Auditory scene analysis by echolocation in bats

Cynthia F. Moss\textsuperscript{a)\textsuperscript{b)\textsuperscript{b)}}
Department of Psychology, Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland 20742

Annemarie Surlykke\textsuperscript{b)\textsuperscript{b)}}
Center for Sound Communication, Institute of Biology, Odense University, SDU, DK-5230 Odense M., Denmark

(Received 30 October 2000; revised 27 June 2001; accepted 5 July 2001)

Echolocating bats transmit ultrasonic vocalizations and use information contained in the reflected sounds to analyze the auditory scene. Auditory scene analysis, a phenomenon that applies broadly to all hearing vertebrates, involves the grouping and segregation of sounds to perceptually organize information about auditory objects. The perceptual organization of sound is influenced by the spectral and temporal characteristics of acoustic signals. In the case of the echolocating bat, its active control over the timing, duration, intensity, and bandwidth of sonar transmissions directly impacts its perception of the auditory objects that comprise the scene. Here, data are presented from perceptual experiments, laboratory insect capture studies, and field recordings of sonar behavior of different bat species, to illustrate principles of importance to auditory scene analysis by echolocation in bats. In the perceptual experiments, FM bats (\textit{Eptesicus fuscus}) learned to discriminate between systematic and random delay sequences in echo playback sets. The results of these experiments demonstrate that the FM bat can assemble information about echo delay changes over time, a requirement for the analysis of a dynamic auditory scene. Laboratory insect capture experiments examined the vocal production patterns of flying \textit{E. fuscus} taking tethered insects in a large room. In each trial, the bats consistently produced echolocation signal groups with a relatively stable repetition rate (within 5\%). Similar temporal patterning of sonar vocalizations was also observed in the field recordings from \textit{E. fuscus}, thus suggesting the importance of temporal control of vocal production for perceptually guided behavior. It is hypothesized that a stable sonar signal production rate facilitates the perceptual organization of echoes arriving from objects at different directions and distances as the bat flies through a dynamic auditory scene. Field recordings of \textit{E. fuscus}, \textit{Noctilio albiventris}, \textit{N. leporinus}, \textit{Pippistrellus pipistrellus}, and \textit{Cormura brevirostris} revealed that spectral adjustments in sonar signals may also be important to permit tracking of echoes in a complex auditory scene. © 2001 Acoustical Society of America. [DOI: 10.1121/1.1398051]

PACS numbers: 43.80.Lb, 43.80.Ka, 43.66.Ba, 43.66.Lj, 43.66.Mk [WA]

I. INTRODUCTION

Auditory perception goes beyond the detection, discrimination, and localization of sound stimuli. It involves the organization of acoustic events that allows the listener to identify and track sound sources in the environment. For example, at a concert, individuals in the audience may be able to hear out separate instruments, or differentiate between music played from different sections of a symphony orchestra. At the same time, each listener may also follow a melody that is carried by many different sections of the orchestra together. In effect, the listener groups and segregates sounds according to similarity or differences in pitch, timbre, spatial location, and temporal patterning, to perceptually organize the acoustic information from the auditory scene. Auditory scene analysis thus allows the listener to make sense of dynamic acoustic events in a complex environment (Bregman, 1990).

Auditory scene analysis shares many perceptual phenomena with visual scene analysis, facilitating the application of Gestalt principles to its study. For example, both visual and auditory scene analyses involve the segregation and grouping of information to develop a coherent representation of the external sensory environment. The auditory and visual systems both perform operations such as object recognition, figure–ground segregation, and stimulus tracking.

The phenomenon of auditory stream segregation in humans appears to follow some fairly simple laws, namely the effect depends on the frequency similarity of groups of tones and the rate at which stimuli are presented. Similar laws also apply to apparent motion in vision, a phenomenon in which spatially separated images that flash in sequence give rise to an observer’s perception of a moving light. As with melodic motion in auditory streaming, visual apparent motion depends on the similarities among the images that flash and the temporal separation of flashes (Körte, 1915).

Perceptual grouping and segregation of sounds plays an important role in auditory scene analysis. In human psychoacoustic studies, a subject listening to tones that alternate slowly between low and high frequencies (e.g., 300, 2000, 400, 2100, 500, 2200 Hz) reports hearing the individual tones in the sequence. However, when the stimulus presen-
Auditory Stream Segregation

FIG. 1. (A) Schematic illustration of alternating tone frequencies over time that give rise to the perception of two pitch streams (after Bregman, 1990). (B) Schematic illustration of alternating echo delays over time that are hypothesized to give rise in the bat to the perception of two sonar target distance streams.

The acoustic parameters a bat uses to determine the features of an object is conveyed by the features of the echoes (Moss and Schnitzler, 1995). For example, a bat can discriminate the size of an object from the intensity of echoes (Simmons and Vernon, 1971), the velocity of a target from the Doppler shift of the echoes (Schnitzler, 1968), and the shape of an object from the spectrum of the echoes (Simmons et al., 1974). Differences in arrival time, intensity and spectrum of echoes at the two ears encode the location of an object in azimuth (Blauert, 1997; Shimozawa et al., 1974; Simmons et al., 1983) and elevation (Batteau, 1967; Grinnell and Grinnell, 1965; Lawrence and Simmons, 1982; Wotton et al., 1995). The third dimension, the distance between the bat and an object (target range), is determined from the time delay between the outgoing sound and the returning echo (Hartridge, 1945; Simmons, 1973). Together, these cues provide the bat with information to form a three-dimensional (3D) representation of a target and its position in space. While research over the last 30 years has elucidated the echo features that bats use to localize and discriminate sonar targets, we know little about how these acoustic features are perceptually organized in the representation of an auditory scene. Here, we present a conceptual framework to advance our understanding of the bat's perceptual organization of sound for the analysis of complex auditory scenes.

II. DETAILING THE PROBLEM: AUDITORY SCENE ANALYSIS BY ECHOLOCATION

The analogy between auditory scene analysis in bats and visual scene analysis in other animals is particularly strong, because sonar targets are auditory objects, and the echolocating bat uses spatial information carried by sonar echoes to identify and track these auditory objects. Many species of bat use echolocation to hunt small insect prey in the dark. This is a daunting perceptual task, given the acoustic environment in which the bat must operate. The bat produces sonar vocalizations and processes information in returning echoes to detect, track, and intercept insects that may be only a few millimeters in diameter. To accomplish this task, the bat must sort out its own sonar vocalizations from returning echoes. In addition, it may encounter echoes from multiple targets (several insects in a cluster, branches, walls, etc.), as well as signals produced by other bats in close proximity. To successfully intercept the insect and avoid obstacles, the bat must organize acoustic information collected from multiple
sonar targets arriving from different directions and at different times. This problem is illustrated in Fig. 2(A).

Figure 2(A) (above) shows a bat pursuing an insect that is flying in the vicinity of trees. This figure contains six panels, each illustrating a slice in time, and the bat’s position relative to the insect and trees changes over time. Below the illustrated panels is a schematic showing the relative timing of sonar pulses (solid bars) and target echoes (individual trees and insect echoes are represented by filled triangles, circles, and x’s). Bottom shows separate plots of the time delay between each sonar pulse and returning echo depends on the distance between the bat and the reflecting object, plotted separately for each of the objects. (B) A schematic that illustrates the changing echo delays for each of the reflecting objects. The right y-axis shows the stages of insect pursuit. B-I refers to buzz one and B-II refers to buzz two. As the bat flies closer to the trees and insect, the delays shorten. The echo amplitudes are arbitrarily set at fixed values that do not change with distance but durations and signal intervals were taken from a pursuit sequence recorded in the field. Only when the bat has passed an object does the amplitude decrease rapidly to reflect the low SPL radiated in the backward direction. The rate of delay changes over time for each of the reflecting objects as a distinct ridge with a particular slope. In this display, one can visually identify and track the returning echoes from the trees and insect over time.

