ANIMAL BEHAVIOUR, 1999, **58**, 37–44
Article No. anbe.1999.1128, available online at http://www.idealibrary.com on IDEAL

Song types as fundamental units in vocal repertoires

WILLIAM A. SEARCY*, STEPHEN NOWICKI† & SUSAN PETERS†
*Department of Biology, University of Miami
†Department of Zoology, Duke University

(Received 12 March 1998; initial acceptance 1 June 1998;
final acceptance 2 March 1999; MS. number: A8152)

We investigated whether song types function as fundamental units of song variation in song sparrows, *Melospiza melodia*. As the size of a male song sparrow's repertoire increases, so does the mean similarity of his song types, as measured by the sharing of minimal units of production (MUPs). It follows that if MUP similarity is important perceptually, then small repertoires (of dissimilar song types) may be functionally equivalent to large repertoires (of similar song types). We performed two experiments to test whether MUP similarity is important perceptually to male song sparrows. Both experiments used a habituation/recovery design, in which recovery in response at a switch in stimuli is used to gauge the subject's perception of the similarity of the stimuli. The results of both experiments indicate that the level of perceived similarity between pairs of songs does not depend on their level of MUP similarity, within the range of MUP similarities found between song types. Songs with high enough MUP similarity to be judged as variants of the same song type are, however, perceived to be much more similar than are any two song types. The results are compatible with a categorical model of song type perception.

'A critical issue in the study of communication is to determine the unit or units of analysis' (Hauser 1997).

Song types are widely viewed by researchers as fundamental units for assessing song complexity in many species of birds. Numerous studies have measured song repertoire size as the number of song types sung by males, and related this measure to male mating success (Howard 1974; Yasukawa et al. 1980; Searcy 1984) or reproductive success (McGregor et al. 1981; Lambrechts & Dhondt 1986; Hiebert et al. 1989). Other studies have experimentally assessed the effects of repertoire size on female preferences (Searcy 1984; Baker et al. 1986) or on territory defence against other males (Krebs et al. 1978; Yasukawa 1981). Repertoire size also has been used extensively in comparative studies (Kroodsma 1977; Read & Weary 1990, 1992). Relatively little attention, however, has been paid to the question of whether song types are fundamental units to the birds that use them. In this paper, we continue our investigation of this question in song sparrows, *Melospiza melodia*, focusing on the response of males to differing degrees of similarity between song types.

The question of whether songbirds recognize song types as units has direct relevance to understanding the evolution of song characteristics. If elaboration or variability in song has been favoured by sexual selection, as has often been hypothesized (Catchpole 1980; Andersson 1994), then it is important to know how birds measure variability, so that we can identify the trait on which selection acts most directly. Furthermore, knowing the relevant dimensions may provide insight into whether the trait is arbitrary or is correlated with something females or other males may benefit from assessing.

In some species of songbirds, males assemble a limited number of syllable types in an open-ended number of combinations, and for these species, syllable repertoire size has been used as the primary measure of song complexity (Catchpole 1980, 1986). In those species in which the number of combinations of syllables or notes appears to be finite, researchers have used the number of such combinations (termed 'song types') as the primary measure of complexity. A major problem arises with this latter practice when variability is found within song types, that is, when variation exists in the syllable or note type sequence found in different renditions of the 'same' song type (Kroodsma 1982). If variability exists within what we recognize as song types, then it may be that the species in question does not employ the same categories as we do or does not recognize song type categories at all.

Song sparrows provide a good example of this categorization problem. Male song sparrows are usually described as singing repertoires of 5–20 song types, but these song types have been defined subjectively by human observers, and observers acknowledge that considerable variation occurs within song types (Nice 1943;

Correspondence: W. A. Searcy, Department of Biology, University of Miami, Coral Gables, FL 33124, U.S.A. (email: wsearcy@miami.edu). S. Nowicki and S. Peters are at the Department of Zoology, Duke University, Durham, NC 27708, U.S.A.

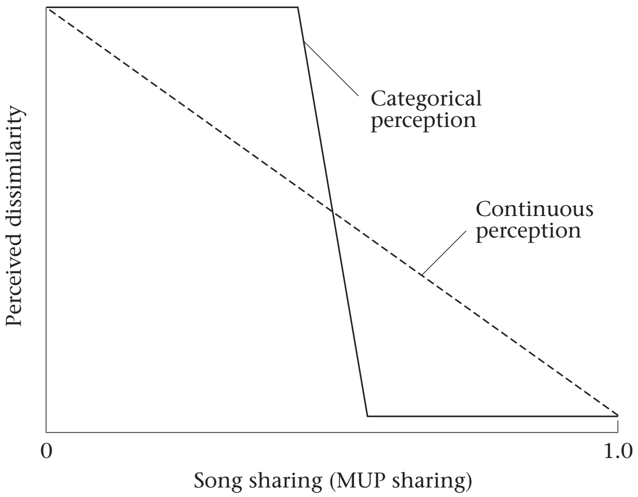

Figure 1. Continuous perception (dashed line) and categorical perception (solid line) models of song type perception. In the continuous perception model, perceived dissimilarity is a smooth, continuous function of MUP similarity; in the categorical perception model, perceived dissimilarity is a step-function of MUP similarity. Perceived dissimilarity (rather than perceived similarity) is graphed to match with our behavioural measure (response recovery).

Borror 1965; Mulligan 1966). Song sparrows have been shown to perceive and respond to within-song type variation (Stoddard et al. 1988; Searcy et al. 1995). Given that variation exists within song types and that this variation is perceived by song sparrows, how can we be sure that song types are not completely artificial categories without any meaning to song sparrows?

