Perceptual classification based on the component structure of song in European starlings

Timothy Q. Gentner[a] and Stewart H. Hulse
Johns Hopkins University, Department of Psychology, Baltimore, Maryland 21218

(Received 15 March 1999; revised 17 September 1999; accepted 14 March 2000)

The ability to recognize individuals based on their vocalizations is common among many species of songbirds. Examining the psychological and neural basis of this functionally relevant behavior can provide insight into the perceptual processing of acoustically complex, real-world, communication signals. In one species of songbird, European starlings (*Sturnus vulgaris*), males sing long and acoustically complex songs composed of small stereotyped note clusters called motifs. Previous studies demonstrate that starlings are capable of individual vocal recognition, and suggest that vocal recognition results from the association of specific motifs with specific individuals. The present study tests this possibility by examining how variation among the motifs that comprise a song affect its discrimination and classification. Starlings were trained, using operant techniques, to associate multiple songs from a single male starling with one response, and songs from four other male starlings with another response. The level of stimulus control exerted by motif variation was then measured by having subjects classify three sets of novel song bouts in which motifs from the training songs were systematically recombined. The results demonstrate a significant, and approximately linear, relationship between song classification and the relative proportions of familiar motifs from different singers that compose a bout. The results also indicate that the motif proportion effects on song classification are primary to retroactive interference in the recall for specific motifs, and independent of any biases due to the syntactic organization of motifs within a bout. Together, the results of this study suggest that starlings organize the complex vocalizations of conspecifics by memorizing large numbers of unique song components (i.e., motifs) that are then associated with different classes. Because individual starlings tend to possess unique motif repertoires, it is likely that under natural conditions such classes will correspond to individual identity. Thus, it is likely that perceptual processing mechanisms similar to those described by the results of the present study form the basis for individual vocal recognition in starlings. © *2000 Acoustical Society of America.* [S0001-4966(00)04806-2]

PACS numbers: 43.66.Gf, 43.80.Lb [DWG]

INTRODUCTION

Most songbirds studied to date are able to recognize conspecific (i.e., same species) individuals based on their songs. Song-based vocal recognition is well documented among male territorial songbirds (see Stoddard, 1996 for review), and more recent examples of this behavior have been observed in both female songbirds (Wiley *et al.*, 1991; Lind *et al.*, 1997; O'Loghlen and Beecher, 1997) and in nonterritorial songbirds (Gentner and Hulse, 1998). The function of vocal recognition varies depending on behavioral context, helping to modulate the complex social dynamics among territorial males, and perhaps influencing mate choice and/or fidelity in females. Among some nonterritorial species vocal recognition is thought to function in the modulation of dominance hierarchies (Eens, 1997). From a proximate standpoint, the widespread capacity for vocal recognition observed among songbirds implies a number of constitutive cognitive processes. First, the animal must be able to discriminate one song from another. Second, the animal must form associations between certain songs and external referents, such as the individuals singing certain songs or locations in space from which certain individuals sing. For species that sing multiple songs, the problem is made more difficult as either multiple songs must be associated with the same referent, or some categorically unique feature must be linked to the set of songs sung by a given individual. In the present study, we examine the first two of these underlying cognitive processes, by investigating how variation in the component structure of male European starling, *Sturnus vulgaris*, song affects its discrimination and classification.

Male starlings tend to present their songs in long episodes of continuous singing referred to as bouts. Song bouts, in turn, are composed of much smaller acoustic units referred to as motifs (Adret-Hausberger and Jenkins, 1988; Eens *et al.*, 1991), which in turn are composed of still smaller units called notes. Notes can be broadly classified by the presence of continuous energy in their spectrogram representations, and although several notes may occur in a given motif, their pattern is usually highly stereotyped between successive renditions of the same motif. One can thus consider starling song as a sequence of motifs, where each motif is an acoustically complex event. The number of unique motifs that a male starling can sing (i.e., his repertoire size) can be quite large, and consequently different song bouts from the same male are not necessarily composed of the same set

---

[a] Electronic mail: tim@drozd.uchicago.edu

of motifs. This broad acoustical variation in their song provides several potential cues that starlings might use when learning to recognize the songs of an individual conspecific, and while maintaining that recognition over time. One straightforward recognition mechanism is the association of specific motifs with specific singers. Although some sharing of motifs does occur among captive males (Hausberger and Cousillas, 1995; Hausberger, 1997), the motif repertoires of different males living in the wild are generally unique (Adret-Hausberger and Jenkins, 1988; Eens *et al.*, 1989, 1991; Chaiken *et al.*, 1994; Gentner and Hulse, 1998). Thus, learning which males sing which motifs can provide a discriminative cue for song classification.

Data from operant studies of individual vocal recognition in starlings support the idea that recognition is based at the level of the motif. Starlings trained to recognize individual conspecifics by one set of song bouts can readily generalize correct recognition to novel song bouts from the same singers (Gentner and Hulse, 1998; Gentner *et al.*, 2000). However, when these novel song bouts have no motifs in common with the training songs, and when song exposure outside of the operant apparatus is restricted, recognition falls to chance (Gentner *et al.*, 2000). Likewise, starlings trained to discriminate among pairs of motifs will reverse the discrimination when transferred to the same motif sung by the opposite individual, and perform at change when transferred to novel motifs sung by the training singers (Gentner, 1999). This failure to generalize correct recognition to songs composed of novel motifs, or to single novel motifs, is inconsistent with the use of individually invariant source and/or filter properties (so called ''voice characteristics'') for individual vocal recognition in starlings.

Based upon the above results, it appears that starlings learn to recognize the songs of individual conspecifics by attending to information contained at (or below) the level of the motif, and then by associating distinct sets of motifs with individual singers. If this is true, then once recognition is learned it should be possible to control it systematically by varying the proportions of motifs in a given bout that come from two ''vocally familiar'' males. That is, if a song bout contains more familiar motifs from male A than male B, then the bout ought to be recognized more often as having been sung by male A than by male B, and vice versa. The manipulations detailed in the present study provide a test of this hypothesis.

I. GENERAL METHODS

A. Subjects

Eight male European starlings served as subjects for this study. Subjects were wild-caught on a farm 30 miles North of Baltimore, MD. Prior to the start of the study subjects were housed in a mixed sex aviary, along with ~50 other starlings, in large flight cages that each contained three to six birds of the same sex. During the course of the study, the subjects were housed in individual cages in a single sex aviary. The photoperiod in the aviaries followed the natural seasonal change in sunrise and sunset times at this latitude. Throughout the course of the experiment, all birds were maintained on a restricted diet of Purina Start and Grow (Purina) such that each subject weighed 85% of their *ad libitum* body weight prior to testing each day. The birds had access to water at all times. All of the subjects were naive to the stimuli and operant procedures used in this experiment. The same eight subjects participated in each of the three experiments, and the sequence of experimental procedures (1, 2, 3, see later in this article) was the same for each subject.

B. Apparatus

Experimental sessions were conducted inside a sound-attenuating test chamber (IAC Model AC-3) fitted with a custom-made operant response panel (see Gentner and Hulse, 1998, for complete description). Briefly, the panel contained three horizontally aligned response buttons (keys) centered above an opening through which the subjects gained controlled access to a food hopper. Two 10 W incandescent lamps provided indirect illumination of the test chamber. Stimuli were presented through a speaker (Bose model 101) mounted above and behind the response panel, out of view of the subject. A PC-type computer controlled the stimulus presentation, monitored response contingencies, and collected the data for each session. Prior to testing, the maximum sound level of the acoustic stimuli inside the test chamber was set at 70 ± 2 dB (A) SPL. The same apparatus was used in each of the three experiments.

