

# Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences

Scott A. MacDougall-Shackleton,<sup>a)</sup> Stewart H. Hulse,<sup>b)</sup> Timothy Q. Gentner, and Wesley White<sup>c)</sup>

Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218

(Received 23 July 1997; revised 14 February 1998; accepted 2 March 1998)

Like humans, animals that use acoustic stimuli to perceive their world ought to be able to parse the auditory scene into functionally significant sounds. The ability to do so ought to have significant adaptive value when, for example, an animal can identify the sounds of a predator among other natural noises. In earlier work it was shown that a species of songbird, the European starling, can identify excerpts of both its own song and songs from other avian species when the songs are mixed concurrently with other natural signals. In this experiment it is demonstrated that starlings can segregate two synthetic pure-tone sequences when the sequences differ in frequency. Taken together, the experiments show that at least one nonhuman species is capable of auditory scene analysis both for natural and for non-natural acoustic stimuli. This suggests in turn that auditory scene analysis may be a general perceptual process that occurs in many species that make use of acoustic information. © 1998 Acoustical Society of America. [S0001-4966(98)03606-6]

PACS numbers: 43.66.Gf, 43.66.Lj [RHD]

## INTRODUCTION

People readily parse concurrent overlapping sounds into separate functional auditory objects identified by a unique set of acoustic cues such as spatial location, spectral composition, and pitch (Bregman, 1990). Because one can attend to just one auditory object at a time, selective attention is important for auditory scene analysis, as in the familiar example of the “cocktail party effect” (Bregman, 1990; Cherry and Taylor, 1954). The processes involved in auditory scene analysis have been studied over the past 25 years or so, and the topic remains a very active area of research in human audition (Bregman, 1990; Bregman and Campbell, 1971; Darwin and Carlyon, 1995; Darwin *et al.*, 1995; Dowling *et al.*, 1987; Hartmann and Johnson, 1991; Newman and Jusczyk, 1996; Wood and Cowan, 1995). Here we report prototypical auditory stream segregation by a nonhuman animal, a songbird. European starlings (*Sturnus vulgaris*) were found, like humans, to segregate and organize perceptually a serial tone pattern into two subpatterns on the basis of differences in tone frequency.

There are few reports of auditory scene analysis by nonhuman animals. Yet the capacity to parse the auditory scene into significant objects—a predator’s rustle amidst other forest sounds, for example—would surely convey significant evolutionary advantage. To our knowledge only two previous reports have appeared. First, Hulse *et al.* (1997) found that European starlings could be trained to discriminate and identify a sample of one species’ bird song presented concurrently with a sample of another species’ bird song. Such species identification also held when the test songs were pre-

sented amidst the added distractors provided by a recording of the dawn chorus, the cacophonous mixture of sounds of many songbirds singing early on a spring morning (e.g., Henwood and Fabrick, 1979; Kacelnik and Krebs, 1983; Sticker *et al.*, 1996). Second, Wisniewski and Hulse (1997) discovered that European starlings could not only learn to discriminate between many samples of the songs of two individual starlings, starling A and starling B, but most important, could maintain that discrimination when song samples from a third individual, starling C, were added to both A and B stimuli, and when additional songs from starlings D, E, and F were added as further background distractors. To be sure, discrimination accuracy by the test birds decreased as more and more distractors were added to the background mixture, but individual identification was still well above chance even with the target A and B songs mixed concurrently with song samples from four other starlings. The basis on which the starlings were performing the scene analysis remains to be identified, but the process appears to be robust when European starlings listen to natural sounds like bird song.

Although it seemed sensible to begin a study of auditory scene analysis in nonhuman animals by using stimuli with ecological significance, the available data do not establish the process as a general auditory capacity even within the single species studied to date. The question is whether or not the process is somehow unique to natural signals like bird song, or if scene analysis holds more generally for other acoustic events, such as non-natural arbitrary stimuli. The present experiment was directed to that latter issue.

## I. THEORY

The theoretical strategy we chose to study the problem was based on a variation of the experimental procedure used

<sup>a)</sup>Now at Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ. Electronic mail: Shackles@princeton.edu

<sup>b)</sup>Electronic mail: hulse@jhu.edu

<sup>c)</sup>Now at Department of Psychiatry, Johns Hopkins University Medical School, Baltimore, MD.

