101. Lilly, John C. 1967. Dolphin Vocalization in Proc. Conf. on Brain Mechanisms Underlying Speech and Language. F. L. Darley, Ed. A Symposium held at Princeton, N.J. 1965. Grune and Stratton, New York City. P. 13-20

8

DS

Reprinted from Brain Mechanisms Underlying Speech and Language
Copyright Grune & Stratton, Inc., 1967
Printed in U.S.A.

Dolphin Vocalization

JOHN C. LILLY
Communication Research Institute
Miami, Florida

WE HAVE BEEN PURSUING some research with one of the cetaceans, *Tursiops truncatus*, the bottlenose dolphin, in captivity.[240] This animal has a brain approximately the size of ours. As he ages, his brain passes ours; as an adult he has about 10 to 20 per cent more cerebral cortex than we have. About 98 per cent of his brain surface is cortical, as compared to our 96 per cent. The density of the cells in corresponding areas of cortex is close to ours. The number and kind of major connections between cortical areas correspond to those that we have. There are some quantitative differences, however. For example, his visual inputs are about one-tenth those of ours, 120,000 fibers/eye, for our 1,200,000. However, on the ear side, he has 2¼ times the number of fibers we have. His hearing frequency spectrum is approximately five times ours. His usable spectrum, in terms of complex pattern hearing, is something of the order of five times ours in frequency.

If we compare his highest and lowest frequencies and their ratio, it comes out about the same as the ratio for our speech band. If we accept the telephone speech band as the one that carries the most essential meanings for us, that is, from 300 cycles to about 3,500, and you now multiply those frequencies by the ratio of the speed of sound in air to that in water,[5] you come out with the bands that the dolphins apparently use in their intraspecies communication (1500-17,500 cycles).

We have made several thousands of measurements of the ranges which they cover in exchanging whistles one with the other and have found that about 90 per cent of the lowest frequencies that they choose run around 6 kilocycles whereas the highest frequencies of the fundamental that they use run around 24 kilocycles. They can use higher intrinsic pulse frequencies (for sonar) than these (though not higher pulsing rates). The upper limits that we have worked with run around 300 kilocycles. However, most of the energy seems to peak between 40 and 120 kilocycles in this band.

The vocal versatility of a given individual is rather extreme. They can "mimic" simple tunes that are played to them, a few notes each. They can mimic the variations in the human voice (not all the variations but certain aspects of the human voice) extremely well, through rather long passages when newly played to them.

In a recent issue of *Science*[243] I published a paper dealing with two of the physically measurable aspects of this mimicry. If one makes up a set of "nonsense" syllables so as to avoid the problem set up by using "meaning" in the transmissions, one can then furnish the animal with a set of sonic human stimuli, voice given, including all the complexity of our speech, with very large numbers of variations in various parameters of the sounds.

We chose a set of seven vowels (ē/, ĭ/, ā/, ĕ/, ä/, ō/, o͞o/) and two diphthongs (aĭ/ and oi/). We chose eleven consonants: r/, l/, z/, v/, ch_ or _tch/, w/, m/, n/, t/, k/, and s/. We combined these in vowel-consonant and consonant-vowel pairs and randomized a set of these in the usual fashion with a set of random numbers. Then we divided them into groups, randomly ordered as to the number of nonsense syllables in each group, in groups from one to ten (see Table 1.).

This list was then read to the dolphin in air by the human operator standing by the tank in which the dolphin was resting (Fig. 11). Within a matter of 15 minutes, a dolphin who had been exposed to a number of kinds of different sonic emissions and who had some operant conditioning training in this area, picked up the rules of this particular experiment and proceeded to put out *matched numbers of bursts* of sound which matched the numbers that the human had just given. For example, if the human said, "ez, ot, ir," the dolphin came back with three corresponding bursts of sound. I do not attempt to mimic the dolphin mimicking the human. This is a very difficult thing for us to do; their usual pitch runs from about 500 to 1000 pulses/second, whereas the human operator's pitch is running from about 125 to 300.

Under very special conditions we have obtained direct mimicking of pitch down to as low as 250. This is not a usual performance.

Let us pay attention to the bursts of sound themselves and the inter-burst silences, their durations and the numbers in the bursts. In the first experiment the dolphin reproduced the number that the human put out in 206 different emissions; he mimicked the number within 91 per cent of the correct value. In the second experiment he was up to 92 per cent. In the third experiment he was running 98.5 per cent. In the fourth

TABLE 1.

