90. Lilly, John C. 1965. "Vocal Mimicry in Tursiops. Ability to Match Numbers and Duration of Human Vocal Bursts." Science Vol. 147 (3655). P. 300-301

2

Science - 16 Jul 1965. 90

Vocal Mimicry in Tursiops: Ability to Match Numbers and Durations of Human Vocal Bursts

Abstract. *In addition to its normal underwater sonic communication path, the dolphin* (Tursiops truncatus) *can be trained to emit sounds from the blowhole opened in air. By proper rewarding (positive reinforcement) and evocative techniques, such vocal emissions can be changed from the natural patterns. One such group of new sounds is said to resemble the human voice ("vocal mimicry"). Apects of these sounds which are physically determinable, specifiable, and demonstrable are the similarities in numbers of bursts of sound emitted by man and dolphin and in durations of successive emissions. In 92 percent of the exchanges the number of bursts emitted by* Tursiops *equalled, ±1, the number just previously emitted by a man in sequences of one to ten bursts.*

In the course of experiments on *Tursiops truncatus*, it was found possible to induce individual dolphins to change the mode of vocalization from that used naturally underwater to another type of emission in air, and to modify the natural patterns of sound (*1*). Certain aspects of the results are in essential agreement with the observations of Aristotle (*2*). Some subjective judgments claiming detection of resemblance of some of the sounds made in air to human voice sounds were reported (*1*). In this paper I present some of the physically measurable aspects of two sets of sequential sounds (human, *Tursiops;* human, *Tursiops;* and so on) in which physical resemblances can be detected. Such results suggest vocal mimicry of some specifiable physical features of the voice of the human investigator.

Individual *Tursiops truncatus* (Montagu) were isolated for at least 24 hours and restricted to a tank of water 2.3 m square and 67 cm deep. On the side of the tank was a sidearm constructed of plexiglass, 2.3 m long, 40 cm wide, and 45 cm deep, and containing 15 to 25 cm of water. On a cue from the investigator, the animal would spontaneously enter this area and strand itself.

The animal could either stay in the tank or enter the sidearm on its own initiative. To obtain high-quality recordings of sonic emissions from the human operator and from *Tursiops*, most experiments were done when the animal was in the sidearm. The animal terminated the experiment by returning from the sidearm to the tank. The walls of the tank room were acoustically treated to reduce echos.

Tursiops was induced (trained) to make a series of sounds immediately after a human made a sonic emission consisting of several bursts. The technique employed was that of operant conditioning, with immediate positive reinforcement after each *Tursiops* emission, regardless of whether resemblances were detected (*3*). After initial successes, only emissions resembling those of the man were rewarded; detectable resemblances were thus selected and encouraged. Later, reinforcement was "delayed" or "random," and finally, for short runs, no formal reinforcement was needed.

If, in a given session, *Tursiops* did not give the desired responses, the operator would say so loudly and leave his site beside the sidearm (negative reinforcement); the resulting "transaction deprivation" seemed to improve performance in the next session. Generally, the operator was absent for 5 minutes before returning for another attempt.

Sounds were recorded in two magnetic-tape channels (*4*) with two air microphones (*5*), one beside the man and the other by *Tursiops*.

Three separate series of experiments were conducted with human speech sounds: (i) A set of words and phrases chosen from ordinary speech were used by the human operator; three animals were used and several hundred experiments were conducted over a 4-year period. (ii) Lists of consonant-vowel and vowel-consonant syllables were read aloud by the man; the same three animals were used as in (i), and 130 experiments extended over 3 months. (iii) Various rearrangements of the lists presented by a taped program were played back to one of the *Tursiops* used in experiments (i) and (ii); 11 such experiments were conducted. This report concerns the second group.

During the experiments with free human speech (i) the recordings showed some correspondence between the num-

