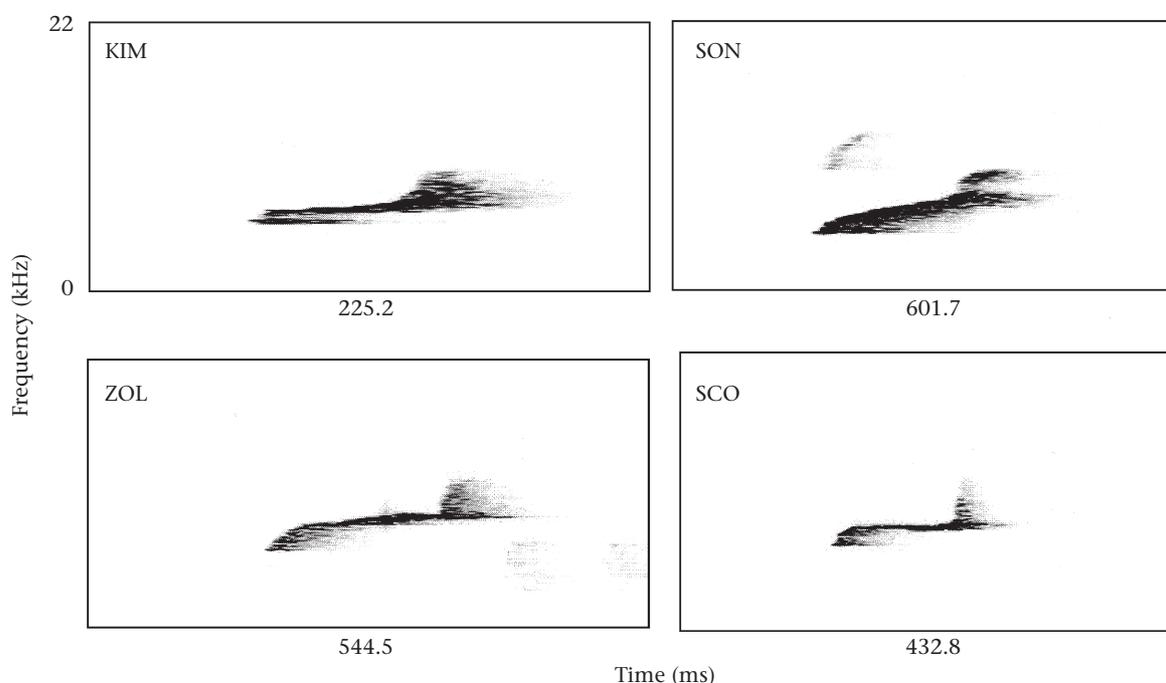


Figure 3. Representative spectrograms of isolation peeps from four adult female squirrel monkeys of the same social group (spectrograms were generated using a 44.1 kHz sampling rate and 512-point FFT with Hamming windowing).

Quantifying 'Signature Information' in a Shared Call Type

The quantification of whistles permitted the discrimination of individuals based upon subtle variations in whistle type 2 contour. Table 4 presents the results from the application of discriminant analysis using individual subject as the grouping factor, on the three principal components generated from the correlation coefficients of the whistle measurements on 86 type 2 whistles recorded from five individuals from the same social group. The percentage of whistles correctly classified to each individual was significantly higher than expected by chance (Table 4). To assess the potential effect of data pooling (Leger & Didrichsons 1994) on statistical inferences, we conducted discriminant analysis on a random subset of 16 whistles for the three individuals for which we had sufficient sample size and found no differences in results from the original analysis.

These data demonstrate that while these dolphins do not possess signature whistles, 'signature information' or vocal 'cues' are found in the subtle variations of the most predominant and shared whistle type contour, similar to that found in other nonhuman species, such as squirrel monkeys. In fact, the application of our quantitative technique to squirrel monkey isolation peeps further supports the efficacy of this technique (Table 6, Fig. 3). Subtle variations in the contours of isolation peeps from four individuals within one captive social group of squirrel monkeys were reliably discriminated using the three principal components generated by our technique. In addition, McCowan's method was superior to the human observer method in attributing whistles to individual dolphins. While some individuals were distinguishable by both methods (i.e. AVA, BRI, CHE; Tables 4, 5), our quantitative technique was able to discriminate the whistles of two individuals (i.e. NOR and SAD; Table 4) that the human observer method was incapable of discriminating (Table 5).

