A COMPARISON OF SONG SYLLABLE PERCEPTION BY FIVE SPECIES OF BIRDS

Michel L. Kreutzer
Robert J. Dooling
Susan D. Brown
Kazuo Okanoya
University of Maryland

ABSTRACT: Perception of song syllables taken from natural song sparrow and swamp sparrow songs was examined in five different species: song sparrows (Melospiza melodia), swamp sparrows (Melospiza georgiana), zebra finches (Taeniopygia guttata catanotis), canaries (Serinus canaria), and budgerigars (Melopsittacus undulatus). Using operant conditioning techniques, we trained these birds to discriminate among sparrow song syllables. By testing the birds on all possible pair-wise combinations of syllables, we generated "similarity" matrices for these song syllables from the birds' response latencies to detect changes in a repeating background of syllables. A number of different statistical techniques were used to examine species differences in perception including factor analysis of an inter-species correlation matrix, cluster analysis of the similarity matrices, and individual-differences multidimensional scaling (MDS) of the similarity matrices from each species. The spatial map of these stimuli produced by MDS revealed the perceived relations among these syllables for each species. Several acoustic measurements taken from these syllables were correlated with stimulus coordinates of the syllables in multidimensional space. These results demonstrate that natural vocal signals can be effectively used as stimuli in psychoacoustic experiments and the outcome of such experiments are likely to reveal robust species differences in perception.

Much is know about song learning in two congeneric species of sparrow, the song sparrow (Melospiza melodia) and the swamp sparrow (Melospiza georgiana) (Marler & Peters, 1981; Marler, 1987). It is also known that these two species exhibit innate preferences for the learning of species-specific songs (Marler & Peters, 1977, 1989). But the boundary conditions of these innate preferences are likely complex. Male swamp sparrows, for instance, can learn song sparrow notes from hybrid syllables in which song sparrow notes are inserted in swamp sparrow syllables and presented in swamp sparrow temporal patterns (Marler, 1987).

Recent studies have begun to focus on perceptual processing in these two species (Marler & Peters, 1989). A perceptual predisposi-
tion to learn conspecific song is most likely innate in swamp sparrows and perhaps also in song sparrows (Dooling & Searcy, 1980). Because they sing and defend their territories, male song birds have more often been the subject of studies on theresponsiveness to natural, synthetic, and modified vocalizations. But using copulation solicitation displays, it has recently been shown that female song sparrows and female swamp sparrows were also sensitive to both species-specific phonology and temporal patterns (Searcy, Marler, & Peters, 1981).

We recently applied operant conditioning procedures to study basic auditory perception (Okanoya & Dooling, 1988a) and the perception of song syllables by both species of sparrow (Okanoya & Dooling, 1990). These two species of sparrow are slightly different in terms of absolute auditory sensitivity which parallels species differences in the long term average power spectra of the song (Okanoya & Dooling, 1988b). Furthermore, females of both species perceived the differences between normal and modified swamp sparrow and song sparrow syllables with swamp sparrows tending to be somewhat more discriminating than song sparrows (Okanoya & Dooling, 1990).

In the present experiments, we tested other species of small birds on these same syllables. By comparing the results of sparrows tested on sparrow syllables with other species tested on the same syllables, it is possible to arrive at an indication of whether the perception of song syllables by sparrows is in any way “special.” If sparrows and other species of birds all perceive sparrow song syllables in a similar way, then we can conclude that the perceptual mechanisms used by sparrows in perceiving song syllables are not special but are probably due instead to general auditory-system processing mechanisms common to all birds.