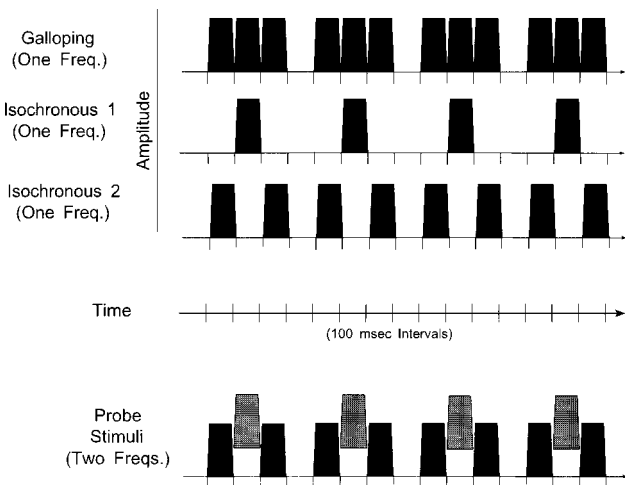


FIG. 1. Schematic diagram of the experimental plan. Each filled quadrangle at the top of the figure represents a time-amplitude plot of a single-frequency sine tone (e.g., 1000 Hz). The sine tones are arranged to form three stimulus sequences with which the starlings were trained initially, a galloping stimulus and two isochronous stimuli, isochronous 1 and isochronous 2. The sequence below the time line represents a Probe stimulus. With the probe stimuli, there are two frequencies in the sequence, one (say, 1000 Hz) represented by the filled quadrangle, and one (say, 1710 Hz) represented by the stippled quadrangle. Note that although the overall temporal structure of the probe stimulus has a galloping pattern, the use of two frequencies turns the probe stimuli into two intermixed isochronous stimuli—on the assumption that stream segregation takes place.

originally by Van Noorden to study stream segregation in humans (Van Noorden, 1975). The basics of the strategy are outlined in Fig. 1.

Subjects, European starlings in this case, are first trained to make one response (peck, say, the left key for food in an operant task) in response to a galloping stimulus. The galloping stimulus (displayed at the top of Fig. 1) is so termed because the pause introduced after every third tone adds an accent that makes the stimulus sequence seem to “gallop” like the hoof beats of a race horse to human listeners—especially at relatively fast tempos. At the same time, the starlings are trained to make another response (peck, say, the right key) when they hear *either* the isochronous 1 or isochronous 2 stimuli. During initial discrimination training, all stimuli are presented at a single frequency; therefore the birds learn the discrimination solely on the basis of the temporal structure of the stimuli.

Once this initial discrimination has been learned, occasional probe stimuli (displayed at the bottom of Fig. 1) are then introduced during ongoing trials with the training stimuli. Unlike the training stimuli, the probe stimuli contain *two* frequencies arranged as shown in the figure, but the temporal structure of the probe stimuli is just like that of the galloping stimuli. Note that the frequency changes in the probe stimuli provide the basis on which auditory stream segregation might take place because the frequency changes restructure the probe stimuli, potentially, into one or the other of the two isochronous stimuli. That is, if stream segregation does occur, and if selective attention holds for starlings as it does for humans in auditory scene analysis (a tenable proposition that nevertheless remains to be tested directly in future research with starlings), the birds ought to

hear the probe stimuli as sometimes the isochronous 1 stimulus and sometimes the isochronous 2 stimulus. They should therefore peck the key associated with the isochronous stimuli. Presumably, furthermore, the tendency to respond as if the probe stimuli are isochronous should increase as the size of the frequency change in the probe stimuli increases. That is so because the tendency for stream segregation to occur is a function (at least for humans) of the magnitude of the frequency difference between the two streams (Van Noorden, 1975). If, however, streaming fails to occur, the birds ought to peck the key associated with the galloping stimulus because the probe stimuli have a galloping temporal structure.

We now turn to the experiment that tested these theoretical predictions.

## II. METHOD

### A. Subjects

The subjects were ten adult male European starlings (*Sturnus vulgaris*) of unknown age caught in the wild near Baltimore, MD. The starlings were kept in a mixed aviary containing approximately 50 starlings of both sexes, zebra finches (*Taeniopygia guttata*), and ring doves (*Streptopelia risoria*). The starlings were fed commercial poultry starter (Start and Grow, Purina, Inc., St. Louis, MO) supplemented periodically with raw spinach. None of the birds had prior experience with operant conditioning procedures. During the experiment, the birds' access to food was controlled to ensure that body weight was approximately 85% of *ad libitum* weight before each experimental session.

Three starlings failed to complete the experiment: one died and two failed to reach the criterion for initial baseline discrimination learning. Thus seven birds completed the experiment and are included in the analysis.