C. Stimuli

1. Song recording

Recordings of six male European starlings were used to generate all the stimuli for this experiment. The procedures for obtaining digital song recordings from male starlings and for manipulating those songs on a computer have been detailed elsewhere (Gentner and Hulse, 1998). Briefly, a minimum of 0.5 h of song was recorded from each of the five males, while housed individually in a large sound-attenuating chamber. During recording, males had visual and auditory access to a female starling. The same female was used to induce song from all the males. All the songs were recorded on digital audiotape (16 bit, 44.1 kHz) using the same microphone (Sennheiser ME66-K6), downloaded to a computer, and high-pass filtered at 250 Hz to remove extraneous low-frequency background noise. Of the six males recorded, only five were used to generate the stimuli for the baseline training. The songs of the sixth male were used only as the ''unfamiliar'' songs during experiment one (see below). None of males whose songs were used to generate the stimuli served as subjects in the operant testing. The males we recorded for these stimuli were taken from an area in which the birds used as subjects were also obtained. However, the subjects were caught nearly three years after the stimulus birds. General familiarity with the training songs is possible, but unlikely, as subjects caught at this site a year earlier failed to recognize these same songs (Gentner *et al.*, 2000). Regardless of this possibility, the effects reported here are based only on the classifications learned in the base-

TABLE I. Baseline stimulus set configurations showing correct responses to the different bouts from each singer. Letters denote singer identity, numbers denote specific song bouts from that singer (e.g., A1 refers to bout number ''1'' from male ''A''). Note that songs A1, A2, B1, and B2 occur in both set 1 and set 2

	Peck left (INDIV)	Peck right (MULT)
Set 1	A1, A2, A3, A4, A5, A6, A7, A8,	B1, B2, C1, C2, D1, D2, E1, E2
Set 2	B1, B2, B3, B4, B5, B6, B7, B8	A1, A2, C3, C4, D3, D4, E3, E4

line training, and any facilitated recognition due to stimulus preexposure would not bias the results except to increase the acquisition rate of the baseline task.

2. Baseline training stimuli

We used two different stimulus sets for the baseline discrimination. Half of the subjects were trained with each set. Each stimulus set consisted of eight exemplars sampled from the songs of a single bird, and eight exemplars sampled from the songs of four other birds (two per bird). Each exemplar was 15 ± 0.5 s of continuous song taken from either the beginning, middle, or end of a song bout. Table I shows the detailed configurations of the two baseline stimulus sets. Many of the exemplars sampled from the beginning of a song bout included whistles, along with other ''warble'' motifs [i.e., ''variable'' motifs, rattles, and high-frequency motifs; see Adret-Hausberger and Jenkins (1988) and Eens (1991) for nomenclature of motif classes]. Those sampled at later time points in a bout comprised only warble song motifs. Earlier data indicate that recognition is easily learned with this length of a song bout sample, and is unaffected by the relative position of the sample within a bout and the broader motif classes it may or may not contain (Gentner and Hulse, 1998).

D. Procedure

1. Shaping

Subjects were trained to work the operant apparatus through a combination of auto-shaping and shaping procedures that represented successively closer approximations to the baseline training task (see Gentner and Hulse, 1998, for a full description). The baseline training procedure used a one-interval choice design (Hulse, 1995; Macmillan and Creelman, 1991). In the task, a peck to the center key initiated a trial by starting the playback of a randomly selected stimulus exemplar. After the exemplar played, pecks to either the left or the right key led to positive reinforcement or punishment depending on the key with which that stimulus was associated. Half of the stimuli were associated with the right key and the other half were associated with the left key (see Table I). Correct responses (e.g., a left key peck following a stimulus associated with the left key) were reinforced with 2.5-s access to the food hopper. Incorrect responses (e.g., a left key peck following a stimulus associated with the right key) were punished with a 6-s time-out during which the house lights were extinguished and the food hopper remained inaccessible. In addition to causing a time-out, an incorrect response to a playback stimulus initiated a correction trial sequence in which the same exemplar repeated on all subsequent trials until the bird responded either appropriately or not at all. The interval between successive trials was 2 s. In the event that a bird failed to respond within 5 s following the completed presentation of a given exemplar, the trial ended without a time-out, and the computer waited for a center key peck to begin the next trial. During the first 12 s of each stimulus presentation responses to the keys had no effect, thereby increasing the amount of song to which subjects were exposed. This, in turn, helped limit potential biases for the animal to solve the task using only the first portion of each exemplar. The first response following this 12-s ''observation-period'' was reinforced.

2. Baseline training

The subjects' task was to classify the songs of an individual male starling on one key, and those of four other conspecific males on the other key. For example, one subject was reinforced for pecking the left key each time it heard a song from bird A, and for pecking the right key each time it heard a song from either bird B, C, D, or E. Another bird was reinforced for pecking the left key each time it heard a song from bird B and for pecking the right key each time it heard a song from bird A, C, D, or E. In each case, we refer to the key associated with the songs of the single bird as the *individual* (INDIV) *key*, and the key associated with the songs from multiple birds as the *multiple* (MULT) *key*. Likewise, the songs associated with each key are referred to as ''INDIV'' and ''MULT'' stimuli respectively (see Table I). The association of the INDIV and MULT stimuli with either the left or right response key was counterbalanced across subjects. Two different stimulus sets were used to train the baseline task (see Table I). Half the subjects ($N=4$) were trained with one set, and half ($N=4$) were trained with the other; thus, each subject was exposed to only one baseline training set.

The distribution of correct responses is dependent upon the extent to which the subjects are capable of recognizing the stimuli associated with each of the two response keys. To solve this task subjects must discriminate among the different stimulus exemplars and associate each one with the proper response. It is important to note, however, that the baseline training task does not require subjects to employ a categorical solution strategy, such as ''this is (or is not) a song from bird A.'' Rather, the subjects only need to learn that one ''set'' of songs is associated with one response, and perhaps that the complement to this set is associated with the other response. Importantly, abstract category formation with respect to the baseline stimuli is not a prerequisite to the experimental manipulations that follow the baseline training. When we refer to classification or recognition in discussing the results of these experiments, we mean explicitly that the subjects discriminated a stimulus exemplar and correctly associated it with an operant response. The present study examines these discriminative and associative behaviors.

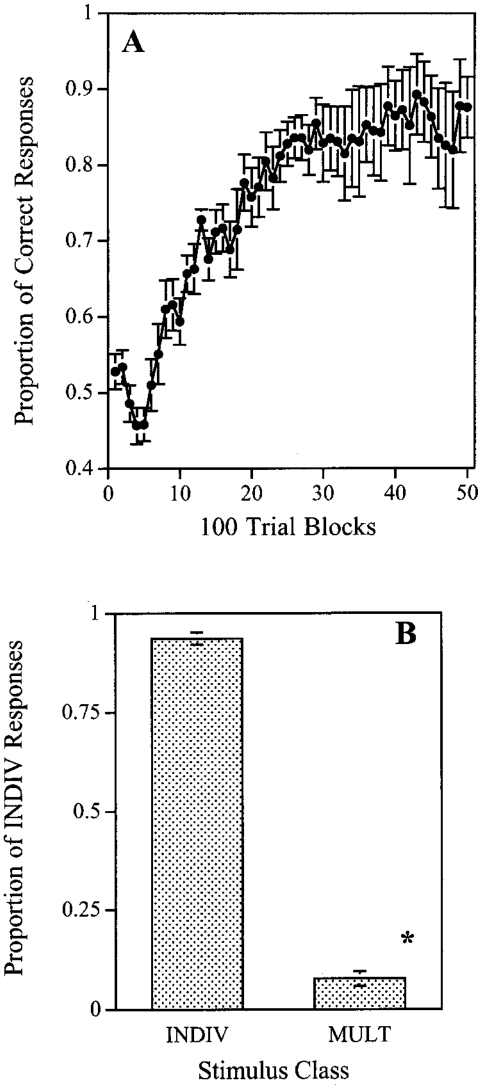

FIG. 1. (a) Baseline acquisition curve. Mean proportion of correct responses made to all the baseline stimuli over the first 50 blocks of trials. (b) Asymptotic performance on the baseline task. Mean proportion of responses made to the key associated with the INDIV stimuli following the presentation of both INDIV and MULT baseline exemplars. (* Significant difference between the means, $p < 0.0001$.)