Three lines of evidence support the idea that song types are units meaningful to song sparrows. First, Podos et al. (1992) showed that song types can be defined objectively using a statistical clustering algorithm. The similarity between any two songs produced by one male is characterized in terms of the proportion of ‘minimal units of production’ (MUPs) shared by the songs. MUPs are either single notes or groups of notes that always occur together and in sequence (Podos et al. 1992). Cluster analysis of MUP similarities produces categories of songs highly similar to those formed by human observers when subjectively grouping songs into song types. Few songs cannot be readily classified with this method, contrary to what would be expected if continuous variation existed between song types within repertoires (Podos et al. 1992). Second, Stoddard et al. (1992) used operant conditioning to show that song sparrows associate different variants of the same song type with each other. Third, Searcy et al. (1995) used a habituation/recovery playback paradigm to show that male song sparrows in the field respond to song types from the same male as being more dissimilar than are variants of the same song type.

The results to date are compatible with two different models of how perception of songs by song sparrows maps onto song similarity as measured by MUP sharing (Fig. 1). In one model, perceived similarity is a smoothly increasing function of MUP sharing; this model is analogous to what Studdert-Kennedy et al. (1970) termed ‘continuous perception’. Continuous perception accounts for the result that song sparrows judge similarity to be higher among variants than among song types (Searcy et al. 1995) and for the result that song sparrows associate variants of the same song type together (Stoddard et al. 1992). The continuous perception hypothesis makes the additional prediction that song sparrows should judge some pairs of song types to be more similar than other pairs, depending on the degree of MUP sharing within the pairs. This prediction is particularly significant in light of the fact that Podos et al. (1992) found that the mean similarity between song types increases with the number of song types in the repertoires of male song sparrows. If MUP similarity increases with increasing repertoire size and MUP similarity between song types is important to perception, then small repertoires (of dissimilar song types) may be functionally equivalent to large repertoires (of similar song types), weakening the rationale for using song repertoire size as a central measure of song complexity.

In the second model, perceived similarity is a step function of MUP similarity (Fig. 1). The step represents the boundary between what song sparrows do and do not consider to be the same song type. This model is analogous to what Studdert-Kennedy et al. (1970) term ‘categorical perception’ (but see Discussion). The categorical perception hypothesis predicts that any pair of song types (from the same male) would be perceived by song sparrows as being equally distinct; judgments of perceived similarity would not change with MUP similarity as long as two songs were different enough to be judged separate song types. If this prediction is correct, then the correlation between song repertoire size and MUP similarity is unimportant, and song repertoire size is supported as a fundamental measure of song complexity.

We test the relationship between perceived similarity and MUP similarity across song types. We start by using a single-speaker, habituation/recovery design (Searcy et al. 1995) in which the recovery of a subject’s response when the playback stimulus switches from one song type to another is used to measure the perceived dissimilarity of the two songs. With this experimental design, we test the prediction of the continuous perception model that perceived dissimilarity is a decreasing function of MUP similarity between song types. Previous habituation/recovery experiments with swamp sparrows, *Melospiza georgiana* (Nelson & Marler 1989) and red-winged blackbirds, *Agelaius phoeniceus* (Searcy et al. 1994) support the assumption that recovery at a switch in stimuli during territorial playback is proportional to the dissimilarity of the pre- and postswitch stimuli. We also present results using a new, two-speaker habituation/recovery playback design, which allowed us to control additional factors, extraneous to the treatment of interest, that may have affected the subjects’ response.

EXPERIMENT 1

Methods

This experiment employed the single-speaker, habituation/recovery design of Searcy et al. (1995). We played 60 min of a base song type from a single speaker placed well within a male's territory. After 60 min, playback switched to 6 min of a second song type recorded from the same source male. The sole response measure was distance of the subject to the speaker, averaged over 3-min blocks. Searcy et al. (1995) found that subjects habituate to the base song type, showing a significant increase in distance to the speaker between the second 3-min time block and the 20th (and last) 3-min block before the switch. Recovery in response to the switch in stimuli is measured as a decrease in mean distance of the subject to the speaker after the switch.

Trials were carried out during June and July of 1993 and 1994. Subjects in this and the subsequent experiments were free-living male song sparrows holding territories within 10 km of the Pymatuning Laboratory of Ecology, Crawford County, Pennsylvania. We performed all trials in the mornings, between 0600 and 1200 hours.

We placed the single speaker face up on the ground, and set out poles or flagging at distances of 4 and 8 m on both sides of the speaker to aid in estimating distances. We noted distance to the speaker in five categories: 0–2, 2–4, 4–8, 8–16 m and greater than 16 m. We timed trials using a stopwatch started at the first playback song, and noted distances on flow charts divided into 5-s blocks. We calculated mean distance to the speaker for every 3-min block as in Searcy et al. (1995). Songs were played from a Marantz PMD 221 tape recorder over a Nagra DSM speaker. Amplitudes were set at 92–94 dB measured at 1 m from the speakers.