FIG. 2. (A) Illustration of a bat pursuing a prey item in the vicinity of trees. Each panel represents a new slice in time, each separated by a fixed interval. In each panel, the bat’s position relative to the trees and insect changes, as the bat pursues its prey. Below the drawing is a schematic that displays the relative timing of sonar pulses (solid bars) and target echoes (individual trees and insect echoes are represented by filled triangles, circles, and x’s). Bottom shows separate plots of the time delay between each sonar pulse and returning echo depends on the distance between the bat and the reflecting object, plotted separately for each of the objects. (B) A schematic that illustrates the changing echo delays for each of the reflecting objects. The right y-axis shows the stages of insect pursuit. B-I refers to buzz one and B-II refers to buzz two. As the bat flies closer to the trees and insect, the delays shorten. The echo amplitudes are arbitrarily set at fixed values that do not change with distance but durations and signal intervals were taken from a pursuit sequence recorded in the field. Only when the bat has passed an object does the amplitude decrease rapidly to reflect the low SPL radiated in the backward direction. The rate of delay changes over time for each of the reflecting objects as a distinct ridge with a particular slope. In this display, one can visually identify and track the returning echoes from the trees and insect over time.
with target distance, but the figure does illustrate how echo amplitude decreases rapidly after the bat flies past an object. It is important to note how the increase in sound repetition rate (and the decrease in signal duration) in the terminal phase defines and “sharpens up” the ridges corresponding to the insect and the three trees. Just as the layout of a graphic display can aid the identification of patterns in echo delay changes over time, we propose that the bat’s auditory system perceptually organizes the signals arriving from multiple objects at different distances to create a readable display, which then permits tracking and capture of insects in a dynamic and complex acoustic environment.

Moreover, echolocation is an active system, i.e., the bat transmits the very sound energy it uses to probe the environment. As a bat flies toward a target, changes in the repetition rate, bandwidth, and duration of its sonar emissions occur, and these dynamic vocal production patterns are used to divide the bat’s insect pursuit sequence into different phases: search, approach, and terminal buzz [Griffin et al., 1960; Webster, 1963; see Fig. 2(B)]. Given that the perceptual organization of sound depends on the acoustic features of signals, changes in the bat’s sonar vocalizations certainly impact its analysis of an auditory scene. While the active modulation of sonar vocalizations in response to changing echo information complicates our task of understanding scene analysis by echolocation, the bat’s vocal production patterns also provide us with a window to the acoustic information the bat is actively controlling as it maneuvers through the environment.

How does the bat’s sonar system operate to coordinate spatial acoustic information from a complex environment to avoid obstacles and intercept prey? Here, we describe some experimental data showing that the bat assembles acoustic information over time, a requirement for spatial tracking of moving targets and the analysis of auditory scenes. We also discuss sonar vocalization patterns of bats operating in a dynamic auditory scene. These vocalization patterns provide some insights to the bat’s active control for scene analysis by echolocation.

IV. STUDY 1. PSYCHOPHYSICAL EXPERIMENTS

FM bats, such as Eptesicus fuscus, receive brief acoustic flashes of a changing environment as they fly. In order for the FM bat to organize acoustic information from sonar targets in a dynamic scene, it must assemble echo information over time. E. fuscus must therefore use working- or short-term memory to track changing echo information about sonar targets.

Field and laboratory observations of the bat’s increasing repetition rate during insect approach and capture might suggest that echolocating bats integrate acoustic information over time; however, these data have not been taken as the definitive demonstration of echo integration, because echo delay-dependent changes in vocal production patterns do not explicitly inform us about the animal’s perception of echo sequences. To experimentally address this question, we have conducted some perceptual experiments, and the data demonstrate that Eptesicus fuscus can indeed assemble information about changing echo delay to discriminate sonar targets. These findings suggest that the bat’s perceptual system meets the minimum requirements for analysis of a dynamic auditory scene.

A. Animals

Five FM bats of the species E. fuscus served as subjects in the different perceptual tasks. The animals were collected from private homes in Maryland during the summers of 1995 and 1996 and housed in a colony room at the University of Maryland, College Park. The temperature in the colony room was maintained at approximately 27 °C and the day/night cycle was reversed, with lights out between 7:00 a.m. and 7:00 p.m. Bats were given free access to water and maintained at about 85% of ad lib body weight. Food was available only as a reward during behavioral experiments, which were carried out 6–7 days/week over a period of 18 months.

B. Apparatus and target simulation

Behavioral experiments took place in a large (5.6 × 6.4 × 2.5 m) carpeted room, whose walls and ceiling were lined with acoustic foam (Sonex) that reduced the amplitude of ultrasonic reverberation by a minimum of 20–30 dB below what would be present if the room surfaces were hard and smooth.

Each bat was trained to rest at the base of an elevated Y-shaped platform (1.2 m above the floor) and to produce sonar sounds. The bat’s sonar sounds were picked up by a vertically oriented 1/8 in. Bruel & Kjaer condenser microphone (model 4138) that was centered between the arms of the platform at a distance of 17 cm from the bat. The bat’s
Echolocation sounds were amplified, bandpass filtered at 20–99 kHz (Stewart filter, model VBF7), digitized with 12-bit accuracy at a rate of 496 kHz controlled externally by a Stanford Research Systems function generator, electronically delayed (custom DSP board, SPB2-Signal Data, installed in a 486 computer), attenuated (PA-4, Tucker-Davis Technologies), low-pass filtered (Krone-Hite), and broadcast back to the bat through a custom electrostatic loudspeaker (designed by Lee Miller, University of Odense, Denmark). The loudspeaker, positioned in front of the microphone 0.5 cm lower than the microphone grid and 15 cm from the bat, was powered by a dc amplifier (Ultrasound Advice) and had a frequency response that was flat within 3 dB between 25 and 95 kHz [see Figs. 3(A),(B)]. Placement of the microphone behind the speaker eliminated feedback in the system, and the signals recorded were not distorted by the presence of the speaker (see Wadsworth and Moss, 2000).

The total gain of the system feeding into the DSP board was approximately 60 dB, to bring the peak–peak amplitude of most bat sonar sounds to a level just below the 12-bit limit of the processor for maximum signal-to-noise ratio (S:N). Digital attenuators (PA-4, Tucker Davis Technologies) permitted adjustment of the playback level of the sounds returning to the bat’s ears, which was set at approximately 80 dB SPL (rms) for all experiments.

Each sound produced by the bat resulted in a single sonar signal playback that simulated an echo from a target positioned directly in front of the bat, whose distance was determined by an electronic delay controlled by the experimenter. The shortest echo delay used was 4.43 ms, corresponding to a target distance of 76 cm, and the longest echo delay was 4.93 ms, corresponding to a target distance of 85 cm.

Before each experiment, a calibration routine was run to test each of the components of the target simulator. The electrostatic loudspeaker broadcast a linear 1-ms 10–100-kHz frequency modulated sweep that was picked up by a condenser microphone (QMC) positioned on the test platform. The signal received by the microphone was amplified, filtered (Stewart VBF-7, 10–99 kHz), and delivered to the DSP board. The arrival time and power spectrum of the FM sound picked up by the microphone were measured and compared against standard values. Experimental data were collected only when the delay and power spectrum of the calibration signal matched the standard values, a 0.29-ms delay when the microphone was positioned 10 cm from the speaker (one-way travel delay, 2.9 ms/m) and a relatively flat spectrum [±3 dB at 25–95 kHz; see Fig. 3(B)].

C. Behavioral task: Sequential versus random echo delay sequences

Each of the bats was trained in a two-alternative forced-choice experiment to discriminate between two stimulus sets.
of echo delays. Each stimulus set contained the same set of six echo delays, and the bat received only one echo playback for each sonar emission. In one echo set, the delays systematically increased or decreased (stimulus “S”), and in the other echo set, the sequence of delays was randomized (stimulus “R”). The echo stimulus set repeated until the bat made its response (see below). The step size and direction of echo delay change was manipulated. The delay step size of echoes in stimulus set S was unequal in some experiments to ensure that the bat was discriminating the pattern of echo delay change, and not simply between variable and fixed delay steps in echo sets S and R. A schematic of these echo sets is illustrated in Fig. 3C. Table I lists the different echo delay conditions tested in these experiments.

