

Supplemental Data

Spontaneous Motor Entrainment to Music

in Multiple Vocal Mimicking Species

Adena Schachner, Timothy F. Brady, Irene M. Pepperberg, and Marc D. Hauser

Supplemental Experimental Procedures

Case Studies: Methods

Subjects

Subjects 1 and 2 were (respectively) an African Grey Parrot (“Alex”), and a Sulphur-Crested Eleanora cockatoo (“Snowball”). Alex’s lifelong owner, co-author I.M.P., reports no previous training for rhythmic movement. The cockatoo was also unlikely to have been trained; neither his current nor previous owners reported training and were not professional animal trainers. The current owner reports spontaneous rhythmic movement in response to music. Also tested were 9 cotton-top tamarin monkeys, socially housed in Harvard’s Cognitive Evolution laboratory. Many of these subjects had been exposed to music [1]; however, they had never been trained to move rhythmically to auditory stimuli.

Stimuli

Subject 1: Stimuli consisted of novel, natural rhythmic music created by author A.S. using Garageband (Apple, Inc) and composed from a variety of percussion loops included in the software library. Tracks were created at two tempos: 120 bpm and 150 bpm, and ranged from 2 to 3 minutes in length (depending on tempo, e.g. faster tempo was shorter).

Subject 2: Stimuli consisted of natural rhythmic music in the form of popular human music. The song in video 1 (Everybody, Backstreet Boys) had been heard by the subject many times before (>100 times); the other three (Another One Bites the Dust, Queen; Come on Christmas, Ringo Starr; I Want a New Drug, Huey Lewis and the News) were novel or had only been heard a small number of times before (<3 times), and spontaneous response was reported on the first hearing. These songs were selected as the result of informal testing, during which the owner played a wide variety of music, noted songs that evoked response, and videotaped the bird’s response to those songs at a later time.

Cotton-top tamarins: Stimuli consisted of seven segments of stimuli to which Subjects 1 and 2 had entrained (each between one and three minutes), concatenated into a medley with five seconds of silence between each segment. Minute-long click tracks at 90 and 120 beats per minute were also included in the medley to test the possibility that the tamarins could not parse the beat out of complex stimuli, but might entrain to simple periodic sounds. Two different orders of the medley were created, each 15.5 minutes long. All subjects were exposed to both orders of the medley for a total of two sessions. Each segment of the medley was analyzed separately, and treated as a separate trial.

Procedure

The subject was placed in a comfortable location that allowed complete freedom of movement. The stimuli were played at a normal volume (~65 dB at the subjects' location), and the subjects' behavior was video recorded (using Panasonic mini DV video cameras). Human rhythmic movement was explicitly eliminated during testing of Subject 1: When human experimenters were present in the room, they remained still throughout the session and did not engage in any rhythmic movement. In the case of Subject 2, human movement was never consistently present throughout the trial, and was not present for the majority of time within each trial. Because any human rhythmic movement was minimal, and because entrained movement has also been observed and verified in the absence of any human movement in this subject [2], the effect of any human rhythmic movement was likely to have been minimal during these trials.

Online Database Analyses: Methods

Details on Overrepresentation of Vocal Nonmimics Relative to Vocal Mimics

Vocal nonmimicking species are much more highly represented on YouTube than are vocal mimics, producing a greater number of results both in searches for animal terms alone and for animal terms + “dancing” (Animal terms alone: Vocal mimics 1,535,551 hits; Vocal Nonmimics 2,921,122 hits. Animal terms + “dancing”: Vocal mimics 58,539 hits; Vocal Nonmimics 128,115 hits). This overrepresentation of vocal nonmimicking species increases our chance of detecting entrainment in these species should it exist.

Search Terms

We chose the search term “dancing”, as in pilot searches it had more results than related terms when combined with sample animal terms (“dancing” + “dog”, “bird”, “monkey” resulted in 8,410, 7,050, and 9,240 hits respectively, for a total of 24,700; the same animal terms with “dance” totalled 22,210; with even fewer results for “groove” “rhythm” etc). For each search, the term “dancing” was combined with the animal-related search term, and the top 50 results (as sorted by relevance) were recorded and categorized.

Searches proceeded in four steps. We first searched broadly for a wide variety of animals commonly in contact with humans (including common pets and domesticated animals) and then iteratively conducted more specific searches following up on any searches in which we found evidence of entrainment. Specifically, in the second step we followed up on evidence of entrainment in birds by performing more specific searches, one search for each order of species within neaves. Search terms were selected to be the most commonly used name for the group, (e.g. “parrot” for parrots, “hawk” for diurnal birds of prey) using information from the Tree of Life taxonomy project (www.tolweb.org). These searches found evidence of entrainment in parrots, and so in the third step we followed up with additional more specific searches for groups in which we found evidence of entrainment (cockatoo, macaw, lovebird, parakeet, budgerigar, African grey).

Lastly, we also performed a number of theoretically-driven searches, including: great ape species, who are closely related to humans but vocal nonmimics; additional vocal mimicking species such as elephants; and for species which are closely related but differ in their status as vocal mimics (e.g. oscine and suboscine bird species; see Table S1 for all exact search terms).