For playbacks, we used songs recorded in Dutchess County, New York. We made recordings either in the field using a Nagra 4.2L tape recorder and a Sennheiser MKH 816 microphone, or in the laboratory using a Marantz PMD 221 tape recorder and a Tandy 1070A microphone. We chose 20 pairs of song types recorded from five males to give a range of similarities as measured by sharing of MUPs. Eight song types were used in two pairs and 24 in one pair only. The pairwise similarity measure is Jaccard's coefficient of correlation:

$$(CC_j)_{\mathrm{a,b}}=c/(c+u_{\mathrm{a}}+u_{\mathrm{b}})$$

where a and b are the two songs being compared, c is the number of MUPs common to both a and b, and u_{a} and u_{b} are the number of MUPs unique to a and b, respectively (Podos et al. 1992). For the analysis we modified CCj by truncating the longer song to match the length of the shorter, thus emphasizing the earlier portions of the song. Songs were not truncated for playback. The observers were blind to the similarity of the two song types on a given playback tape during all trials.

Figure 2 shows three examples of pairs of song types differing in degree of similarity, together with one pair of songs classified as variants of the same type. Note that the classification of pairs of songs as types or variants is not determined by a set criterion for the correlation coefficient; rather, the classification is based on the outcome of a cluster analysis performed on the songs of each male (see Podos et al. 1992 for more details).

We played each of the 20 pairs of songs to four different males at four separate sites, producing a total sample size of 80 individuals, each tested once. For statistical analysis, we averaged recovery across the four males presented with a pair of songs and used the number of stimulus sets as the sample size (McGregor et al. 1992). In some analyses we combined the results from these 20 stimulus sets with results from previous experiments (Searcy et al. 1995), in which we used the same playback design to present 12 stimulus sets with a base song type and a variant of that same song type and 12 sets with a base song and a second song type from the same male. These stimulus sets were also presented to four males each and the results averaged. Following Searcy et al. (1995), we used the increase in mean distance to the speaker between the second and the 20th 3-min time block to measure habituation. We used the decrease in mean distance to the speaker between the last 6-min preswitch and the 6-min postswitch to measure recovery, as this was the most powerful measure of recovery examined in Searcy et al. (1995). We previously showed that recovery was greater for the tapes with a switch to a new song type than for those with a switch to a new variant (Searcy et al. 1995), but we have not previously related recovery to MUP similarity.

Results

Distance to the speaker increased from a mean (± SE) of 3.6 ± 0.4 m during the second 3-min time block to 6.7 ± 0.6 m during the 20th 3-min block, and this measure of habituation was significant ($t_{19}=7.01$, $P<0.0001$). The mean recovery in distance to speaker between the last 6-min preswitch and the 6-min postswitch was 2.5 m, which was significantly greater than 0 (paired $t=6.42$, $N=20$ tapes, $P<0.0001$; Fig. 3). There was a slight negative correlation between recovery and similarity (Pearson's $r=-0.188$), but the relationship was not significant ($P=0.428$) and the R^2 was low (0.035).

When we added the 12 additional sets containing between-type switches (Searcy et al. 1995) to the 20 used in experiment 1, for a total of 32 stimulus sets and 128 trials, the relationship between recovery and similarity between song types still was not significant ($r=-0.118$, $N=32$, $P=0.519$; Fig. 4), and the R^2 (0.014) was even lower than in the previous analysis. The relationship between recovery and similarity for switches between variants was positive ($r=0.537$), rather than negative as predicted, but this relationship too was not significant ($N=12$, $P=0.072$). The mean (± SE) recovery for switches between song types (2.5 ± 0.3, $N=32$) was significantly greater than that for switches between variants (0.4 ± 0.5, $N=12$; unpaired $t=-3.31$, $P=0.0019$).

EXPERIMENTS 2 AND 3

We found no relationship between recovery and similarity between song types in the above analysis, but we were not completely confident of this result because of doubts about the sensitivity of the single-speaker design employed both in experiment 1 and by Searcy et al. (1995). With the single-speaker design, we attempt to