E. Statistical analysis

For each subject, we recorded a single response (or lack thereof) and the stimulus presented on every trial. Each response was coded as either ''correct'' or ''incorrect,'' depending upon the particular key with which the exemplar presented on a given trial was associated. Alternatively, responses could be coded as either ''INDIV'' or ''MULT'' depending upon the key that was pecked. Using these two coding strategies, performance could be expressed as the probability of responding correctly to any stimulus event [as in Fig. 1(a)], or as the probability of making either an INDIV or MULT response to any stimulus event [as in Fig. 1(b)]. Each subject's performance during the initial baseline training was examined in blocks of 100 trials. Performance during each of the separate experiments was analyzed as a single block of trials that contained all of that subject's responses to the test and baseline stimuli. Performance on the baseline task was assessed against a 95% confidence interval around chance, and with a one-way repeated measures analysis of variance (ANOVA) to detect any significant difference between the responses made to INDIV and MULT baseline stimuli. The significance level (α) was set at 0.05, two tailed. Data from correction trials (i.e., the nonrandom trial immediately following an incorrect response) and trials in which the subject made no response were not included in any of the statistical analyses, but did contribute to the total number of baseline acquisition trials shown in Fig. 1(a). Data from the test conditions were analyzed using multi-way repeated-measures ANOVAs, *t*-tests, and Fisher's PLSD *post hoc* comparisons.

F. Baseline training results and discussion

All of the subjects quickly learned the baseline task, discriminating between the songs of an individual male starling (INDIV stimuli) and those of four other male starlings (MULT stimuli). As a group, the subjects were responding above 70% correct after 768 trials [approximately 1200 trials including correction and no-response trials, see Fig. 1(a)]. Figure 1 shows the acquisition curve for baseline training, along with the mean levels of performance at asymptote. The subjects' performance improved significantly over the first 50 blocks of training [$F(6,49)=16.042$, $p<0.0001$] and there was no significant difference in the rate of acquisition between subjects trained on the two stimulus sets [$F(1,49)=1.064$,NS]. The mean (± SEM) percent correct score over the last five blocks of baseline trials was 93.6 ± 1.6 for the INDIV stimuli and 92.3 ± 1.8 for the MULT stimuli. Likewise, at asymptote [Fig. 1(b)], subjects were correctly classifying the stimuli into their appropriate classes (INDIV and MULT). This is reflected in the significant difference between responses (either left or right key pecks) to the stimuli in the INDIV class and those in the MULT class [$F(1,7)=726.857$, $p<0.0001$, see Fig. 1(b)].

The results of the baseline training demonstrate that subjects were able to discriminate accurately among and classify correctly the different exemplars in each of the two training sets. Moreover, within each training set, we have established two discriminable classes of songs, those associated with the INDIV response and those associated with the MULT response. In the experiments that follow, we examine the perceptual mechanisms underlying this classification by testing the hypothesis that it is based on the varying proportions of familiar motifs in a song bout. Although the baseline data alone do not demonstrate that the subjects categorized the stimuli according to individual identity, the robust discrimination and classification is a sufficient basis for the manipulations that follow. The extent to which their behavior in the operant apparatus is related to the more ecologically relevant task of individual vocal recognition is taken up in the general discussion.

II. EXPERIMENT 1: SERIAL POSITION EFFECTS

A. Introduction

The hypothesis that song classification is based on the varying proportions of familiar motifs in a song bout is, in essence, a claim that motifs are perceptually distinct and memorable auditory events. For recognition to occur in this manner, starlings must parse song bouts into their constituent motifs and then recall with which response each motif has been associated. Under such conditions, motifs that are more easily heard or those that are more easily remembered would exert greater control over their behavior. Therefore, prior to a direct test of our hypothesis, we consider in experiments 1 and 2 how such a motif-based classification system is affected by other cognitive processing mechanisms that, in turn, might influence the salience of specific motifs.

One way that specific motifs might elicit differential control in song classification is through their location in the sampling interval. Because the motifs in a song bout are sung in serial order over time, the recall of specific motifs, and thus decisions about how to classify a song bout at any given time point in the sampling interval, may be affected by constraints in the memory of the classification system. Among humans, the ability to recall a list of words often appears as a characteristic U-shaped function. This is the well-known serial-position effect in which items placed at the beginning and end of a list are more easily recalled than those in the middle (Baddeley, 1990)—the so-called ''primacy'' and ''recency'' effects, respectively. Recency effects are well established in the animal learning literature, but the existence of primacy effects remains more equivocal (see Deacon and Rawlins, 1995; Wright and Rivera, 1997). To the extent that starlings are subject to serial-position effects during song classification, motifs occurring at either the beginning or the end of a sampling interval may exert differential control over their decision to consider that sample (i.e., that bout) as belonging to one class or another. In the first experiment, we investigate how serial-position effects modulate the song classification behavior learned in the baseline training.

B. Methods

1. Stimuli

The test stimuli for experiment 1 were constructed from four pairs of baseline stimulus exemplars. Two pairs of exemplars were chosen from each baseline training set, and the same two pairs were used to test all animals trained with that set. Each pair comprised one INDIV exemplar and one MULT exemplar (set 1: A1–D1 and A2–E1; set 2: B1–A2 and B2–D3, see Table I). Each exemplar was chosen at random with the constraint that it had been accurately classified (above 90% correct) by all subjects during the last five blocks of the baseline training. Using digital editing software (SoundDesigner II, Digidesign-Avid) each baseline exemplar was divided into three ~ 5-s segments, taking care that the point of division between any two adjacent segments fell at a natural transition between motifs. The test stimuli were made by combining the song segments from each pair of INDIV and MULT exemplars according to the six patterns shown in Fig. 2. Given the constraints that the original relative position of each song segment (either first, second, or third) is always maintained, the six patterns in Fig. 2 account for all possible three-element permutations. In addition to the 12 stimuli constructed using the two pairs of INDIV and MULT baseline exemplars (six test stimuli per pair), we constructed two more sets of stimuli by substituting 5-s song segments taken from the song bouts of a novel (i.e., unfamiliar) male starling (see Sec. I C 1). In all, there were 24 different stimuli for each of the two baseline stimulus sets. Each test stimulus was ~ 15 s long.

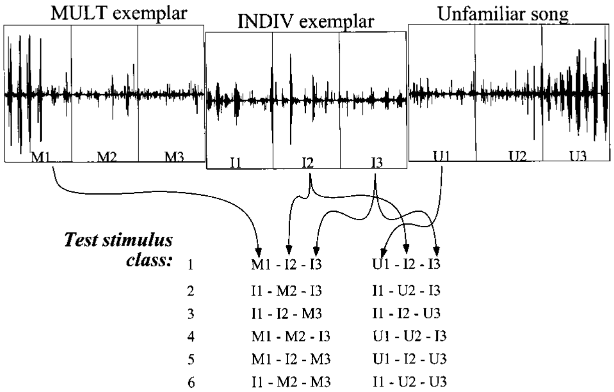

FIG. 2. Strategy for constructing the experiment 1 test stimuli from two baseline stimulus exemplars and an unfamiliar song bout. Non-overlapping 5-s song segments were taken from each of the three stimuli and recombined following the six patterns shown.

The test stimuli used in experiment 1 varied along three dimensions: (1) the location at which a song segment from one singer was inserted into an INDIV exemplar—either the beginning, middle, or end; (2) the source of those inserted song segments—either from a familiar (i.e., MULT) or unfamiliar singer; and (3) the number of song segments inserted into a given exemplar—either 1 or 2.

In addition to examining primacy and recency effects in vocal recognition, the test stimuli for the first experiment also allow us a brief look at the effect that variation in the overall proportion of INDIV and MULT motifs in a particular song bout has on the classification of that bout. For half of the test stimuli one segment from a MULT song (or unfamiliar) bout was combined with two segments from an INDIV bout. For the other half of the stimuli, one segment from an INDIV song bout was combined with two segments from a MULT (or unfamiliar) bout. To the extent that song classification is under the control of the familiarity associated with each of the motifs in a bout, one would expect the test stimuli that contain more INDIV segments than MULT (or unfamiliar) segments to elicit significantly more responses to the INDIV key than those for which the proportion of MULT (or unfamiliar) and INDIV motifs is reversed.