	Vowel or Diphchong	r	l	z	v	ch- -tch	w	m	n	t	k	s
ē	ē	ēr	ēl	ēz	ēv	ētch		ēm	ēn	ēt	ēk	ēs
		rē	lē	zē	vē	chē	wē	mē	nē	tē	kē	sē
ĭ	ĭ	ĭr	ĭl	ĭz	ĭv	ĭtch		ĭm	ĭn	ĭt	ĭk	ĭs
		rĭ	lĭ	zĭ	vĭ	chĭ	wĭ	mĭ	nĭ	tĭ	kĭ	sĭ
ā	ā	ār	āl	āz		ātch		ām	ān	āt	āk	ās
		rā	lā	zā	vā	chā	wā	mā	nā	tā	kā	sā
ĕ	ĕ	ĕr	ĕl	ĕz	ĕv	ĕtch		ĕm	ĕn	ĕt	ĕk	
		rĕ	lĕ	zĕ	vĕ	chĕ	wĕ	mĕ	nĕ	tĕ	kĕ	sĕ
ä	ä	är	äl		äv	ätch		äm	än	ät	äk	äs
		rä	lä	zä	vä	chä	wä	mä	nä	tä	kä	sä
ō	ō	ōr	ōl	ōz	ōv	ōtch		ōm	ōn	ōt	ōk	ōs
		rō	lō	zō	vō	chō	wō	mō	nō	tō	kō	sō
o͞o	o͞o	o͞or	o͞ol	o͞oz	o͞ov	o͞otch		o͞om	o͞on	o͞ot	o͞ok	o͞os
		ro͞o	lo͞o	zo͞o	vo͞o	cho͞o	wo͞o	mo͞o	no͞o	to͞o	ko͞o	so͞o
aī	aī	aīr	aīl	aīz	aīv	aītch		aīm	aīn	aīt	aīk	aīs
		raī	laī	zaī	vaī	chaī	waī	maī	naī	taī	kaī	saī
oi	oi	oir	oil	oiz	oiv	oitch		oim	oin	oit	oik	ois
		roi	loi	zoi	voi	choi	woi	moi	noi	toi	koi	soi

Randomized sets of vowel-consonant and consonant-vowel pairs used to test dolphin mimicry. (From Lilly.[23])

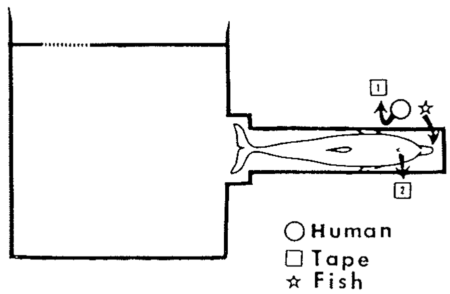

Fig. 11. Schema of experimental configuration. The Tursiops is in the recording position in the side-arm. The investigator (O) is standing beside the side-arm; his voice is recorded through microphone #1 and tape channel #1. The emissions from the blowhole of the Tursiops are recorded through microphone #2 and tape channel #2. When a food reward ("fish"*) is used, it is either manually given to *Tursiops* or by means of a mechanical feeder triggered outside the tank room. The fiberglass tank (2.5 x 2.5 meters) has a door (........) opening into other tanks. The transparent side-arm is approximately 2.5 meters long by 0.5 meters wide by 0.5 meters deep.

experiment it dropped to 93. By the fifth experiment he was down to 98, and by the sixth experiment he did 98.7 per cent. At the seventh session he failed to perform at all (Fig. 12).

The same list was used in each experiment; we started at a different point in the list each time and alternated in direction for the 198 different items. I do not know whether it was possible for the dolphin to memorize any portion of this list. We did our best to prevent this contingency. No portion of it was gone over any more than any other portion. The emissions, in duration, matched within plus or minus 50 per cent of the duration of the human in 96 per cent of the emissions (Fig. 13). The inter-burst silences matched similarly. Thus, we have a physically determinable series of events with high levels of interinvestigator agreement.

The usual experiment was terminated by the dolphin leaving the experimental situation. He could swim away to his home tank, rest, and then return. Usually he broke the experiment at the end of 10 minutes; at this time the emissions were coming at a rate of about one per second. He then took a five minute break and came back for another ten minutes of work. Experiments varied in duration from 12 minutes to half an hour. We staged three or four such experiments a day when the human operator could stand the pace. The dolphin apparently would "loaf" through them very well. The novelty factor was absolutely essential to elicit this kind of performance. The dolphin stopped this kind of experiment and refused to go on. But if we changed the list to another set of sounds, he would start again with similar accuracies as he had in the first list.