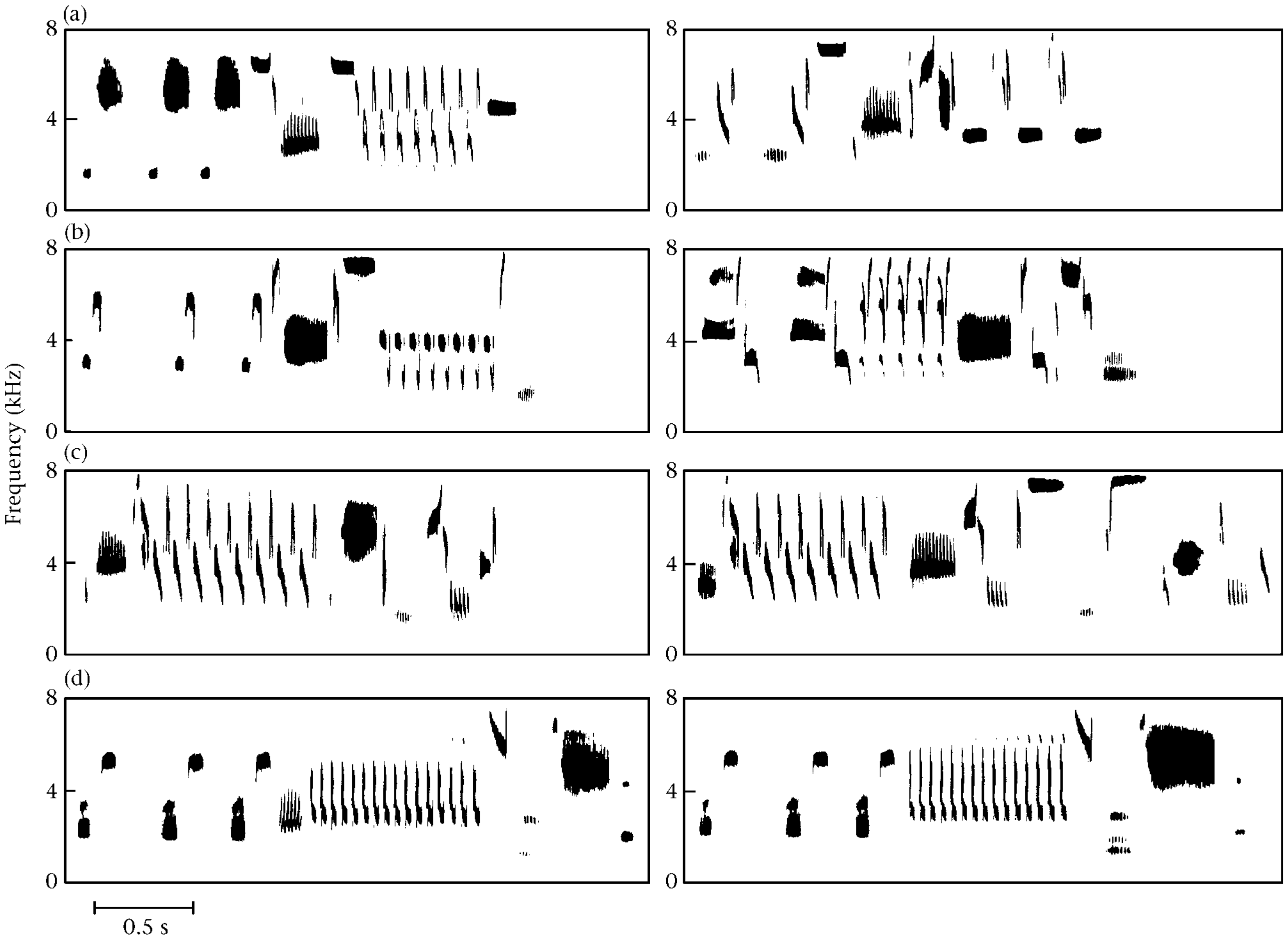

Figure 2. Sound spectrograms of three pairs of song types (a–c) and one pair of variants (d). The three pairs of song types (from experiment 1) were chosen to illustrate zero similarity (a), low similarity (b), and moderate similarity (c). The pair of variants was used in experiment 2. Note that the classification of pairs of songs as type or as variants is determined by cluster analysis of each male's songs rather than by a fixed similarity criterion.

explain variation in recovery measured across males by variation in song similarity. In our experience, a great deal of variation exists in the responsiveness of different individuals to playback of the same stimulus. Variation also exists in the responsiveness of a given individual to a given stimulus depending on context, that is, depending on what else is going on in the immediate environment at the time of playback. The single-speaker design does not control for either between-individual variation in response or for within-individual, between-context variation, and these two sources of variation distract from finding significant relationships between the variables of interest. Accordingly, we devised a new two-speaker design, in which comparisons are made within individuals and within contexts. In experiment 2 we tested the power of this new design by using it in an attempt to replicate an earlier result from the single-speaker design, showing greater recovery for switches between song types than for switches between variants (Searcy et al. 1995). In experiment 3, we used the two-speaker design to test whether recovery is greater for switches between dissimilar song types than for switches between similar song types.

Methods

We placed two speakers within the subject's territory, 16 m apart. The base song type was played from both speakers for 60 min at the rate of one song/10 s, with the songs from the two speakers alternating without overlap (i.e. offset by 5 s). After 60 min, both speakers switched to new stimuli. For example, speaker 1 might switch to stimulus R1 and speaker 2 to stimulus R2. The subject, thus, was asked to judge which stimulus, R1 or R2, was more dissimilar to the base song type.

In experiment 2, one recovery stimulus was a variant of the base song type and the other was a second song type from the same male's repertoire. We used 12 stimulus sets in this experiment, each presented to one subject. We determined the recovery stimulus to be played through each speaker for each trial by flipping a coin. The trials were performed during June 1995. We used the same song types and variants in this experiment that were used by Searcy et al. (1995). These were recorded in Dutchess County, New York, from six males. Recordings were made as in experiment 1.

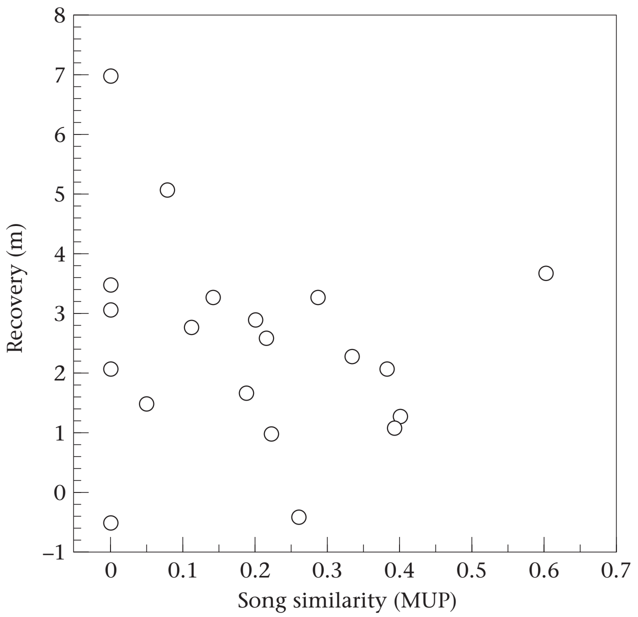

Figure 3. Recovery in distance to the speaker after a switch in song types graphed against the similarity of the preswitch and postswitch song types. Each point represents the mean of four trials performed with one of the 20 stimulus sets in experiment 1. Similarity is measured in terms of sharing of minimal units of production (MUPs).