2. Procedure

Once performance on the baseline stimuli reached an asymptote—as assessed by five consecutive 100-trial blocks in which the total percent correct during each block was above 85%—the rate of reinforcement for correct responses was reduced from 100% to 70%. The observation period was then set equal to the length of the stimulus on a given trial, and performance was again allowed to reach an asymptote. During an experimental session, all 24 test stimuli derived from a given baseline stimulus set were presented randomly without replacement on 50% of the trials until the pool of test stimuli was exhausted. Every 48 trials the pool of test stimuli was replenished. During each session, subjects continued to hear the baseline exemplars on the remaining 50% of the trials. All responses to the test stimuli were reinforced with food 70% of the time, regardless of which key the subject pecked, but only if the subject pecked a key. ''No-responses'' to the test stimuli were never reinforced or punished. Each animal completed one 2-h session per day until it had responded at least 50 times to each test exemplar.

Our method of partial nondifferential reinforcement for the test stimuli served two purposes. First, because reinforcement was independent of the response choice on any given trial, subjects could not learn to associate a particular response with a given test stimulus. Therefore, provided that accurate classification of the baseline stimuli is maintained during the experiment, the proportion of right and left key pecks made to the different test stimuli (i.e., the response generalization gradient) reflects the feature space used to solve the baseline task. That is, systematic variation along relevant perceptual dimensions among the test stimuli should lead to corresponding variation in the probability with which a subject responds to either key. Second, because the subjects continued to receive reinforcement for responses to the test stimuli, responding to them did not extinguish. This allowed us to examine responses to a very large number of test stimuli over an extended period.

C. Results

The subjects maintained very accurate discrimination of the baseline stimuli over the course of the first experiment, correctly responding to the INDIV and MULT baseline stimuli ~90% of the time. The probability of pecking the INDIV key in response to an INDIV baseline exemplar was significantly different from the probability of making that same response to one of the MULT baseline exemplars [$F(1,7)=547.062$, $p<0.0001$]. At the same time, not all of the test stimuli were responded to in the same manner. Figure 3 shows the proportion of responses made to the INDIV key following the presentation of the six different test stimuli along with the baseline stimuli for reference. The differences among the mean proportion of INDIV responses associated with each of the six different types of test stimuli were significant [$F(5,35)=6.854$, $p<0.0001$]. There were no significant differences between subjects trained on different baseline stimuli, in either the overall level of response to the test stimuli [$F(1,6)=0.420$, NS] or the pattern of responses distributed among the different test stimuli [$F(5,30)=0.115$, NS]. Thus, for subsequent analyses, we pooled the data for all subjects.

Most of the variation among responses to the test stimuli was due to the number of INDIV song segments that were present in a given test stimulus, and whether or not an inserted song segment was familiar (i.e., from a MULT exemplar) or unfamiliar. These main effects are shown in Fig. 4. The proportion of pecks to the INDIV key was significantly higher for the test stimuli that had only one MULT or unfamiliar song segment inserted in them, than for those test stimuli that had two MULT or unfamiliar song segments inserted [$F(1,7)=220.115$, $p<0.0001$]. In addition, the proportion of INDIV responses made to the test stimuli that had an unfamiliar song segment (or segments) inserted in it were significantly higher than those for the test stimuli that had a MULT song segment (or segments) inserted in them [$F(1,7)=1333.353$, $p<0.01$]. This familiarity effect was most pronounced among the test stimuli in which two song segments had been inserted, as shown by the significant in-

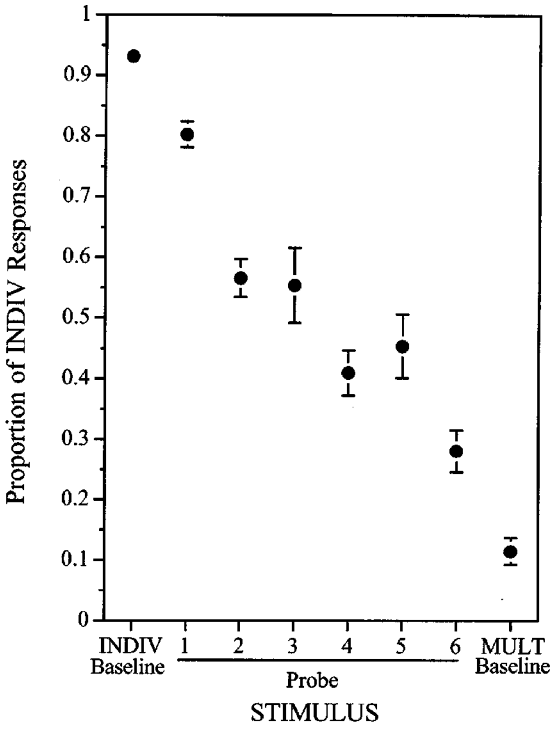

FIG. 3. Mean (±SEM) proportion of INDIV key responses as a function of stimulus class for experiment 1. The means show the data for all the exemplars in each test stimulus class (see Fig. 2) and the baseline stimuli.

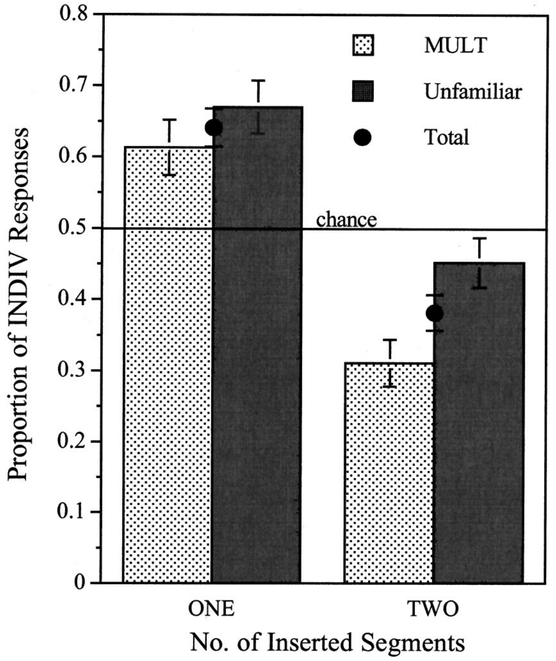

FIG. 4. Mean (±SEM) proportion of INDIV responses as a function of the test stimuli in experiment 1. Means show the data for all exemplars in which either one or two MULT song segments were inserted, and those in which either one or two unfamiliar song segments were inserted.

teraction between the number of inserted song segments and the familiarity (either MULT or unfamiliar) of those inserts $[F(1,7)=19.545,\ p<0.01$; see Fig. 4].

In contrast to the above results, varying the location at which a song segment was inserted into a test stimulus produced a limited effect on recognition. From Fig. 3, one can see that insertions at the first position (test stimuli 1 and 6, see Fig. 2) had smaller effects on classification than did insertions at the second and third positions. *Post hoc* comparisons among the different test stimuli show that responding to the type 1 test stimuli was significantly higher than that to types 1 and 2 ($p<0.05$, Fisher's PLSD), whereas all comparisons among test stimuli 4–6 were not significant.[1]

D. Discussion

The stimuli used in the first experiment were designed to test the idea that motifs at different locations in a sampling interval (e.g., those in the first few seconds) have differential effects on the discrimination and subsequent classification of a song bout. Because serial-position effects are well documented in both human and nonhuman memory tasks (Baddeley, 1990; Deacon and Rawlins, 1995; Wright and Rivera, 1997), we reasoned that similar effects in songbirds might make certain motifs at either the beginning or end of a sampling interval easier to remember. Thus, these motifs may exert greater control during classification decisions that follow the presentation of a song bout. The results of the first experiment suggest, however, that this is not the case. Varying the location of inserted motifs within a song bout had relatively little effect on classification. Among the test stimuli in which only a single segment from a MULT or unfamiliar song was inserted (nos. 1–3, see Fig. 1), insertions at the first segment had the smallest effect on recognition. This is inconsistent with the notion that primacy effects exert control over song discrimination behavior, because such effects should render insertions at the first portion of the song more salient, and thus have a large effect on classification. In addition, the fact that we observed no difference between substitutions made at the second and third segments is inconsistent with a strict criterion for recency effects. However, substitutions at both the second and third positions did have more substantial effects on recognition than substitutions at the first position. Therefore, one might consider this a modest demonstration of recency effects. The trend for more terminal motifs to exert greater control over discrimination in the present task may reflect the ability for starlings to hold sensory traces of different motifs in a short-term memory store. If, as is reasonable to assume, this memory store decays over time and is subject to retroactive interference, those motifs having the closest temporal proximity to the point at which the recognition choice is measured will exert the greatest control. As a caveat to the observed serial-position effects, note that our method of baseline training, in which subjects were forced to withhold responses until most of the stimulus had been presented, may have biased the observed lack of proximity effects. A better test for general primacy effects in songbirds would examine the recall for each motif independently, rather than as a function of a single classification judgment weighted by all the motifs in a sampling interval. For our immediate purposes, however, it was necessary to quantify serial-position effects with respect to performance in the present task.