METHOD

Animals

For this experiment we had available to us only female sparrows: two song sparrows and two swamp sparrows. The remaining birds were four budgerigars (Melopsittacus undulatus) (1 female and 3 males), two German Roller canaries (Serinus canaria) (1 male, 1 female), and four zebra finches (Taeniopygia guttata (catanotis) (2 females, 2 males). The four sparrows reported here were the same birds tested previously (Okanoya & Dooling, 1990). These birds were hand-reared at the Rockefeller University Field Research Center in New York and were excluded, on the basis of their sex, from a study on song learning. The budgerigars, canaries, and finches were obtained from local commercial pet dealers or bred in the laboratory. During
the behavioral testing, the birds were housed in an aviary at the University of Maryland and maintained under a day/night cycle correlated with the season.

Apparatus

The apparatus used for training and testing the birds has been described previously (Okanoya & Dooling, 1988a, 1990). The birds were tested in a small wire cage (15 × 15 × 15 cm) placed in a sound attenuation chamber (IAC-1: 39 × 58 × 35 cm). A standard pigeon grain hopper was mounted on one side of the cage. A response panel consisting of two sensitive microswitches with attached light emitting diodes (LEDs) was mounted just above the hopper opening. The birds could close the microswitch by striking the LED with their beaks.

All experimental events including stimulus presentation, response measurement, and reinforcement contingencies were controlled by an IBM-AT microcomputer. The song syllables were stored on the hard disk of the computer in digital form, output at a sampling rate of 20 kHz through a digital-to-analog converter (Data-transla- tion DT2801A) and low-pass filtered at 10 kHz for anti-aliasing. A loud speaker mounted inside the sound proof chamber transduced the stimuli. Song syllables were presented at a peak sound pressure level of 72 dB(A) SPL at the bird’s head as measured by a General Radio 1982 sound level meter.

Stimuli

The song syllables used in this experiment were taken from original field recordings of full length songs of both species made by Dr. Donald Kroodsma while he was at the Rockefeller University. Songs of swamp sparrows typically consist of a repetition of a single syllable whereas the songs of song sparrows are more complex, consisting of 3 to 6 different syllables arranged in a characteristic syntactical pattern (Marler & Peters, 1989). Since swamp sparrow songs consist of simple repetition of a single syllable, we selected syllables from song sparrow songs which occurred in the repetitive sections of the song.

In all, ten song syllables of roughly equal length (100-160 ms) were selected from both song and swamp sparrow songs. These syllables are shown in Figure 1 and are the same set used to test the sparrows in an earlier paper (Okanoya & Dooling, 1990). We did not conduct an exhaustive acoustic analyses of these syllables. But, a total of 32 acoustic measures were taken from these syllables for the purpose of finding a correlation between the acoustic characteristics of these complex stimuli and species differences in perception. In addition to duration, bandwidth, lowest frequency, highest frequency,
and peak frequency, we also measured a number of additional spectral measures. We computed the long term average power spectrum over the entire duration of each syllable using ILS-PC. Then, we examined the distribution of energy in these syllables in selected bands from 1.0 kHz to 10.0 kHz and expressed the result as a percent relative to the total energy in the syllable (e.g., percentage of energy in the 1-2 kHz band).

Training and Testing Procedure

The training and testing procedures used in these experiments have also been described in detail elsewhere (Okanoya & Dooling, 1988a, 1988c, 1990) and thus are only briefly summarized here. Birds were food deprived overnight and for several hours in the morning. These birds were trained to peck one key (observation key) repeatedly during the repetitive presentation of one syllable (background) and to

FIGURE 1. Sonograms of the set of natural song sparrow and swamp sparrow syllables used in this experiment.
peck the other key (report key) when another syllable (the target) was presented instead of the background. Syllables were presented at a rate of 5 per s to mimic the temporal pattern of swamp sparrow songs. A peck on the report key during a change in the repeating syllable was reinforced by access to food. Once the birds learned this response, they were tested on the set of syllables consisting of five song sparrow syllables and five swamp sparrow syllables (Figure 1).