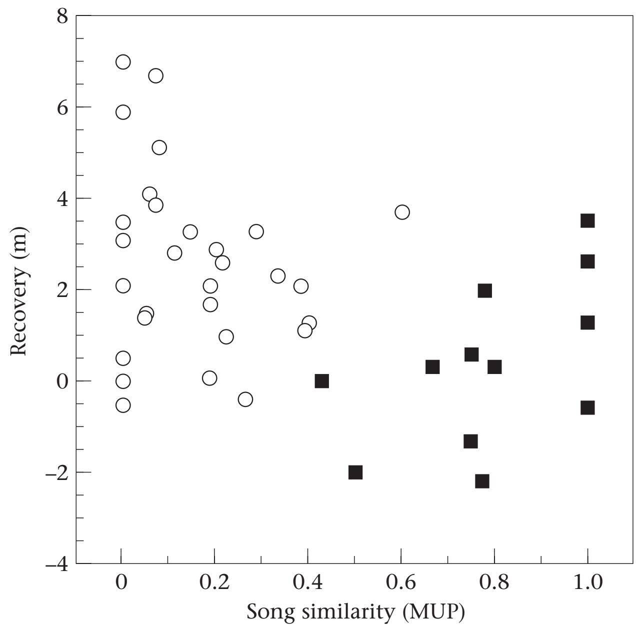

Figure 4. Recovery in distance to the speaker for switches between song types (○) or between variants (■). The song type results include the 20 stimulus sets used in experiment 1 and 12 sets used in Searcy et al. (1995); the variant results are from 12 tapes used in Searcy et al. (1995).

In experiment 3, both recovery stimuli were additional song types from the repertoire of the same male as the base song type. For each stimulus set, one of the recovery song types was chosen because it had a high MUP similarity with the base song type (the ‘similar song type’), and the other was chosen because it had a low MUP similarity with the base song type (the ‘dissimilar song type’). We used 18 sets of three song types, recorded from eight males in Crawford County, Pennsylvania. Recordings were made using Marantz PMD 221 and Sony TCM 5000EV tape recorders and either a Sennheiser ME88 shotgun microphone or a Sony ECM 170 microphone in a Sony PBR 330 parabola. We used local, Pennsylvania songs in these experiments rather than New York songs because we had recently shown that male song sparrows in our population discriminate between these two classes of songs, responding slightly, but significantly, more strongly to local songs (Searcy et al. 1997). We see no obvious reason why discrimination between similar and dissimilar song types should differ for local and foreign songs, but to control for this possibility we used local songs in this final experiment.

Each of the 18 sets of songs in experiment 3 was presented to four different males. We tested only four males (out of 72) in the neighbourhoods where the stimulus songs were recorded, with intervals of 2–4 years between recording and playback, so it is unlikely that subjects had direct familiarity with the playback songs in more than one or two cases. Trials were run during June and July 1996, and May and June 1997. Again, the flip of a coin determined which stimulus played from which speaker in each trial, and in this experiment the observers were blind to which recovery stimulus was similar and which dissimilar to the base song.

In both experiments 2 and 3, we used a Sony TCM D5M stereo tape recorder and two Nagra DSM speakers for playbacks. Amplitude was set at 85–87 dB measured at 1 m. In experiments 2 and 3, the observers had to keep track of the subject's distance to both speakers simultaneously, with poles and flagging set out at measured distances from each speaker to facilitate distance estimations during trials, and with the same distance categories and averaging methods as in experiment 1 using a single speaker. We again use mean distance to the speaker in the last 6-min preswitch minus the distance in the 6-min postswitch to measure recovery.

Results

Experiment 2

For the first 20 time blocks, when both speakers played the same base song, there was no significant difference in the distance of approach to the two speakers (NS in all cases; Fig. 5). This result demonstrates that there was no biased tendency to approach the speaker eventually switching to a new song type over the speaker eventually switching to a new variant, or vice versa. After the switch in stimuli, subjects approached the new type rather than the new variant (Fig. 6). Approach between the last 6-min preswitch and the first 6-min postswitch was significantly greater for the speaker switching to a new song type than for the speaker switching to a new variant (t=4.42, N=12 pairs, P=0.001). We believe this result reflects attraction to the new song type rather than repulsion from the new variant, because in single-speaker experiments, males are

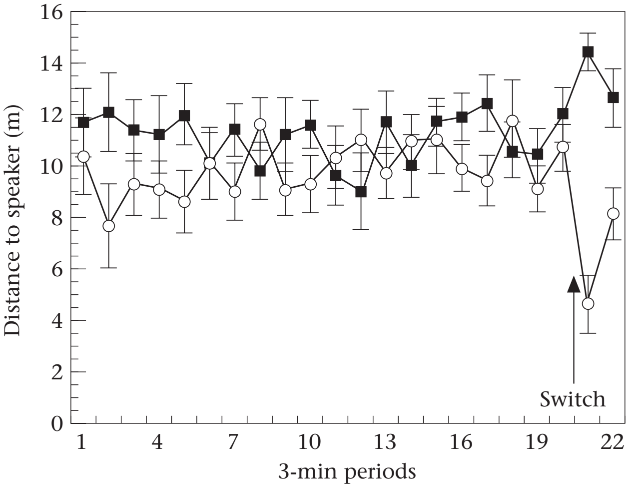

Figure 5. The time course of response during the two-speaker playback experiments in which one speaker switched to a variant (–■–) of the base song and the other speaker switched to a new song type (–○–). The arrow indicates the point at which playback switched from the base song to the recovery stimulus. Error bars are standard errors.