### B. Apparatus

During experimental sessions, each starling was tested inside an IAC, Inc. (Bronx, NY) model AC-3 sound attenuation chamber (80 cm×60 cm×60 cm interior dimensions) containing a speaker, response panel, and two feeders. The response panel formed one wall of a 28 cm×20 cm×30 cm stainless steel weld-wire cage that housed the bird. Two 2.8-W light bulbs behind a translucent screen on the rear wall indirectly illuminated the chamber.

The response panel was a 26 cm×30 cm sheet of 2-mm-thick stainless steel equipped with three response keys and two 4.5 cm×6 cm openings that provided controlled access to food hoppers. The keys were 2-cm-diam plastic disks attached to microswitches; keys were spaced 6 cm apart in a horizontal row. Each food hopper opening was 4.5 cm below the left- and rightmost keys. A Bose 101 loudspeaker was mounted above and behind the response panel. A PC microcomputer generated sound stimuli from its hard disk drive through a Data Translation model 2801 D/A converter. The stimuli were then fed to a Crown model D75 amplifier and then to the speaker in the test chamber. The computer also monitored the response keys, house lights, and food hoppers, and collected data.

Stimuli for <u>Baseline Discrimination</u> Between Isochronous and Galloping Tone Patterns		
Isochronous Patterns	Galloping Pattern	Stimulus Frequency
(1) X - X - X - X - (2) X - - - X - - - X - - -	X X X - X X X -	1000 Hz
(1) A - A - A - A - (2) A - - - A - - - A - - -	A A A - A A A -	1050 Hz
(1) B - B - B - B - (2) B - - - B - - - B - - -	B B B - B B B -	1710 Hz
(1) C - C - C - C - (2) C - - - C - - - C - - -	C C C - C C C -	4538 Hz

Stimuli for <u>Test Probes</u>	
Pattern	Frequency Difference
A X A - A X A - A X A -	1050 Hz and 1000 Hz = 50 Hz
B X B - B X B - B X B -	1710 Hz and 1000 Hz = 710 Hz
C X C - C X C - C X C -	4358 Hz and 1000 Hz = 3538 Hz

FIG. 2. Details of the stimulus patterns used in the experiment. The birds were trained initially to discriminate between one galloping and two isochronous baseline patterns. On any trial, the tones in a given stimulus pattern were of constant frequency, but over the course of training, the birds were exposed to patterns containing four frequencies: 1000, 1050, 1710, and 4538 Hz. The stimuli for the test probes, shown at the bottom of the figure, were combinations of the 1000-Hz tone with the other three frequencies. All probe stimuli had a galloping temporal structure.

### C. Stimuli

The stimuli were created from repeated sine tones with frequencies of 1000, 1050, 1710, or 4538 Hz. Each tone was 100 ms in duration and began and ended with a 10-ms linear amplitude ramp. Successive tones either followed one another immediately or were separated by intertone intervals that varied according to experimental conditions (Fig. 2). Tones and intertone intervals repeated to form a stimulus that was 10 s in total duration. Each 10-s stimulus began and ended with a 2-s linear amplitude ramp.

The tones and intertone intervals were combined to form one stimulus type that had a galloping rhythm and two other stimulus types that had an isochronous rhythm (Figs. 1 and 2). The galloping stimulus consisted of repeated sequences of a group of three consecutive 100-ms tones separated by a 100-ms intertone interval. One isochronous stimulus (isochronous 1) consisted of 100-ms tones separated by 300-ms intertone intervals. The other isochronous stimulus (isochronous 2) consisted of alternating 100-ms tones and 100-ms intertone intervals. Both the galloping stimulus and the isochronous stimuli for the initial *baseline discrimination* task contained tones of a single frequency (Fig. 2). For *probe stimuli* used for later testing, each stimulus contained 1000-Hz tones and tones of either 1050, 1710, or 4538 Hz (Figs. 1 and 2). Thus, the probe stimuli consisted of sequences containing small (5%), intermediate (71%), or large

(354%) frequency changes within the sequence.

The stimuli were created with SoftSynth software (Peabody Conservatory, Baltimore, MD) on a Macintosh IIx computer, digitized with 16-bit accuracy at a 20 kHz sampling rate, and stored in PC format on computer disk for on-line retrieval. The stimuli were played through the speaker at a level of 69 dB SPL as measured by a Rion sound level meter placed at a location approximately at the level of a bird's head in the test chamber.