The bat learned to report whether it perceived a systematic increase or decrease in target distance by crawling down the left arm of the platform to indicate an “S” (sequential) response or if it perceived a random presentation of the same echo delays by crawling down the right arm of the platform to indicate an “R” (random) response. The presentation of the random and sequential echo sets followed a pseudorandom schedule (Gellerman, 1933), and the bat’s response (S or R) was recorded. For each correct response, the bat received a food reward (a piece of mealworm), and for each incorrect response, the bat experienced a 10–30 s time-out. No correction trials were introduced. Each test day included 25–50 trials per bat.

D. Results
1. Behavioral performance

The data from two bats that successfully completed testing under all conditions are plotted in Fig. 4. (Three out of the five bats trained in the task fell ill during the experiment, and their data sets are incomplete.) The conditions are separated into different panels and identified by number 1–5 (see Table I for stimulus details on each condition). Note that in some instances, the bat’s discrimination of echo patterns in one condition transferred immediately to a new pattern, and performance remained at or above 75%-correct following the introduction of a new stimulus set. These data show that the bat can integrate echo sequences along the dimension of delay (Morris and Moss, 1995, 1996), and they lay the foundation for future studies that can directly probe the phenomenon of stream segregation along the delay axis in echolocating bats. Such experiments will determine whether the bat hears out delay streams from complex echo delay patterns.

Table I. Summary of stimulus conditions in the sequential versus random echo delay discrimination task. Values refer to electronic echo delays in ms. Acoustic travel time from the bat to the microphone and from the loudspeaker to the bat adds 0.93 ms to these values for the total echo playback delays presented in the sequential conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Decreasing:</th>
<th>Condition</th>
<th>Increasing:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.00 3.90 3.80 3.70 3.60 3.50</td>
<td></td>
<td>3.50 3.60 3.70 3.80 3.90 4.00</td>
</tr>
<tr>
<td>Condition</td>
<td>Decreasing:</td>
<td>Condition</td>
<td>Increasing:</td>
</tr>
<tr>
<td>2</td>
<td>4.00 3.95 3.90 3.85 3.80 3.75</td>
<td></td>
<td>3.75 3.80 3.85 3.90 3.95 4.00</td>
</tr>
<tr>
<td>Condition</td>
<td>Increasing:</td>
<td>Condition</td>
<td>Increasing:</td>
</tr>
<tr>
<td>3</td>
<td>3.50 3.60 3.70 3.80 3.90 4.00</td>
<td></td>
<td>3.50 3.55 3.75 3.82 3.90 4.00</td>
</tr>
</tbody>
</table>

FIG. 4. Performance of two individual bats (M-6 and G-6) trained to discriminate between the sequential and random playbacks. Breaks between the data points appear when experimental trials were not run over consecutive days. Each block shows data that come from a single condition, indicated by numbers 1–5 above the x-axis (see Table I). Criterion for successful discrimination was 75%-correct performance over a minimum of 3 days.

V. STUDY 2: BEHAVIORAL STUDIES OF ECHOLOCATING BATS OPERATING IN A COMPLEX AUDITORY SCENE

A complete understanding of auditory scene analysis should specifically address the acoustic problems that must be solved by the species under study. The use of biologically relevant stimuli in the study of auditory scene analysis has not been widely applied, but this approach has been adopted...
by Hulse and colleagues working on the perceptual organization of sound in the European starling (Hulse et al., 1997; Wisniewski and Hulse, 1997). They studied the European starling’s perception of conspecific song and found evidence for stream segregation of biologically relevant acoustic signals by the starling. Similarly, in bats, one can study the phenomenon of scene analysis in the context of biologically relevant acoustic tasks. In particular, the bat’s pursuit and capture of insect prey in a dynamic acoustic environment requires analysis of the auditory scene using echoes of its own sonar vocalizations. Thus, studies of vocal production patterns in echolocating bats can provide insights to the animal’s perception and control of the auditory scene (Wadsworth and Moss, 2000). Indeed, the echolocating bat’s perception of the world builds upon sonar returns of its vocalizations. Given that the repetition rate and frequency separation of sounds affect stream segregation in human listeners, the vocal production patterns of the bat would be expected to influence its perceptual organization of space using sonar echoes. Below, we describe high-speed video and sound recordings of the echolocating bat as it flies through a dynamic auditory scene to intercept insect prey tethered in open laboratory space and in the vicinity of echo clutter-producing leaves and branches.

### A. Animals

Seven FM bats of the species *Eptesicusfuscus* served as subjects in the laboratory insect capture studies. Four of the animals were collected from private homes in Maryland during the summers of 1997–1999 and three from a cave in Ontario, Canada in the winter of 2000. Bats were housed in the University of Maryland bat colony room under conditions comparable to those of the bats in the psychophysical experiments. Food was available only as a reward during behavioral experiments, which were carried out 4–5 days/week over a period of 27 months.

### B. Behavioral studies of echolocation behavior in free flight

Big brown bats (*Eptesicusfuscus*) were trained to capture tethered whole mealworms (*Tenebrio molitor*) in a large flight room (6.4 × 7.3 × 2.5 m) lined with acoustical foam (Sonex). Their vocalization behaviors were studied under open-space and clutter conditions. In the open-space condition, no obstacles were in the vicinity within 1 m of the insect target; however, the walls, ceiling, and floor of the flight room prevent us from creating a truly open space environment. In the clutter condition, the tethered mealworm was suspended 5–90 cm from leafy house plants placed in the calibrated space of the behavior room (see video processing methods below). The separation between target and clutter was manipulated between trials: close (5–15 cm), intermediate (20–30 cm), and far (>40 cm). A schematic of the setup for these experiments is presented in Fig. 5.

Experiments were carried out using only long-wavelength lighting (Plexiglas #2711; Reed Plastics and Bogen Filter #182) to eliminate the use of vision by the bat (Hope and Bhatnagar, 1979). Mealworms were suspended at a height of about 1.5 m above the floor by monofilament line (Trilene Ultra Thin, 0.1-mm diameter) within a 5.3-m target area in the center of the room. A mealworm was suspended at a randomly selected location within the target area, and then the bat was released in a random direction to orient to the target area and find the mealworm. So that the bat would not memorize the target area, the mealworm was suspended outside the target area 50% of the time, and those trials were...
not recorded. Once each bat achieved a consistent capture rate of nearly 100% in open-space conditions (typically within 2 weeks of introduction to the task), audio and video recordings of its capture behavior began.

1. Video recordings

Two gen-locked (frame synchronized), high-speed video cameras (Kodak MotionCorder, 640×240 pixels, 240-Hz frame rate, and 1/240-s shutter speed) were positioned just below the ceiling in the corners of the flight room. A calibration frame (Peak Performance Technologies) was placed in the center of the room and filmed by both cameras prior to each recording session. The high-speed video cameras were used to record target position, bat flight path and capture behavior. The resulting images were used in calculation of the three-dimensional positions of the bat, target, and microphones.

The video buffer contained 1963 frames, allowing for recording of 8.18 s of data at 240 frames/s. Using an end-trigger on the video, we captured the behavior leading up to and just following successful and unsuccessful insect captures.

2. Audio recordings

Echolocation signals were recorded using two ultrasonic transducers (Ultrasound Advice) placed within the calibrated space. In some trials, the microphone output was amplified (Ultrasound Advice) and recorded on the direct channels of a high-speed tape recorder (Racal Store-4 at 30 in. per second). An FM channel of the tape recorder was used to record TTL sync pulses corresponding to the start of each video frame and gated to the end of video acquisition. In other trials, the microphone signals were amplified, bandpass filtered (10–99 kHz, 40-dB gain, Stewart, VBF-7) and recorded digitally on 2 channels of an IoTech Wavebook 512 at a sample rate of 240 kHz/channel. The Wavebook, controlled by a Dell Inspiron 7000 laptop computer, had 16 Mbytes of random access memory (RAM) and was set to record 8.18 s prior to the trigger; the trigger was set to simultaneously stop the audio and video acquisition. The experimenter triggered the system on each trial after the insect capture was attempted and/or accomplished. This approach allowed us to record 8.18 s of audio data that corresponded precisely to the video segment for a given trial.

