Composition of Biosonar Images for Target Recognition by Echolocating Bats

JAMES A. SIMMONS,1 PRESTOR A. SAILLANT,1 JANINE M. WOTTON,1 TIM HARESIGN,1 MICHAEL J. FERRAGAMO1 AND CYNTHIA F. MOSS2

1 Brown University and 2 Harvard University

(Received and accepted 27 March 1995)

Abstract—Echolocating bats can recognize flying insects as sonar targets in a variety of different acoustic situations ranging from open spaces to dense clutter. Target classification must depend on perceiving images whose dimensions can tolerate intrusion of additional echoes from other objects, even echoes arriving at about the same time as those from the insect, without disrupting image organization. The big brown bat, Eptesicus fuscus, broadcasts FM sonar sounds in the 15–100 kHz band and perceives the arrival-time of echoes with an accuracy of 10–15 ns and a two-point resolution of 2 μs, which suggests that perception of fine detail on the dimension of echo delay or target range is the basis for reconstructing complex acoustic scenes and recognizing targets that are embedded in these scenes. The directionality of the bat's sonar sound is very broad, making it impossible to isolate echoes from individual targets merely by aiming the head and ears at one object instead of another. Consequently, segregation of targets must depend on isolating their echoes as discrete events along the axis of delay. That is, the bat's images must correspond to impulse responses of target scenes. However, the bat's sonar broadcasts are several milliseconds long, and the integration time of echo reception is about 350 μs, so perception of separate delays for multiple echoes only a few microseconds apart requires deconvolution of spectrally-complex echoes that overlap and interfere with each other within the 350-μs integration time. The bat's auditory system encodes the FM sweeps of transmissions and echoes as half-wave-rectified, magnitude-un squared spectrograms, and then registers the time that elapses between each frequency in the broadcast and the echo, effectively correlating the spectrograms. The interference patterns generated by overlap of multiple echoes are then used to modify these delay estimates by adding fine details of the delay structure of echoes. This is equivalent to transformation of the spectrograms into the time domain, or deconvolution of echo spectra by spectrogram correlation and transformation (SCAT). However, while deconvolution overcomes integration time, the bat's receiving antennas reverberate for about 100 μs, smearing the echoes upon arrival. The bat overcomes this problem by receiving echoes from different directions than the transmitted sound, which radiates from the mouth. The broad range of antenna reverberations common to the emission and echoes thus cancel out, leaving only narrow elevation-dependent differences, which in fact appear in the bat's images. The SCAT algorithms successfully recreate images comparable to those perceived by the bat and provide for classification of targets from their glint structure in different situations.

Keywords—Biosonar, Echolocating bats, Target-ranging, Target recognition (ATR), Localization, Classification, SCAT algorithms.

1. INTRODUCTION

A radar or sonar system broadcasts signals into the environment to excite objects into returning echoes. The reflected signals reveal the location and nature of the objects as parts of a scene to be depicted in images. Effective localization and recognition of targets depends on faithful rendering of the features of each object in the corresponding component of the image, which entails extracting the information associated with each object from echoes and routing that information to the appropriate part of the image. Echoes, however, undergo a variety of transformations in amplitude, delay, and spectrum from aspects of the environment not directly related to the characteristics of the individual objects themselves,

Acknowledgements: Work supported by ONR, NIMH, NSF, NIH, DRF, Raytheon, and McDonnell-Pew

Requests for reprints should be sent to James A. Simmons, Department of Neuroscience, Box 1953, Brown University, Providence, RI 02912, USA.
which can distort or disrupt an object's representation. These effects compete with effective localization and recognition of targets.

For example, the spectrum of echoes is affected by the shape of a sonar or radar target and is commonly used to represent shape in the images. In fact, target recognition frequently is described entirely in terms of classifying echo spectra. However, the echo spectrum is also affected by the target's distance (frequency-dependent propagation losses), by the target's motion (Doppler shifts and changes in reflectivity associated with aspect angle), by the target's location in the transmitting and receiving beams (frequency-dependent beam-widths), and by interference among overlapping echoes reflected by concomitant clutter (frequency-dependent reinforcement and cancellation of signals). How can any particular target be located in a complex multiple-target situation with so many converging influences on echoes? Furthermore, how can the shape of any one target, and even its existence apart from clutter, be registered in the face of these competing influences on the echo spectrum? In fact, because the target's aspect angle profoundly affects the spectrum of echoes, targets in effect have aspect-dependent shapes, so how can an object be identified at all?

Observations of animals that perceive their surroundings by sonar reveal that high levels of sonar performance routinely can occur in very complex situations where multiple competing effects on echoes at first glance might seem to overwhelm the contributions of individual targets to echoes. Our paper describes target imaging by one species of echolocating bat to illustrate the technological potential of biologically-inspired approaches to image formation in sonar or radar systems. To make this account work, we have to bring together material on a variety of aspects of echolocation because the bat's sonar exhibits an unusual degree of systems integration concealed by dispersal of important features across several stages of the imaging process.

2. BIOSONAR SYSTEMS

2.1. Echolocating Bats and Cetaceans

Echolocating bats (Griffin, 1958; Neuweiler, 1990; see reviews in Popper & Fay, 1995) and cetaceans (Au, 1993) use a biological sonar system to perceive targets in complex acoustic environments in “real time”. Sonar-guided behaviors include tracking and identification of prey during interception, recognition of different target types in clutter and reverberation, and control of navigation through obstacle-strewn surroundings. Judging from the behavior of bats and cetaceans, the images derived from successive broadcasts are combined into an evolving real-time perceptual reconstruction of the locations and identifying features of numerous objects—virtually a sonar “motion-picture” of the immediate environment. Critical aspects of the performance of biosonar systems are especially well-documented in bats due to greater success in overcoming practical impediments to behavioral observations at night than to observations underwater. These observations reveal that echolocating bats are capable of locating and identifying sonar targets even when they are imbedded in very complex surroundings where the prevailing conditions foster the generation of competing effects on echo features.

2.2. Target Localization and Recognition by Bats

With the guidance of echolocation, bats routinely accomplish the very demanding tasks of detecting, tracking, identifying and intercepting flying insects, and also avoiding obstacles to flight, such as vegetation (Griffin, 1958; Simmons, 1989; Neuweiler, 1990; Schnitzler & Kalko, in press; Simmons et al., 1995). In extreme cases, insectivorous bats can pursue targets into highly cluttered surroundings—for example, when chasing an insect that flies through vegetation. Figure 1 shows a little brown bat catching a moth in a cluster of branches after having detected it and initiated pursuit while the insect is in the open (Simmons et al., 1995). The numbered images of the bat (100 ms apart at positions 1–8) are connected to the corresponding images of the moth by dashed lines. At first, the bat is approaching the moth (position 1), which flies along an erratic path while the bat comes even closer (positions 2–4). The bat seizes the moth (position 5) and then flies out of the cluster of branches with the moth in its mouth (positions 6–8). To follow the moth through the clutter and catch it while it is deeply embedded in the surrounding vegetation is a formidable task. In Figure 1, the distance to the moth is similar to the distances to the cluster, and the targets are of comparable reflective strength, so the echoes from the vegetation bracket those from the insect, making it difficult to see how the bat’s performance is possible to achieve.

The behavior of bats in realistic situations such as shown in Figure 1 demonstrates simultaneous localization and recognition of several objects present together in a complex acoustic scene. (The bat has to capture the insect while avoiding serious collisions with any of the surrounding objects). Using more controlled obstacle-avoidance tasks, numerous studies of bats flying through arrays of vertically-stretched wires demonstrate that accurate localization of several objects, simultaneously present at about the same distance, is routinely possible for
many different kinds of bats (Griffin, 1958; Schnitzler & Henson, 1980). In more controlled interception tasks conducted to assess target recognition, echolocating bats can intercept airborne mealworms while rejecting airborne spheres or disks that have about the same overall target strength and are presented either sequentially or simultaneously (Griffin et al., 1965; Simmons, 1989; Simmons & Chen, 1989).
Moreover, three-dimensional motion-analysis of video recordings shows that bats can select and capture one of several identical targets (each one a mealworm on a thread) even when all are constantly located equidistant from the bat so that their echoes arrive at about the same time and constitute mutual clutter (Simmons et al., 1995). All of these experiments reveal that echolocating bats can locate targets while classifying them, even when several objects are present simultaneously at about the same distance. Furthermore, not only can the bat untangle the overlapping echoes to perceive the objects at their correct locations, but the bat’s correct selection of the insect shows that its biosonar images must incorporate a dimension representing the target’s fine structure apart from overall cross-section or reflective strength (Simmons & Chen, 1989; Kober & Schnitzler, 1990; Moss & Zegaeski, 1994).

2.3. Experimental Assessment of Image Content

The targets in Figure 1 are arrayed at different distances from the bat, so their echoes arrive at different times (5.8 ms of delay per meter of range). To isolate one object from other objects requires isolating its echoes from other echoes, even if the echoes from different objects overlap to some extent. In Figure 1, the bat seems likely to receive echoes from the moth in close temporal proximity to echoes from the branches. This raises the question of how good the bat is at tasks requiring segregation of closely-spaced echoes to perceive just one particular object’s reflection. A series of tests have been carried out with the big brown bat, Eptesicus fuscus, to answer this question, and the results are surprising. Eptesicus is an insectivorous species of bat (see Kurtz & Baker, 1990) that uses wideband ultrasonic sonar sounds to perceive much of its surroundings (see details below), and it has been used extensively for studies of image content in echolocation.