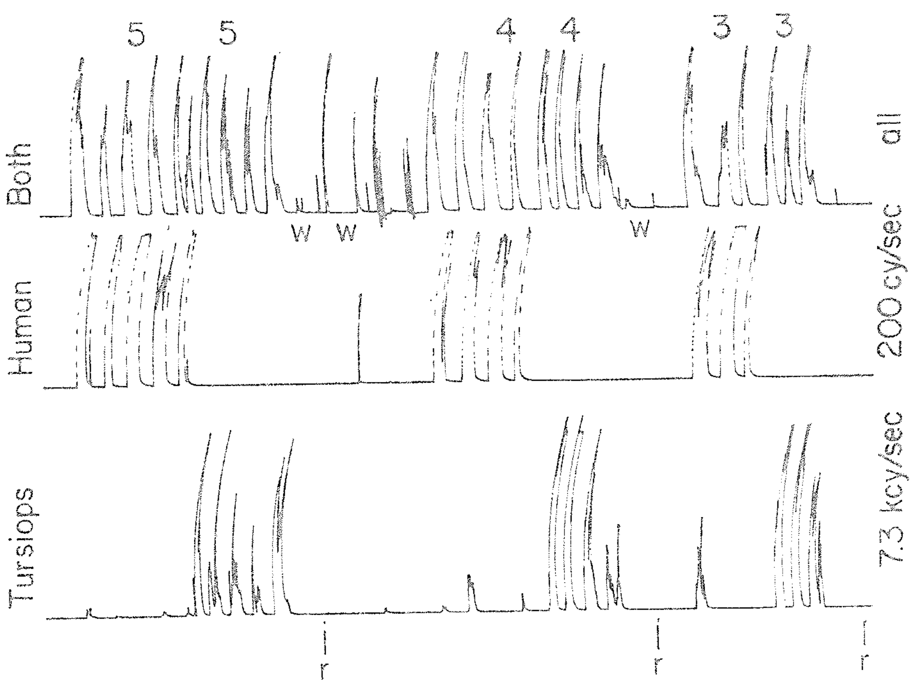

Fig. 1. Three typical vocal exchanges, man-dolphin. Analysis and graphic presentation of a portion of a magnetic tape recording without cutting or editing (real time, continuous). Five, four, and three bursts in each of three human presentations are matched exactly by those of the dolphin. In the middle and bottom traces the two voices are separated for graphic purposes by two narrow pass band filters (Spencer-Kennedy) and displayed separately. To cover the wide amplitude range (40 db) an automatic gain control circuit was applied to the combined signals, and the resulting signal is displayed in the uppermost trace. Food reinforcement was used at the times indicated by *r*; *w* indicates water splashes. The duration of this segment of record is 25 seconds.

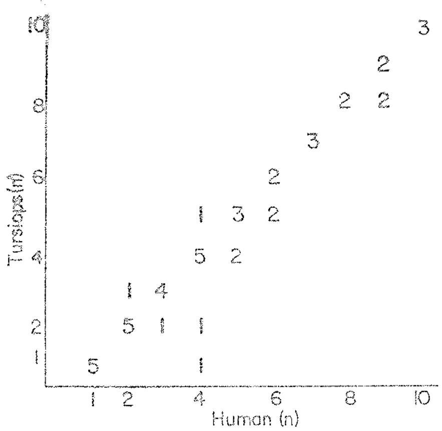

Fig. 2. Distribution of the numbers of bursts in each of 40 human emissions and in each immediately consequent dolphin emission. The number of bursts in each human emission (n) is on the abscissa; the number in the consequent dolphin emission (n') is on the ordinate. The instances of exact equality ($n = n'$) are on a 45° line (starting at $n = 1$ with five instances and running up to $n = 10$ with three instances). The two instances in which the dolphin added one burst are at $n = 2$ and $n = 4$. In seven instances it deducted one; in one instance, deducted two; and once deducted three. In no instance did the dolphin fail to reply to the human emission.

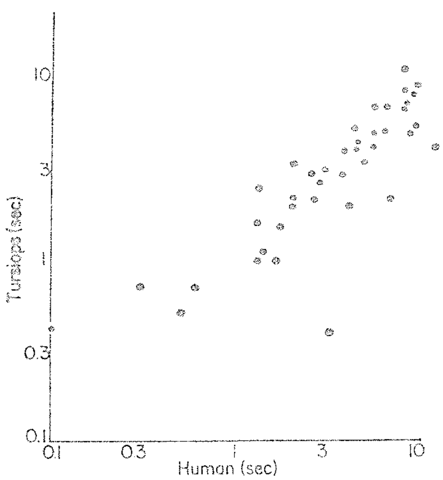

Fig. 3. Distribution of durations of human emissions and of each consequent emission by the dolphin. The duration of each human emission is on the abscissa; of each dolphin emission, on the ordinate: each is a log scale. Exact equality positions are on a line at 45°. Within each emission, the time from the beginning of one burst to the beginning of the next burst, averaged over the whole series, is 0.84 second for the human and 0.78 second for the dolphin's replies. Some longer emission times by the dolphin are accounted for by more bursts per emission (seven instances); in two instances of shorter emissions, the number of bursts was also reduced; in other instances, variations in duration of bursts and of silences between bursts accounted for the differences.

ber of sonic bursts emitted by the man and the number of immediately consequent bursts emitted by *Tursiops*. There was fair accuracy in the replies to up to ten or more syllables if the operator's delivery was sufficiently well differentiated. To check these results a list of nonsense syllables (ii) was devised to clarify the phenomena by removing items carrying meaning for man and introducing systematic variations in the set of stimuli.