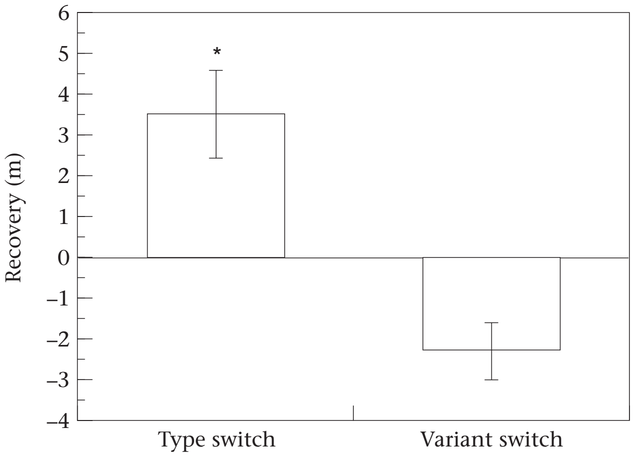

Figure 6. Mean (±SE) recovery in the two-speaker experiments in which one speaker switched to a variant of the base song type and the other speaker switched to a new song type. Recovery is measured as the difference in distance to the speaker between the 6-min preswitch and the 6-min postswitch. Subjects on average approached the new song type while moving away from the new variant. $^*P<0.001$.

attracted to rather than repulsed by new variants (Searcy et al. 1995).

Experiment 3

As in experiment 2, both speakers played the same base songs for the first 20 time blocks, and again there was no significant difference in the distance to the similar speaker compared with the distance to the dissimilar speaker for any time block (NS in all cases; Fig. 7). Approach measured over 6-min was slightly greater for similar songs than for dissimilar songs, but this difference was not significant ($t=0.523$, $N=18$ pairs, $P=0.607$; Fig. 8).

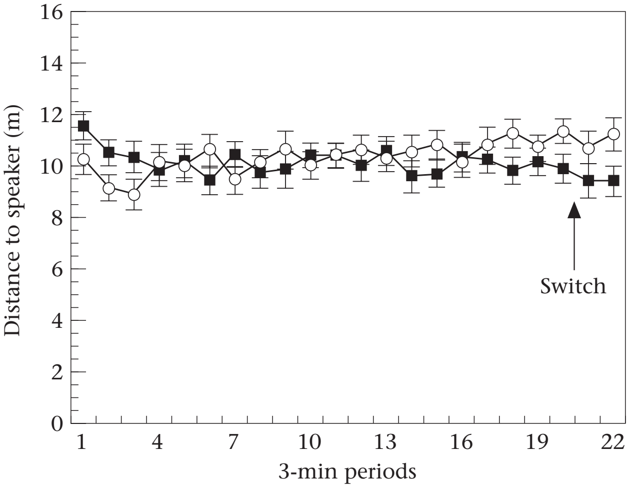

Figure 7. The time course of response during the two-speaker experiments in which one speaker switched to a song type that was similar (–■–) to the base song type and the other speaker switched to a dissimilar (–○–) song type. Scales are the same as in Fig. 5. The arrow indicates the point at which playback switched from the base song to the recovery stimulus.

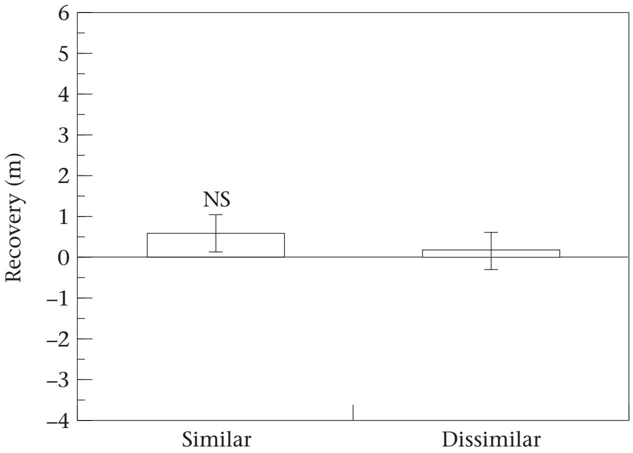

Figure 8. Mean (±SE) recovery for the two-speaker experiments in which one speaker switched to a song type similar to the base song type and the other speaker switched to a dissimilar song type. Scale on *Y* axis is the same as in Fig. 6.

DISCUSSION

Our experiments provide strong evidence that the degree of similarity perceived by male song sparrows between pairs of song types does not depend on the level of MUP similarity between the songs. This conclusion is supported first by the results of the single-speaker experiments, showing that the magnitude of recovery to a switch in song types does not correlate with the degree of MUP similarity (Figs 3, 4). We had some doubts, however, about the robustness of this result, because of the variability in response we noted across subjects and playback contexts. Accordingly, we tested the hypothesis again using our two-speaker design, which controls for these sources of variability by making within-subject and within-context comparisons. Using the two-speaker design, we were able to replicate our earlier result (Searcy et al. 1995) showing a difference in recovery for switches between song types compared with switches between variants using just 12 trials (Figs 5, 6). Searcy et al. (1995) required 96 trials to show this difference using the single-speaker design, indicating that the two-speaker design is indeed more powerful. Nevertheless, in the two-speaker experiment, we again found no evidence that the magnitude of recovery is predicted by the degree of MUP similarity between song types (Figs 7, 8).