Figure 2 illustrates three perceptual capabilities of the big brown bat that are relevant for understanding how this species recognizes targets. In Figure 2A, the bat broadcasts a sonar sound towards a target located at a particular range. In the tests described here, the targets are located at distances of roughly 50 cm, for echo delays of about 3 ms. The most basic capability is the bat’s accuracy for determining echo delay. How small a change in the arrival-time of echoes can be perceived from one echo to the next by the bat? By jittering the echoes over small increments of delay (\(\delta t\) in Figure 2B1) and training the bat to detect jittering echoes (+ stimulus) as opposed to stationary echoes (−stimulus), the bat’s image of an echo along the scale of delay can be traced (shaded peak in Figure 2B2). In quiet ambient laboratory conditions, Eptesicus perceives echo-delay changes as small as 10–15 ns (yes, nanoseconds). Furthermore, at a known echo signal-to-noise ratio (\(2E/N_0\)) of 36 dB, the bat’s acuity is about 40 ns, which corresponds...
roughly to the limit for ideal-receiver performance (Simmons et al., 1990). This means that the width of the image trace in Figure 2B2 is very sharp indeed. It is thus plausible that the bat could distinguish an insect in the presence of dense clutter (Figure 1) because its images are capable of registering differences in echo delay on a very fine scale.

The second perceptual capability illustrated in Figure 2 is also related to the bat’s performance in Figure 1. How close together do two echoes have to be for one of them to mask the presence of the other? That is, as two nearly-simultaneous echoes become closer together in arrival-time, they eventually will merge into just one sound. What is the delay separation at which merging occurs? It is clear that this is a question about clutter interference, which is another aspect of the bat’s perceptual capabilities. By training the bat to detect an echo at a specific delay, and then adding an additional echo at a series of different delays (δt in Figure 2C1), the time separation where detection of this particular echo fails to occur can be measured. The bat’s task is to determine whether both echoes are present (+stimulus in Figure 2C1) or just one echo (−stimulus). Experiments with Eptesicus reveal that the delay separation required for clutter interference to occur is about 350 μs (Simmons et al., 1989). This means that the clutter-interference zone (shaded area marked integration time in Figure 2C2) is about ±350 μs wide. If two targets are closer together than 6 cm, their echoes will merge together into just one sound, and it would seem that the presence of one object is obscured by the other. However, in Figure 1, the bat successfully pursues and captures the moth in a situation where the moth is virtually always much closer than 6 cm in range to some part of the vegetation. How can the bat perceive the moth if its echoes have merged with echoes from some of the branches at each stage of the flight?

The third perceptual capability illustrated in Figure 2 is resolution of closely-spaced echoes. This experiment probes inside the clutter-interference zone (or integration time in Figure 2C2) to determine if the bat can tell whether there are two reflected signals present even after the sounds have merged together in an echo-detection task. At delay separations shorter than 350 μs, one echo indeed masks the other, but the bat’s task was simply to detect one sound in the presence of the other. Might the bat detect that there is only one sound and yet still perceive that there are two distinct arrival times contained in that sound? While this possibility might seem contradictory, the answer is not self-evident because the clutter-interference experiment (Figure 2C2) measures something quite different from resolution itself. In this third experiment, the stimuli consisted of jittering echoes that alternated not only between two values of delay (δt in Figure 2D1) but also between two echo structures—one echo or two echoes. The bat’s task is to detect the jittering echoes (+stimulus) as opposed to stationary echoes (−stimulus), while the delay separation of the two echoes presented together in alternation with one echo presented alone is used to block perception of the jitter (δt). This method traces the bat’s image of a double echo (shaded area in Figure 2D2), and the critical issue is whether there is just one peak in this image or two peaks. Experiments reveal that Eptesicus can perceive an image containing two separate peaks for double echoes at different delays for delay separations as small as 2–10 μs (Saillant et al., 1993; Simmons, 1993). This shows that the merging of echoes in a clutter-interference task (Figure 2C1–2) only evaluates whether separate sounds are detected by the bat, not whether separate delays are perceived. The bat’s ability to perceive two echoes as having different delays even when they have merged into just one sound is vital for explaining the performance shown in Figure 1. As the moth moves through the vegetation, the bat can still perceive that there are reflections arriving at its ears from the moth when the echoes returned by the branches effectively have merged with the echoes from the moth. Even though the bat in Figure 1 really receives one long stream of echoes each time it broadcasts a sound, it can perceive subdivisions of this stream with sufficient resolution to follow the moth in the presence of the branches. This paper now examines the use of echo-delay images of this kind for recognizing targets.

2.4. Biosonar Systems as Models for Echo-Processing Strategies

The high performance achieved by biosonar systems suggests that they could offer insights into novel signal-processing techniques directed at achieving efficient target localization as well as recognition across different situations (i.e., in the open as well as in clutter). The strategies used by insectivorous bats to cope with multiple influences on each echo parameter ought to be especially instructive because these kinds of problems are well-defined acoustically, as in the example of converging influences on the echo spectrum given above, and the bat’s solution ought to be discernible if the proper observations are made. The animal may even have found an approach that does not require a “solution” to this “problem”. To the bat, the seemingly separate functions of localization and classification of targets may prove to be part of the same process, without any real conflict at all. Here, we examine how the big brown bat processes the arrival times of sonar echoes to represent the information they contain about the target’s shape at different ranges and aspect angles in
the presence of competing influences from the propagation path and spreading losses, from the presence of clutter, and from the bat's external ears.

3. CHARACTERISTICS OF BIOSONAR TRANSMISSIONS

3.1. Echolocation by the Big Brown Bat

Figure 3 schematically illustrates the process of echolocation by *Eptesicus*. The bat broadcasts a sonar sound (arrow 1 in Figure 3), which is picked up through the bat's ears at the moment of transmission (arrow 2) as well as radiated into the environment (arrow 3) to impinge on nearby objects (the target, a moth) to form an echo (arrow 4). The reflected echo returns to the bat's ears (arrow 5) and is picked up at a slightly later time according to the target's range. After the broadcast sound and the echo stimulate the ears, the bat's auditory system serves as the sonar receiver, processing the directly-heard transmission and the delayed echo as a pair of signals to represent the target as an image (arrow 6) depicting range and other distinguishing features. For economy of explanation, Figure 3 shows only the insect, although the insect can be imbedded in vegetation, as Figure 1 demonstrates. *Eptesicus* is one example of a species of bat that typically hunts for prey in relatively open areas but is nevertheless capable of flying into more complex surroundings to complete its interception (Neuweiler, 1990; Fenton, 1995). This is an important capability biologically because a significant proportion of night-flying insects (e.g., moths, lacewings) are themselves equipped with high-frequency hearing and can detect the sonar broadcasts of an approaching bat in time to take evasive action, which often leads to flight into surrounding vegetation.

3.2. FM Biosonar Signals

Figure 4 shows spectrograms of the frequency-modulated (FM) echolocation sounds emitted by *Eptesicus*. To obtain accurate recordings of these transmitted waveforms, the signals actually impinging on the target were recorded by training bats to capture mealworms suspended on a fine filament (Gaillant et al., in prep.). On some flights, the mealworm was replaced with a small microphone to pick up the signals sent out by the approaching bat. The biosonar sounds in Figure 4 contain several downward-sweeping harmonics (FM₁, FM₂, FM₃, and sometimes even FM₄) that together cover the frequency band from 15–25 kHz up to about

---

**FIGURE 3.** Diagram of the process of echolocation by the big brown bat. The bat broadcasts a sonar sound that both impinges on the bat's ears (arrow 1) and travels out to impinge on the target (transmission, arrow 3). The target (a moth) contains two prominent glints (A and B) that each return a replica of the broadcast to the bat's ears (echo, arrow 4). The transmission is a broadband FM signal with a sharp autocorrelation function (ACR) that is well-suited to target ranging, and the target's glint structure is conveyed by the target impulse response (with peaks A and B corresponding to the glints). However, upon reception by the external ears (the bat's antennas), both the transmitted sound (arrow 2) and the returning echo (arrow 5) reverberate within the external ear to create a series of peaks (numbered 1–4) in the external-ear impulse response for the transmission and a separate series of peaks for each glint (peaks A₁–A₄ and B₁–B₄) in the echo. The severe arrival-time ambiguity created by these multiple peaks is removed by the directional dependence of the external-ear impulse response, which shifts the timing of one of the reverberation peaks (peak 3, marked *) relative to the others (see also Figure 10). Using the external-ear impulse response for the transmitted sound as a kind of "mask" for simplifying the target + external-ear impulse response for the echo, a SCAT range image is formed that depicts the target's two glints (peak A₁ and peak B₁) at their correct distances, with an additional "glint" (peak A₃, marked *) contributed by the elevation of the target in relation to the external-ear. The other peaks (peaks A₂, A₄, B₂, B₃, B₄) are largely suppressed by the summarizing action of the SCAT algorithms.
100 kHz in each signal. The bat shortens its sounds (from about 8 ms in Figure 4a to 0.5 ms in Figure 4f, g) as it approaches a target to keep the duration slightly shorter than the two-way travel time of echoes. As shown in Figure 2A, the path-length from the bat’s mouth out to the insect and back to its ears becomes filled with the sound. When the broadcast signals shorten below about 1 ms, the harmonics also slide downward in frequency (Figure 4d–g), but the signals retain their broad bandwidth due to the inclusion of more of the higher-harmonic sweeps (FM3, FM4). The two most important features of these signals are their broad bandwidth (they contain ultrasonic frequencies from about 15–20 kHz to 100 kHz which span 2–3 octaves) and the fact that this bandwidth is in the baseband (the sounds have a center-frequency of 60 kHz, which actually is smaller than their overall bandwidth of 80 kHz; see Simmons & Stein, 1980; Menne & Hackbarth, 1986). The broad baseband bandwidth imbues the signals with especially good qualities for measuring the arrival time of echoes if the bat can somehow use the information the sounds theoretically contain.