Tursiops truncatus is an animal that varies

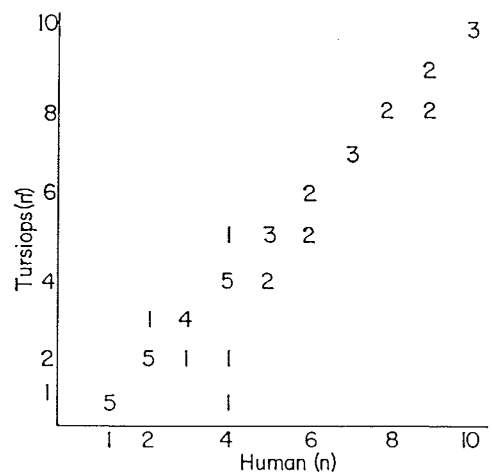

Fig. 12. Distribution of the numbers of bursts in each of 40 human emissions and in each immediately consequent dolphin emission. The number of bursts in each human emission (n) is on the abscissa; the number in the consequent dolphin emission (n′) is on the ordinate. The instances of exact equality (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. (From Lilly.[23]) Figs. 12-14 copyright 1965 by the American Association for the Advancement of Science.

from about 100 Kg. to about 200 Kg. in weight. The brain varies in the newborn from 685 Gm., in one case we have, to 1800 Gm. in the older animals, with all the variations in between. Brain weight increases linearly with body length from birth to full development.

These animals have a long beak ("bottlenose"). The true porpoises do not have this beak and also do not have this size brain. Their brain is comparable in size to that of the chimpanzee.

A usual dolphin's brain weighs 1600 Gm.; the usual man's brain weighs 1400, including the cerebellum in each case. With regard to cortical weights, the dolphin's is approximately 15 per cent more than that of man. The gyri and sulci are more complex. This is true even in the newborns and also in the fetal brain, as we have seen recently.

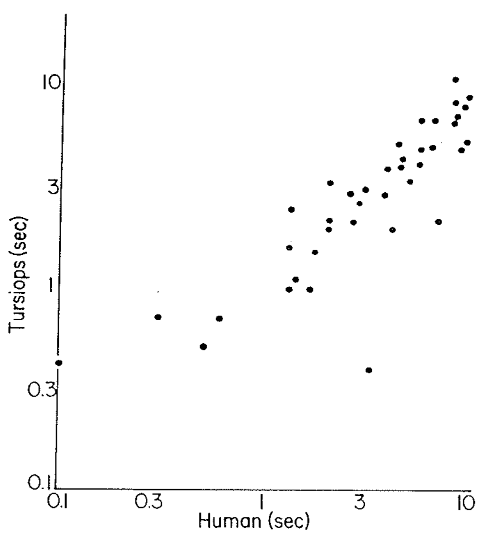

Fig. 13. Distribution of durations of human emissions and of each consequent emission by the dolphin. The duration of each human emission is on the abscissa; of each dolphin emission, on the ordinate: each is a log scale. Exact equality positions are on a line at 45°. Within each emission, the time from 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. (From Lilly.[213])

Do not make the mistake of thinking that *Tursiops* brain is the largest of its kind. The brain of a killer whale is much larger and far more complex. One we collected weighed 4500 Gm. for a 17-foot female *Orcinus orca*.

We are developing an atlas of the dolphin, *Tursiops'*, brain, in collaboration with Yakovlev. This atlas is scheduled for publication in 1967 by Pergamon Press. There are now 2500 slides from which templates are being prepared. The basic portions of the brain, below cortex, are now in hand. For instance, the archeocortex recently has been gone over, and the findings agree with those of Filimonoff,[94] who recently looked at *Delphinus delphis'* brain. The temporal lobes in these animals are very much larger than ours. Currently we are estimating their area as being something of the order of 2½ times the area of ours. However, their occipital poles are very much smaller than ours.

Those who have studied human cortices microscopically agree that the dolphin has the same number of layers of cells as does man, and the cellular density is the same, within a first approximation. The areas of differentiation of cortex are currently being done in our laboratory.