### D. Procedure

#### 1. Shaping

Birds were trained initially to peck the keys in the apparatus by attaching a small amount of food directly to the center key. In the first 2-h shaping session, pecks to the center key were rewarded with access to one of the food hoppers for 2 s. In the next session on the following day, pecks to the center key followed by pecks to the left or right key were rewarded with 2-s food access at the left or right hopper, respectively. In the next session, baseline discrimination began.

#### 2. Baseline discrimination

In the baseline discrimination task, the starlings were tested in the apparatus for 2-h sessions, 6 days per week. Birds initiated a trial by pecking the center key. This resulted in playback of a 10-s tone stimulus with either the galloping ( $p=0.50$ ) or the isochronous ( $p=0.50$ ) stimulus structure. Within the isochronous stimuli, the isochronous 1 and 2 stimuli appeared with equal likelihood. Pecks to any of the keys during the 10 s while the stimulus played had no consequence. The first peck following the 10-s observation period led either to 2-s access to food or to a 12-s time-out period during which the house lights were turned off. Left or right pecks, counterbalanced across birds, were rewarded with food following a galloping stimulus or either of the two isochronous sound stimulus, respectively. Incorrect pecks led to the time-out period, and the trial then repeated until the bird performed correctly. If the bird did not peck any key for 10-s following the listening period, the trial ended and the stimulus was repeated on the next trial. There was a 250-ms interval between trials.

The birds were tested in two groups. In the first group ( $n=2$ ), baseline stimuli constructed with all four frequencies were used ( $p=0.25$ ) beginning with the first session of baseline training. In the second group ( $n=5$ ), the four frequencies were introduced one by one in an order partially counterbalanced across birds. When a bird demonstrated statistically significant acquisition of the task with one frequency, stimuli at another frequency were added to the pool of baseline stimuli. This was repeated until birds were discriminating baseline stimuli at all four frequencies.

For the second group of birds, once a bird reached an acquisition criterion of 80% correct or better for three consecutive sessions, the reward probability for a correct response was reduced over four or five sessions so that only 80% of correct responses led to food reward. This was to prepare the birds for nonreinforced probe trials in later probe

sessions. Once a bird performed at the 80% correct criterion for three consecutive sessions at this reward probability, probe sessions began. For the first group of birds, probe sessions began without the reduction in reinforcement levels for the baseline stimuli.

### 3. Probe sessions

During probe sessions, some baseline discrimination trials were replaced with probe trials using the three galloping probe stimuli that varied in the size of the frequency differences between successive tones (Figs. 1 and 2). On probe trials, the three galloping probe stimuli were sampled randomly ( $p=0.33$ ). In the first group of birds, 20% of the trials in a probe session were galloping probe trials, and the remainder were baseline trials. One hundred per cent of correct responses to baseline stimuli were rewarded. Pecks to either the left or right key in response to a probe stimulus were rewarded. In the second group of birds, 10% of the trials in a probe session were galloping probe trials, and 90% of correct responses to baseline stimuli were rewarded. Pecks to either the left or right key in response to galloping probe stimuli led neither to reward nor punishment. Probe sessions were continued in both groups until each bird had responded to at least 100 of each type of probe stimulus in the first group of two birds, and until each bird had responded to at least 20 (range 20–87) of each probe type in the second group of five birds.

### E. Data analysis

The probability that a bird responded to the galloping and isochronous stimuli as if they were isochronous was determined for each daily session during initial baseline training. The data averaged over the last three sessions of baseline training were used to assess discrimination at the end of training, prior to the introduction of probe stimuli. Accurate discrimination was indicated by a high probability of a response on the key associated with isochronous stimuli when an isochronous stimulus occurred, and a corresponding low probability of a response on the isochronous key when a galloping stimulus occurred. The birds never failed to respond, and correction trials were excluded from the analysis.

For sessions that included probe trials, the mean probability that a bird responded to a probe stimulus as if it were an isochronous stimulus was calculated over all probe trials for that bird, and then averaged across birds. Also, for comparison purposes with baseline performance, mean performance with baseline stimuli across birds during probe sessions was calculated.

A preliminary analysis revealed no significant differences between the data for birds in the first and second groups that were tested with somewhat different procedures, so data from the two groups were combined for further analysis.