3.3. The Transmitted Beam

The larynx and vocal-tract of Eptesicus generate the biosonar sounds, which are projected out through the open mouth. For the frequencies emitted by Eptesicus (see Figure 4), the mouth acts as an aperture approximately 9 mm in diameter (Hartley & Suthers, 1989), which produces a relatively broad beam. In azimuth, the transmitted beam is pointed straight ahead of the open mouth with a 6-dB
beamwidth of about 40–60° (±20–30°) at representative frequencies of 40, 60 and 80 kHz (Figure 5A), while in elevation the beam is pointed downward by 5–25° (below the eye-nostril plane of the bat's head) with a 6-dB beamwidth of about 50–70° (±25–35°) (Figure 5B). However, the horizontal 6-dB beamwidth is fully 100° (±50°) at 25 kHz, which is even broader (vertical beam data at 25 kHz are not published). In this respect the transmitted beam of *Eptesicus* is different from the narrow broadcast beam used by marine mammals (Au, 1993) and also of bats which broadcast constant-frequency (CF) echolocation sounds (Neuweiler, 1990; Popper & Fay, 1995).

The bat's transmitted signal has a broad enough beam to ensonify much of the hemisphere of space to its front with at least the first harmonic of its sounds, especially at the lower-frequency end (Figure 5). Moreover, a surprisingly large operating zone—a conical volume with an angular width of ±25–35° on an axis pointed to the bat's front—is probed with the full effective bandwidth of the harmonics in the sounds ±3–6 dB. When the bat aims its head at the insect (Figure 1), it ensures that the target is ensonified with the strongest incident sound, but the bat also strongly ensonifies objects located over a wide sector of space around the insect, so vegetation and other objects (branches in Figure 1) are sure to reflect strong echoes, too. The practical consequence is that the bat cannot segregate echoes from several targets located in different directions just by mechanically scanning with its sounds; instead, all objects located to the bat's front and many objects located to the sides will return appreciable echoes of each emission. Moreover, the bandwidth or spectrum of the sound impinging on objects located anywhere within the frontal cone of ±25–35° is similar, so the echoes returning to the ears will in principle be fairly similar spectrally, too. Localizing objects or distinguishing among different objects by their azimuth or elevation must therefore depend on the signal-processing capabilities of the bat's auditory system for segregating overlapping echoes because the zone of ensonification is too large to permit separation of targets just from scanning with the head or from the direct effect of their position in the broadcast beam on the spectrum of echoes.

4. CHARACTERISTICS OF ECHOES

4.1. Echo strength

The transmitted sound of *Eptesicus* has an amplitude of 95–105 dB SPL (p. to p.) at a point 10 cm in front of the bat's mouth (a convenient place to pick up the sounds with a microphone), and spreading losses reduce the strength of the sound impinging on the target by 6 dB per doubling of target range beyond this standard distance. In addition, an appreciable signal, attenuated by only 20–25 dB, radiates backward and stimulates the ears to generate the reference signal for processing of subsequent echoes (see below). Insectivorous bats, like *Eptesicus*, that frequently hunt for flying prey in open spaces, broadcast intense sounds to detect prey at reasonable distances in spite of the relatively high absorption of ultrasonic sound by the atmosphere (Griffin, 1958; Neuweiler, 1990; Fenton, 1993; Schnitzler & Kalko, in press). After being reflected by the target, the echo wavefront returning to the bat also undergoes spreading losses, but their magnitude depends upon the target's geometry. For individual insect-sized targets, which approximate point-sources, the echo expands spherically to add a further 6 dB per doubling of range on the return journey. Thus, for point sources of echoes, spreading losses on both traverses of the path combine to be 12 dB per doubling of range. Many other objects have an extended geometry and their echoes spread out differently, however. For example, long, thin wires used in obstacle-avoidance experiments reflect echoes that expand cylindrically around their longitudinal axis and spherically around their radial axis to yield a further 3 dB of loss per doubling of range on the return journey (Griffin, 1958). For echoes from wires, spreading losses thus total 9 dB per doubling of range. A large, flat surface yields a planar echo wavefront that only adds 6 dB on the entire return journey to losses that have already occurred on the outward journey. For flat targets, spreading losses thus total 6 dB per doubling of range plus 6 dB. More complex, and realistic, extended surfaces, such as dense vegetation or the ground, are composed of numerous discrete reflecting sources whose echoes can combine in complex ways when picked up at different locations. Consequently, the peak amplitudes of echoes from extended surfaces often do not weaken appreciably over a considerable span of distances (Griffin & Buchler, 1978).

Echolocating bats operate chiefly at ultrasonic frequencies where atmospheric absorption of sound adds significantly to spreading losses. Attenuation by the atmosphere is roughly 1 dB/m at 30 kHz, 2 dB/m at 50 kHz and 3 dB/m at 100 kHz (Griffin, 1958, 1971; Suthers, 1970; Pye, 1980; Lawrence & Simmons, 1982). As a result of the cumulative effect of atmospheric attenuation over both the outgoing and return journeys (twice the target's range), echoes from targets at long range are eventually weakened more by absorption of energy than by spreading losses, which makes it a major limitation on the maximum operating range of echolocation. Atmospheric absorption also colors the spectrum of echoes.
progressively as target range increases, thus discriminating against higher frequencies to a greater extent at long ranges than short ranges. At a range of 2 m, for instance, echoes are 10 dB weaker at 100 kHz than they are at 30 kHz due to absorption, while at 50 cm echoes are only about 2 dB weaker at 100 kHz than at 30 kHz. Coloration of the echo spectrum is a significant effect that competes with the contribution of the target's shape to the echo spectrum (see below).

The various sources of attenuation incurred during sound propagation and echo formation determine the maximum operating range for echolocation. *Eptesicus* actually can detect a sphere with a diameter of 4.3 mm as far away as 3 m, and a sphere with a diameter of 19.1 mm as far away as 5 m (Kuck, 1982). At these distances, spreading losses deduct about 58–68 dB from echo strength, atmospheric attenuation deducts about 6–25 dB, and the target's size deducts a further 26–38 dB. These small spheres yield fairly naturalistic measurements of distances of detection for insects because the target strength of flying insects approximates that of spheres in this size range (Simmons & Chen, 1989).

### 4.2. Echo Delay

The arrival time of echoes is retarded by 5.8 ms/m of range in air, and echolocating bats perceive the distance to objects from the delay of their echoes (Simmons, 1973, 1989). As sonar signals, the echolocation sounds of *Eptesicus* are surprisingly well-suited to target ranging. The conventional index of this function is the autocorrelation function (ACR) of the broadcast waveform (Simmons & Stein, 1980; Menne & Hackbarth, 1986). Figure 3 illustrates the sharpness of the ACR for an *Eptesicus* signal as a marker for the time-of-occurrence of echoes. The ACR produced by *Eptesicus* has a single central peak about 14–16 μs wide, which is approximately the reciprocal of the 60-kHz center frequency of the sound as a whole (see frequency span of spectrograms in Figure 4). The upper half of this main peak is about 7–8 μs wide, which is narrow enough to permit registration of echo delay with an accuracy of a fraction of a microsecond at signal-to-noise ratios higher than about 10 dB (Menne & Hackbarth, 1986; Simmons et al., 1990). In addition, there are only two prominent side-peaks; these are 6–10 dB lower than the main peak and located about 30 μs away. Their presence reflects minimal ambiguity about which peak actually is the main peak in the ACR, an indication that the time-of-occurrence of echoes can by judged with little error from accidental selection of the wrong peak, except at low signal-to-noise ratios. The separation of the main and principal side peaks in the ACR is approximately the reciprocal of the median or center frequency of just the first harmonic in the sound (FM<sub>1</sub> in Figure 4) which is 30–35 kHz, not the sound as a whole because the first harmonic determines the intrinsic base periodicity of the waveform (Simmons & Stein, 1980). The ACR reveals that the bat's signals are good raw material for fashioning sonar images if the bat can utilize the information implied by their broad bandwidth and high frequencies to determine the arrival time of echoes (Menne & Hackbarth, 1986).

### 4.3. Glints in Targets and Echoes

When the broadcast sound impinges on the target (see arrow 3 in Figure 3), it reflects or scatters in different directions depending on the size and geometry of the object. The sound that returns back in the direction of the bat (see arrow 4) carries all the information that will be available for describing the object. The interaction of the sound with the object forms the echo, and the bat has to run this interaction in reverse to reconstruct what the object must have been from the echo. Sonar is thus an example of an inverse problem in which the characteristics of targets are determined by a process of reverse inference using the rules of scattering and propagation to guide reconstruction. The success of the imaging process depends upon the method used by the sonar receiver for describing the difference between the broadcast and reflected waveforms, particularly for attributing part of this difference to individual targets rather than to the propagation-path or to other properties such as direction or clutter. One crucial aspect of this target description is the nature of the bat's reference or template for the emitted sound used to recognize echoes. Another is the nature of the bat's internal model for targets that supports this description. In a broad sense, understanding how the bat's sonar works largely means identifying the metric dimensions in the bat's target model because the dimensions of this model are the dimensions of the bat's images.