A habituation phase preceded each testing phase. During this phase neither keylight was illuminated and the background sound was repeated until the bird either pecked the observation key four times or refrained from responding on the report key for 10 s. Once either of these conditions was met, the LEDs on both observation and report keys were illuminated and the testing phase began. This procedure was highly effective in decreasing spontaneous responding.

During the test phase, a peck on the observation key initiated a random waiting interval of 1-7 s. Following this interval, a peck on the observation key resulted in the replacement of the background repeating syllable with the target syllable. A response to the report key within 3 s from the beginning of the repeating target stimuli was reinforced with a 2 s (budgerigars) to 4 s (other species) access to food. About 20% of the trials were sham trials in which the target stimulus was the same as background stimulus. A response on the report key during a sham trial or during a waiting interval was punished with a 16 s timeout period during which lights in the test chamber were extinguished but the repeating sound continued.

The testing phase continued until the background stimulus was paired with every other stimulus in the set three times. Testing proceeded a row at a time (i.e., same background stimulus) until all possible combinations within the row were exhausted. Following a 10 s silent interval, another background stimulus was then randomly selected and a new habituation phase begun. This was continued until all stimuli had served as a background and as a target. This procedure produced a total of three latency matrices for analysis.

A median reaction time matrix was constructed out of the three latency matrices. The response latency required to detect changes in the repeating background syllable was taken as an index of perceptual similarity between the background and the target syllables (Okanoya & Dooling, 1988b). The upper and lower triangular halves of the matrix were averaged and log-transformed and the diagonal discarded to produce a single half-matrix of response latencies for each bird. For some of the following analyses, the half-matrices for the individuals of each species were then averaged to produce a single half-matrix for each species. An example of such a log-transformed half-matrix is given in Table 1.
|       | SG1 | SG2 | SG3 | SG4 | SG5 | SW1 | SW2 | SW3 | SW4 | SW5 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|       | X   | 312 | 302 | 306 | 304 | 302 | 311 | 306 | 303 | 303 |
|       |     | X   | 322 | 300 | 307 | 297 | 307 | 301 | 296 | 300 |
|       |     |     |     | 306 | 297 | 299 | 304 | 295 | 301 | 292 |
|       |     |     |     |     | 300 | 301 | 298 | 295 | 302 | 302 |
|       |     |     |     |     |     | 301 | 299 | 295 | 298 | 301 |
|       |     |     |     |     |     |     | 304 | 308 | 310 | 312 |
|       |     |     |     |     |     |     |     | 311 | 314 | 315 |

**Statistical Procedures**

Four procedures were used to assess the differences among species in the perception of sparrow syllables. First, the log-transformed half-matrices for each bird were used to construct a matrix of interindividual correlations. Principal components (Varimax rotation of all components) were computed for this matrix using SYSTAT. The purpose of this analysis was simply to provide a visual representation of the similarity in response patterns (i.e., between half-matrices) among individuals and among species. The spatial arrangement of the individual birds on the two highest rotated loadings in two-component space provides one way of characterizing the differences in response patterns among individuals.

Another way of characterizing differences in response latency matrices is through cluster analyses. To this end, a single half-matrix for each species was obtained by averaging the half-matrices of individual members of the species. Each average half-matrix was then submitted to a hierarchical, complete linkage, cluster analysis (Alldenderfer & Blashfield, 1984). The results of these cluster analyses indicate the extent to which each species perceived the differences between swamp and song sparrow syllables and the similarity among sparrow syllables from the same species.

In a third procedure, the half-matrices for each species were analyzed by a multidimensional scaling analysis (MDS) program, SIND-
SCAL (Shepard, 1980). MDS arranges stimuli in multidimensional space so that stimulus similarity (as determined by response latencies) is represented by spatial proximity. Species weights provide an indication of how well the arrangement of syllables in multidimensional space accounts for the data from each species.