### III. RESULTS

During baseline training, all starlings achieved a high probability of pecking the response key associated with isochronous stimuli when isochronous stimuli occurred, and a

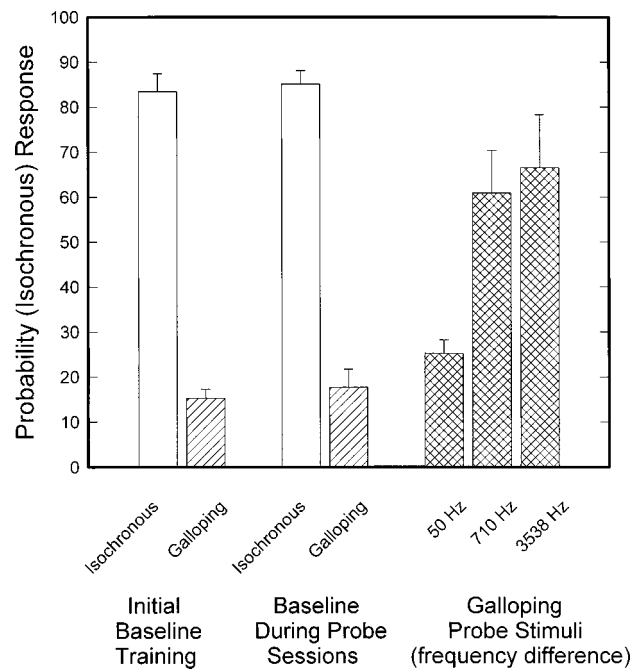


FIG. 3. The two left pairs of columns show the average probability (+SEM) that a starling pecked the key associated with the isochronous stimuli for isochronous and galloping stimuli during initial baseline training and on trials with baseline stimuli during sessions that included probe trials. The three columns at the right of the figure show the average probability (+SEM) of pecking the isochronous key when galloping probe stimuli occurred containing sine tones that differed in frequency by 50, 710, or 3538 Hz.

correspondingly low response probability of pecking the isochronous key when galloping stimuli occurred. As Fig. 3 shows, that was true both for mean probabilities over the last 3 days of initial baseline training and for baseline trials during probe sessions. To be sure, the initial baseline discrimination task was difficult, requiring a mean of 8296 trials (range 5368–13 459) for the birds to reach criterion. Nevertheless, once acquired, the birds maintained stable performance on the baseline task at a high level of accuracy.

Most important, when frequency changes were introduced within the galloping stimuli on probe trials, there was a progressively higher probability of responding to a probe stimulus as if it were isochronous the larger the frequency change between successive tones in the probe stimuli. For a 50-Hz frequency change there was little change in the probability of a response to the isochronous key, but as the frequency difference increased from 50 to 710 Hz to 3538 Hz, the probability of pecking the isochronous key when the probe stimulus occurred approached that for the baseline isochronous stimuli. Therefore, when frequency differences between tones within the probe stimulus were sufficiently large on probe trials, the starlings apparently reorganized the galloping stimulus perceptually into one (or two) isochronous stimuli (Fig. 1), attended to one or the other, and responded accordingly—an example of auditory stream segregation at work.

Statistical analysis amply supports these conclusions. The starlings first learned accurate discrimination of the baseline stimuli and then maintained that discrimination during sessions with probe trials. A two-way within-subject

analysis of variance (ANOVA) on the probability of pecking the isochronous key for isochronous and galloping stimuli during initial baseline training and over baseline trials during probe sessions yielded a significant difference for isochronous as compared with galloping stimuli,  $F(1,6) = 361.70$ ,  $p < 0.001$ . No other effects were significant, indicating no reliable evidence for a change in discrimination performance on baseline stimuli during sessions with probe trials.

For large frequency differences on galloping probe trials, the probability of producing responses associated with isochronous stimuli was very similar to that observed for baseline isochronous stimuli. A one-way within-subject ANOVA on the probability of pecking the isochronous key that included (a) data for baseline stimuli during probe sessions and (b) the three frequency-difference conditions yielded a significant difference among the means for conditions,  $F(4,24) = 24.557$ ,  $p < 0.001$ . Subsequent Tukey's tests showed significantly greater ( $p < 0.05$ ) probability of pecking the isochronous key during baseline trials as compared with the galloping probe stimuli incorporating a small 50-Hz frequency difference. However, with the larger 710- and 3538-Hz frequency differences on galloping probe trials, the difference between the probability of an isochronous response on baseline trials with isochronous stimuli and on probe trials with galloping stimuli was not significantly different ( $p$ 's  $> 0.05$ ).