In the generic example shown in Figure 3, the target is a flying moth. Most of the insects pursued by bats have linear dimensions of several millimeters to several centimetres. For *Eptesicus*, the broadcast wavelengths range from 3.4 mm at 100 kHz to about 20 mm at 15–20 kHz, so targets typically fall in the wavelength-to-size regime of scattering where the spectrum of echoes varies conspicuously across frequency due to target circumference and geometry. In air, targets are inelastic bodies whose impedance is so hugely different from air that they reflect or scatter the incident sound without absorbing significant energy to be re-radiated later. Consequently, targets in air do not resonate elastically at specific ultrasonic frequencies the way many hollow targets do in water; instead their echoes are composed largely of replica-
like reflections from different surfaces and discontinuities (glints) on their bodies. By themselves, the materials making up solid targets in air do not selectively absorb certain frequencies, they reflect equally well across the ultrasonic range. Variations in the spectrum of echoes from small targets occur because the "echo" from a "target" actually consists of multiple reflections from different glints at slightly different delays that add together and interfere with each other. (Two glints in the same object at different distances return discrete replicas of the incident sound at delays separated by 5 μs/mm of range separation.) This interference spectrum conveys information about the spacing of the glints. Targets in air can be recognized chiefly by the locations of their constituent glints, especially along the axis of target range (Simmons, 1989; Simmons & Chen, 1989). For the moth in Figure 3, several prominent body-parts, such as the head (A) and wings (B), contribute disproportionately to the total acoustic return, so the echo is dominated by two or three discrete replicas of the incident sound arriving at time separations determined by the range separation of the head and wings. In turn, the echo interference spectrum has peaks and notches at specific frequencies depending upon the size of these time separations (Simmons & Chen, 1989; Kober & Schnitzler, 1990; Moss & Zagaeski, 1994).

Because the sizes of insects encountered by bats are typically several millimeters to several centimeters, the time separations of the reflected replicas from their glints fall in the range of 0 to 150 μs, while the bat's sonar sounds are always much longer in duration (typically several milliseconds; see Figure 4). The overall echo from an insect thus is really composed of several distinct overlapping signals added together slightly offset in time. These overlapping signals reinforce and cancel each other to produce an interference spectrum whose peaks and notches are related to the reciprocal of the time-separation of the reflected replicas, and thus the range separation of the glints. This is the principal means by which the spectrum of echoes returning to the bat from objects ensonified in air comes to represent the shape of the target.

4.4. Describing Glint Structure with the Target's Impulse Response

Figure 3 traces the echo returning to the bat from the moth (arrow 4), and, just above, the impulse response of the target. Other representations of the target's reflective properties exist, including the transfer function and a variety of time-frequency distributions, but the impulse response is economical because it directly depicts the time separation of the reflected replicas from different parts of the target. In Figure 3, the ACR of the incident sound contains a single, prominent peak flanked by just two side-peaks, while the impulse response associated with the echo from the moth contains two prominent peaks (A, B) that correspond to the arrival times of the replicas reflected by the target's glints, chiefly the head (A) and wings (B). In this particular example, the time-separation of these two principal peaks is about 120 μs, which is equivalent to a range separation of about 2 cm—a reasonable spatial extent for the head and wings of a flying moth. The principal peaks in the impulse response mark the occurrence of the echoes from the largest glints, while the smaller side-peaks mark a combination of the intrinsic temporal structure of the incident sound (corresponding to the side-peaks in the ACR) and the minor glint structure of the target (corresponding to weaker reflections from smaller body-parts such as the insect's legs; see Simmons & Chen, 1989). The impulse response is a useful form for the image to take because it captures those features of the target's geometry that are arrayed along the axis of distance, or range. In the simple situation diagrammed in Figure 3, the resulting image portrays the principal glints in the insect, while in the more complex situation illustrated in Figure 1, the image would have to portray the entire acoustic scene in depth well enough to locate and identify the moth while locating the branches, too.

5. RECEPTION OF TRANSMISSIONS AND ECHOES

5.1. The Receiving Beam

Figure 6 shows the head and ears of the big brown bat. The external ears are the receiving antennas for the bat's sonar. Each ear is a directional receiving antenna with two quite different components to its directionality: First, in overall gain there is a broad receiving beam pointed slightly to one side that is complementary to the broadcast beam in its function. Second, during reception, echo fine-structure under-
goes very complicated directionally-dependent changes apart from overall gain that have to be understood in the context of image-formation. In *Eptesicus* (Figure 6), the principal structure in the external ear is the *pinna*—a tilted, curved surface about 14 mm long and 7 mm wide that forms a partial horn pointed forward and off to the side. The pinna acts as a moderately directional receiver with its acoustic axis also pointed forward and off to the side, as might be expected from its orientation on the head, and this axis shifts with frequency. Figure 7 shows directional receiving beams for the left and right ears of *Eptesicus* at frequencies of 25, 40, 60, and 80 kHz. The receiving acoustic axis is aimed about 60° to the side at 25 kHz, 45° to the side at 40 kHz, 40° to the side at 60 kHz, and about 25° to the side at 80 kHz (estimated from Jen & Chen, 1988; Wotton et al., 1995). Furthermore, the 6-dB width of the receiving beam changes with frequency—it is over 150° (±75°) at 25 kHz, 70° (±35°) at 40 kHz, 50° (±25°) at 60 kHz, and 50° (±25°) at 80 kHz (Figure 7). Thus, the gain pattern of the ear, like that of the broadcast, is rather coarse, with zones up to 50–70° and even over 100° wide having the same gain within ±3–6 dB. Taking the two ears together, the bat's hearing sensitivity is approximately omnidirectional over a conical volume extending from the front to fully 100–140° around the bat's head (Figure 7). At 25 kHz, this zone is at least 180° wide.

5.2. Directionality of the Whole Biosonar System

The bat's left ear points to the left and the right ear points to the right (Figure 7), but the broadcast beam points directly to the front (Figure 5A). The broadcast and receiving beam-widths also change with frequency, as does the direction of the receiving beam. The combination of frequency-dependent changes in three features—the width of the broadcast beam, the direction of the acoustic axis of the ear's receiving beam, and the width of the receiving beam about this axis—yields a surprisingly simple overall gain directionality for the *sonar system as a whole*. Figure 8 shows the sum of the transmitted and receiving beams in the horizontal plane at 25, 40, 60 and 80 kHz. For the right ear, echoes are received predominantly from the right side of the head, while, for the left ear, they are received predominantly from the left side of the head. Whereas the ear alone points decidedly to the side, the sonar system points to the front and side together. At most frequencies, the sensitivity of each ear extends in azimuth from about 10° on the contralateral side to about 25–40° on the ipsilateral side. Over this span of azimuths, at 40, 60, and 80 kHz the directional gain is within ±3 dB. At 25 kHz, sensitivity is much broader, extending from 30° contralateral to 65° ipsilateral. At 40–80 kHz, a target located off to the side strongly stimulates the ear on that side, while it more weakly stimulates the ear on the other side (Figure 8). At 25 kHz the same target substantially stimulates both ears unless it is very far to the opposite side. If only one target is present, this broad directionality is an asset because the target cannot escape being tracked by moving away from the center of the beam, but, by the same token, in complex surroundings a target cannot be isolated from surrounding clutter by aiming the sonar system as a whole in one direction or another because the coverage of the system is too broad. During the maneuver shown in Figure 1, the bat surely receives
5.3. Antenna Elements in the External Ear

The pinna of the bat’s ear (Figure 6) is not really a sound-gathering “dish” because its dimensions (14mm by 7mm) are too small—they are about the same as the wavelength of the bat’s sounds (3.4–20 mm). Nevertheless, it does seem chiefly responsible for the directional gain of the bat’s external ear (Jen & Chen, 1988; Wotton et al., 1995). There is another prominent structure in the external ear, however. This is the tragus (see Figure 6), a blade-shaped projection about 8 mm long and 3 mm wide located in front of the pinna, and just in front of the opening to the ear canal. The tragus lies parallel to the inside of the pinna and is situated in a position that is strongly reminiscent of the secondary reflector found in many antenna systems, although, like the pinna, it is not large relative to the wavelengths of the sounds and therefore cannot strictly be a reflector. The presence of a whole series of features besides the horn-like pinna—including the tragus, a series of horizontal ridges on the inner surface of the pinna, and the walls of the ear canal itself (into which the pinna and tragus both merge)—makes the external ear a more complicated structure with more antenna elements than just a partial horn. This complexity is manifested in fine-structure to the external ear’s sound-reception properties that is not well-captured in the broad directional gain patterns of Figures 7 and 8.