Table 5. Classification of type 2 whistles ($N=86$) from five subjects from the same social group using the human observer method (predominant whistle type for each subject is in bold)

Whistle type 2 category	Number of whistles categorized to each subject				
	AVA	BRI	CHE	NOR	SAD
2A	23				
2B		2			
2C				4	
2R1			1	26	5
2R2			1	2	
2R3			13	7	
2R4			1	1	
Total whistles	23	2	16	40	5

Table 6. Discriminant analysis results on the percentage of isolation peeps ($N=83$) correctly classified to four adult female squirrel monkeys from one social group using McCowan’s method (predominant number of classified isolation peeps for each subject is in bold)

	Number of peeps categorized to each subject				Total	% Correct	χ^2
	KIM	SON	SCO	ZOL			
KIM	10	0	0	0	10	100.0	<0.00001
SON	0	27	1	9	37	73.0	<0.00001
SCO	0	0	15	0	15	100.0	<0.00001
ZOL	0	6	1	14	21	66.7	<0.00001

DISCUSSION

The Fallacy of the Signature Whistle Hypothesis: a Comparative Perspective

The results from this study demonstrate conclusively that 12 bottlenose dolphins from three different social groups do not produce signature whistles in contexts of isolation. These results are inconsistent with the signature whistle hypothesis. We replicated both the approach and methods of historical and current studies reporting signature whistles to ensure that differences in results could not be due to methodological differences in these studies. Instead of obviously different signature whistle contours, we found that dolphins produce one shared predominant whistle type that contains individual variability in contexts of isolation, a rise whistle type previously termed whistle type 2 (McCowan & Reiss 1995b). These results are consistent with those of past studies conducted prior to the advent of the signature whistle hypothesis (Dreher & Evans 1964; Dreher 1966; Burdin et al. 1975; Kaznadzei et al. 1976). In many of these studies, investigators have reported a predominant rise whistle type in contexts of separation and more socially interactive contexts in both captive and free-ranging populations of dolphins.

In addition, we have shown that this predominant shared whistle type contains ‘signature information’, and thus is similar to the contact calls found in other species. For example, subtle contour variations in dolphin type 2 whistles and squirrel monkey isolation peeps

distinguished individuals in both species (Tables 4, 6, Figs 2, 3). This individual contour variation in squirrel monkey isolation peeps has never been interpreted as representing individually distinctive ‘signature call types’ but rather ‘signature information’ within a call type (Symmes et al. 1979). Thus the subtle variation in the contours of type 2 whistles across individual dolphins should not be considered representative of different signature whistles, which are defined as categorically different whistle types that function as labels or names. Rather, such individual acoustic variation within a whistle type is more appropriately considered ‘signature information’, defined as individual differences in ‘voice’ that permit individual and group recognition. Conversely, if dolphins are to be described as having signature whistles based upon such subtle acoustic variation, then squirrel monkeys must also be described as having ‘signature peeps’ that function as labels or names as purported for dolphins. The most conservative interpretation is that both dolphins and squirrel monkeys produce a shared call type in contexts of isolation that contains acoustic information on individual and/or group identity (McCowan et al. 1998; McCowan & Newman 2000), which is common among all other nonhuman animal species studied to date.

The differences between the results of our studies and those reporting signature whistles in dolphins are due to two principal factors: differences in the methods used for collecting and for categorizing whistles (Tyack 1986;

Caldwell et al. 1990; McCowan 1995; Janik & Slater 1998). First, during data collection, signature whistle proponents frequently assume whistles to be the signature whistles of a particular dolphin without appropriate empirical evidence (Tyack 1986; Caldwell et al. 1990; Janik 2000). For example, in the most recent study to date (Janik 2000), wild dolphins were reported to be positively identified as vocalizers at an interanimal distance of greater than 26 m, but the whistles recorded were implicitly assumed to be these dolphins' signature whistles without any empirical testing. However, dolphins produce a large and complex whistle repertoire (McCowan & Reiss 1995a, b; Janik & Slater 1998; Janik 1999), and thus, determining first which one of the many whistle types each dolphin produces is actually its own signature whistle would be necessary to conclude 'whistle matching' in dolphins.