In contrast to the modest position effects observed during the first experiment, much more dramatic and general effects on discrimination behavior were elicited by variation in the proportion of familiar motifs composing a bout. The subjects were better at correctly classifying test stimuli in which two out of the three song segments were from the INDIV singer than those in which one out of three song segments was from the INDIV singer. The tendency to respond based on motif proportions was observed for all of the test stimuli, regardless of the exact sequence in which the song segments composing each exemplar were presented, and is consistent with the idea the starlings learn to classify conspecific songs by associating sets of motifs with different singers.

Finally, the fact that test stimuli composed of unfamiliar and INDIV motifs were easier to classify than those composed of MULT and INDIV motifs suggests that subjects recognized motifs in all the baseline training songs. Had subjects simply been listening for only INDIV motifs (or MULT motifs) in a pool of otherwise unclassified ''noise,'' the differences between the MULT and unfamiliar song segments would not have affected the classification of the test stimuli. As it was, hearing a MULT song segment in one of the test exemplars was more distracting than hearing an unfamiliar song segment in the same location (see Fig. 4). Thus, it appears that subjects had learned to recognize a substantial portion of the motifs from all, or at least many, of the baseline exemplars, both INDIV and MULT.

III. EXPERIMENT 2: MOTIF SYNTAX EFFECTS

A. Introduction

Another way in which some motifs may exert comparatively stronger (or weaker) control over classification responses is through their syntactic relationship to other motifs in a given bout. Earlier experiments have demonstrated that starlings are sensitive to the syntactical structure of familiar motif sequences in male song bouts, such that randomizing the order of motifs in a familiar bout lowers the recognition of that bout (Gentner and Hulse, 1998). The syntactic structure of a song bout may influence the perception of its constituent motifs in two different ways. First, the position of each motif relative to the other motifs in that bout may be important. For instance, the leading motifs in a sequence may facilitate, or ''prime,'' the recognition of trailing motifs. Receivers may also learn to recognize larger ''chunks'' or ''sub-sequences'' of motifs as single objects, such that violating the sequence of motifs in the chunk would lead to decreased recognition of the constituent motifs. Second, the position of each motif relative to the overall temporal structure of the song bout may be important. Because the motifs in a bout are presented over an extended period of time, when the subjects are trying to recognize a given exemplar, they may listen for specific motifs at specific points in the sampling interval. Violations of the second, more global, syntax rules may affect the perception of individual motifs

TABLE II. The 12 motif patterns used to generate the test stimuli for experiment 2. Letters and numbers denote particular motifs drawn from the INDIV or MULT baseline exemplar, respectively. Patterns marked with an asterisk have more motifs from one or the other baseline exemplars, and were presented in the form shown below and in another form (not shown) in which the first two motifs in the sequence (denoted by the underscore) were removed. The notation used here should not be confused with that used in Table I.

Baseline stimuli pair

INDIV:	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	Q	r
MULT:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Test stimuli

Stimulus type	Sequence																	
I	a	2	c	4	e	6	g	8	i	10	k	12	m	14	o	16	q	18
II	1	b	3	d	5	f	7	h	9	j	11	l	13	n	15	p	17	r
III*	a	b	3	4	e	f	7	8	i	j	11	12	m	n	15	16	q	r
IV*	1	2	c	d	5	6	g	h	9	10	k	l	13	14	o	p	17	18
V	a	b	c	4	5	6	g	h	i	10	11	12	m	n	o	16	17	18
VI	1	2	3	d	e	f	7	8	9	j	k	l	13	14	15	p	q	r
VII	a	8	c	18	e	3	g	15	i	16	k	9	m	2	o	7	q	17
VIII	11	b	4	d	1	f	5	h	13	j	6	l	14	n	10	p	12	r
XI*	a	b	4	10	e	f	16	5	i	j	14	2	m	n	8	7	q	r
X*	15	9	c	d	13	1	g	h	12	3	k	l	11	17	o	p	18	6
XI	a	b	c	18	13	8	g	h	i	15	10	6	m	n	o	5	14	2
XII	9	16	4	d	e	f	12	7	11	j	k	l	1	17	3	p	q	r

and thus recognition of the song. In the second experiment, we vary both these syntactical features in an attempt to make some of the motifs in a given bout more or less salient, and thereby bias classification of the test stimuli.

B. Methods

1. Stimuli

For the stimuli used in experiment 2, we selected one pair from exemplars from each baseline stimulus set (set 1: A3 and C2; set 2: B3 and C3, see Table I). As in the first experiment, each pair comprised one INDIV exemplar and one MULT exemplar that had been very accurately recognized (above 90% correct) by all subjects at asymptotic performance on the baseline procedure. The same pair of stimuli was used for all the subjects trained with each baseline set. In addition, the baseline exemplars chosen for experiment 2 were different than those used to generate the stimuli for experiment 1. Table II shows the patterns used to generate the test stimuli for experiment 2. We purposely chose pairs of INDIV and MULT baseline exemplars that contained the same number of motifs, and parsed each exemplar, using digital editing software (SoundDesigner II, Digidesign-Avid), according to the natural divisions between motifs. Each test stimulus was created by combining half of the motifs from an INDIV exemplar with half of the motifs from the MULT exemplar with which it had been paired. Each test stimulus was ~15 s long.

In all, 12 different primary types of test stimuli were created. As shown in Table II, test stimulus types I and II contained the same number of INDIV and MULT motifs, but the sequence of motifs was arranged so that no two INDIV or MULT motifs appeared in immediate succession. At the same time, the relative position of each motif with respect to its original position in the baseline exemplar was maintained. Stimulus types I and II are complements of one another, such that all of the motifs omitted in type I are contained in type II, and vise versa. Stimulus types III–VI followed a similar pattern to that of I and II except for the following difference. In stimulus types III and IV the motifs from the INDIV and MULT exemplars were organized into consecutive runs of two motifs, so that two motifs from the INDIV exemplar were followed by two motifs from the MULT exemplar. In stimulus types V and VI, the motifs from each of the INDIV and MULT baseline exemplars were organized into consecutive runs of three. Again, for all stimulus types I–VI the relative position of each motif with respect to its original position in the baseline exemplar was maintained. Stimulus types VII–XII follow the general pattern of types I–VI, respectively, except that the relative position of the MULT motifs (with respect to the original baseline stimulus) was not maintained. For example, in type VII stimuli the sequence of INDIV motifs is identical to that for type I, but the sequence of MULT motifs is dramatically altered from its original order in the baseline exemplar (see Table II). The motif placed at each location in types VII–XII was chosen at random (without replacement) from all the MULT motifs in a given baseline exemplar, with the constraint that a given motif could not occupy the same location in the both the baseline and test exemplar. In addition, none of the MULT motifs in stimulus types VII–XII were in an appropriate position relative to other MULT motifs from that same song (e.g., the motif sequence 4–5–6 does not appear anywhere in stimulus type XI or XII).

If the original syntax of motifs is important, then recognition should improve as one looks from stimulus type VII to XII, because these syntax cues have been removed from the MULT motifs and progressively strengthened among the INDIV motifs. In addition, if subjects are expecting to hear certain motifs at certain points in the presentation of an exemplar (e.g., near the beginning), then recognition of the

INDIV motifs ought to be better for types I–VI than for types VII–XII because these cues are missing from the MULT motifs in the later stimuli. Lastly, for stimulus types III, IV, IX, and X the sequences shown in Table II bias the test stimuli so that exemplars in these classes have more INDIV motifs than MULT motifs (or vise versa). For these stimuli, we created an additional version of each exemplar in which the first two motifs were omitted so that the proportions of INDIV and MULT motifs in each exemplar were equal. In all, each subject was exposed to 16 different test stimuli created from a particular pair of INDIV and MULT baseline exemplars.