Comparisons among the probe stimuli themselves showed that the probabilities of pecking the isochronous key for the 710- and 3538-Hz probe stimuli were both significantly greater by Tukey test than the probability of an isochronous response to the 50-Hz probe stimuli ( $p$ 's  $< 0.001$ ). Interestingly, there was no significant difference between the probability of pecking the isochronous key given a 50-Hz probe stimulus, and the probability of pecking the isochronous key given a galloping baseline stimulus ( $p > 0.05$ ). This indicates that a 50-Hz frequency difference on galloping probe trials was insufficient to cause stream segregation to occur. It would be interesting to know from future research the frequency difference on galloping probe trials that would just be sufficient for stream segregation to occur.

#### IV. DISCUSSION

These data, to our knowledge, provide the first observation of auditory stream segregation in nonhuman animals of non-natural, arbitrary (pure tone) stimuli based on frequency differences. The interpretation of our results is straightforward (Bregman and Campbell, 1971; Bregman, 1990; Van Noorden, 1975). Gestalt principles hold that sounds of common frequency will group together. The stimulus structures we used on probe trials provided the possibility that two separate streams, marked by different temporal structures, could form on the basis of common frequencies within streams, but different frequencies between streams. The data show that grouping did indeed occur, as measured by a shift in the starlings' perception of the temporal structure of a stimulus sequence from galloping to isochronous with the introduction of frequency differences between adjacent tones in the stimuli. Furthermore, grouping tendencies should be more pronounced the larger the frequency difference be-

tween two stimulus groups. Once again, the data support this prediction. The tendency of the starlings to respond to the galloping sequence as if it were isochronous was an increasing function of the frequency disparity between the tones of the two putative isochronous streams within the galloping stimulus.

The process of selective attention is one of the hallmarks of stream segregation. Observers can attend to one or the other of two ongoing streams, but not to both at the same time. Although we have no direct evidence that selective attention was taking place between the high- and low-frequency streams on probe trials, the data are not out of line with that idea. Presumably, the starlings were hearing one of the two isochronous streams as an isochronous 1 stimulus on some probe trials, and as an isochronous 2 stimulus on others (Figs. 1 and 2)—or perhaps shifting attention between the two stimuli within a trial—but we do not know for sure. This issue could be settled by introducing a third response contingency into baseline training. Starlings would be trained to peck one key for a galloping stimulus, a second key for an isochronous 1 stimulus, and a third key for an isochronous 2 stimulus. During subsequent probe trials with galloping stimuli containing sufficiently different frequencies, the starlings should always show a low probability of responding to the key associated with galloping stimuli, but might well distribute their responses from probe trial to probe trial between both isochronous keys—on the assumption that selective attention between the two streams would vary from trial to trial (or even from time to time within a trial until the moment when a response decision was made). A bias to respond to the stimulus stream associated with one isochronous key as compared with the other could be introduced in many ways and might be interesting to explore from the point of view of selective attention, and so would the introduction of spatial location as a cue correlated with different streams.

#### A. Auditory scene analysis as an adaptive phenomenon

The present results join our earlier observations with natural stimuli, birdsong in particular, to support the claim that auditory stream segregation is a general process in at least one avian species and, we venture to guess, in others (Hulse *et al.*, 1997; Wisniewski and Hulse, 1997). Whether or not auditory stream segregation is a general process with boundaries extending beyond avians (and humans) remains to be determined. However, the potential significance of the process for evolutionary development in species that uses acoustic information certainly suggests that auditory scene analysis, in some form, is probably quite common.

For example, although not cast specifically in terms of scene analysis, there are suggestions that auditory object formation in the sense of scene analysis has been observed in green tree frogs (e.g., Gerhardt, 1989, 1992; Gerhardt and Klump, 1988) and in chorusing insects (e.g., Otte, 1974), among other species, where the identification of a potential mate in a "noisy" background is of singular importance. The problem arises not only when individuals must be identified in a chorus of other individuals of the same species, but

also when important sounds are “degraded” by the context (e.g., forest or city sounds) in which they occur (Dabelsteen *et al.*, 1993). McGregor and Dabelsteen (1996), in their discussion of communication networks in birds, clearly implicate scene analysis as a process in avian auditory interactions. Any circumstance in which auditory communication (or the detection of functional auditory objects) must take place in an acoustic environment that is full of ecologically relevant noise provides a likely place to look for scene analysis at work. An especially interesting comparison would be to examine auditory objects that require long as compared with short time constants for integrating functionally useful information, e.g., bird song lasting several seconds or more as compared with a single loud alarm call. The former may be more amenable to streaming and scene analysis than the latter.