Second, in the categorization of recorded whistles, signature whistle proponents have either visually sorted the spectrograms themselves (Tyack 1986; Caldwell et al. 1990 for review) or have specifically instructed other human observers to look for 'contour stereotypy' as the criterion for classifying whistles into types (Sayigh et al. 1990; Janik & Slater 1998; Janik 1999, 2000). Investigators categorizing the whistles themselves have classified what they term 'partial', 'deleted' and even 'repetitively looped' whistles into one signature whistle type (Tyack 1986; Caldwell et al. 1990; Sayigh et al. 1990). Yet no evidence suggests that 'partial', 'deleted' or 'looped' whistles should be categorized into the same whistle type (for a detailed discussion of this issue see McCowan 1995; McCowan & Reiss 1995a). In addition, researchers who ask naïve blind observers to visually inspect and classify whistles frequently use a small subset of the whistle data set as the means to justify categorizing the predominant remainder of data set themselves (e.g. see Janik & Slater 1998).

Both of these methods are problematic because the investigators who collected the recordings may be unwittingly aware of the identity of the whistler during their categorization of whistles, which would bias classification of whistles towards signature whistles (Tyack 1986; Janik & Slater 1998; Caldwell et al. 1990 for review). In addition, the criterion used by the 'objective' human observers to sort whistles into types is frequently based upon on a subjective five-point similarity/dissimilarity measure (1=highly dissimilar, 5=highly similar). A consistent score of 3 or higher is frequently considered sufficiently similar to categorize two whistles into the same whistle type (Sayigh et al. 1990; Janik 2000). Yet a median score of 3 suggests that whistle similarity is ambiguous, and thus whistles are neither especially similar nor dissimilar. Although visual inspection of spectrograms has been used historically for categorizing animal signals, this type of categorization scheme represents a subjective quantification of an even more subjective qualitative technique (that is marginal at best), and should not be thought to hold the power of more objective quantitative methods.

Furthermore, categorizing call types based upon 'stereotypy' is problematic for any species, and especially

for dolphins considering the reports on how captive dolphins imitate and perceive whistles (Richards et al. 1984; Reiss & McCowan 1993; Ralston & Herman 1995). These independent studies have demonstrated that dolphins perceive and imitate whistles based upon relative whistle contour, not stereotypical whistle contour. Whistle contours shifted, expanded or compressed with respect to frequency and/or time have been documented as perceived similarly by dolphins (Ralston & Herman 1995). In this only study to date on dolphin whistle perception, highly trained dolphin subjects initially categorized whistles based upon absolute frequency and temporal parameters. But once the subjects understood that there were no restrictions on how they should categorize whistles, they categorized whistles based upon relative whistle contour instead. Imitation studies fully support the saliency of relative whistle contour in that younger and adult dolphins spontaneously (without any training) and consistently imitate, even across years, novel computer-generated whistles by compressing, expanding or shifting their imitations with respect to frequency and/or time (Reiss & McCowan 1993). These reports suggest that classifying whistles based upon stereotypical production of whistle contours should be considered inappropriate for dolphins.