Practically the whole vocalization apparatus, except that portion in the larynx, is innervated by the seventh cranial nerve. They have an intranarial pair of vocalization apparatuses, one on each side. Each one is innervated by about 30,000 fibers. If we added up all the nerve fibers we have for our vocalization, we seem to come out with about the same order of magnitude numerically. On the output side, they come up to where we are, and on the input side somewhat better than we are, in terms of vocalization and the hearing side.

Figure 14 shows the results of one of the experiments in which the animal is mimicking certain aspects of our voices. Another portion of a record showed ten responses—ten nonsense syllables given by the human, immediately followed by the dolphin giving ten sounds in response to those ten. The performances are really remarkable to listen to.

These experiments demonstrate some rather dramatic differences between this species and the mimicking birds. The mimicking birds, as I understand it from those who have worked with them, and I have not, will mimic extremely well, probably pronouncing much better than the dolphin does, but they will not stick to the task nor use the degree of concentration that the dolphin has.

The diligence with which a dolphin will spend minutes working with you, at very high speed, in this kind of exchange is startling. The only other animal species I know

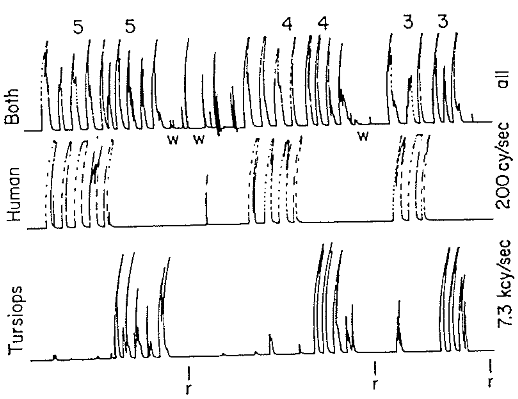

Fig. 14. Three typical vocal exchanges, man-dolphin. Analysis and graphic presentation of a portion of a magnetic tape recording without cutting or editing (real time, continuous). Five, four, and three bursts in each of three human presentations are matched exactly by those of the dolphin. In the middle 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. (From Lilly.[243])

of that will do this is the human species; I know that my four year old daughter can give a performance very similar to that of the *Tursiops*. However, she and I both flunk out at about five or six nonsense syllables on a new list, whereas the dolphin maintains his accuracy right up to ten in that list. (Additional references: [234-238,241,242,244-246].)

CHAIRMAN MILLIKAN: Dr. Geschwind.

DR. NORMAN GESCHWIND: Returning to Dr. Thorpe's presentation, I would like to comment on Konorski's theory about why human beings can repeat and why primates cannot. The structure of which Konorski spoke, the arcuate fasciculus, is one which runs from the postero-superior temporal lobe to the lower frontal lobe. The best evidence seems to indicate that this tract runs backward around the posterior end of the Sylvian fissure and then runs forward in the lower parietal region, eventually reaching the frontal lobe. I agree with Konorski about the particular type of aphasia which results from a lesion of this structure. He calls it "audio-verbal" but it has a more classical name, which is "conduction aphasia." It is a distinctive and not uncommon syndrome.

Konorski based his views on the failure of macaques to repeat on a diagram which was published by Bonin and Bailey.[40] In their diagram of the long corticocortical connections in the nonhuman primate, you do not see the arcuate fasciculus; that is, the posterosuperior temporal region and the lower frontal region are not connected by a pathway which runs back around the Sylvian fissure and then forward in the lower parietal lobe. However, as was pointed out by these same authors in a later publication, the reason for this was that the technic they were using for showing the connections was the technic of strychnine neuronography. With this method you know where the fiber begins and you know where it ends, but you have no way of knowing what pathway it follows since all you can pick up is the firing at the end of the fiber. In drawing their earlier diagram, instead of drawing the pathway to follow that of the known arcuate fasciculus, Bonin and Bailey had drawn a more or less straight line running through the temporal lobe and ending up in the frontal lobe. It is clear, however, from their diagram that the two regions mentioned by Konorski are connected in the macaque. Bonin and Bailey pointed out, furthermore, that it was quite likely that the pathway ran in the lower parietal lobe, but they could neither definitely confirm nor reject this on the basis of the technic that they were using. In conclusion, I agree that the arcuate fasciculus is very important in man for repetition, but I do not believe we can argue that a monkey fails to repeat because he lacks this pathway. I think it is clearly present in the nonhuman primate, and we must therefore seek another explanation for the failure of repetition in these animals.