5.4. Impulse Response of the External Ear

Figure 9 illustrates measurement of the sound-receiving properties for the bat’s external-ear displayed in the time and frequency domains (Wotton et al., 1995). Here, the “input” to the ear (Figure 9A) is an FM test sound (sweeping from 100 kHz to 10 kHz over a duration of 0.5 ms) delivered from a loudspeaker located straight ahead of the bat (0°H, 0°V), and the “output” (Figure 9B) is this same FM sweep recorded by a small microphone placed at the site of the ear-drum. The transfer function (Figure 9C) for the sound reaching the ear-drum has two sharp notches near 33 kHz (N_1) and 87 kHz (N_2) and one broad peak at about 60–65 kHz (P_1). These local frequency-domain features are created by interference between several overlapping replicas of the input sound that reach the ear-drum at slightly different times. The origin of
multiple path-lengths through the external ear is reverberation within its structures, as revealed in the impulse response (Figure 9D). These path-lengths are associated with throughput delays indicated by a series of 3–6 peaks spaced 10–20 μs apart (P₀–P₇ in Figure 9D). Crudely put, the sound passing down the external ear bounces around several times on its way to the ear-drum. This delivers the sound at several staggered arrival times rather than one arrival time, thus smearing its waveform over a span of roughly 100 μs (Figure 9D) and introducing sharp notches and peaks in the spectrum as a result of interference (Figure 9C). A single replica reflected by a target’s glint would thus be stretched and smeared to consist of up to half-a-dozen similar replicas of the same reflection by the time it reaches the inner ear. One obvious problem is that the 100-μs time scale of the external-ear’s reverberation in steps of 10–20 μs is about the same as the time-scale for the spacing of replicas reflected by a target such as an insect (see above), which might seem to defeat any possibility of sharp registration of echo delay or recognition of target structure if this is the goal of signal-processing in echolocation. We have to address this problem to account for target recognition by bats.

5.5. Directionality of the External Ear Impulse Response

Several of the finer details in the external ear’s sound-reception properties are directionally dependent. The two prominent notches in the transfer function (N₁, N₂ in Figure 9C) shift downward in frequency as the sound-source moves downward in elevation from about 0° (horizontal plane) to −45° (Wotton et al., 1995). However, these two notches remain fixed at about the same frequency when the sound source moves up in elevation from 0° to +45°. Instead, as the source moves up in elevation, the amplitude of the spectral peak (P₁ in Figure 9C) changes systematically. These effects are in the frequency domain; comparable changes occur in the time domain as sound-source elevation changes, especially for one of the peaks in the impulse response (P₀ in Figure 9D). Figure 10 shows this complex effect in detail because it seems to play an important role in echolocation. The plot in Figure 10 shows a three-dimensional surface depicting the impulse response of the ear for sounds arriving from elevations from +45° to −45° in elevation (all at 0° azimuth). As indicated in Figure 9D, the impulse response contains a series of peaks (P₀–P₇) spaced at intervals of about 10–20 μs. The first peak in Figure 9D (P₀) corresponds to the arrival of the sound by the shortest path to the ear-drum, and subsequent peaks (P₁–P₇) correspond to progressively longer reverberation delays produced by the ear’s antenna elements. The reverberation delays for most of these peaks (P₀, P₁, P₂, P₃, P₄, P₅ in Figure 9D) are relatively independent of elevation, but the delay of one particular peak (P₄ in Figure 9D) does change systematically with elevation (sloping line marked d in Figure 10). The timing of this fourth peak in the impulse response (P₄) is shortest at elevations above the horizon and becomes progressively longer as the source moves to elevations below the horizon. The total shift in this example is about 20 μs from +45°–45° to −40–45°, or about 0.25 μs°° (Wotton et al., 1995). In addition, the height of the preceding peak

![External-ear impulse response at different elevations](image)

FIGURE 10. Impulse responses for the external ear of the big brown bat for passively-delivered sounds arriving from a source at elevations of +45° to −45°. The timing for only one of the peaks (d) is strongly dependent on elevation.
(P_e in Figure 9D) changes with elevation, increasing smoothly for elevations from about −10° to +25°. Smaller changes also occur with azimuth. A movement of the sound source by 10° further to the ipsilateral side in azimuth (e.g., from 0° to 10° or from 20° to 30°) produces about the same effect as a decrease in elevation by about 2–3° (Wotton et al., 1995). Thus, embedded in the complex series of reverberation delays created by the external ear is one particular peak that varies in its delay as the elevation of the sound-source changes. This change in delay is important for suppressing the smearing effect of the whole series of reverberations on the bat’s ability to perceive the arrival times of echoes.

5.6. Reception of the Broadcast by the External Ear

The bat receives both the broadcast signal and the echoes through its external ears (Figure 3). Even so, the 100-μs temporal dispersion of sounds by the impulse response of the ears (Figure 9D) ought to stretch and smear both the received version of the transmission and the received version of the echoes to preclude determination of echo delay with an accuracy or resolution better than 100–200 μs (equivalent to 2−4 cm of target range). After all, if both the reference signal and the echoes are smeared, how can the time that elapses between them be measured with great accuracy? The bat copes with the external ear’s temporal smearing of sounds by receiving the outgoing broadcast from one direction and the echoes from other directions.

The broadcast sound is picked up directly by the external ears at the moment of emission (arrow 1 in Figure 3) and then passed down through the ear canal to the ear-drum. The bat uses this direct pick-up to create a reference for recognizing subsequent echoes of that sound (Simmons et al., in press; see below). In Figure 3, the ACR of the broadcast, with its single, sharp main peak, passes through the external ear (arrow 2) to become a series of closely-spaced peaks corresponding to the series of reverberation delays imposed by the impulse response of the external ear (shown as peaks 1, 2, 3, 4 in Figure 3 and representing the impulse-response peaks in Figure 9D). Expressed in the time domain, this is the reference representation for the broadcast that the bat uses to compare with echoes.

5.7. Reception of Echoes by the External Ear

When echoes are received by the bat’s external ears, their waveforms are changed relative to the broadcast according to the location of the target in relation to the directional gain of the ears (they already have been modified by the target’s location in the broadcast beam and by the target’s distance, size and shape). They are also changed by the reverberant properties of the external ear to have a series of reverberation delays stretching for about 100 μs in steps of 10–20 μs (Figure 9D). The crucial fact, however, is that the target is located out in front of the bat, in the extreme case with clutter all around it, while the effective source of the directly-received broadcast sound is not.

The broadcast sound radiates from the bat’s open mouth and reaches the bat’s ears from the direction of the mouth. This direction is about 25° “inboard” relative to straight ahead for each ear and also about 35–45° below the horizontal plane of the head because the ears are located higher than the mouth. That is, the ears are placed above the eye–nostril plane while the mouth is placed below. The external-ear impulse response is different for the broadcast than the echo because these two sounds arrive from different directions. The external-ear impulse response has a specific series of peaks (P_a–f in Figure 9D) associated with each particular direction. The reference signal for the broadcast arriving at the ear-drum is “stamped” with the impulse-time features of the direction from the ear to the mouth (elevation about −35 to −45° in Figure 10). In contrast, echoes, which arrive from targets mostly in different directions than the mouth, are stamped with different impulse-time features associated with those directions rather than the direction to the mouth (e.g., elevation about 0° in Figure 10). Consequently, when the external-ear impulse response is calculated for an echo from a target in front of the bat rather than just for a sound coming from some particular direction, the impulse response is somewhat simpler due to both the “input” signal (the broadcast) and the “output” signal (the echo) for the impulse-response calculations passing through the external ear.

Figure 11 shows impulse responses for echoes arriving from targets located at elevations between +45° and −45°. These echo impulse responses contain a series of 4–6 prominent peaks that generally shift in time as elevation changes, in contrast to the series of peaks in Figure 10 that mostly do not shift with elevation. This condition is an improvement in the sense that, now in Figure 11, target elevation is represented quite robustly in the images because most of the peaks move to different times at different elevations, but there are still too many peaks. The arrival time of a single echo is stretched over a span of about 80–100 μs in steps of 10–20 μs, which smears the unambiguous registration of delay over several possible values and seems to preclude a delay accuracy better than 80–100 μs. Furthermore, the target’s glint structure, which is represented by the time spacing of replica reflections at separations of roughly 10–100 μs, would be
observed by the series of peaks due to the external ear alone. How can the bat perceive the target’s structure through the external ear’s repeated reverberation? Also, given the 350-μs integration time of the bat’s hearing, how can the impulse response of the target be determined at all?

First, to fully describe the echo waveform reaching the ear-drum, the impulse response of the external ear has to be combined with the impulse response of the target. While the target impulse response displays the spacing of glints along the range axis (A, B in Figure 3), the external-ear impulse response displays the timing of reverberations that take place as the sound passes down into the ear canal through the structural elements of the antenna (shown as peaks 1,2,3,4 in Figure 3). The composite impulse response for the echo waveform at the ear-drum shown in Figure 3 (target + external ear) contains two sets of peaks—a strong peak (A1) for the first target glint, followed by a series of smaller peaks (A2,3,4) for the external-ear reverberation associated with that first glint, a strong peak (B1) for the second target glint, and a series of smaller peaks (B2,3,4) for the external-ear reverberation associated with the second glint. This very complicated delay structure mixes information about the target’s shape from the target’s impulse response with information about the target’s location from the directional impulse response of the external ear. However, similarly complicated information (peaks 1,2,3,4 in the external-ear impulse response; Figure 3) was also imposed on the waveform of the broadcast sound passing through the external ear on its way to forming the reference template, with two differences—first, the target’s glints are not part of this reference signal (no A, B in external-ear impulse response), and second, the reverberation-delay of one of the peaks (P4 in Figure 9D; peak 3 in Figure 3, marked with * in Figure 3) is different for the echo than the transmission because the target is usually not located in the same direction from the ears as the bat’s mouth. This seemingly small difference emerges to become one of the most salient features of the bat’s images.