### **1. Noise versus distraction in auditory stream segregation of functional auditory objects**

In the foregoing connection, we wish to make a distinction between noise as it is usually conceived in psychoacoustics, and noise—we prefer the term *distraction*—as it is conceived in the analysis of functional, perhaps ecologically relevant, auditory stimuli. The distinction is but tangentially relevant for the simple two-stream segregation phenomenon we report here, but it has important implications for auditory scene analysis broadly construed.

In psychoacoustics, noise generally refers to a broadband spectrum of some kind, at some level, of which white noise is a prime example. Such noise has random structure through time. Psychoacoustics is replete, of course, with studies of auditory perception modulated by such noise. In auditory scene analysis, however, “noise” is perhaps best conceived as the auditory scene—a collection of auditory objects that have potential functional or ecological significance (Bregman, 1990). In that context, auditory objects are different from noise; that is, when parsed from the auditory scene by the auditory system, they have a location, a pitch, a spectral structure, perhaps a name, and so on, in a real world. Perhaps most important, auditory objects are usually acoustic signals that flow through time with a predictable serial or temporal structure from moment to moment, such as the syntax of speech or of bird song. That is not to say that auditory objects cannot be treated as noise in the psychoacoustic sense. One can obtain frequency spectra and their associated levels for any collection of sounds. A critical difference, however, is that a pool of auditory objects provides an opportunity for active *selective attention* among individual objects in the pool, and this is important for the perception of functional, ecologically significant auditory events in a real world.

### **2. Central versus peripheral factors in auditory scene analysis**

Although we stress the role of central factors associated with attention in auditory scene analysis, we do not intend to say that some of the principles of basic psychoacoustics might not also play a significant role. For example, Hartmann and Johnson (1991) have stressed the potential impor-

tance of peripheral channeling, especially frequency and spatial location, in the development of stream segregation and Klump (summarized in Klump, 1996) has stressed the role of comodulation masking release in hearing signals in noise. Klump and Langemann (1995) showed that European starlings demonstrated comodulation masking release for signals in background noise for rates of amplitude modulation of the noise not unlike those found for background sounds in the birds’ natural habitat (50 Hz). Also, Schwartz and Gerhardt (1989) found spatially mediated release from masking in treefrogs. We do claim, however, that selective attention, especially of functional auditory objects, implies a central process in the acoustic domain—based especially on Gestalt grouping principles or other high-level processes—that plays a significant role in auditory scene analysis. That is especially true for natural signals in the everyday world in which both human and nonhuman animals live.

### **B. Neurophysiology and models of auditory scene analysis**

Finally, we note briefly that the demonstration of robust auditory scene analysis in a nonhuman animal provides a ready source of neurophysiological information with which to constrain models of auditory scene analysis that are currently under rapid development (e.g., Beauvois, 1996; Beauvois and Meddis, 1991; McCabe and Denham, 1997; Todd, 1996; Wang, 1996). The songbird brain has been under close scrutiny for some time, especially because of significant structural and chemical changes that take place annually in a neural song control system unique to songbirds (e.g., Ball and Hulse, 1998). Consequently we know a great deal about the neural circuitry that is involved in the learning and production of bird song, and advances are forthcoming rapidly in understanding the neural basis of song perception. Therefore, research designed to study directly the neural control of auditory scene analysis in songbirds may be especially helpful in formulating models of how auditory scene analysis takes place.

## **V. CONCLUSIONS**

We have demonstrated that European starlings show auditory stream segregation for sequences of pure tones that differ in frequency, and that the magnitude of the segregation effect is a function of the size of the frequency difference between the streams. When combined with earlier data from starlings that have demonstrated segregation of ecologically relevant auditory signals (birdsong), the present data support the conclusion that auditory scene analysis plays a general role in the auditory perception of at least one avian species. It seems reasonable that the scene analysis process is potentially significant for many nonhuman species that must parse the world into significant auditory events. Moreover, the demonstration of robust auditory scene analysis in a nonhuman animal provides a potentially fruitful source of neurological information that should facilitate the construction of formal models of scene analysis.

## ACKNOWLEDGMENTS

The research was supported by National Science Foundation Grant No. IBN 9317868 to Stewart H. Hulse.