2. Procedure

Following completion of the first experiment, all subjects were returned to the baseline procedure (at 70% reinforcement) and allowed to stabilize at asymptotic performance. Subjects then began the second experiment. The procedure for experiment 2 was identical to that described for experiment 1, except that different test stimuli were presented. During this experiment, one subject developed a strong position bias for all the stimuli, and his data were excluded from the analysis.

C. Results

The subjects maintained accurate classification of the baseline stimuli over the duration of the second experiment. They responded correctly to the INDIV and MULT baseline stimuli ~90% of the time (mean for INDIV=89.5%±3.8%, mean for MULT=91.7%±1.6%). Accordingly, the difference between the proportion of INDIV responses made to the INDIV and the MULT baseline stimuli was significant [$F(1,6)=249.281$, $p<0.0001$]. At the same time, the mean proportion of INDIV responses to the test stimuli (62.6% ±1.3%) was significantly less than that for the INDIV baseline stimuli ($t=-23.075$, $p<0.0001$, two-tailed), and significantly greater than that for the MULT baseline stimuli ($t=36.035$, $p<0.0001$, two-tailed). This fact suggests that subjects recognized the test stimuli as different from the baseline stimuli. In addition, the overall level of responding to the test stimuli was not significantly different among the two groups of subjects trained on different baseline stimulus sets [see Table I; $F(1,5)=0.382$, NS], and the distribution of responses among the different test stimuli did not differ between these groups [$F(19,95)=1.649$, NS]. Thus, for subsequent analysis, the data were pooled across all subjects.

Although the test stimuli were treated differently than the baseline stimuli, the subjects appear to have responded to all the test stimuli in the same manner. We observed no significant differences among the mean proportion of INDIV responses made to the different test stimuli [$F(19,114)=0.711$, NS].

D. Discussion

The second experiment examined song classification when the syntax of motifs composing a song bout varied in two ways. First, the position of each motif relative to other motifs from the same singer was varied. Second, the position of motifs relative to the overall temporal structure of a bout was varied. In both these cases, we failed to find any significant effect on recognition performance. The subjects generally responded in the same way to all of the test stimuli by pecking the INDIV key approximately 63% of the time. The test stimuli for this second condition each contained an equal number of INDIV and MULT motifs, but varied according to the syntactic structure of the MULT motifs in a given exemplar. We hypothesized that as syntactical cues were removed from the MULT motifs, they would become less distracting, and it would therefore bias the subjects to respond to the test stimulus as if it was an INDIV bout. The data are not consistent with this hypothesis. None of the syntactic variations among the test stimuli elicited any corresponding variation in the subjects' responses.

Although we found no support for the role of syntactical cues in the present context, it should be pointed out that this is a very strong test of that hypothesis. In fact, earlier data do support the role of syntactical cues in more straightforward recognition tasks where starlings are better at recognizing naturally ordered strings of motifs compared to randomly ordered strings (Gentner and Hulse, 1998). The present task was more complex in that we attempted to use the absence of syntactical cues to make some of the motifs in a bout less distracting. In this context, where sets two of motifs are, in a sense, competing for a classification decision, removing syntactical cues from one set of motifs has no detectable effect on the ability to recognize the other motifs in that bout.

IV. EXPERIMENT 3: MOTIF PROPORTION EFFECTS

A. Introduction

As detailed in the general introduction, the existing evidence suggests that starlings rely on variation in the proportions of familiar motifs in order to classify conspecific songs. In the third experiment, we test this hypothesis directly by constructing a large corpus of stimuli in which the percentage of familiar motifs taken from two different training singers is systematically varied from song bouts composed of 10% singer 1 motifs and 90% singer 2 motifs to bouts composed of 90% singer 1 motifs and 10% singer 2 motifs. We then observe how the subjects classify these composite songs. The results of the first two experiments provide important information regarding the effects that serial position and motif syntax have on the classification of conspecific song. In the third experiment, this information is used to build appropriate controls into the set of test stimuli.

B. Methods

1. Stimuli

For the stimuli used in experiment 3, we again selected two pairs of baseline exemplars, one pair from each set of baseline stimuli (set 1: A4 and B2; set 2: B4 and E3, see Table I). Each pair contained one INDIV baseline exemplar and one MULT baseline exemplar. The baseline exemplars chosen for experiment 3 were different than those used in the first two experiments, and the same pairs of baseline stimuli were used for all subjects trained on a given set (see Table I). Each of the four chosen baseline exemplars was parsed into

TABLE III. Examples of motif patterns used to generate the test stimuli for experiment 3. Different letters and numbers denote unique 2–3 motif segments taken from one of the INDIV or MULT baseline exemplars, respectively. Note that for all of the sequence patterns, the original position of each segment in the baseline exemplar has been maintained. The notation used here should not be confused with that in Table I.

	Baseline stimuli									
INDIV:	a	b	c	d	e	f	g	h	i	j
MULT:	1	2	3	4	5	6	7	8	9	10

	Test stimuli									
Stimulus class	Sequence									
I	a	b	3	d	e	f	g	h	i	j
I	a	b	c	d	e	f	g	8	i	j
II	a	2	c	4	e	f	g	h	i	j
II	a	b	c	d	e	6	g	h	9	j
III	a	2	3	d	5	f	g	h	i	j
III	a	b	c	d	e	6	g	8	i	10
IV	1	2	c	4	5	f	g	h	i	j
IV	a	b	c	d	e	6	7	h	9	10
V	1	2	3	4	5	f	g	h	i	j
V	a	b	c	d	e	6	7	8	9	10
VI	a	b	3	d	e	6	7	8	9	10
VI	1	2	3	4	5	f	g	8	i	j
VII	a	2	c	4	e	6	7	8	9	10
VII	1	2	3	4	5	6	g	h	9	j
VIII	a	2	3	d	5	6	7	8	9	10
VIII	1	2	3	4	5	6	g	8	i	10
IX	1	2	c	4	5	6	7	8	9	10
IX	1	2	3	4	5	6	7	h	9	10

ten smaller segments which themselves contained two to three whole motifs. Using digital editing software, the segments from each pair of baseline exemplars were then combined in varying proportions, so that each test stimulus contained ten segments. Table III shows representative patterns for the different classes of test stimuli. Nine different classes of test stimuli were created in which the percentage of INDIV and MULT motifs in a given exemplar varied from 90% to 10% in increments of 10%. For example, in the class I test stimuli the ratio of INDIV:MULT segments was 9:1, in the class II stimuli the ratio of INDIV:MULT segments was 8:2, and in the class III stimuli the ratio of INDIV:MULT segments was 7:3. This pattern of combining song segments from the baseline exemplars meant that for classes I–IV and VI–IX, either the MULT or INDIV motifs, respectively, made up the minority portion. These eight classes of test stimuli were further subdivided into two types of exemplars, those in which the minority motifs occurred somewhere within the first five segments (i.e., the first half of the bout), and those in which the minority motifs occurred somewhere within the last five segments. The remaining class of test stimuli (class V: 50% INDIV–50% MULT) were composed of five segments from the INDIV baseline exemplar followed by five segments from the MULT baseline exemplar, or vise versa. Each test stimulus was ~ 15 s long.

The total number of possible permutations for this stimulus set is very large. Therefore, for each stimulus class, a subset of all the possible permutations was selected to achieve a reasonable number of test stimuli that met the criteria outlined above and also covered the range of possible permutations in that class. For stimulus classes I, IV, VI, and IX, six different exemplars per class were constructed. For classes II, III, VII, and VIII, 12 different exemplars per class were constructed. There were only two exemplars in class V. Each subject was exposed to 74 different test stimuli. For all of the test stimuli in experiment 3, the location of each segment with respect to its position in the original baseline exemplar was maintained (see Table III).

2. Procedure

Following completion of the second experiment, all subjects were returned to the baseline procedure (at 70% reinforcement) and allowed to stabilize at asymptotic performance. Subjects then began the third experiment. The procedure for experiment 3 was identical to that described for experiment 1, except that different test stimuli were presented.