That male song sparrows do not judge the degree of dissimilarity between song types based on MUP similarities does not rule out the possibility that they are responsive to some other aspect of similarity/dissimilarity. Falls et al. (1990) found that eastern meadowlarks, *Sturnella magna,* respond more strongly to switches between song types that humans judge to be dissimilar than to switches between song types judged to be similar. Similarly, Searcy et al. (1994) found that response recovery is greater in red-winged blackbirds for switches between song types dissimilar in length and trill rate than for switches between song types similar in these two features. For song sparrows, such features of ‘cadence’ as tempo, phrase structure and dominant frequency may be used to judge similarity (McArthur 1986). In addition, song sparrows may weigh particular MUP types more than others in judging song similarity, whereas our similarity measure weighs all MUPs equally. We should also point out that our results on song categorization in song sparrows come from experiments with male subjects only. Female birds in some instances have been shown to make more subtle discriminations in responding to song than do males (Searcy & Brenowitz 1988; Ratcliffe & Otter 1996). It is possible, then, that females attend to variation in MUP similarity among song types even though males do not.

Our results accord better with the categorical model for the perception of MUP similarities than with the continuous model (see Fig. 1). Consistent with the categorical model, MUP similarity between song types seems not to be important to perception. In addition we have some evidence that MUP similarity between variants is perceptually unimportant, in that recovery in single-speaker experiments was not negatively correlated with MUP similarity between variants (Fig. 4); this result also supports the categorical model. In terms of the graphical model, these two sets of results suggest that the lines relating perceived similarity to MUP similarity are flat in both the song variant and the song type regions of the graph (Fig. 1). We also have strong evidence that any two song types are perceived as being substantially more dissimilar than are two variants of the same song type, indicating that there is a step in the line between the two regions. We have not determined how abrupt this step is.

Our categorical model for song type perception is not completely analogous to categorical perception as originally proposed for human speech perception (Studdert-Kennedy et al. 1970) and later applied to perception of certain stimuli in nonhuman animals (Kuhl & Miller 1978; Nelson & Marler 1989; Ehret 1992; Wyttenbach et al. 1996). Categorical perception typically refers to the sorting into two categories of stimuli differing along a single physical dimension, as in distinguishing between ‘da’ and ‘ta’ based on the time lag between the plosive consonant and voicing. In our work, however, we have not been focusing on just two categories of stimuli but on a whole class of categories. We might have concentrated on just two song types, and asked how well song sparrows discriminate between those two and a series of synthesized intermediates, but this would have limited the generality of our results, and also would have been artificial, in that the variants of a song type produced by a song sparrow are usually not intermediates between that song type and any specific second type (Podos et al. 1992; Nowicki et al. 1994). Instead, we used many different song types as bases for comparison, and asked how dissimilar our subjects judged those base song types to be from stimuli that differed to varying degrees. Not only are our procedures different than in classical categorical perception, so too are our conclusions: specifically, the conclusion that our subjects treat all song types from the same male as equally distinct, regardless of MUP similarity, has no analogy in terms of traditional categorical perception categories.

Our results demonstrating lack of attention to MUP similarity between song types are important in showing that the positive correlation between repertoire size and mean MUP similarity found in song sparrow repertoires (Podos et al. 1992) does not imply an important constraint on repertoire complexity. Song types from large repertoires have higher MUP similarity than song types from small repertoires, but as song sparrows do not attend to this measure of similarity, large repertoires ought to be judged as more complex overall than small repertoires. Our results thus support the validity of song repertoire size as a measure of song complexity.

Acknowledgments

We thank Cindy Hogan, Preston Few, Matt Lovern, Jeff Podos and Melissa Hughes for help with the fieldwork. Steve Tonsor and the staff of the Pymatuning Laboratory of Ecology provided important logistical support. We thank the Pennsylvania Game Commission for allowing access to some of the sites used in the experiments. Financial support was provided by NSF through grants IBN-9408360 to S.N. and IBN-9523635 to W.A.S.

References

Andersson, M. 1994. *Sexual Selection*. Princeton, New Jersey: Princeton University Press.

Baker, M. C., Bjerke, T. K., Lampe, H. & Espmark, Y. 1986. Sexual response of female great tits to variation in size of males’ song repertoires. *American Naturalist*, **128,** 491–498.

Borror, D. J. 1965. Song variation in Maine song sparrows. *Wilson Bulletin*, **77,** 5–37.

Catchpole, C. K. 1980. Sexual selection and the evolution of complex songs among European warblers of the genus *Acrocephalus*. *Behaviour*, **74,** 149–166.

Catchpole, C. K. 1986. Song repertoires and reproductive success in the great reed warbler *Acrocephalus arundinaceus*. *Behavioral Ecology and Sociobiology*, **19**, 439–445.

Ehret, G. 1992. Categorical perception of mouse-pup ultrasounds in the temporal domain. *Animal Behaviour*, **43**, 409–416.

Falls, J. B., Dickinson, T. E. & Krebs, J. R. 1990. Contrast between successive songs affects the response of eastern meadowlarks to playback. *Animal Behaviour*, **39**, 717–728.

Hauser, M. D. 1997. *The Evolution of Communication*. Cambridge, Massachusetts: MIT Press.

Hiebert, S. M., Stoddard, P. K. & Arcese, P. 1989. Repertoire size, territory acquisition and reproductive success in the song sparrow. *Animal Behaviour*, **37**, 266–273.