More generally, call structure including the degree of stereotypy may vary with behavioural context for any given species. For example, one would expect more stereotypy in calls produced by individuals in contexts of isolation than in the same call types produced during close social interaction (Shannon & Weaver 1949; McCowan et al. 1999). By producing highly stereotypical signal contours under isolation contexts, whether they are individually distinctive or not, signallers can ensure effective signal reception across distance or through other types of 'noisy' communication channels (also see Owings & Morton 1998). Signals and/or portions thereof can be masked by different amounts and types of noise and thus stereotypy may ensure that what is received is more easily perceived and interpreted by the recipient, even if the intended message is similar in multiple contexts. Categorically similar calls can vary from highly stereotypical to more variable in acoustic structure and such acoustic variation is frequently dependent upon the context of its production, as found in other mammalian species (Insley 1992; Norcross & Newman 1993; Brown & Farabaugh 1997; Jennings et al. 1997). Therefore, the stereotypic dolphin whistles that signature whistle proponents describe as signature whistles are likely to be more stereotypical versions of the same whistle types used in contexts other than isolation and by multiple individuals across different social groups, as demonstrated in this and our previous studies (McCowan & Reiss 1995a, b).

Furthermore, papers reporting signature whistles have presented exemplars of isolation whistles to illustrate the lack of variability in signature whistle types. Yet in comparing the spectrograms of isolation whistles across many of the papers reporting signature whistles, one finds striking similarities in the signature whistle contours of different individuals from both similar and different populations (Tyack 1986; Sayigh et al. 1990; Janik &

Slater 1998; Janik 1999, 2000; reviewed in Caldwell et al. 1990). In fact, many of the whistle types described as signature whistles in these studies show remarkable similarity in relative whistle contour to those reported in our own studies (McCowan & Reiss 1995a, b, 1997). For example in Janik & Slater (1998), signature whistle type A is very similar to our whistle type 5. Signature whistle type B is very similar to our whistle type 6. Signature whistle type C is very similar to our whistle type 2, and signature whistle type D is very similar to our whistle type 3 (see Fig. 1; McCowan & Reiss 1995a, b; Figure 1 in Janik 1999). We have previously shown that all of these whistle types are shared across multiple individuals from several different captive social groups. They therefore should not be considered 'signature whistles'. Furthermore, this sharing of whistle types cannot be explained by the hypothesis that dolphins imitate each other's signature whistles, as recently claimed by Janik (2000), because the predominant whistle type in both isolation (Table 1) and socially interactive contexts (McCowan & Reiss 1995a, b, 1997) is the same shared rise whistle type, previously termed whistle type 2.

Finally, signature whistle proponents have speculated that bottlenose dolphins possess obviously categorically different signature whistle contours due to the problem of increased atmospheric pressure on the vocal tract during diving, which is claimed to render the vocal tract incapable of producing reliable vocal cues within whistle types (Tyack 2000). Yet no references are provided to support this claim. In addition, Tyack (2000) argued that the calls of diving marine mammals are unique with respect to this problem by offering squirrel monkeys as a representative terrestrial mammal in which reliable vocal cues are found within call types. In contrast, we have provided clear evidence that dolphins not only produce a shared predominant contact call in isolation like squirrel monkeys, but that the isolation/contact calls of both species contain vocal cues that identify the vocalizer (Tables 4, 6) and in dolphins, the social group (McCowan et al. 1998). Whether dolphins, like squirrel monkeys (McCowan & Newman 2000), use vocal cues to identify specific individuals and/or group members is not yet known and will require appropriately designed playback experimentation.

Categorizing Dolphin Whistles

Our data also suggest that the use of quantitative methods is important in elucidating the structure and organization of animal vocal repertoires, and specifically the whistle repertoires of bottlenose dolphins. Human observers do not represent an unbiased filter through which one can classify the vocalizations of nonhuman animal species. There is no evidence that the human visual system is exactly matched to any nonhuman animal auditory system, and certainly not that of dolphins or squirrel monkeys. Although we may never be completely unbiased in classifying signals, we can maximize objectivity by using the least subjective method.

Although the human observer and contour similarity methods agreed on the classification of a small subset of

whistles in this study, we must be very cautious in concluding that the human observer method alone is an acceptable method. As data sets become larger and more variable, we would expect these two methods to significantly diverge (see McCowan 1995 for a discussion of this issue). The categorization of the over 1000 whistles from four captive bottlenose dolphins in Janik & Slater (1998) was validated using the classification of only 104 whistles (10% of the entire data set) by five naïve human observers who were blind to the identity of the vocalizer and context. Only one human observer, the primary author himself, categorized the entire data set (Janik & Slater 1998). No validation of Janik's own classification of the entire data set was provided in this report. Thus the most effective and reliable method for objectively classifying all of the vocalizations in a large and variable data set is an appropriately designed quantitative method because it removes such observer subjectivity.