C. Data analysis

1. Video processing methods

A commercial motion analysis system (Peak Performance Technologies, Motus) was used to digitize both camera views with a Miro DC-30 Plus interface. The Peak Motus system was also used to calculate the three-dimensional location of points marked in both camera views. Digitization was to 1/4-pixel resolution using magnification. The digitization procedure resulted in 2560×1920 lines of resolution. The video image spanned less than 6 m horizontally, so that 1/4-pixel resolution corresponded to approximately 0.4 mm. Accuracy of the system was within ±0.5% over a calibrated volume extending approximately 2.2×2.2 m across the room and 1.6 m vertically. The three-dimensional space calibration frame provided 25 control points for direct linear transformation (DLT) calibration. The calibration procedure produced a mean residual error of 1.0 cm in each coordinate for the 25 control points.

The video position data for the bat, the target, and clutter-producing plants (when present) were entered in a database for each trial. The bat’s position with respect to the microphones was also measured, and a correction factor for the sound travel time from the bat to the microphone was used to accurately record the vocalization times.

2. Audio processing methods

Recordings of the bat’s sonar vocalizations in the laboratory insect capture studies were processed following two distinct methods. The sounds recorded on the Racal Store 4 tape recorder were played back at 1/4th the recording speed (7.5 in. per s) and digitized using a National Instruments board AT-MIO-16-1 with a sampling rate of 60 kHz per channel, resulting in an effective sampling rate of 240 kHz per channel. Custom software (LABVIEW) trimmed the digitized audio data to begin with the first and end with the last frame of each trial’s video segment, and output files were exported to a signal-processing program (SONA-PC®). Using SONA-PC, a fast Fourier transform (FFT) was performed over 256 points per time step, with 16–20 points being replaced in each time step, and the signals were displayed spectrographically. The onset time, duration, and start- and end frequencies of the first harmonic of the emissions were marked with a cursor on the display and entered in a database for further analysis. Data acquired digitally with the IoTech Wavebook were displayed as time waveforms and spectrograms (256-point FFT) using MATLAB. The onset time and duration of the signals were measured using the time waveforms, and frequency measurements were taken from the spectrograms. Audio and video data were merged in a single analysis file in order to associate vocal behavior with motor events. MATLAB animation permitted dynamic playback of the bat’s position data and corresponding vocalizations, enabling detailed study of the bat’s behavior under different task conditions (programming by Aaron Schurger). Examples can be found at http://batlab.umd.edu/insect_capture_trials.

D. Results

Three hundred and twenty trials were run in the open-space condition and 170 trials in the clutter condition. Bats successfully intercepted tethered insects at a rate of 95% in the open space and 80% in the presence of echo clutter-producing plants, when the target was 20–90 cm from the clutter. Detailed video and sound analyses were carried out for a subset of trials, 53 open-space trials and 107 clutter trials. Three-dimensional plots of the bats’ flight paths in representative trials from selected open-space and clutter conditions are shown in Fig. 6. Each data point (open circle) along the bat’s flight path indicates the occurrence of a sonar vocalization. The asterisks in the clutter trials denote the position of the branches. The final target position is shown with
a star, and in the case of moving target trials, the path of the insect is illustrated with a fine line. All trials presented in Fig. 6 resulted in successful target capture. Consistent with earlier reports (e.g., Griffin, 1953, 1958; Cahlander et al., 1964; Webster, 1963), the interval between the bat’s sonar vocalizations decreased as the bat approached the insect, but in general, the decrease was not continuous. In fact, most trials contained groups of sounds that occurred with relatively stable intervals, interrupted by a longer pause, and followed by another group of sounds, sometimes with the same stable interpulse interval (IPI). We will refer to the sound groups with stable intervals as sonar strobe groups.

Figure 7 shows the spectrograms of sounds produced by the bat as it approached and intercepted the tethered insect target under the same set of open-space (left panels) and cluttered (right panels) conditions as presented in Fig. 6. Note in all of the trials that the bat’s sonar sounds sometimes occurred in groups, interrupted by gaps. The groups of sounds contained signals with relatively stable intervals (less than 5% variance in the interval between sounds). This is displayed in Fig. 8, which plots the interpulse interval of successive sounds in the same four trials relative to the time of insect capture. In the example of the open-space moving target trial, the bat extends the production of these sound strobe groups with short IPIs over 1000 ms before contact, and the intervals remain just below 20 ms over several successive strobe groups. Processing these groups of sounds with relatively fixed repetition rate might facilitate the bat’s analysis of sonar scenes and planning for target capture.

These sonar signal sequences, typical of those recorded in the lab (and under some conditions in the field; see below), include examples of several groups of sounds with stable intervals. Histograms summarizing the interpulse intervals and durations of these sonar sound groups in the open-space and cluttered conditions are presented in Fig. 9. Stable-space IPI groups with intervals smaller than 13 ms were excluded from this summary figure, as they comprise the terminal buzz (see Fig. 11). In both the open and cluttered environments, the strobe groups occurred most frequently with intervals characteristic of the late approach phase of insect pursuit: median = 16.5 ms under open space conditions and 17.4 ms under clutter conditions. The differ-

FIG. 6. Data from selected insect capture trials. Flight paths of bats pursuing tethered insects under open-space (left) and cluttered (right) conditions. Each point (open circle) along the flight path denotes the occurrence of a sonar vocalization. The star shows the end position of the target, which in the upper two examples was stationary and in the lower two examples was moving. In the moving target trials, the path of the insect is illustrated by a thin line. Note that the sonar vocalizations occur at a higher repetition rate as the bat approaches the target. For the clutter trials shown, asterisks denote the position of the branches. The bat successfully intercepted the target in all four trials shown.
FIG. 7. Spectrograms of the sounds produced by the bats in the same trials as shown in Fig. 6. Each set of three panels comes from a single trial, with time wrapping from one panel to the next. Trials on the left come from open-space conditions and on the right from cluttered conditions. The upper panels display data taken when the target was stationary and the lower panels display data taken when the target was moving. Note that the sounds occur in stable IPI groups with a relatively consistent repetition rate, followed by a pause (e.g., clearly demonstrated in the middle panel of the open-space, moving target trial). Note that the panels show sonar sounds starting at approximately 1500 ms prior to target capture.

FIG. 8. Interpulse interval relative to target contact, plotted for the same four trials as shown in Figs. 6 and 7. The plots on the left show data from trials in open space and on the right from cluttered space. The upper plots show data from stationary target trials, and the lower plots show data from moving target trials. Note that the time scale on each panel shows sounds produced by the bat starting at 1500 ms prior to target contact.
ence in distribution of strobe group intervals between open space and clutter conditions was not statistically significant [Fig. 9(A)]. Across behavioral trials, stable IPI groups contained three, four, and sometimes more sounds. The median duration of the stable IPI groups (onset of the first sound to the onset of the last sound in the group) was 43.3 ms for the open space condition and 37.6 ms for the cluttered condition [Fig. 9(B)]. The difference between duration of strobe groups under open-space and cluttered condition was not statistically significant.

Figure 10 shows the incidence of sonar strobe groups for individual trials, displayed separately for open-space and clutter conditions. Individual trials analyzed are listed on the y axis (53 for open-space and 107 for clutter), and time relative to target contact on the x axis. Data points mark each sonar strobe group within a trial, and the position of each data point along the x axis indicates its time of occurrence relative to target contact (time zero). In both open-space and clutter trials, there is a clustering of strobe groups 250–750 ms before target contact. At earlier times relative to target contact, the incidence of sonar strobe groups is much lower in open space than in clutter. It is noteworthy that the incidence of strobe groups is much higher after target contact in the clutter condition than in the open space. In the clutter condition, the bat must negotiate the branches after contact, and this appears to influence the production of sonar strobe groups following insect capture in these trials.