3. Statistics

The statistical analysis followed that given in the general methods, with one exception. In order to balance the structure of the models used to analyze the interaction between the different test stimulus classes and the structure of the stimuli with each class, some of the data from classes with more than six exemplars were not used. For the larger test stimulus classes (II, III, VII, and VIII), we selected six stimuli at random from each of the classes and analyzed the responses across all classes with a repeated measures ANOVA. Identical tests using different exemplar samples from the larger classes yielded results similar to those reported below. The separate repeated measures ANOVAs run on the stimuli within each class used data from all the exemplars.

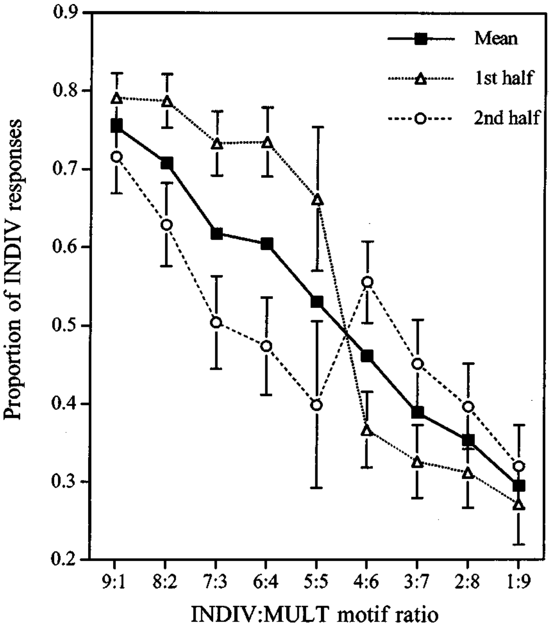

FIG. 5. Mean (±SEM) proportion of INDIV responses as a function of test stimulus class in experiment 3. The mean shows the data from all exemplars in each class (see Table III). Each class is labeled according to the ratio of INDIV:MULT motifs that comprise its exemplars. Data are further divided according to the location of the minority motifs (either the first or last half of the exemplar, see Table III).

C. Results

All of the subjects maintained very accurate classification of the baseline stimuli over the duration of the third experiment, making correct responses to the INDIV and MULT baseline stimuli ~91% of the time (mean for INDIV=91.1%±1.7%, mean for MULT=90.7%±2.8%). Accordingly, the mean proportion of INDIV responses made to the INDIV baseline stimuli was significantly greater than the proportion made to the MULT baseline stimuli $[F(1,7)=568.468,\ p<0.0001]$. From Fig. 5, it is apparent that the proportion of INDIV responses made to each class of test stimuli varied along with the proportions of INDIV and MULT motifs present in each stimulus. The variation among these responses to the different test stimuli was significant $[F(73,438)=12.870,\ p<0.0001]$. As in the first two experiments, performance between the two groups of subjects trained on the different baseline stimulus sets did not differ significantly in either the overall level of response to the test stimuli [$F(1,6)=0.117$, NS], or in the pattern of responding observed among the different test stimuli [$F(73,438)=0.759$, NS]. Subsequent analyses pooled the data from all subjects.

To examine more subtle patterns in the responses made to the different classes of test stimuli we analyzed the variance across stimulus test classes I–IV and VI–IX using a single repeated measures ANOVA (see Sec. IV B). Again, the probability of making an INDIV response varied significantly among the test stimuli $[F(7,49)=47.819,\ p<0.0001]$. We observed no difference in the mean proportion of INDIV responses made to exemplars with substitutions in the first half of the bout and those with substitutions in the last half of the bout [$F(1,7)=1.584$, NS]. However, there was a significant interaction between the location of a substitution and the eight different classes of test stimuli in this analysis $[F(7,49)=9.995,\ p<0.0001]$. This last finding suggests that the location of a substitution (i.e., the minority motifs, see Table III) was important, but only in some of the test stimulus classes (see Fig. 5). To explore this effect further, we ran separate repeated measures ANOVAs on all the test stimuli in each of the different classes. For classes II–VII, the location of the minority motifs had a significant effect on the mean proportion of INDIV responses made to a given stimulus. For classes II–V, subjects were significantly more likely to make an INDIV response when the MULT motifs appeared in the first half rather than the second half of the stimulus. For classes VI and VII, this trend was reversed, with subjects more likely to make an INDIV response when INDIV motifs appeared in the second half rather than the first half of the stimulus. These results can be understood as position effects, and are consistent with the modest recency effects observed in the first experiment. Recall that in experiment 1, motifs in the terminal portion of the sampling interval exerted greater control over the subjects' responses than did those in the initial portion. The serial-position effects for experiment 3 can be seen as the difference between the dotted and dashed curves in Fig. 5, and the motif proportion effect as the slope of each curve.

D. Discussion

In the third experiment, we examined how recognition was effected by variation in the relative proportions of familiar motifs composing a song bout. To do this, we created several sequences of motifs (i.e., song bouts) in which the relative proportions of motifs from two of the baseline-training singers were systematically varied. We then observed the pattern of responses to these bouts. Classification of the test stimuli was strongly controlled by the proportions of familiar motifs in each exemplar. When a bout contained more motifs from the INDIV singer, the subjects tended to classify that song as an INDIV bout. When a bout contained more motifs from the MULT singer, subjects tended to classify that song as a MULT bout. As Fig. 5 shows, the generalization gradient between INDIV and MULT responding closely follows the variation in relative proportions of motifs from the two singers contained in each exemplar. In addition, within each class of test stimuli, all of the different permutations elicited similar proportions of INDIV responses. This last fact suggests that no single motif exerted any greater control over recognition than did any other. By extension, then, it appears that large subsets of motifs were associated with each response.

The data from the third experiment also provide converging support for the manner in which motifs at different locations of the sampling interval control recognition. In general, motif substitutions made in the later half of a bout exerted greater control over recognition than those in the first half of the bout. As with the recency effects observed in the first experiment, those seen in experiment 3 are likely due to the short delay between motifs presented in the second half of the sampling interval and the point at which the subject makes a choice on any given trial. Interestingly, these serial-position effects are not consistent across the different test stimulus classes, but instead are centered around those exemplars in which the INDIV:MULT ratio is closest to one, and at a maximum when the ratio equals one (see Fig. 5). This pattern suggests two conclusions. First, subjects appear to integrate acoustic information over the time span of several motifs. Combined with the negative results from experiment 2, in which we observed no differences in responding to any of the test stimuli, we can estimate a lower bound on this integration window of approximately four to six motifs. Because even those exemplars in which a single minority segment appeared in the first half of the stimulus, and those with insertions at the first position in experiment 1, were treated differently than the baseline stimuli, the upper bound to the integration widow appears to be at least 10 s. Second, because the serial-position effects are inversely related to the bias in motif proportions (see Fig. 5), and at a maximum when the motif proportion bias is minimized (INDIV:MULT ratio$=1$), the stimulus control they elicit appears to be secondary to that elicited by the ratio of familiar motifs.

V. GENERAL DISCUSSION

The results of this study suggest that when starlings are compelled to classify conspecific songs, they do so by memorizing large numbers of unique song components (i.e., motifs), and then organize subsets of these motifs into separate classes. As a cognitive strategy, classifying songs according to their component structure represents a parsimonious method of dealing with these complex acoustic signals under more natural conditions. Because individual starlings tend to possess unique motif repertoires, disjoint sets of motifs will generally correspond to individual identity. Therefore, attending to the motif structure will capture a significant portion of the individual variation in the signal, which, of course, is a requisite to individual vocal recognition. Given this potential source for individual variation, it is tempting to consider the possibility that perceptual processing mechanisms similar to those described here serve as the basis for individual vocal recognition, at least in starlings. Although we suggest that this is likely to be the case, it is important to remember that individual vocal recognition requires more than the discrimination and classification of song. Like most real-world behaviors, individual vocal recognition is likely to involve complex multi-modal processing mechanisms that integrate nonarbitrary visual and auditory information, as well as a level of categorical reference that we have not addressed here. Nonetheless, it is difficult to imagine how individual vocal recognition might occur without the discrimination and classification of song, and our data suggest a likely behavioral mechanism for such cognition.