DR. GIAN FRANCO ROSSI: A very simple question. I was impressed by the similarity between some of the expression of human language and parrot language. I wonder if you know of any evidence of the existence of brain dominance in birds. Has anybody tried to make lesions, on one side only, of the brain of the birds to see whether and in what way language was affected and, above all, to see whether the possible language impairment occurred only or chiefly following lesion of one side of the brain?

DR. THORPE: No, as far as I know, that has not been done.

DR. H. W. MAGOUN: With respect to Dr. Thorpe's remarks on vocalization in birds, experimental studies on a range of experimental animals, which include birds, carnivores, and subhuman primates, indicate that vocal and related mimetic responses involved in the expression of affective states are managed by a mechanism in the middle brain stem. This subcortical mechanism for faciovocal expression is present in man as well as in animals. Midbrain lesions in man may impair such behavior without impairment of his speech. Conversely, widespread bilateral cortical injury in man may be followed by a pathologic exaggeration of laughter and crying, interpreted as release of lower functions from higher inhibition.[254]

It has sometimes been proposed that man's capacity for speech has developed from this subcortical mechanism for nonverbal communication, but no intimate relationship is known between this deep-lying mesencephalic mechanism, present widely through the animal kingdom, and the topographically distant cortical region for speech, which has only appeared with the relatively recent evolution of associational cortex in the human brain. In keeping with their phylogenetic differences, these two mechanisms for communication display widely differing maturation times in the ontogeny of the human infant. The older, more stereotyped, subcortical emotional mechanism is already functional at birth. By contrast, activity of the cortical mechanism in understandable speech only develops between one and two years after birth and its capacities in written language are not gained until the child is five or six years of age.

These observations seem to oppose the view that man's capacity for speech evolved from the abilities for emotional vocalization present in lower animals. On the contrary, man's communication by symbols, both vocal and written, appears to represent an entirely novel functional increment related to the acquisition of associational cortex in front of the face and hand parts of the motor area in the case of speaking and writing, and around the cortical sensory areas for audition and vision in the case of recognition of spoken and written language. Man's capacities for communicating by symbolic language are unique also in depending upon neural mechanisms which develop only in the dominant one of the two cerebral hemispheres, rather than bilaterally. One can conclude that there are two unrelated central neural mechanisms for vocal expression in vertebrates: one for nonverbal affective communication, widely present in the animal brain stem, and a second for verbal communication, present only in the lateral neocortex of the brain of man.

DR. THORPE: I think it is very important to distinguish in birds those vocalizations which can be produced without any practice or training, which are not learned from those which can be modified by experience.

I did not speak about what we usually designate "call notes" because, in general, these do not require learning for their expression. You can rear birds in auditory isolation and they will come out with the perfectly normal call notes of their species, and there is no impairment.

Where I think you can say that a learned language does come in in bird communication is in the context of song. If you rear a bird such as the chaffinch in auditory isolation, it comes out with only an extremely simple sequence of notes, resembling the full song in only a few respects. If it is kept in isolation until it is about 14 months old, then it can never develop its song further. It retains the restricted kind of vocalization as its "song" for the rest of its life. So that all the fine structure of the song vocalization is the result of learning from conspecifics during the period from the time it leaves the nest through the first autumn but, particularly, again the following spring. Then it is that all the fine detail is put in, and then it is that the individual characteristics of the vocalization appear. This seems to me to be very similar to what we call language. In many species the songs are in fact signals acquired in such a way as to be characteristic of the individual. They are recognized by other individuals as characteristic, as identifying that particular individual.

Coming to the question of language, it seems to me that if you use these three characteristics I mentioned—propositional, syntactic, and purposive—it is difficult to find an animal in which we can say quite definitely that all three exist together, but you can find examples of these characteristics separately in a great variety of animals.

After all, the dance of the honeybee is propositional in that it does give precise information about direction and distance of a good supply. The songs of these shrikes I was describing to you are syntactic. There is, it seems, careful organization of the different phrases in a particular order, as characteristic of the pair. And, as I said, there is a lot of evidence that much communication in animals is purposive. So it seems to me that we can find good examples of all these three characteristics of language, but we cannot at the moment say quite definitely that they are all present in one particular animal.

CHAIRMAN MILLIKAN: Are there other comments or questions? Does that conclude your commentary, Dr. Thorpe?