As the bat gets closer to a target, it carefully adjusts its vocalizations to avoid overlap between sonar pulses and target echoes. In addition, the bat appears to adjust the interval between sounds to allow time for reception of relevant target echoes before producing the next sound. Just prior to capture, the bat produces sounds at a very high rate. This group of sounds, produced at a rate higher than about 80 sounds/second and reaching approximately 150/second just prior to capture, is called the terminal buzz [see Fig. 2(B)]. In some trials, the interval between the bat’s sounds dropped abruptly from about 15–18 ms down to 6–8 ms, omitting the buzz I component from the terminal sequence. The terminal sequences recorded in the laboratory always contained a buzz II component (intervals less than 8 ms), and summary data on buzz II are presented in Fig. 11.

Figure 11(A) plots buzz onset and offset (relative to target contact) and total buzz II duration for insect capture sequences in open space (moving and stationary targets) and in the presence of clutter (far, intermediate, and close to the target). While moving insects were presented in both the open-space and clutter conditions, the data are excluded from the clutter summary in this plot, because physical limitations of our setup only allowed us to introduce moving targets at the intermediate and far target–clutter separations.

Buzz onset time differed across conditions ($F = 5.37; p < 0.001$), with the smooth motion open-space condition yielding significantly earlier onset time than stationary open
space and clutter trials. The buzz onset time in the open-space moving target trials was earliest (mean = −265.3 ms, SEM 19.8 ms). The shortest target–clutter separation tested (5–15 cm) gave rise to the latest buzz onset time relative to contact (mean = −188.7 ms, SEM = 12.2 ms). The intermediate target–clutter separation (20–30 cm) and the far target–clutter separation (>40 cm) yielded similar buzz onset times (intermediate mean = −202.85 ms, SEM = 13.1 ms, and far mean = −195.8 ms, SEM = 10 ms), which did not differ statistically from the open-space stationary buzz onset time (mean = −213.9 ms, SEM = 4.7 ms).

The offset time of the buzz relative to target contact also differed significantly across conditions ($F = 6.27; p < 0.001$). It was earliest for the near-clutter condition (mean $= −66.2$ ms, SEM = $3.8$ ms), and decreased systematically from the intermediate ($−53.7$ ms, SEM = $3.1$ ms) to the far ($−44.79$ ms, SEM = $3.9$ ms) clutter conditions, and again slightly for the open-space stationary condition ($−38.08$ ms, SEM = $2.9$ ms). The mean open-space moving target trial buzz offset time relative to target contact was $−54.6$ ms (SEM = $4.7$ ms), similar to the mean for the intermediate target–clutter separation.

The mean duration of the buzz shortened systematically from the open-space moving target trials (mean = 210.8 ms, SEM = 17.2 ms) to the open-space stationary trials (mean = 175.8 ms, SEM = 4.8 ms) and across the far (151.06 ms, SEM = 10.2 ms), intermediate (149.2 ms, SEM = 13.5 ms), and near (122.5 ms, SEM = 14.6 ms) clutter–target trials. This overall change in buzz duration was statistically significant ($F = 6.23, p < 0.001$). Post hoc analyses show that the significant contrasts were between open-space motion and all three target–clutter conditions (Tukey comparisons, $p < 0.05$). By shortening the duration of the terminal buzz with target clutter, the bat shortened the time period when it was likely to experience a mixing of pulses and echoes from closely spaced sonar vocalizations.

Figure 11(B) shows that the target distance at which the bat initiated the terminal buzz was longer in open space than in clutter ($F = 19.77, p < 0.001$). In open space the mean distance was 61.55 cm (SEM = 2.75 cm) and 64.69 cm (SEM = 2.27 cm) for moving and stationary trials, respectively. As the target was placed closer to the clutter, the buzz onset distance became shorter, suggesting that the bat made its decision to intercept the target at closer range under these conditions. The mean target distances at buzz onset in cluttered space were 51.17 cm (SEM = 2.57 cm), 42.92 cm (SEM = 2.21 cm), and 33.77 cm (SEM = 3.0 cm) for the far, medium, and near conditions, respectively. Post hoc comparisons across conditions show that the buzz onset distance differed significantly across all five conditions (Tukey test, $p < 0.01$).

The difference in the bat’s distance to the target when it terminated the buzz was not statistically significant across conditions ($F = 2.36, p > 0.05$); however, the distance traveled by the bat during the buzz was different across all conditions ($F = 12.42, p < 0.001$). Distances traveled over the course of the buzz were 86.64 cm (SEM = 7.78 cm) in the open-space moving condition, 65.77 cm (SEM = 3.94 cm) in the open-space stationary condition, 51.43 cm (SEM = 3.3 cm) in the far-clutter condition, 49.93 cm (SEM = 4.49 cm) in the medium-clutter condition, and 36.24 cm (SEM = 5.73 cm) in the near-clutter condition.

VI. FIELD STUDIES OF ECHOLOCATION BEHAVIOR

The complex and dynamic acoustic environment of the echolocating bat requires perceptual organization of sound stimuli to segregate and track signals from different objects and other animals in the environment. Observations from field recordings suggest that the bat actively controls the information it extracts about the auditory scene by modulating the spectral and temporal characteristics of its sonar vocalizations. Here, we report on field recordings from several species of echolocating bat, with the goal of using sonar signal design and temporal patterning as a window to the bat’s perceptual control of the spatial information sampled from the environment.
A. Field sites

Echolocating bats inhabit both temperate and tropical environments, and sound recordings of foraging bats were taken at several different sites around the world. *Noctilio albiventris* and *N. leporinus* (Noctilionidae) and *Cormura brevirostris* (Emballonuridae) were recorded on Barro Colorado Island in Panama in October, 1999 in collaboration with E. Kalko and M. E. Jensen. The *Noctilio* species flew low over a calm water surface in a small bay. They were using echolocation to capture fish swimming just below the water’s surface and mealworms (set out by experimenters) floating on the water’s surface. The *Commura brevirostris* was recorded while it hunted aerial insects in a small clearing in the tropical rain forest covering the island.

The vocalizations of *E. fuscus* were recorded in the summers of 1999 and 2000 as they foraged insect prey in Greenbelt, Maryland. At this field site, *E. fuscus* flew along the edge of a forest that abutted a meadow. Under moonlight, two bats were sited that flew 3–5 m above the ground and 2–6 m from the forest edge, and on two occasions we recorded several bats simultaneously at this location. The echolocation behavior of *Pipistrellus pipistrellus* [45-kHz phonic type (Barratt et al., 1997)] was recorded in Montel (Midi-Pyrénées, southern France) in collaboration with Lee A. Miller *et al.* The area was under and close to a small bridge over a creek with brinks covered with shrubbery. Several pipistrelle bats hunted simultaneously in this area.

B. Recording methods

All recordings used microphones that operate in the ultrasonic range. In Panama, recordings were taken with a battery operated ¼-in. model 40BF G.R.A.S. microphone, am-

---

FIG. 11. (A) The onset and offset (relative to target contact) and duration of the terminal buzz II produced by bats just prior to insect capture under open space (moving and stationary) and stationary target in clutter trials (near, 5–15 cm; medium 20–30 cm, and far >40-cm target-clutter separations). (Error bars show one standard error of the mean.) (B) The bat’s distance to the target at the time it initiated and ended the terminal buzz across the same five conditions. Also shown is the distance traveled by the bat while producing the terminal buzz sequence in the five insect capture conditions.
A/D converted on-line of a 2-m thin rod. The microphone signals were continuously supplied. In all cases the microphone was mounted on the end and amplified with a Larsen Davis battery-operated power supply in a ring buffer. The Wavebook had 128 Mbytes of random access memory (RAM) and was set at 3–6 s pretriggering time and 1–2 s post-triggering time. The Wavebook was controlled by a laptop computer (AST Ascentia P series, IBM Thinkpad 600 or Dell Inspiron 7000). When we picked up strong signals on a bat detector (D240 Pettersson Electronics), we began signal acquisition with the Wavebook, storing the 3–6 s preceding and the 1–2 s following trigger. At all recordings sites, the data storage system was battery operated, resulting in a low noise floor.