Finally, a fourth procedure was used in conjunction with the multidimensional scaling analysis described above. In all, we measured 32 spectral, temporal, and intensive acoustic measures from these sparrow syllables. These measures were then correlated with the stimulus coordinates from MDS. To the extent that these measures were significantly correlated with syllable location in multidimensional space, they provide a potential explanation of the stimulus map in acoustic terms.

RESULTS

The two-component solution from the intercorrelation of the 14 matrices is shown in Figure 2. An interesting pattern emerges from

![Figure 2](image)

**FIGURE 2.** Two-dimensional Principle Component Analysis of the interindividual correlation matrix. All four sparrows as well as C1, B1, Z1, Z2 are females. Birds B2, B3, C2, Z3 and Z4 are males.
this analysis. First, while there are clear species differences among canaries, song sparrows, and swamp sparrows, there are relatively small individual differences within each of these species—the differences among individuals of the same species are quite small compared to the differences among individual budgerigars and zebra finches.

Dendrograms from the cluster analyses of the data from all species combined showed that these 10 syllables were clustered into two groups corresponding fairly well to the species of sparrow. One of the song sparrow syllables is in among the cluster of swamp sparrow syllables. These results are shown in Figure 3. Separate cluster analyses were also performed on the data from each species. These analyses revealed that only swamp sparrows clearly showed two major clusters of syllables corresponding to the species of sparrow which produced them. Song sparrows intermixed only one syllable of each type. Dendrograms from all the other species showed somewhat more of an intermixing of swamp and song sparrow syllables.

The two-dimensional spatial representation obtained by a SIND-SCAL analysis of the response latencies matrices of each species is shown in Figure 4a. The variance accounted for by this two-dimensional solution was 55% with the first and second dimensions accounting for 33% and 22%, respectively. The species weights (i.e., one weight for each species) for this solution are shown in Figure 4b.

In two-dimensional stimulus space, the 10 sparrow syllables are separated roughly into two groups by species along the first dimension. This arrangement follows closely the grouping that was evident from a cluster analysis on the data from all five species (Figure 3).
FIGURE 4 (a). Two-dimensional spatial representation of the eight sparrow syllables by the five species. Swamp sparrow syllables are on the left and song sparrow syllables are on the right. (b) Species weights corresponding to this stimulus space for the five species—budgerigars(B), canaries(C), zebra finches(Z), swamp sparrows(SW), and song sparrows(SG).
The relation among the species weights for this solution reflects the relative amount of variance in each species' data that is accounted for by the distances between stimuli in the spatial map. The species weights for song and swamp sparrows are very similar. Furthermore, the first dimension of the stimulus space accounts for more variance in the two sparrows' data than for the other three species. It is this dimension that separates the 10 syllable types by species. The second dimension accounts for more variance in the data of budgerigars and zebra finches than for the other three species. The characteristics of the stimuli represented by this dimension are less obvious from visual inspection alone.

**Correlation between Syllable Coordinates in Perceptual Space and Acoustic Measures**

In an effort to explain the stimulus space in acoustic terms, we measured a total of 32 acoustic characteristics of these syllables. This was not intended to be an exhaustive analysis. Rather, we selected measures which made sense from examination of sonographs of these syllables and which other investigators have used to characterize the differences between swamp and song sparrow songs (Marler & Pickert, 1984). These characteristics ranged from note durations, duration and direction of frequency sweeps, inter-note durations, and syllable durations to a variety of spectral measures including total energy in selected bandwidths. Many of these measures are also correlated to some degree with one another (e.g., syllable duration and internote-intervals).

We correlated these acoustic characteristics with location of stimuli in multidimensional space. Of these 32 measures, only five were correlated with stimulus coordinates on either the first or second dimension and all of these were spectral measures. Four measures were significantly correlated with stimulus coordinates along the first dimension. One is the total amount of energy in the region of 2-4 kHz and the other three are related to the bandwidth of these syllables (i.e., bandwidth, the lowest frequency, and the highest frequency). Only one measure, the amount of energy in the region of 2-3 kHz, was significantly correlated with stimulus coordinates along the second dimension. Some temporal measures such as syllable duration, internote-intervals, and percent of constant frequency portions of the syllables (the inverse of percent of sweep portions) are notable for their lack of correlation with the stimulus coordinates in perceptual space. These results are given in Table 2.