Although the operant task detailed here is an idealization of vocal recognition in more natural contexts, the two behaviors share a functional equivalence in that they require subjects to associate specific acoustic events with external referents. Under natural conditions, these referents may be locations in space or the visual images of individual conspecifics, whereas in the operant task, the referent is a response button at a given location. We designed the operant contingencies so that classification covaried with individual identity, with the intent that the subjects would engage functional vocal recognition mechanisms to solve the task. In fact, previous data support this idea by suggesting that behaviors observed under similar conditions involve substantially more than the rote memorization of arbitrary stimuli. Correct classification acquired with these same stimulus sets will readily transfer to novel song bouts containing both familiar motifs, and those composed of entirely novel-motifs sung by the training singers (Gentner and Hulse, 1998). Whereas the transfer to songs containing novel instances of familiar motifs holds for all starlings tested to date (Gentner and Hulse, 1998; Gentner *et al.*, 2000), the transfer to songs composed of entirely novel motifs sung by the training singers obtains only when the subjects have had experience with the singers prior to any exposure in the operant apparatus (Gentner *et al.*, 2000). In other words, individual vocal recognition can be demonstrated in starlings, but only after the subjects have had direct experience with the songs that different birds sing. Therefore, when individual vocal recognition has been established in an operant context, it appears to be essentially a problem in the classification of familiar songs. It does not appear to be mediated by the use of more global acoustic properties imparted to all, or some subset of, the motifs sung by a given individual (so called ‘‘voice-characteristics’’). The results of the present study strongly suggest that the discrimination and classification of familiar song is controlled by the relative proportions of motifs from different singers comprising a bout.

Based on the results of this study, we can begin to outline a minimum set of cognitive processes that are likely to serve as the basis for individual vocal recognition in starlings. First, starlings need to be able to dissociate strings of motifs (i.e., songs) from irrelevant background noise and the songs of other individuals occurring simultaneously. This is analogous to the cocktail party effect in human listeners (Cherry and Taylor, 1954), and is a problem in auditory stream segregation (Bregman, 1990; Wisniewski and Hulse, 1997; MacDougall-Shackleton *et al.*, 1998). Once a relevant stream has been segmented from others, that song must be parsed into its constituent motifs, and then sets of motifs that occur in a single bout must be grouped into separate classes. Novel motifs occurring in subsequent song bouts must then be assigned to different classes, perhaps based on their proximity to already familiar motifs. Only after a suitable ‘‘library’’ of motif classes has been established can accurate recognition take place. Within this context, decisions about which singer produced a given song bout are then based on judgments about the relative proportions of motifs from each class that occur within some sampling interval. These putative processes require a mechanism for consolidating neural representations of each motif into a long-term memory store, and suggest a memory retrieval/decision mechanism (or mechanisms) in which similarity judgments between new and stored motifs are made. The results of this study suggest that the later mechanism is integrating similarity judgments made for many motifs in order to reach a decision about individual identity. Although the proximate mechanisms that give rise to these, and in fact most, cognitive process remain largely unknown, we hope that the present study will provide at least some of the groundwork necessary for their subsequent investigation.

ACKNOWLEDGMENTS

The work reported here was supported by NSF Grant No. IBN 9317868 to Stewart Hulse. We thank three anonymous reviewers and D. Wesley Grantham for helpful comments on earlier drafts of the manuscript.

[1] Recall that for each of the two baseline stimulus sets, two pairs of INDIV and MULT exemplars were used to construct the test stimuli for experiment 1. This dimension in the design provided a within-subject control for effects restricted to a single exemplar from the baseline stimulus set, as opposed to those involving more general stimulus control. Including exemplar information in the analysis yielded a significant effect for the serial-position test stimuli [$F(2,14)=8.841$, $p<0.005$]. Closer inspection of the data revealed that this effect was due to a drop in performance when a single substitution was made at the second segment in one of the test stimuli (type 2, Fig. 1). Rather than arguing for reliable serial-position effects, this suggests that in this particular exemplar, motifs located in the central region exerted greater control over the discrimination than did motifs at this location in the other test stimuli.

Adret-Hausberger, M., and Jenkins, P. F. (**1988**). ‘‘Complex organization of the warbling song in starlings,’’ Behaviour **107**, 138–156.

Baddeley, A. (**1990**). *Human Memory: Theory and Practice* (Allyn and Bacon, Needham, MA).

Beecher, M. D. (**1988**). ‘‘Kin recognition in birds,’’ Behav. Genet. **18**, 465–482.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge).

Chaiken, M. L., Böhner, J., and Marler, P. (**1994**). ‘‘Repertoire turnover and the timing of song acquisition in European starlings,’’ Behaviour **128**, 25–39.

Cherry, E., and Taylor, W. (**1954**). ‘‘Some further experiments upon the recognition of speech, with one and with two ears,’’ J. Acoust. Soc. Am. **26**, 554–559.

Deacon, R. M. J., and Rawlins, J. N. P. (**1995**). ‘‘Serial position effects and duration of memory for non-spatial stimuli in rats,’’ J. Exp. Psychol. **21**(4), 285–292.

Eens, M. (**1997**). ‘‘Understanding the complex song of the European starling: An integrated approach,’’ Adv. Study Behav. **26**, 355–434.

Eens, M., Pinxten, R., and Verheyen, R. F. (**1989**). ‘‘Temporal and sequential organization of song bouts in the starling,’’ Ardea **77**, 75–86.

Eens, M., Pinxten, R., and Verheyen, R. F. (**1991**). ‘‘Organization of Song in the European Starling—Species-Specificity and Individual-Differences,’’ Belgian J. Zool. **121**(**2**), 257–278.

Gentner, T. Q. (**1999**). ‘‘Behavioral and neurobiological mechanisms of song perception among European Starlings,’’ unpublished Ph.D. Thesis, Johns Hopkins University.

Gentner, T. Q., and Hulse, S. H. (**1998**). ‘‘Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*,’’ Anim. Behav. **56**, 579–594.

Gentner, T. Q., Hulse, S. H., Bentley, G. E., and Ball, G. F. (**2000**). ‘‘Individual Vocal Recognition and the Effect of Partial Lesions to HVc on Discrimination, Learning, and Categorization of Conspecific Song in Adult Songbirds,’’ J. Neurobiol. **42**(1), 117–133.

Hausberger, M. (**1997**). ‘‘Social influences on song acquisition and sharing in the European starling (*Sturnus vulgaris*),’’ in *Social Influences on Vocal Development*, edited by C. Snowden and M. Hausberger (Cambridge U.P., Cambridge).

Hausberger, M., and Cousillas, H. (**1995**). ‘‘Categorization in birdsong: From behavioral to neuronal responses,’’ Behav. Processes **35**(1–3), 83–91.

Hulse, S. H. (**1995**). ‘‘The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals,’’ in *Methods in Comparative Psychoacoustics (Vol. 10)*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser Verlag, Basel, Switzerland).

Lind, H., Dabelsteen, T., and McGregor, P. K. (**1997**). ‘‘Female great tits can identify mates by song,’’ Anim. Behav. **52**, 667–671.

MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., and White, W. (**1998**). ‘‘Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences,’’ J. Acoust. Soc. Am. **103**, 3581–3587.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U.P., Cambridge).

O'Loghlen, A. L., and Beecher, M. D. (**1997**). ‘‘Sexual preferences for mate song types in female song sparrows,’’ Anim. Behav. **53**, 835–841.

Stoddard, P. K. (**1996**). ‘‘Vocal recognition of neighbors by territorial passerines,’’ in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Comstock/Cornell, Ithaca).

Wiley, R. H., Hatchwell, B. J., and Davies, N. B. (**1991**). ‘‘Recognition of individual males' songs by female Dunnocks: a mechanism increasing the number of copulatory partners and reproductive success,’’ Ethology **88**, 145–153.

Wisniewski, A. B., and Hulse, S. H. (**1997**). ‘‘Auditory scene analysis in European starlings (*Sturnus vulgaris*): Discrimination of song segments, their segregation from multiple and reversed conspecific songs, and evidence for conspecific song categorization,’’ Journal of Comparative Psychology **111**(4), 337–350.

Wright, A., and Rivera, J. (**1997**). ‘‘Memory for auditory lists by rhesus monkeys,’’ J. Exp. Psychol. **23**(4), 441–449.