Towards a Broader Perspective of Dolphin Whistle Communication

In previous studies we have shown that bottlenose dolphins possess a large and diverse whistle repertoire that changes substantially over development (McCowan & Reiss 1995a, b; 1997). This repertoire appears to be learned in a manner similar to that observed for bird song and human languages, including learning through imitation, overproduction or babbling, and attrition (Reiss & McCowan 1993; McCowan & Reiss 1995b, 1997; McCowan et al. 1999). We have also demonstrated that sequences of whistles by individual dolphins have the potential to contain a large amount of information (McCowan et al. 1999). Indeed, infant dolphins even appear to 'babble' sequences of whistles that eventually become more organized as they become adults (B. McCowan, L. R. Doyle & S. F. Hanser, unpublished data). Thus if we are to unravel the functions of dolphin whistles and the nature of dolphin vocal communication in comparison to that of other species, we will need to take a broader perspective that concentrates on the repertoire structure and sequential organization of dolphin whistle communication.

Acknowledgments

This research was supported in part by the Marine World Foundation, the California Regional Primate Research Center of University of California, Davis and the National Institutes of Health (contract grant numbers: HDO79776-02, RR001690), and funds from the City Council of the City of New York to the New York Aquarium. We would like to thank the volunteer staffs at the former Marine World Africa U.S.A., Six Flags Marine World and the New York Aquarium who assisted in data collection and analysis. Special thanks to Stacie Hooper for assisting in data collection at Six Flags Marine World and to Kim Barlowe and Evelyn Neuteufel at the Osborne Laboratories for their assistance as observers. We also would like to thank the Marine Mammal Department training staffs at the

former Marine World Africa U.S.A., Six Flags Marine World and the New York Aquarium who were responsible for the care and husbandry of the animals. The research presented here was described in Animal Research Protocol No. D9905, approved on 31 May 1999 by the Institutional Animal Care and Use Committee of Six Flags Marine World.