Second, the impulse responses in Figure 3 cannot simply be made by directly displaying the peaks in the raw waveforms of the sounds: The broadcast sound is several milliseconds long, while the reverberation-delays introduced by the ear span only 100 μs, and that in steps of only 10–20 μs. Similarly, the glint separations reflected by most targets are only 10–100 μs. Moreover, the integration time for echo reception is 350 μs (Figure 2C1–2). The numerous delayed replicas of the original broadcast signal reaching the ear-drum will add together to interfere with each other over the duration of the whole sound. Consequently, a frequency-domain description of these waveforms is better suited to the actual signals than a time-domain description. Thus, the reference signal for the broadcast at the ear-drum actually has peaks and notches in its spectrum at frequencies that correspond to the transfer function of the ear (Figure 9C) rather than peaks at times that correspond to the impulse response (Figure 9D). Similarly, the overall echo at the ear-drum has peaks and notches in its spectrum that relate to both the target’s glint structure and the external ear’s reverberation-
delays. Subsequent transduction of these sounds by the bat's inner ear captures these frequency-domain peaks and notches rather than the time-domain peaks for further processing in the brain.

6. BIOSONAR SIGNAL-PROCESSING

6.1. The Bat's Gain Control for Echoes

After being received by the external ear, sound travels down the ear canal to the ear-drum which vibrates in response to sound-pressure waves. These vibrations are conducted along the very small bones of the middle ear to the inner ear, where the mechanical motion is transduced into cellular bioelectric events and, ultimately, neural discharges. The middle ear is the site of at least one stage of control over the amplitude of stimulation delivered to the inner ear, and, in bats, this mechanism regulates the amplitude of sonar echoes. The middle-ear bones are equipped with two muscles which contract to attenuate the transmission of sound along the bones to the inner ear. In many animals, these middle-ear muscles contract in response to loud sounds to exert a protective effect on hearing sensitivity and the receptors of the inner ear. In echolocating bats, these muscles also contract just prior to vocalization, reaching a level of maximum contraction during the sonar broadcast and then relaxing thereafter (Henson, 1970; Suga & Jen, 1975). The period of relaxation lasts for at least 5–8 ms and thus intrudes into the critical interval of time following the broadcast when echoes are received from targets at distances up to at least 1–1.5 m from the bat. As would be expected, the bat's hearing thresholds change during this interval, becoming progressively less sensitive by about 30 dB as the echo delay shortens from 6 ms to 1 ms (Kick & Simmons, 1984; Simmons et al., 1992). It is not presently known what other mechanisms contribute to the full range of observed changes in echo sensitivity, but middle-ear muscle contractions synchronized to vocalization seem certain to cause a large part of it, particularly over the critical target distances from 10–20 cm out to 1–2 m.

The rate of reduction in hearing sensitivity for echoes amounts to 11–12 dB per halving of target range, and nicely matches the rate of increase in echo strength from a small (insect-sized) target, which amounts to 12 dB per halving of target range (see above). Consequently, the amplitude of echoes is perceived as being constant in its level relative to threshold (constant sensation level) as the bat approaches from a distance of 1 m or more to a distance of only 17 cm. This change in sensitivity is a range-related gain-control that operates by linking the change in hearing sensitivity at different echo delays with the change in delay at different target ranges. In effect, the amplitudes of echoes are stabilized during reception to eliminate those changes in strength that occur as a consequence of changes in distance, leaving changes in strength caused by the target's variation in cross-section to pass through and be detected.

6.2. Auditory Spectrograms of FM Transmissions and Echoes

The central problem to be considered here is how the bat's auditory system creates the images the bat perceives given the constraint that FM sweeps are represented by spectrograms with a relatively long integration time of 300–400 µs. Figure 12 is a diagram which models the principal computational stages required to convert the raw time-series waveform of a sonar emission and two overlapping echoes first into a spectrogram and then into an "A-scene" sonar image depicting the delay of both echoes along the same scale of time. The model's first stage is its "cochlea" (Saillant et al., 1993), which uses 81 band-pass filters in parallel to transform the raw input waveforms into hyperbolically-scaled spectrograms for further processing. This component of the model emulates the most critical features of the bat's inner ear and then generates "neural discharges" registering successive frequencies in the FM sweeps of the sounds. The model's remaining stages are two parallel pathways—a spectrogram correlation system for determining the time separation between the spectrogram of the emission and the spectrogram of echoes (Altes, 1980, 1984), and a spectrogram transformation system for converting the pattern of peaks and notches in the spectrogram of overlapping echoes (Altes, 1984) into an estimate of the time-separation of the merged, interfering replicas. The model's algorithms combine these two operations into a single imaging process that runs in parallel (SCAT algorithms; Saillant et al., 1993).

6.3. Spectrogram Correlation to Determine Echo Delay

Figure 12A shows the raw waveforms of the input signals—a sonar transmission with a duration of 2 ms and two overlapping echoes (A and B). The delay of the first echo (t_A) is 3.7 ms, and the delay of the second echo (t_B) is only 60 µs larger. This short delay separation (δt = 60 µs) results in the two echoes merging to form just one spectrogram in Figure 12B. The amplitude of the echo spectrogram at different frequencies contains peaks and notches caused by the interference taking place within the 350-µs integration time of the spectrograms. To form an image, the SCAT model first determines the arrival time of the
FIGURE 12. Diagram of the SCAT process for echo spectrogram processing and deconvolution of echo spectrograms. (A) Waveform of 2-ms FM sonar emission and two overlapping echoes (A + B) separated by 60 μs. (B) 81-channel SCAT spectrogram of transmitted and reflected waveforms, with spectrogram delays (tA1–tA6) showing time that elapses between broadcast of sonar sound and return of echoes. (C) Alignment and averaging of spectrogram delays to determine overall spectrogram delay (indicated as tA). (D) Shape of spectrum for overlapping echoes, with notches and peaks caused by interference, used to set the amplitudes of basis vectors. (E) Cosine-phase basis vectors specified by the frequency of each channel and triggered by average spectrogram delay (tA). (F) Echo-delay ("A-Scope") sonar image formed by summing basis vectors. Image contains delay for echo A originally from spectrogram delays and echo B from summation and cancellation of different basis vectors weighted by the amplitude in each frequency channel of the spectrogram. (After Saillant et al., 1983 and Simmons et al., in press.)
compound echo (A + B) from the time intervals between the spectrogram of the emission and the spectrogram of the echo at different frequencies (that is, the horizontal time displacement of the echo spectrogram to the right of the emission spectrogram in Figure 12B). We use the term spectrogram delay (shown as \( t_{A1} - t_{A5} \) in Figure 12B) for these individual emission-to-echo intervals and average spectrogram delay for the overall delay (\( t_A \)) to distinguish them from group and phase delays of the echo. Spectrogram delays are extracted using delay-lines that register the time-of-occurrence of each frequency in the emission and then compare it with the time-of-occurrence of the corresponding frequency in the echo. In effect, the delay-lines store the frequency-by-frequency shape of the spectrogram for the emission and slide it to the right in Figure 12B until it lines up with the shape of the spectrogram for the echo—a process equivalent to correlation of the echo and emission spectrograms (Altes, 1980). To register the occurrence of one specific frequency in the broadcast sweep, an “event” travels along each delay-line from one delay tap to the next, and the relative position of similar events across all the delay-lines preserves the shape of the broadcast FM sweep as the events propagate along the delay-lines.

At each frequency, the spectrogram delay of the echo (\( t_{A1} - t_{A5} \)) is represented by the specific delay tap in the delay-line that is active at the same moment that the echo arrives. This “moment” is judged by detecting coincidences between events taking place at the delay taps and events triggered by the incoming echo. This coincidence-delay process has a physiological equivalent in the responses of delay-tuned neurons located within the big brown bat's auditory cortex (Dear et al., 1993; see Simmons et al., in press). The spectrogram delays (\( t_{A1} - t_{A5} \)), which are represented by the active delay taps in different delay-lines, are then averaged across all the delay-lines in Figure 12C to estimate the average spectrogram delay of the echo as a whole (\( t_A \)). This overall delay value is obtained from measurements of the model's equivalent of the timing of neural discharges and represents the distance to the object that contains the two glints; it is usually interpreted to be the distance to the nearer of the two glints, \( t_A \) (Simmons et al., 1990; Saillant et al., 1993; Simmons, 1993). In the absence of noise, all 81 channels normally register their delay estimates at the same delay value (same delay-line tap at each frequency) provided the shape of the FM sweep in the echo is not changed by phase-shifts occurring during echo-formation. The addition of noise to the echo merely broadens the distribution of active delay taps around the mean value, and, at high levels, noise sometimes displaces the mean, too. Registration of echo delay is very precise by this method when averaged across a number of parallel delay-lines, and *Eptesicus* is known to be very accurate at determining the delay of echoes to within 10–15 ns from the timing of neural discharges (Simmons et al., 1990; see Figure 2B1–2).