**DISCUSSION**

Most of what we know about the perceptual basis of selective vocal learning in sparrows comes from song learning experiments of
TABLE 2
Correlation Coefficients of Acoustic Properties with Stimulus Coordinates

<table>
<thead>
<tr>
<th>Property</th>
<th>Dimension I</th>
<th>Dimension II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>0.670*</td>
<td>-0.183</td>
</tr>
<tr>
<td>High Frequency</td>
<td>0.884**</td>
<td>0.144</td>
</tr>
<tr>
<td>Low Frequency</td>
<td>0.801**</td>
<td>0.495</td>
</tr>
<tr>
<td>Energy 2-4kHz</td>
<td>-0.822**</td>
<td>0.132</td>
</tr>
<tr>
<td>Energy 2-3kHz</td>
<td>-0.429</td>
<td>-0.810**</td>
</tr>
<tr>
<td>Syllable Duration</td>
<td>-0.106</td>
<td>0.035</td>
</tr>
<tr>
<td>Internote-interval</td>
<td>0.221</td>
<td>0.235</td>
</tr>
<tr>
<td>Percent CF</td>
<td>0.256</td>
<td>0.561</td>
</tr>
</tbody>
</table>

*p<.05  
**p<.01

Marler and his colleagues (Marler & Peters, 1989). Field playback studies with male sparrows (Peters, Searcy, & Marler, 1980), copulation/solicitation studies with female sparrows (Searcy et al., 1981), and heart rate recordings (Dooling & Searcy, 1980) have also been quite useful in refining the understanding of perceptual selectivity for conspecific song in these sparrows. Still, none of these measures have addressed the issue of whether swamp and song sparrows perceive their song syllables differently than any other avian species might perceive them. In other words, the issue of whether sparrows might be specialized for the perception of conspecific song syllables has not been explored. The present experiments describe an approach that might be useful in answering this question and, by comparing perception of song syllables in swamp and song sparrows and three other species of small birds, provides some preliminary evidence.

The results of these experiments show that sparrows emerge as different from these other species in the perception of sparrow syllables. First, the results of a Principle Components analysis on the intercorrelation matrix of all 14 birds revealed that both species of sparrows are similar to one another and to canaries in the consistency with which individual birds perceived the differences among the 10 sparrow syllables. This is in contrast to the results of both zebra finches and budgerigars who showed large interindividual differences. Perhaps this difference is related to the fact that complex, stereotyped, tonal syllables are a common feature of canary (Güttinger, 1985), swamp sparrow, and song sparrow song. Simple tonal syllables are a less common feature of the long rambling budgerigar song and do not occur at all in zebra finch song (Farabaugh, Brown & Dooling, submitted; Immelmann, 1969).

Reasoning along these same lines, the syllables in this experi-
ment were presented in a fixed temporal pattern approximating that found in normal swamp sparrow song. This patterning also bears some resemblance to normal canary song and to subunits of normal song sparrow song. Budgerigars and zebra finches, on the other hand, have a considerably different arrangement in the temporal pattern of their songs.

Thus, temporal patterning in the presentation of these syllables could have a differential effect on the perception of these syllables by the five species. However, the suggestion that sparrows and canaries may have more in common in the perception of these syllables than do the other species must be tempered by the small sample size. A proper test of this hypothesis would require testing many more sparrows and extending these tests to both males and females of each species.

The results of cluster analyses on the data for each bird revealed that only swamp sparrows showed two complete clusters of syllables corresponding perfectly to species type. These results support a variety of earlier studies showing that swamp sparrows are more selective than song sparrows in the perception of conspecific syllables (Dooling & Searcy, 1980; Marler, 1987; Okanoya & Dooling, 1990). The four other species tested in this experiment all showed a slight intermixing of swamp and song sparrow syllables by a cluster analysis.