DR. THORPE: I would like to ask Dr. Lilly in regard to the imitative sounds of the porpoises, what the sound spectrograph shows. The very striking feature, of course, about the vocal imitations of a bird like the Indian mynah (*Gracula religiosa*) is the extraordinary precision of its imitations. The vocalization is so good that you can distinguish individual human voices and vocal inflections; the bird can imitate, say, the same thing in the voices of two or three different human beings it knows. When you look at the sound spectrograms, you find that the hubs of the vowels are present in almost exactly the same relation as they are in the voice of the model. In all respects, this is an extraordinarily precise imitation. I may give an anecdote to show how precise it is. I have had mynahs which can produce, I think one can say perfectly, every phoneme in the English language, and many other phonemes as well. It is easiest to train the mynahs by having them taught by a person. You can train them by tape, but they do not learn so quickly or so securely that way. I had one of my mynahs trained by a lady who was very keen on doing this. I asked her, in order to get some sounds which I had not had from a mynah before, to teach the bird to say, "I just saw a zebra."

She taped all the training sessions, so I could study the details of the learning process. The mynah came out saying this very well. "But," she said, "you know, it is curious that it says 'I just saw a *debra*.'" Actually the teacher is a Hungarian who came to England as a girl, a refugee, and now speaks absolutely perfect English, but she does occasionally mispronounce the "z." She noticed the mistake at once when the mynah said "debra" and not "zebra." That is just a little anecdote indicative of the extraordinary precision of their imitation. But one finds it again and again in one's training experiments.

Mynahs learn a very great deal which they do not often utter. Yet they have phrases which are particular favorites, which they will repeat constantly. They can store away a very great number of sounds but it is often difficult to get them to reproduce their full vocabulary.

I would like to know what the sound spectrogram shows in the case of the dolphin vocalizations, because that seems to me to be the real criterion for good imitation of the human voice. Also I would like to know is there anything known about the range of discrimination of the dolphin's hearing.

DR. LILLY: I think the answer to that is, as the speech people keep saying, that the spectrograph is a very poor measure, a very poor judge of how well anything is mimicked. One can pick up, of course, with the mynahs and the parrots, the basic pitch and various formants of things. Also one can do this to a certain extent with the dolphin, but we have many technical problems here; one is that the dolphin's voice goes to extremely high frequencies, and his hearing apparently falls off very rapidly in the very low frequencies but continues way beyond ours.

To get the best enunciation from a dolphin, we found that we had to chop off the fourth formant very sharply at about 80 db./octave, using very sharp filters, in order not to confuse him with the presence and absence of the fourth formant because apparently his receiving side has such a high amplitude that he tends to overemphasize it in what he puts back. When we cut off the fourth formant, we then find he will mimic very well our formants 2 and 3, as measured on the sound spectrograph. He cannot do very well either with our basic pitch, formant 1 region; in other words, he cannot work down in that region well, except by changing his repetition rate which, of course, does not give you the resonances in that region. In other words, he is copying formants 2 and 3 but putting them at characteristic frequencies two to four times our characteristic frequencies.

This means that, with the ordinary speech spectrograph and with the filters used, you cannot see very well the resemblances. You must widen out the pass band of the analyzing filter, and you must do other things to the spectrograph before you can, as it were, make a one-to-one correlation between these two vastly different regions in frequency.

I do not feel that the dolphin is as good a mimic in our hearing and speech frequencies as are the mynah birds or the human being. This is obvious to anybody who listens to the tapes. As I said before, the pitch is way too high; practically everything is way too high.

All we are saying is that there are certain aspects in which he can mimic extremely accurately with new material over long periods of time, in very complex sequences, which the mynah bird or the parrot cannot do. The dolphin has abilities not matched by the bird, and the bird has abilities the dolphin cannot match. The two are very, very different animals and both very, very different from us. Can Dr. Thorpe's mynahs produce on first exposure without practice 10 nonsense syllables immediately after a human utterance of them, and then 9, 3, 7, 2, etc., at an average rate of 1 per second for stimuli and responses and latencies between human and bird of 0.5 second?

CHAIRMAN MILLIKAN: Dr. Thorpe, do you have any final comment?

DR. THORPE: No, I don't think so, except to say that some birds also can imitate long and complex new sequences.

CHAIRMAN MILLIKAN: Dr. Hirsh will now discuss "Information Processing in Input Channels for Speech and Language."