If the processing of echo information were to stop at this point, the range image would depict just the overall distance to the target. No distinction would be made about the distances to the two glints. Further information about the target is contained in the shape of the spectrum of the overlapping echoes (Figure 12D), and one widely-accepted hypothesis is that the bat classifies targets in terms of the spectral coloration supplied by the peaks and notches at different frequencies (Neuweiler, 1990; Schmidt, 1992). In terms of distance, the bat would just perceive the overall spectrogram delay. The bat does not appear to stop at this point, however, because it perceives the delays associated with both parts of the two-glint target in the same image (Figure 2D1–2).

6.4. Spectrogram Transformation to Determine Delay Separation

The capacity to deconvolve overlapping echoes that have been merged into one spectrogram (Figure 12A, B) depends upon being able to translate the pattern of peaks and notches at different frequencies in the echo spectrum (Figure 12D) into an estimate of the delay separation required to create these peaks and notches by interference. In effect, the SCAT algorithms assume that all features of the echo spectrum that differ from the broadcast spectrum arise from overlap and interference of replicas of the broadcast, with no contribution from elastic phenomena, and that the target is modeled in terms of the glint structure required to create the time-series of these replicas. It is not sufficient to know just the frequency values for spectral features, however; deconvolution requires knowledge of the values for the periods of the frequencies corresponding to the tops of the peaks and the bottoms of the notches. To be complete, deconvolution also requires knowledge about the detailed shape of the ridges in the echo spectrogram in the vicinity of the peaks and notches.

The most complete implementation of the frequency-to-period knowledge required for deconvolution is reconstruction of the waveform of echoes at each frequency within the integration-time window for convolution. This can be achieved even after spectrograms have been formed (that is, after convolution) by using the spectrogram delays at individual frequencies (\( t_{A1} - t_{A5} \) in Figure 12C) as time-marking events to trigger the start of oscillatory signals, or basis vectors, that represent the original echo frequencies themselves. That is, each delay-time is used to register the arrival time of one specific frequency in the echo, and the moment a coincidence
between the echo and the emission at that frequency is detected, the basis vector begins to oscillate. In the SCAT model, the basic vectors (Figure 12E) are cosine functions with durations sufficient to cover the interval of time across which the glint structure of echoes is to be reconstructed (usually the integration time of several hundred microseconds). However, the model is robust and works reasonably well for any periodic function used as basis vectors, even squarewaves (Saillant et al., 1993). The horizontal slices in the spectrograms (Figure 12B) correspond to the frequency channels of the SCAT model, and each channel is tuned to a specific frequency in the emission or echo. Each channel then produces its own basis vector at a frequency that matches (or is proportional to) its original tuned frequency (Figure 12E). The amplitude of the basis vector in each frequency channel is adjusted according to the shape of the spectrum for the echo (from Figure 11D to Figure 12E). That is, the amplitudes at different frequencies in the echo spectrogram are used to determine the amplitudes of different frequencies in the basis vectors.

Once the basis vectors begin to oscillate in their corresponding frequency channels, the next stage in the imaging process is simple. The arrival times of the overlapping echoes (A and B) are reconstructed by summing the basis vectors across all 81 parallel channels (Figure 12F) to form an average basic waveform. This average waveform is the image of the target’s glint structure along the axis of echo-delay or target range. Due to the reinforcement and cancellation of peaks and troughs in the basis vectors across channels, the original arrival times of the echoes from the two glints (tA and tB) appear as positive-going peaks in the resulting image even though no corresponding well-resolved pair of events registers their arrival-time separation in the raw echo-waveforms. Moreover, because the basis vectors are aligned to start oscillating in cosine phase at the echo-delay value specified by the delay-lines (tA), the entire image is displayed in absolute units of echo-delay or target range. SCAT thus amounts to a serial analysis-and-synthesis process carried out through the intermediary of a spectrogram representation, which permits the required serial operations to occur largely in parallel.

6.5. Time-stretching of Basis Vectors and their Reconstructed Images

One particularly significant feature of the SCAT model is the capacity to extract estimates of spectrogram delay (for echoes A + B together) and delay separation (from echo A to echo B) using processing elements that have different time scales. It is convenient to introduce this feature of the SCAT model by initially equating the frequency of each basis vector with the center frequency in each frequency channel of the spectrogram. In this case, the reconstructed image (Figure 12F) has the same time-scale as the original signals; that is, the time between the first and second delay estimates is 60 μs in the time-series signal formed by summation of the basis vectors, just as it was 60 μs in the original ultrasonic echoes [δt in the echoes (Figure 12A, B) equals δt in the image (Figure 12F)]. However, this requires the basis vectors to be oscillations at ultrasonic frequencies of 20–100 kHz, which is physiologically implausible for neurons to achieve. Oscillatory responses observed in the mammalian auditory system typically have frequencies of 100 Hz to at most about 1–2 kHz (Langner & Schreiner, 1988; Langner, 1992) but never anywhere near the ultrasonic frequencies of the bat’s sounds themselves.

An alternative is to scale the frequencies of the basis vectors to be lower than the original ultrasonic signals, keeping the frequencies of the basis-vector oscillations proportional to the center frequencies of the band-pass filters rather than equal to them (Saillant et al., 1993). In this case, while the ultrasonic frequencies in emissions and echoes are scaled hyperbolically from 20 to 100 kHz in Figure 12B, the frequencies of the basis vectors extend hyperbolically from some preset fraction of 20–100 kHz in Figure 12E. This fraction is a scaling factor for the frequencies of the basis vectors; it lowers the frequencies in the time-series waveform that reconstructs the image. Given that physiologically-plausible bursts of oscillatory neuronal responses can occur at frequencies up to about 1 kHz, scaling factors of \( \frac{1}{20} \) to \( \frac{1}{100} \) might be expected in the bat’s equivalent of the SCAT basis vectors. Of course, the scaling factor correspondingly lengthens the time spacing of the image components for the glints when it lowers the frequencies of the basis vectors, which preserves the time-bandwidth product of the original ultrasonic signals in the new representation. In Figure 12, the time interval between the ultrasonic echoes (δt in Figure 12A, B) is only 60 μs, but the corresponding time-interval between the first and second delay estimates in the time-series signal created by adding the basis vectors together (δt in Figure 12F) could be anywhere from 1.2 ms (for a scaling factor of \( \frac{1}{20} \)) to 6 ms (for a scaling factor of \( \frac{1}{100} \)). In the example in Figure 12, the original echo-delay separation is 60 μs, while the separation of the peaks in the reconstructed image is about 2.3 ms, so the scaling factor is about \( \frac{1}{38} \). This scaling factor is within the range actually observed physiologically in bats: neuronal responses recorded in *Eptesicus* show time-stretching of ultrasonic waveform events from the range of tens of microseconds to the range of a millisecond or two, so some kind of process
analogous to the basis vectors seems to be present (Simmons et al., in press). These longer time intervals might realistically be represented by the timing of successive neural responses several milliseconds apart whereas the original 60-µs interval could not be represented directly by two successive neural responses only 60 µs apart.

6.6. SCAT Images Derived from External-ear Inputs

The impulse responses for echoes shown in Figure 11 contain peaks spread over a span of 80-100 µs. The presence of these multiple impulse-response peaks ought to create ambiguity about which of these peaks constitutes the “real” arrival time of echoes, but no comparable repetitive ambiguity is evident in the bat’s perception of echo arrival time (Simmons et al., 1990). In effect, the bat’s images of arrival time lack any sign of the series of peaks caused by external-ear reverberation. However, the type of image formed by the SCAT algorithms differs from a true impulse response in that repetitive peaks spaced at fixed intervals (in the case of the external-ear reverberation, these intervals are 10–20 µs) are replaced by a single peak that summarizes the multiple spacings in a compressed format. A SCAT image is thus a simplified version of the impulse response—it shows a set of base intervals between the returning echoes from different glints rather than each interval individually. In effect, the SCAT image depicts the delay-axis scales in the echoes without being burdened with all the details.

Figure 13A illustrates SCAT images over a delay span of 0–55 µs for the same set of echoes arriving from different target elevations (+25° to −25°) as for the full echo impulse responses shown in Figure 11. These SCAT images are aligned at zero time (0 µs) which is the objective arrival time of the echoes. Besides a sharp “main peak” at 0 µs (which was subtracted from these SCAT images to show only side peaks), there are two prominent features to these SCAT images. The first feature is a broad “side peak” located at delays of 25 µs to 45–50 µs at different elevations. The second feature is a lower-amplitude ripple with a spacing of 10–15 µs that spreads across the SCAT images in Figure 13A. This ripple is a residue of the strong 10–20 µs repetitive reverberation in the impulse responses across all elevations in Figures 10 and 11. The SCAT process largely suppresses the repetitive character of these multiple peaks, but it does not suppress it totally. The prominent side peak slopes downward to the right in Figure 13A from about 25–30 µs at an elevation of +25° to about 35 µs at 0°, finally reaching 40–50 µs at −25°. It emerges in the SCAT images as their dominant characteristic representing the difference between the external-ear reverberations encountered by the emission directly picked up from the mouth and the external-ear reverberations encountered by the echoes from different elevations. In the impulse responses of Figure 10, a distinct spacing of 25–45 µs.