- Ball, G. F., and Hulse, S. H. (1998). "Bird song," *Am. Psychol.* **53**, 37–58.
- Beauvois, M. W. (1996). "Computer simulation of auditory stream segregation in alternating-tone sequences," *J. Acoust. Soc. Am.* **99**, 2270–2280.
- Beauvois, M. W., and Meddis, R. (1991). "A computer model of auditory stream segregation," *Quarterly Journal of Experimental Psychology* **43**, 517–541.
- Bregman, A. S. (1990). *Auditory Scene Analysis*. (MIT, Cambridge).
- Bregman, A. S., and Campbell, J. (1971). "Primary auditory stream segregation and perception of order in rapid sequences of tones," *J. Exp. Psychol.* **89**, 244–249.
- Cherry, E. C., and Taylor, W. K. (1954). "Some further experiments upon the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **26**, 554–559.
- Dabelsteen, T., Larsen, O., and Pedersen, S. B. (1993). "Habitat-induced degradation of sound signals: Quantifying the effects of communication sounds and bird location on blur ratio," *J. Acoust. Soc. Am.* **93**, 2206–2220.
- Darwin, C. J., and Carlyon, R. P. (1995). "Auditory grouping," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 387–424.
- Darwin, C. J., Hukin, R. W., and Al-Khatib, B. Y. (1995). "Grouping in pitch perception: Evidence for sequential constraints," *J. Acoust. Soc. Am.* **98**, 880–885.
- Dowling, W. J., Lung, K. M-T., and Herrbold, S. (1987). "Aiming attention in pitch and time in the perception of interleaved melodies," *Percept. Psychophys.* **41**, 642–656.
- Gerhardt, H. C. (1989). "Acoustic pattern recognition in anuran amphibians," in *The Comparative Psychology of Audition: Perceiving Complex Sounds*, edited by R. J. Dooling and S. H. Hulse (Erlbaum, Hillsdale, NJ), pp. 175–197.
- Gerhardt, H. C. (1992). "Multiple messages in acoustic signals," *The Neurosci.* **4**, 391–400.
- Gerhardt, H. C., and Klump, G. M. (1988). "Masking of acoustic signals by the chorus background noise in the green treefrog: a limitation on mate choice," *Animal Beh.* **36**, 1247–1249.
- Hartmann, W. M., and Johnson, D. (1991). "Stream segregation and peripheral channeling," *Music Perception* **9**, 155–184.
- Henwood, K., and Fabrick, A. (1979). "A quantitative analysis of the dawn chorus: temporal selection for communicatory optimization," *Am. Nat.* **114**, 260–274.
- Hulse, S. H., MacDougall-Shackleton, S. A., and Wisniewski, A. B. (1997). "Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (*Sturnus vulgaris*)," *J. Comp. Psych.* **111**, 3–13.
- Kacelnik, A., and Krebs, J. R. (1983). "The dawn chorus in the great tit (*Parus major*): Proximate and ultimate causes," *Behaviour* **83**, 297–309.
- Klump, G. M. (1996). "Bird communication in the noisy world," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell U.P., Ithaca, NY), pp. 321–338.
- Klump, G. M., and Langemann, U. (1995). "Comodulation masking release in a songbird," *Hearing Res.* **87**, 157–164.
- McCabe, S. L., and Denham, M. J. (1997). "A model of auditory streaming," *J. Acoust. Soc. Am.* **101**, 1611–1621.
- McGregor, P. K., and Dabelsteen, T. (1996). "Communication networks," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell U.P., Ithaca, NY), pp. 409–525.
- Newman, R., and Jusczyk, P. W. (1996). "The cocktail party effect in infants," *Percept. Psychophys.* **58**, 1145–1156.
- Otte, D. (1974). "Effects and functions in the evolution of signaling systems," *Annu. Rev. Ecology and Systematics* **5**, 385–417.
- Schwartz, J. J., and Gerhardt, H. C. (1989). "Spatially mediated release from auditory masking in an anuran amphibian," *J. Comp. Physiol. A* **166**, 37–41.
- Staicer, C. A., Spector, D. A., and Horn, A. G. (1996). "The dawn chorus and other diel patterns in acoustics signaling," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell U.P., Ithaca, NY), pp. 426–453.
- Todd, N. (1996). "An auditory cortical theory of primitive auditory grouping," *Network Computation in Neural Systems* **7**, 349–356.
- Van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.
- Wang, D. (1996). "Primitive auditory segregation based on oscillatory correlation," *Cogn. Sci.* **20**, 409–456.
- Wisniewski, A. B., and Hulse, S. H. (1997). "Auditory scene analysis in European starlings (*Sturnus vulgaris*): Discrimination of starling song segments, their segregation from conspecific songs, and evidence for conspecific song categorization," *J. Comp. Psych.* **111**, 337–350.
- Wood, N., and Cowan, N. (1995). "The Cocktail Party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel?," *J. Exp. Psychol.* **21**, 255–260.