The vowels and consonants, selected from the list of Fletcher (*6*), were arbitrarily divided into groups; both one vowel and one consonant of each group were chosen to represent that group, the final list containing 9 vowels (v) and 11 consonants (c) (*6*). Pronounceable combinations of each vowel and each consonant were composed into *cv* and *vc* pairs (198 syllables).

In the initial training period pairs of syllables were chosen for each transmission and were presented in a regularly varying sequence. The operator read the material aloud from a two-dimensional matrix table, varying the order of presentation by reading from left to right, from right to left, down columns, or up columns. Each animal was drilled on the syllables with these paired sequences. When *Tursiops* gave two bursts for two syllables and rarely gave extra bursts, the experiments of present interest began.

The 198 syllables in the list were rearranged in random sequence and grouped, each group containing one to ten syllables, the number being chosen at random. Each syllable was used four times in such groups. Each session consisted of approximately 40 presentations, with one group in each presentation. Each subsequent session started at the point in the list at which *Tursiops* terminated the previous session.

The tape recordings were analyzed both by a listener and by inkwriter recordings. The listener listened to both human and dolphin voices and counted the bursts of sound emitted by each. The inkwriter records were measured by hand in the usual way. Figure 1 illustrates how the numbers of bursts and the durations were measured. Figure 2 shows the degree of correspondence between the number of bursts in the human emission and the number of bursts in *Tursiops'* response. The accuracy of *Tursiops'* reproduction of numbers of bursts (±1) over an 8-minute period was 92 percent in 48 exchanges. In 71 percent of the total exchanges the number of bursts in the dolphin's response exactly equalled that in the preceding human emission.

Figure 3 shows the variation in duration of each emission, duration of the total human emission being compared with the duration of *Tursiops'* total reply. Ninety-two percent of the observations fall within ±50 percent of equality over the range 0.1 to 10 seconds. The number, as well as the duration, of the bursts affects the total emission time.

The duration of each experiment depended on the dolphin; the durations of individual experiments with reproducible rapid exchanges (averaging one every 8 seconds) were from 8 to 30 minutes. Once started, only rarely did the animal fail to reply to the human voice. Analyses of other aspects of these emissions (formant and pitch characteristics) will be reported elsewhere.

These results show quantitatively something of the ability of *Tursiops* to mimic certain aspects of human vocal emissions. This ability seems to be one of many functions of the large brain (1700 g) of this mammal and entails severe modification of the naturally occurring complex vocalizations of *Tursiops* (*1*). Differences from observations of other animals are striking: even parrots and mynah birds apparently do not give such large numbers of replies and such sustained and accurate performances. To date, only dolphins and humans share this ability.

JOHN C. LILLY

Communication Research Institute, Miami, Florida 33133

References and Notes

1. J. C. Lilly, *Amer. J. Psychiat.* 115, 498 (1958); *Man and Dolphin* (Doubleday, New York, 1961); *Proc. Amer. Philos. Soc.* 106, 520 (1962); *Perspectives Biol. Med.* 6, No. 2 (1963); *Arch. Gen. Psychiat.* 8, 111 (1963); A. Montagu and J. C. Lilly, in *The Dolphin in History* (Clark Memorial Library, Univ. of California, Los Angeles, 1963), p. 31; J. C. Lilly, in *Handbook of Physiology: Environment I* (Amer. Physiol. Soc., Washington, D.C., 1964), p. 741.
2. Aristotle, *Works* (Oxford Univ. Press and Encyclopedia Britannica, Chicago, 1952), vol. 2, p. 62.
3. C. B. Ferster and B. F. Skinner, *Schedules of Reinforcement* (Appleton-Century-Crofts, New York, 1957).
4. Minnesota Mining 3M Wollensak model 1580.
5. Shure model 545 Unidyne III and model 560.
6. H. Fletcher, *Speech and Hearing Communication* (Van Nostrand, New York, 1953), p. 68. The nine vowels were i, I, e, E[ɛ], a, o, u, aI, ɔI; consonants were r, l, z, v, tʃ, w, m, n, t, k, s (international phonetic symbols).

7. Supported in part by grants from NIH, AFOSR, and private donors, especially the G. Unger Vetlesen Foundation. The technical help of A. M. Miller, S. McVay, W. A. Munson, H. McFarland, and Elvar Tursiops is gratefully acknowledged.

1 October 1964