![SCAT Images](image1)

![Behavioral Results](image2)

**FIGURE 13.** (A) SCAT Images for single-glint echoes arriving from a target at elevations from +25° to −25° and passing through the external ear. The images are aligned at 0 µs, which is the estimated arrival time of the echo (main peak subtracted from SCAT images), but they contain a prominent side peak at 25–40 µs depending upon target elevation. There are also lower-amplitude ripples at spacings of 10–20 µs that are residual from the strong reverberation intervals in the impulse responses (Figures 10, 11). (B) The bat’s echo-delay images for a single-glint target at elevations of +15°, 0°, and −15° (see procedure in Figure 2B). These images are aligned at 0 µs, which is the bat’s most likely delay estimate (main peak subtracted from bat’s images), but they contain a side peak at 25–40 µs depending on elevation. Data consist of percentage errors averaged for three bats.
is only barely discernible in the relation between the sloping peak (d) and the series of three fixed peaks that precede it (peaks a, b, c in Figure 9D). The side peak in Figure 13A is the abbreviated expression in the SCAT images of the whole series of peaks at 20–100 μs in the echo impulse responses of Figure 11. Both the SCAT side peak and the echo impulse-response peaks shift in time with changes in elevation (they slope downward to the right), but the SCAT images contain just one side peak rather than 4–6 parallel peaks. The degree of data-compression exhibited by the SCAT images is useful—there is ambiguity only between the main peak and the single side peak in the SCAT images whereas the impulse responses contain a wider range of different ambiguities among all their peaks.

6.7. The Bat's Images of Echoes at Different Elevations

The behavioral experiment described in Figure 2B₁ measures the bat's accuracy for determining echo delay, and the results sketched in Figure 2B₂ trace the image that the bat perceives along the echo-delay axis. This procedure offers a way to evaluate the bat's images alongside the impulse responses in Figure 11 and the SCAT images in Figure 13A. Figure 13B shows the average performance of three big brown bats in the echo-delay jitter procedure of Figure 2B₁, with the loudspeakers delivering the echoes to the bat placed at three different elevations (+15°, 0° and -15°) relative to the bat's head. The data in Figure 13B show the percentage errors made by the bat as an index of the perceived location of the side peak in the images (the main peak in the bat's images is located at 0 μs but is subtracted from Figure 13B to show only the side peak). The bat's images contain a side peak at 25–40 μs according to the elevation of the echo source. Moreover, the location of this side peak corresponds to the side peak in the SCAT images of Figure 13A, not the multiple peaks in the impulse responses of Figure 11. It therefore appears as though the bat, *Eptesicus*, perceives images that resemble the output of the SCAT algorithms, with a simplified peak structure, rather than true impulse responses, with their greater involvement of multiple peaks. Shrinkage of the extended effects of external-ear reverberation to a narrow region around 25–40 μs gives ample room for the bat to perceive the target's glint structure in the critical 10–100 μs range without sacrificing the ability to register the target's elevation in the same images. Experiments with two-glint echoes demonstrate specifically that the bat can perceive both glints—one glint at 0 μs and another glint at delay separations ranging from 2 μs to 30 μs or more without any difficulty (Saillant et al., 1993; procedure shown in Figure 2D₁). Moreover, physiological responses that represent the time spacing of two glints in the range of 0–200 μs appear to register this time separation on an expanded time scale of several milliseconds, as predicted from the SCAT model (see Simmons et al., in press). We can now outline how the bat's images are able to incorporate information about target shape while also containing information about elevation.

6.8. SCAT Images for Echoes from Complex Targets Reaching the Bat's Ear-drum

In principle, the bat's echo-processing strategy can now be summarized abstractly in the time-domain from Figure 3, purely in terms of the sounds delivered to the ear-drum during echolocation. The transmitted sound, with its sharp ACR, impinges on the target to reflect as two prominent replicas from the principal glints, A and B. The impulse response of the target is carried by this compound echo. However, the reference signal reaching the bat's ear-drum has been stretched by the external ear (the external-ear impulse—peaks at 1, 2, 3, 4—in Figure 3), and, similarly, the echo has been stretched, too (the target + external-ear impulse response—A₁,₂,₃,₄ and B₁,₂,₃,₄—in Figure 3). Thus, the external ear has significantly altered the time-domain information available to the bat. When the broadcast reference signal reaching the ear-drum is used as the "input" for determining the target's impulse response while the echo at the ear-drum is used as the "output", all of the external-ear features that are common to both the broadcast reference and the reflected echo are cancelled out of the resulting target impulse response. By operating strictly with the signals reaching the bat's ear-drum, the time-domain image of the target is left to contain just the target's glints (A, B in Figure 3), which is the object of target imaging, and those features of the external-ear impulse response that are not common to both the broadcast and the echo (peaks A₁ and B₁ in Figure 3, both marked with *). The external-ear reverberation that appears most prominently in the SCAT images in Figure 13A is the elevation-dependent peak (P₄ in Figure 9D; see Figure 11) because the target and the bat's mouth are located in different directions. The other peaks in the external-ear impulse response (Pₜ₋ₐ, Pₜ₋ᵣ) are common to both the "input" and the "output" signals and therefore largely disappear from the resulting time-domain images. Moreover, direct measurements of the bat's echo-delay images at different target elevations confirm that such a process is indeed used by the bat (Figure 13B).

7. SUMMARY

The biosonar systems of echolocating bats exhibit
extremely good target localization and recognition performance across a wide variety of situations, including conditions with just a single target present, conditions with a single target embedded in dense clutter, and conditions with multiple targets of nearly identical features. The big brown bat, *Eptesicus fuscus*, broadcasts broadband FM sonar signals (20–100 kHz band) and can both locate and classify sonar targets to guide airborne interceptions and avoid colliding with obstacles. The fluidity with which bats can follow a flying insect while it moves from simple to complex situations, successfully distinguishing the insect from clutter and achieving interception without colliding with the clutter, reveals that the bat's echo-processing mechanisms rapidly adapt to dynamic environments and can deliver high-quality images across these different conditions. Underlying this performance is the crucial perceptual capacity to segment the incoming stream of echoes into estimates of the arrival times of individual reflected replicas of the transmitted sound, even for multiple reflections from different parts (glints) of the same target. To support image formation along the dimension of echo delay or target range, the big brown bat has an echo-delay hyperacuity of 10–15 ns and a two-point echo-delay hyperresolution of about 2 μs (the signals themselves have a Rayleigh resolution limit of about 10 μs). Even if the bat's performance is degraded by fully one or two orders of magnitude due to noise and psychophysical uncertainty prevailing in natural conditions, the resulting echo-delay acuity would still be adequate to locate individual glints to within a few millimeters and thus to perceive the shapes of objects on a scale as fine as several millimeters. However, the underlying mechanisms must nevertheless be capable of 10–15 ns delay accuracy and 1–2 μs delay resolution in quiet, low-uncertainty conditions as a prerequisite for achieving reasonable performance after degradation occurs from the complexity and noisiness of realistic situations.

The bat's auditory system encodes the FM sweeps of sonar transmissions and echoes as spectrograms from the outputs of parallel band-pass filters whose time-constants and smoothing yield an integration time of about 350 μs. The convolution of each transmission and its subsequent echoes to form these spectrograms transposes information about echo-delay separations smaller than 350 μs from the time domain to the frequency domain, where this information is manifested as peaks and notches in the interference spectrum made up of segments of all the signals occurring in successive time-windows that are ±350 μs wide. To form images with an inherent temporal acuity of 1–2 μs or better requires deconvolution of these spectrograms. Furthermore, to express echo-delays and delay separations that are both longer than 350 μs and shorter than 350 μs along the same perceived delay dimension requires coordinated absolute and relative delay computations of a very specific nature. A combination of algorithms—spectrogram correlation to determine absolute delays and spectrogram transformation to determine relative delay separations within the 350-μs spectrogram window (SCAT algorithms)—deconvolves the delay information in spectrograms to recover the required echo-delay axis in absolute units. The underlying physiological mechanisms used by the bat incorporate recognizable delay-lines and coincidence detectors to read out absolute delay, thus achieving spectrogram correlation, and frequency-tuning of neurons with multiple-frequency selectivity is adequate to provide for detection of peaks and notches in interference spectra as well as read-out back to the time-domain. However, the timing accuracy of these physiological responses appears to be in the range of several hundred microseconds to several milliseconds, with a small minority of responses as sharp as 30–100 μs. Physiological temporal resolution is also limited by neural recovery times to 300–500 μs at best. No plausible model for integration of information across neuronal populations is able to approach within 1–2 orders of magnitude of the observed 10–15 μs echo-delay accuracy and 1–2 μs echo-delay resolution using conventional physiological results interpreted at face value. These results are displayed as PST histograms showing the time-locking of neural responses to stimulus occurrences. The time scale of the histograms is real time relative to the stimulus; no consideration has been given to the possibility that the time scale of neural responses might be altered by neuronal computations to represent the timing of stimulus events at a different rate. New physiological results reveal that the bat's auditory system indeed does encode small transmission-echo time shifts and two-glint time separations in the microsecond and tens of microseconds range on time scales that are expanded or magnified relative to the sounds themselves. Time-scale expansions observed so far are in the range of 10–100 times, which accounts for the one to two orders of magnitude missing from physiological "explanations" of the temporal acuity of echolocation. The use of parallel time-series representations having different rates of time, extending from real time to as much as 100 times faster than real time, is a novel signal-processing strategy whose function appears critical for the bat's implementation of SCAT algorithms.

REFERENCES