References

- Brown, E. D. & Farabaugh, S. M.** 1997. What birds with complex social relationships can tell us about vocal learning: vocal sharing in avian groups. In: *Social Influences on Vocal Development* (Ed. by C. T. Snowdon & M. Hausberger), pp. 98–127. Cambridge: Cambridge University Press.
- Burdin, V. I., Reznik, A. M., Skornyakov, V. M. & Chupakov, A. G.** 1975. Communication signals in the Black Sea bottlenose dolphin. *Soviet Physics Acoustics*, **20**, 314–318.
- Caldwell, M. C. & Caldwell, D. K.** 1965. Individualized whistle contours in bottlenose dolphins (*Tursiops truncatus*). *Science*, **207**, 434–435.
- Caldwell, M. C., Caldwell, D. K. & Tyack, P. L.** 1990. Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin. In: *The Bottlenose Dolphin* (Ed. by S. Leatherwood & R. R. Reeves), pp. 199–233. New York: Academic Press.
- Ding, W., Würsig, B. & Evans, W. E.** 1995. Whistles of bottlenose dolphins: comparisons among populations. *Aquatic Mammals*, **21**, 65–77.
- Dreher, J. J.** 1966. Cetacean communication: small group experiment. In: *Whales Dolphins and Porpoises* (Ed. by K. S. Norris), pp. 529–544. Berkeley: University of California Press.
- Dreher, J. J. & Evans, W. E.** 1964. Cetacean communication. In: *Marine Bioacoustics. Vol. 1* (Ed. by W. N. Tavolga), pp. 373–399. Oxford: Pergamon.
- Innsley, S. J.** 1992. Mother–offspring separation and acoustic stereotypy: a comparison of call morphology in two species of pinnipeds. *Behaviour*, **120**, 103–122.
- Janik, V. M.** 1999. Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods. *Animal Behaviour*, **57**, 133–143.
- Janik, V. M.** 2000. Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, **289**, 1355–1357.
- Janik, V. M. & Slater, P. J. B.** 1998. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behaviour*, **56**, 829–838.
- Jennings, D. J., Hayden, T. J. & Kent, J. P.** 1997. Changes in the structure of the peep vocalizations of female domestic chicks (*Gallus gallus domesticus*) with age. *International Journal of Comparative Psychology*, **10**, 57–73.
- Kaznadzei, V. V., Krechi, S. A. & Khakhalkina, E. N.** 1976. Types of dolphin communication signals and their organization. *Soviet Physics Acoustics*, **22**, 484–488.
- Leger, Daniel, W. & Didrichsons, Inta, A.** 1994. An assessment of data pooling and some alternatives. *Animal Behaviour*, **48**, 823–832.
- McCowan, B.** 1995. A new quantitative technique for categorising whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae *Tursiops truncatus*). *Ethology*, **100**, 177–193.
- McCowan, B. & Newman, J. D.** 2000. The role of learning in chuck call recognition by young squirrel monkeys (*Saimiri sciureus*). *Behaviour*, **137**, 279–300.
- McCowan, B. & Reiss, D.** 1995a. Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins, Delphinidae *Tursiops truncatus*: a re-evaluation of the signature whistle hypothesis. *Ethology*, **100**, 193–209.
- McCowan, B. & Reiss, D.** 1995b. Whistle contour development in captive-born infant bottlenose dolphins: role of learning. *Journal of Comparative Psychology*, **109**, 242–260.
- McCowan, B. & Reiss, D.** 1997. Vocal learning in captive bottlenose dolphins: a comparison to humans and nonhuman animals. In: *Social Influences on Vocal Development* (Ed. by C. T. Snowdon & M. Hausberger), pp. 178–207. Cambridge: Cambridge University Press.
- McCowan, B., Reiss, D. & Gubbins, C. M.** 1998. Social familiarity influences whistle acoustic structure in adult female bottlenose dolphins (*Tursiops truncatus*). *Aquatic Mammals*, **24**, 21–40.
- McCowan, B., Hanser, S. F. & Doyle, L. R.** 1999. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, **57**, 409–419.
- Norcross, J. L. & Newman, J. D.** 1993. Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls. *American Journal of Primatology*, **37**, 54.
- Owings, D. H. & Morton, E. S.** 1998. *Animal Vocal Communication: a New Approach*. Cambridge: Cambridge University Press.
- Ralston, J. V. & Herman, L. M.** 1995. Perception and generalization of frequency contours by a bottlenose dolphin (*Tursiops truncatus*). *Journal of Comparative Psychology*, **109**, 268–277.
- Reiss, D. & McCowan, B.** 1993. Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. *Journal of Comparative Psychology*, **107**, 301–312.
- Richards, D. G., Woltz, J. P. & Herman, L. M.** 1984. Vocal mimicry of computer-generated sounds and labeling of objects by a bottlenose dolphin (*Tursiops truncatus*). *Journal of Comparative Psychology*, **98**, 10–28.
- Sayigh, L. S., Tyack, P. L., Wells, R. S. & Scott, M. D.** 1990. Signature whistles of free-ranging bottlenose dolphins (*Tursiops truncatus*): stability and mother–offspring comparisons. *Behavioral Ecology and Sociobiology*, **26**, 247–260.
- Shannon, C. E. & Weaver, W.** 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Symmes, D., Newman, J. D., Talmage-Riggs, G. & Katz Lieblich, A.** 1979. Individuality and stability of isolation peeps in squirrel monkeys. *Animal Behaviour*, **27**, 1142–1152.
- Tyack, P. L.** 1986. Whistle repertoires of two bottlenose dolphins, *Tursiops truncatus*: mimicry of signature whistles? *Behavioral Ecology and Sociobiology*, **18**, 251–57.
- Tyack, P. L.** 2000. Dolphins whistle a signature tune. *Science*, **289**, 1310.