In France the pipistrelle recordings were taken with a linear array of three Bruel & Kjaer ½-in. microphones, each separated by 75 cm. The array was placed horizontally approximately 1.5 m above the ground with the microphones pointing forward and upward at an angle of 45 deg. The sounds were stored on a RACAL instrumentation tape recorder running at 30 ips (flat frequency response up to 150 kHz), and later digitized for analysis using the IoTech Wavebook (see above).

C. Results

The field recordings demonstrated that bats emit signals that, depending on species, allow them to exploit both time- and frequency cues to segregate their own signals and echoes from a complicated background of noise and emissions from other bats. In general, the bat diversity is much greater in the tropics, which is reflected in a broad representation of sonar signal characteristics from this part of the world. In temperate areas, there are not nearly as many bat species. However, acoustically, temperate species may face comparable sonar challenges, since several individuals of the same species may hunt in close proximity in areas with a high insect density. Naturally, the similarity of signals is greatest within the same species, and thus this situation may create confusion for the bat trying to discriminate between echoes from its own emissions and those of the neighbors.

The recordings of the two Noctilio species (N. leporinus and N. albiventris) illustrate just how complex the acoustic environment of the bat can be. Generally, the density of bats on Barro Colorado Island, Panama was very high. Up to 25 Noctilio bats, most of which were N. leporinus, might fly in the same area simultaneously. Thus, we recorded many overlapping sonar signals, and the silent breaks were few and short. N. leporinus emits signals with most energy around 55 kHz (Schnitzler et al., 1994). Our recordings (Fig. 12) show peak energy around 58 kHz, whereas the dominant energy in the sonar signals of N. albiventris was around 74 kHz, close to the values (67–72 kHz) reported earlier (Kalko et al., 1998). Figure 12 also illustrates that the N. albiventris recorded simultaneously emit slightly different frequencies, e.g., approximately 72 and 74 kHz.

Similar to the insect capture recordings of E. fuscus in the laboratory (see Sec. V above and Figs. 6, 7, and 8), vocal sequences recorded in the field included stable IPI sequences. Groups of sonar signals produced at a stable repetition rate in the field again suggest active vocal control to facilitate analysis of the auditory scene. Illustration of sonar IPI sets produced by E. fuscus hunting along a forest edge is shown in Fig. 13. When E. fuscus is in cruising flight, the interpulse intervals are stereotyped (Surlykke and Moss, 2000). In contrast, a stable pulse repetition pattern, followed by short breaks and another group of signals at a stable, higher repetition pattern, clearly reveals active hunting, similar to the stable IPI groups described in our lab recordings of insect capture.

We took sound recordings of E. fuscus and P. pipistrellus hunting in close proximity of neighbors, and these sound sequences suggest active vocal control of the emitted dominant frequency to facilitate the identification and perceptual segregation of echoes from self-produced cries and those from neighbors. In the case of E. fuscus, a combination of the directionality of sonar emissions (Hartley and Suthers, 1989) and of the microphones indicates that bats recorded simultaneously were close in range. In the case of P. pipistrellus’s echolocation behavior, the linear array of micro-
phones allowed for determination of the distance to the bats, and the directionality of the microphones, combined with the voice notes on the tape, ensured that the recordings chosen for analysis were from bats in front of the array. Although flight angle and speed were not measured, Doppler shifts introduced by the bat’s flight velocity would be less than 0.7 kHz at 25 kHz, assuming a flight velocity of 10 m/s. It is noteworthy that both species show evidence of frequency shifts of several kHz when two bats are flying close together. Thus, these frequency shifts are considerably larger than those that could be explained by Doppler shifts introduced by the bat’s flight velocity (see Surlykke and Moss, 2000), and suggest a jamming avoidance response. Such frequency shifts could help segregate information relevant to different individuals (Fig. 14).

Finally, field recordings show that some bat species emit signals with systematic changes in frequency from emission to emission. Figure 15 shows a 500-ms recording of an Emballonurid bat, Cormura brevirostris, which emits triplets of multiharmonic sounds in the search phase. The dominant harmonic of the first signal is at 26 kHz, the second at 29 kHz, and the third at 33 kHz. This may be the bat’s solution in the frequency domain to recognizing its own signals, and thus echoes, in areas with very high bat density and much acoustic activity in the same frequency range.

VII. DISCUSSION

Echolocating bats probe the environment with sonar vocalizations and use information contained in the returning echoes to maneuver through the environment and intercept small flying insect prey in the dark. These extraordinary behaviors lead one to speculate that the echolocating bat’s representation of the environment obtained through sonar compares well with that of human spatial vision. Spatial vision does not arise from passive reception of retinal images but instead builds upon transformational rules that support interpretation of the visual scene (Hoffman, 1998). Similarly, auditory scene analysis utilizes organizational rules that permit the perceptual grouping and segregation of sound sources in the environment (Bregman, 1990). In the case of the echolocating bat, its vocal control over the acoustic environment...
allows the motor system to play directly into its analysis of the auditory scene. We hypothesize that active perceptual processes and active vocal control over echo information obtained from the environment operate in concert, permitting an elaborate representation of a dynamic auditory scene. Here, we have presented examples from perceptual and motor studies that illustrate the active processes that contribute to scene analysis by echolocation.

Over the past several decades, echolocation research has yielded valuable data on sonar performance and acoustic information processing by many different species of bats (see for example, Griffin, 1958; Busnel and Fish, 1980; Moss and Schnitzler, 1995); however, very little attention has been directed toward understanding how echo information may be interpreted by the bat’s auditory system to build a perceptual representation of the auditory scene. The aims of this paper are to highlight the question of higher-level perception in sonar, to provide a conceptual framework for studying this problem, and to report on data that begin to address questions of scene analysis by echolocation in bats.

In this article, we have emphasized temporal processing of sonar signals for spatial analysis of the auditory scene. Certainly many distinct acoustic features contribute to the bat’s perception of the world through sound (e.g., direction, spectrum, amplitude), but we have focused largely on the processing of echo delay for target distance estimation, because this signal parameter plays a central role in the bat’s perceptually guided behavior through three-dimensional space. In particular, we have chosen to present data from echo discrimination experiments, laboratory insect capture studies, and field recordings, to illustrate the role of temporal patterning in perception and action for scene analysis by echolocation in bats.

FIG. 14. (A) Signals from two *E. fuscus* recorded simultaneously. As long as both bats are recorded at high intensity while they are presumably close together (from 0 to around 2.5 s), they emit signals where the main frequency differs by 4–5 kHz. When one of the bats starts veering away and the other closes in on the microphone (as revealed by the reduced intensity of the low-frequency bat, and increased intensity of the high-frequency bat) both bats adjust the frequency of their emissions to approximately the same value, approximately 26 kHz. (B) Same pattern for *Pipistrellus pipistrellus*. In this case an array of three microphones allowed for discrimination between the emissions of the two bats (filled squares and filled circles, respectively) and for determinations of approximate distance between the bats (line curve). In this case the emitted frequencies of both bats are scattered around 48 and 50 kHz, but when the bats start closing in on one another (around 1600 ms) they adjust their frequencies to be separated by approximately 2.5 kHz.

FIG. 15. In the search phase the tropical bat, *Cormura brevirostris* (Wagner’s sacwinged bats, a.k.a. the “do-re-mi-bat”) emits triplets of sonar sounds with main frequency increasing from 26 kHz to 29 kHz and ending at 33 kHz. This very distinct frequency pattern may help the bat to distinguish between its own emissions and those of the many other bats operating in the same frequency range. Background signals from other bats also appear in this figure. Single bat’s signals are identified by the numbers 1, 2, 3. Another advantage of this frequency shift might be to reduce problems with overlap between emission and echo (see the text).
of acoustic parameters will be the focus of future work.