The results of multidimensional scaling and property fitting are interesting for what they reveal about the relative sensitivities of the five species. The species weights of both sparrow species are highest on the first dimension—the dimension that separates song sparrow syllables from swamp sparrow syllables. We take this as evidence that, compared to the other three species, swamp and song sparrows are more sensitive to the acoustic differences between swamp and song sparrow syllables.

An obvious feature of these syllables that would reliably discriminate between the two classes of syllables is the broad spectrum, rapid frequency glide (Type VI Note) present in all swamp sparrow syllables (Marler & Pickert, 1984). Not surprisingly, this feature also contributes heavily to the quantitative acoustic measures of bandwidth, low frequency limit, and high frequency limit in swamp sparrow syllables. It is perhaps not surprising then that these three measures are significantly correlated with stimulus coordinates along the first dimension in multidimensional space. We conclude that this feature probably discriminates most reliably between the syllables of the two species, and that both swamp sparrows and song sparrows are most sensitive to this feature.

The preponderance of energy in these syllables falls in the spectral region of 2-4 kHz. This is generally more true for song sparrow
syllables than swamp sparrow syllables. Again, it is not too surprising that this spectral measure is also significantly correlated with stimulus coordinates on the first dimension.

On the second dimension, the species weight for budgerigars is the highest followed by zebra finches, and then the two sparrow species and the canaries. From other psychoacoustic tests, we know that budgerigars are unusually sensitive to spectral changes in the frequency region around 3 kHz. The amount of energy in the spectral region of 2-3 kHz is correlated with syllable location on the second dimension in stimulus space. It is more difficult to account for differences among the other species on the second dimension because all four species show a similar pattern of spectral resolving power over the frequency range of about 1-5 kHz (Okanoya & Dooling, 1987).

Comparing the absolute auditory sensitivity of these species at high frequencies reveals another potential mechanism for the species differences in perception we observed. Swamp sparrows and song sparrows are 10-15 dB more sensitive to frequencies above 4 kHz than are canaries and zebra finches and 15-20 dB more sensitive than budgerigars (Okanoya & Dooling, 1987). The greater sensitivity of sparrows to high frequencies could contribute to the fact that sparrows, compared to the other species tested, are most sensitive to the differences between the two classes of sparrow syllables. This argument also supports the notion that the rapid, broad, frequency swept Type VI Note provides an important cue for distinguishing among the two classes of syllables.

The final explanation for the species differences observed in this experiment—including firm evidence for differential sensitivity to different song syllable features—will have to come from experiments with synthetic stimuli. The present experiment shows there are species differences in the perception of sparrow syllables and suggests a possible basis for these differences related to differences in basic auditory processes. But, it is likely that other perceptual and attentional factors residing at more central levels are also responsible. A similar conclusion was reached from a recent study comparing budgerigars, canaries, and zebra finches on the perception of conspecific and heterospecific distance calls (Dooling, Brown, Klump, & Okanoya, in press).

Whatever the final mechanism(s) turn out to be, one great advantage of the present procedures in pursuing the issue of species-specificity in song perception is that they can be used with equal facility not only on different species but also on males and females of the same species. Previous attempts to compare male and female perception of songs and song elements have had to rely on dramatically different methodologies for the two sexes. We feel this represents a considerable advance in the study of the perception of songs and song
syllables in birds. Comparing males and females of the same species should prove extremely interesting in species in which only the male sings and for which there is evidence of a motor basis for song perception.

ACKNOWLEDGEMENTS

This research was supported by NIH grant DCD-000198 to RJD and a NATO grant to MLK. Dr. Kreutzer is now at the CNRS-URA 667, University of Paris XIII, France.

REFERENCES