Howard, R. D. 1974. The influence of sexual selection and interspecific competition on mockingbird song (*Mimus polyglottos*). *Evolution*, **28**, 429–438.

Krebs, J. R., Ashcroft, R. & Webber, M. 1978. Song repertoires and territory defence in the great tit. *Nature*, **271**, 539–542.

Kroodsma, D. E. 1977. Correlates of song organization among North American wrens. *American Naturalist*, **111**, 995–1008.

Kroodsma, D. E. 1982. Song repertoires: problems in their definition and use. In: *Acoustic Communication in Birds* (Ed. by D. E. Kroodsma & E. H. Miller), pp. 125–146. New York: Academic Press.

Kuhl, P. K. & Miller, J. D. 1978. Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, **63**, 905–917.

Lambrechts, M. & Dhondt, A. A. 1986. Male quality, reproduction, and survival in the great tit (*Parus major*). *Behavioral Ecology and Sociobiology*, **19**, 57–63.

McArthur, P. D. 1986. Similarity of playback songs to self song as a determinant of response strength in song sparrows (*Melospiza melodia*). *Animal Behaviour*, **34**, 199–207.

McGregor, P. K., Krebs, J. R. & Perrins, C. M. 1981. Song repertoires and lifetime reproductive success in the great tit (*Parus major*). *American Naturalist*, **118**, 149–159.

McGregor, P. K., Catchpole, C. K., Dabelsteen, T., Falls, J. B., Fusani, L., Gerhardt, H. C., Gilbert, F., Horn, A. G., Klump, G. M., Kroodsma, D. E., Lambrechts, M. M., McComb, K. E., Nelson, D. A., Pepperburg, I. M., Ratcliffe, L., Searcy, W. A. & Weary, D. M. 1992. Design of playback experiments: the Thornbridge Hall NATO ARW consensus. In: *Playback and Studies of Animal Communication* (Ed. by P. K. McGregor), pp. 1–9. New York: Plenum Press.

Mulligan, J. A. 1966. Singing behavior and its development in the song sparrow *Melospiza melodia*. *University of California Publications in Zoology*, **81**, 1–76.

Nelson, D. A. & Marler, P. 1989. Categorical perception of a natural stimulus continuum: birdsong. *Science*, **244**, 976–978.

Nice, M. M. 1943. Studies in the life history of the song sparrow. II. The behavior of the song sparrow and other passerines. *Transactions of the Linnean Society of New York*, **6**, 1–328.

Nowicki, S., Podos, J. & Valdés, F. 1994. Temporal patterning of within-type and between-type variation in song repertoires. *Behavioral Ecology and Sociobiology*, **34**, 329–335.

Podos, J., Peters, S., Rudnicky, T., Marler, P. & Nowicki, S. 1992. The organization of song repertoires in song sparrows: themes and variations. *Ethology*, **90**, 89–106.

Ratcliffe, L. & Otter, K. 1996. Sex differences in song recognition. In: *Ecology and Evolution of Acoustic Communication in Birds* (Ed. by D. E. Kroodsma & E. H. Miller), pp. 339–355. Ithaca, New York: Cornell University Press.

Read, A. F. & Weary, D. M. 1990. Sexual selection and the evolution of bird song: a test of the Hamilton-Zuk hypothesis. *Behavioral Ecology and Sociobiology*, **26**, 47–56.

Read, A. F. & Weary, D. M. 1992. The evolution of bird song: comparative analyses. *Philosophical Transactions of the Royal Society of London, Series B*, **338**, 165–187.

Searcy, W. A. 1984. Song repertoire size and female preferences in song sparrows. *Behavioral Ecology and Sociobiology*, **14**, 281–286.

Searcy, W. A. & Brenowitz, E. A. 1988. Sexual differences in species recognition of avian song. *Nature*, **332**, 152–154.

Searcy, W. A., Coffman, S. & Raikow, D. F. 1994. Habituation, recovery and the similarity of song types within repertoires in red-winged blackbirds (*Agelaius phoeniceus*) (Aves, Emberizidae). *Ethology*, **98**, 38–49.

Searcy, W. A., Podos, J., Peters, S. & Nowicki, S. 1995. Discrimination of song types and variants in song sparrows. *Animal Behaviour*, **49**, 1219–1226.

Searcy, W. A., Nowicki, S. & Hughes, M. 1997. The response of male and female song sparrows to geographic variation in song. *Condor*, **99**, 651–657.

Stoddard, P. K., Beecher, M. D. & Willis, M. S. 1988. Response of territorial male song sparrows to song types and variations. *Behavioral Ecology and Sociobiology*, **22**, 125–130.

Stoddard, P. K., Beecher, M. D., Loesche, P. & Campbell, S. E. 1992. Memory does not constrain individual recognition in a bird with song repertoires. *Behaviour*, **122**, 274–287.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S. & Cooper, F. S. 1970. Motor theory of speech perception: a reply to Lane's critical review. *Psychological Review*, **77**, 234–249.

Wyttenbach, R. A., May, M. L. & Hoy, R. R. 1996. Categorical perception of sound frequency by crickets. *Science*, **273**, 1542–1544.

Yasukawa, K. 1981. Song repertoires in the red-winged blackbird (*Agelaius phoeniceus*): a test of the Beau Geste hypothesis. *Animal Behaviour*, **29**, 114–125.

Yasukawa, K., Blank, J. L. & Patterson, C. B. 1980. Song repertoires and sexual selection in the red-winged blackbird. *Behavioral Ecology and Sociobiology*, **7**, 233–238.