The perceptual experiments presented in this paper explicitly examined the fundamental issue of processing spatial information from a dynamic acoustic environment. Changes in the auditory scene, produced by movement of sonar targets and by the bat’s own flight path, are sampled at a low duty cycle by species that use frequency-modulated (FM) echolocation cries, resulting in brief acoustic snapshots of the world. The bat’s auditory system must assemble information over time to build a representation of the dynamic auditory scene. While others (e.g., Hartley and Suthers, 1992; Roverud, 1993, 1994; Roverud and Rabitoy, 1994) have shown that decreasing echo delay can evoke distance-appropriate increases in sonar signal production in trained echolocating bats, these earlier behavioral experiments did not directly study the bat’s perceptual discrimination of echo delay patterns that increase or decrease over time. Here, we have presented psychophysical data demonstrating that the perceptual system of the big brown bat, *Eptesicus fuscus*, supports the integration of acoustic information over time, permitting the discrimination of increasing or decreasing echo delay patterns. Our future experiments will attempt to determine if the bat hears out streams of echoes from an echo-cluttered background that would allow it to segregate and track moving targets against background.

Here, we have also presented data on active vocal control by echolocating bats, operating both in the laboratory and in the field. Unlike the majority of hearing animals that perceive their surroundings by listening passively to acoustic signals generated in the environment, the echolocating bat actively transmits and controls the acoustic information it obtains about the auditory scene. The bat’s active control over the acoustic information it receives from the environment prevents us from conducting the classical psychophysical experiments on auditory scene analysis with bats, like those employed with passively listening subjects. However, the active selection of sonar signal parameters under changing task conditions presents a valuable opportunity to attack the problem of scene analysis from the motor side.

We propose that the bat’s vocal production patterns during insect capture are consistent with the notion that it uses information over successive echoes to build a representation of the world that ultimately guides its behavior. If the bat were to process each echo as a discrete event and modify its behavior based on the spatial information contained in single echoes, insect capture in a dynamic environment would surely fail. Assuming a behavioral response latency of 100–150 ms, the position of the target with respect to the bat would have changed by the time the bat localized and responded to it, and the bat’s motor adjustments would lag behind the movement of the insect. Furthermore, if the bat were to process echo information from each sound before making appropriate motor adjustments, the intervals between sounds should be greater than the shortest behavioral response latency of approximately 100 ms. However, intervals between the bat’s echolocation sounds are typically less than 100 ms, except during the early search phase of insect pursuit or during cruising (Griffin, 1958), suggesting that the bat emits and processes signals in groups with shorter interpulse intervals. In this context, it is noteworthy that the flight path data often show the bat predicting its point of interception with a moving target (example shown in Fig. 6).

Assuming that the general principles of auditory scene analysis apply to the echolocating bat, we hypothesize that the task of organizing acoustic information from a dynamic environment is simplified if the bat probes sonar targets with vocalizations occurring with a relatively stable interval, which we refer to as stable IPI sound groups. In human listening experiments, the interval between successive sounds influences the tendency to hear out auditory streams (Bregman, 1990), and in the bat, the interval between signals in a group may influence the acoustic streams it segregates from objects at different distances and directions.

What other benefits might the bat derive from the production of stable IPI sound groups? Information about target position may be encoded more reliably in the central nervous system under conditions in which the sonar production rate is stable. The echolocating bat’s auditory system contains a population of neurons that shows facilitated and delay-tuned responses to pairs of sounds, simulating sonar emissions and echoes, a response property believed to be related to the encoding of target range (Suga and O’Neill, 1979). Echo-delay-tuned neurons, found in the auditory brainstem (Mittman and Wenstrup, 1995; Dear and Suga, 1995; Valentine and Moss, 1997), thalamus (Olson and Musil, 1992), and cortex (O’Neill and Suga, 1982; Sullivan, 1982; Wong and Shannon, 1988; Dear et al., 1993b), show very weak responses to single FM sounds. However, these neurons respond vigorously to pairs of FM sounds (a simulated pulse and a weaker echo), separated by a delay. Typically, echo-delay-tuned neurons show facilitated responses to simulated pulse–echo pairs over a delay range of several milliseconds. The pulse–echo interval to which an echo-delay-tuned neuron shows the largest response is referred to as the best delay. For example, a neuron may respond maximally to a pulse–echo interval of 9 ms, which corresponds to a target range of 1.5 m. Pulse–echo pairs with intervals of 7 and 11 ms (simulating ranges between 1.1 and 1.9 m) may also evoke a response, but at a lower rate.

Delay-tuned neurons respond phasically to acoustic stimuli simulating FM sonar signals, and typically fire no more than one or two action potentials for each pulse–echo stimulus presentation. In addition, a population of echo-delay-tuned neurons exhibits shifts in the best delay with the stimulus presentation rate (O’Neill and Suga, 1982; Tanaka and Wong, 1993; Wong et al., 1992). This shift in best delay may be several milliseconds, corresponding to a shift in best range of several centimeters. These two observations together, i.e., that the delay-tuned neurons fire one or two action potentials per stimulus presentation at the best delay, and that the best delay may change with stimulus repetition rate, suggest that sonar strobe groups may serve to stabilize the neural representation of the distance of objects in an auditory scene. For example, the repeated production of sounds with stable IPI for a sonar target at a given distance, e.g., 1.5 m from the bat, would consistently stimulate neurons responsive to the combination of a particular echo delay (e.g., 9 ms)
and repetition rate (e.g., 60 Hz) over several successive signals, which would then increase reliability of target distance estimation.

The stable IPI groups are interrupted by longer gaps, and thus the signals in each group occur in clusters, which may impact both perception and motor planning. With regard to perception, the grouping of signals would serve to increase the information carried by echoes in discrete blocks of time. It may be that the bat’s sonar receiver can extract finer detail about the surroundings with the higher density of echoes that result from these groups. In this context, it is noteworthy that the bat increases its relative production of stable IPI groups shortly before target contact, and again just after capture, when a detailed assessment of the surroundings may be particularly important for planning the final attack and reorienting after tucking the head in the tail membrane to take the prey (Fig. 10). The stable IPI groups were also recorded from actively hunting bats but not from animals that were commuting from the roost to hunting grounds. All of this suggests to us that the stable IPI groups may be important for spatially segregating (streaming) the target (insect prey) and background. Moreover, we speculate that the spatial resolution of the auditory scene varies inversely with the interpulse intervals within stable IPI groups. Thus, shorter interpulse intervals, characteristic of the final stages of insect capture, may permit segregation of acoustic streams from closely spaced echo-producing auditory objects.

Naturally, the clustering of signals in groups, which results in greater echo density, also creates time periods when the echo density is reduced. We speculate that the bat may use the IPI gap time to update motor programs, using the acoustic information carried by echoes in the IPI groups. The median duration of the stable IPI groups was about 43 ms for the open-space condition and 38 ms for the cluttered condition, which might approximate the time over which information is assembled for motor program updates.

It is well established that the echolocating bat’s respiration cycle is correlated with its wingbeat cycle, and the intervals between sonar vocalizations occur during inspiration. The coupling of wingbeat, respiration, and vocal production contributes to the temporal patterning of sonar signals (Schnitzler and Henson, 1980; Suthers et al., 1972; Wong and Waters, 2001), and indeed this coupling can be used, in part, to explain the sonar signal groups we report here. However, some bat species do not produce sonar signal groups, but instead show a relatively regular and continuous decrease in interpulse intervals (see, e.g., Schnitzler and Henson, 1980). The fact that some species produce stable IPI groups and others do not suggests differences in the coordination of vocal production and respiration across bat species, even among those that use FM signals to hunt insects in temperate climates (e.g., compare Eptesicus fuscus and Myotis lucifugus, Schnitzler and Henson, 1980). Such differences in the temporal patterning of vocal production naturally results in differences in echo patterning received by the sonar systems of different species, since the vocal production pattern directly impacts the information available to the sonar receiver. We do not know why differences exist in the temporal patterning of sonar signals across species; however, we propose that the stable IPI groups produced by Eptesicus fuscus play into the analysis of the auditory scene in this species.

In our laboratory studies of insect capture behavior, a clear distinction in vocal production patterns across task conditions appeared in the duration of the terminal buzz. When the bats captured targets in the presence of clutter, they shortened the terminal buzz, as compared with the buzz produced under open uncluttered space conditions ($F = 6.22, p < 0.01$). This result is not surprising, when one considers that the bat typically waits to produce its next signal until it has received and processed echoes of interest. In other words, the bat avoids producing signals before all the relevant echoes have returned. In the terminal buzz the interval between signals is so short that there often is not time for all relevant echoes (from target and clutter) to return before the subsequent sound is produced. Thus, in the terminal buzz, the bat likely receives echoes from the clutter following one sonar signal after producing the next signal, and the bat reduces the time when it experiences the mixing of echoes from different vocalizations under clutter conditions by shortening the terminal buzz.

The onset of the buzz differed between stationary (open-space and clutter conditions) and moving target (open space) trials. When the bat pursued moving prey, it typically followed the insect from behind, in the direction of the target’s motion. Thus, the insect was moving away from the bat during the final stage of capture, and the bat’s approach direction presumably resulted in the extended buzz duration in the moving target trials. The buzz onset times relative to target capture were very similar in the open-space stationary and the three different clutter conditions. The consistency of buzz onset time in clutter relative to stationary prey capture suggests that it reflects a motor planning process for target interception, rather than serving as an indicator of target detection or localization. Measures of the distance between the bat and target when it initiated the buzz show that the bat was closer to the target when it entered the final capture sequence in clutter than in open space. Moreover, the closer the target was positioned relative to the clutter, the shorter was its distance to the prey when it began the terminal buzz. This is consistent with the observation that buzzes recorded in the field are shorter than in the lab (Surlykke and Moss, 2000), since natural prey is unpredictable, and motor planning must be delayed until the bat is closer to the point of capture. Given the potential for mixing signals and echoes in the terminal buzz, one might then ask why the bat produces signals at such a high repetition rate. Furthermore, the bat has little time to react to any new information carried by echoes from the terminal buzz signals. However, the terminal buzz does allow the bat to sample information at a very high and stable repetition rate, providing details about the auditory scene that may be enhanced and sharpened, particularly if the information is assembled over time. This is conveyed in Fig. 2(B) by the clear ridges representing changing target distances during the buzz phase of insect pursuit. While the bat has little time to react to information about the target in the terminal buzz, it may use echoes from the terminal buzz to represent spatial information about the background that it still must negotiate following insect capture. The signal fre-
quencies in the terminal buzz drop well below 20 kHz (Sur- 
lykke and Moss, 2000), and this drop in sound frequency 
serves to reduce the directionality of the bat’s sonar signals. 
Using lower-frequency, more omnidirectional signals may 
thus help the bat in sampling background information that 
could ultimately influence its flight path following target cap-
ture. Thus, if one thinks of the buzz as a sampling unit, rather 
than many closely spaced, discrete elements, one can begin 
to appreciate the rich information it may carry for the analy-
sis of the auditory scene.

Field data from several different species highlight the 
perceptual requirements and vocal-motor strategies used by 
echolocating bats to analyze the auditory scene. In particular, 
we have presented sound recordings from Noctilio that illus-
strate the complicated acoustic environment the bat encoun-
ters, raising questions about how the bat sorts a cacophony of 
signals and echoes in an environment where many animals 
are hunting together. While hunting in less crowded condi-
tions, E. fuscus and P. pippistrellus sometimes encounter 
conspicuousness in close proximity, and we presented examples 
here of adjustments in the frequency of sonar emissions pro-
duced by these two species in response to the signals of 
neighbors. Such adjustments in signal spectrum and sweep 
rate may allow an individual bat to separate out echoes from 
its own signals from echoes of signals produced by conspe-
cifics in close proximity. Finally, we presented the signals 
produced by a bat that samples the environment with multi-
harmonic signals with changing fundamental frequencies (C. 
brevirostris, dubbed the “do-re-mi bat”). We have no data to 
speak directly to this species’ perception of the world by 
sonar, but we note that vocal production patterns of C. brevi-
rostris resemble frequency hopping used in wireless commu-
nication systems. Frequency hopping minimizes interference 
from other sources by carving the transmission band into 
separate frequency channels that are active for short intervals 
(about 100 milliseconds) before shifting to a neighboring 
frequency. This allows wireless communication more access 
points in the same area, and we speculate that C. brevirostris 
uses a similar strategy to minimize interference from neigh-
boring echolocating bats.

In conclusion, we have introduced here a conceptual 
framework for the study of auditory scene analysis by 
echolocation in bats. We assert that the principles of auditory 
scene analysis, which have been identified in humans (Breg-
man, 1990), can be applied broadly to understand the percept-
ual organization of sound in other animals, including the 
echolocating bat. In particular, we postulate that the bat’s 
perceptual system organizes acoustic information from a 
complex and dynamic environment into echo streams, allow-
ing it to track spatially distributed auditory objects (sonar 
targets) as it flies. Our hypothesis emphasizes the importance 
of the echolocating bat’s vocal production patterns to its 
analysis of the auditory scene, as motor behaviors and per-
ception operate in concert in the bat’s active sensing system. 
It is our goal that the conceptual framework presented here 
will stimulate new research avenues for unraveling the de-
tails of auditory scene analysis in animal systems.

ACKNOWLEDGMENTS

We gratefully acknowledge the following agencies, 
foundations, and institutions that have supported this work: 
NIMH (No. R01-MH56366), NSF (No. IBN-925825), 
Whitehall Foundation (No. S97-20) and the Institute for 
Advanced Study, Berlin (to C.F.M.), Danish National Research 
Foundation and the Danish Natural Science Research Coun-
tcil (to A.S.). In addition, we thank the following individuals 
for their valued assistance: Kari Bohn for data collection, 
analysis, and figure preparation, Hannah Gilkenson, Aaron 
Schurger, and Willard Wilson for data collection, analysis, 
and processing of the laboratory insect capture experiments, 
Maryellen Morris for running the psychophysical experi-
ments, and Amy Kryjak for data analysis and overseeing the 
complicated management of the many projects reported here. 
Carola Eschenbach and two anonymous reviewers provided 
critical comments on an earlier version of this manuscript. 
Discussions with Hans-Ulrich Schnitzler contributed to ideas 
presented here. We also thank Elisabeth Kalko and Marianne 
Jensen for their valued assistance in the field.

Barratt, E. M., Deaville, R., Burland, T. M., Bruford, M. W., Jones, G., 


Braaten, R. F., and Hulse, S. H. (1993). “Perceptual organization of tempo-
ral patterns in European starlings (Sturnus vulgaris),” Percept. Psych-
ophys. 54, 567–578.

MA).

Bregman, A. S., and Campbell, J. (1971). “Primary auditory stream segre-
gation and perception of order in rapid sequences of tones,” J. Exp. Psy-
chol. 89, 244–249.

York).


for representation of acoustic scenes in auditory cortex of the big brown 

Dear, S. P., Fritz, J., Haresign, T., Ferragamo, M., and Simmons, J. A. 
(1993b). “Tonotopic and functional organization in the auditory cortex of 

the big brown bat,” J. Neurophysiol. 73, 1084–1100.

ratus),” Hear. Res. 120, 69–76.

tion in goldfish (Carassius auratus),” J. Assoc. for Res. in Otologynol., 
online publication.


Griffin, D. R. (1953). “Bat sounds under natural conditions, with evidence 

Griffin, D. (1958). Listening in the Dark (Yale University Press, New Ha-
vem).  


Hartley, D. J. (1992). “Stabilization of perceived echo amplitudes in echolo-
cating bats. II. The acoustic behavior of the big brown bat, Eptesicus fuscus, 