Designation of Species as Vocal Mimics

The vocal mimicry hypothesis predicts that the ability to modify vocal output to more closely resemble auditory input should be the necessary precondition for entrainment. We therefore define mimicry broadly as the ability to learn vocalizations from an external auditory model, since the predictions of the vocal mimicry hypothesis provide no theoretical reason to use a more narrow definition of vocal mimicry. The definition of vocal mimicry used here thus includes species from many sub-types of vocal mimic: those that can learn from only conspecifics, as well as those that learn sounds outside their species' repertoire; those that learn only during a sensitive period as well as open-ended learning species, which mimic new sounds even as adults. The term "vocal mimic" was chosen over the broader term "vocal learner" as the latter is often taken to also include other types of production learning (e.g. producing novel sounds) and contextual learning (e.g. deploying a specific vocalization in a new context) [3] which are not implied in the current work.

Categorization of species as vocal mimics or nonmimics was done as accurately as possible given the current literature. In the case of elephant species, for example, vocal mimicry has been demonstrated in African elephants [4]. Although there is no published data on vocal mimicry in Asian elephants (the species for which we obtained video material), there are multiple anecdotal reports of mimicry in this species [5-7] (as well as reports of observational learning of novel sound production methods [8,9]). In one case, the speech-like vocalizations of an Asian elephant were reportedly analyzed via spectrogram and found to specifically resemble the vocalizations of the elephants' keeper [7]. These anecdotal reports, combined with the existence of published data supporting mimicry in a closely related species (African elephants), led us to classify Asian elephants as capable of vocal mimicry.

Certain species for which no relevant data exist were not given a designation, as there was no way to determine their capacity for vocal mimicry from the literature. Searches for species without a designation resulted in only four videos featuring an animal. None of these four videos contained music or rhythmic movement, and as such could not have contained entrainment.

Categorization of Videos

Each video was categorized along multiple dimensions to determine whether it should be further evaluated for entrainment or ruled out (as it could not contain entrainment), and excluded from more detailed analysis. These ratings were performed by a single coder and then verified through reliability testing (see below).

First, we excluded all videos that were identical repeats of videos found in earlier searches. This led to the exclusion of 1436 videos (out of 5315 total; see Supplemental Spreadsheet for information on each video by search). Second, we excluded all videos that did not contain any nonhuman animal. This led to the exclusion of 2860 videos.

Third, we excluded all videos in which there was no periodic movement by the animal or in which the periodic movement was fake. In order to count as periodic, an evenly-spaced movement must have occurred at least four times concurrently. Fake periodic movement included videos in which: the animals' movement was directly caused by a human; the animals' movement was indirectly caused by a human (following movement of a light or food), the periodic movement was constructed via videoediting, and the periodic movement appeared to be part of an extraneous activity (mating or convulsion), and unrelated to the auditory stimulus. These criteria resulted in the exclusion of 497 videos.

Fourth, we excluded all videos where the audio track appeared to be edited in afterwards. The following criteria were used to determine if audio was edited in: 1) the lack of normal background/environmental noise, 2) pristine audio quality unlikely to have come from a live microphone, 3) comments by the video's poster stating that the music had been added. In addition, videos in which a human was producing the audio live while attending to the animal were excluded, as in this case it was impossible to tell whether the human was simply entraining to the animal's movements, and not vice versa. These criteria resulted in the exclusion of 320 videos.

Fifth, because the remaining number of videos was still too large to examine frame by frame (202 videos), we excluded all videos that did not seem to have any potential to be entrained. This was based on a coder's judgment as to whether the animals' movement was potentially aligned with the auditory rhythm. A random 10% of these ratings were recoded for reliability by a blind second coder, and ratings were found to agree on 96.9% of trials. In the small number of trials that showed disagreement, this was due to inclusion of more files for further analysis by the main coder, not excessive exclusion. Thus, the determination of exclusion was made very conservatively; any video that seemed ambiguous or could possibly show entrainment was included for further analysis. This criterion resulted in the exclusion of 132 videos.

Finally, all videos that had not been excluded (70 videos) were analyzed frame-by-frame using the methods detailed below. Thus, videos were analyzed if they were not a repeat video, had an animal in them, had both rhythmic audio and rhythmic movement, were not obviously faked, and were potentially entrained. The videos were downloaded using www.KeepVid.com, and converted to .mov or .avi video format using www.media-convert.com. Most videos featured unique subjects: The 70 videos that were analyzed frame-by-frame contained 67 unique individuals. There were two subjects with multiple videos: One subject was featured in two videos, and another was featured in three.

Common Methods

Human Beat Perception Data

To determine the frequency of the musical beat along with the range of error present in human entrainment, human observers were instructed to tap on the space bar of the keyboard to the beat for each song with no visual stimulus present. Taps were recorded using Matlab with the Psychophysics Toolbox [10,11]. We computed the modal inter-tap-interval for each human subject on each song (binned into 10ms intervals), and fit a maximum likelihood Gaussian to the histogram of these modes to get an idea of the range of frequencies at which individual humans tapped.

The videos were divided up into seven sets of videos, each designed to last approximately a half-hour to avoid fatigue by human subjects. A total of 50 human observers tapped; each subject tapped with a subset of the musical samples. Each set of videos was coded by the following number of subjects: Set 1 (S1 and S2 videos): 8 subjects; Set 2 (Videos from Search Step 1 of YouTube data set): 8 subjects; Set 3 (Videos from Search Step 2 of YouTube data set): 8 subjects; Set 4 (First subset of videos from Search Step 3 of YouTube data set): 8 subjects; Set 5 (Second subset of videos from Search Step 3 of YouTube data set): 10 subjects; Set 6 (Third subset of videos from Search Step 3, and Search Step 4 from YouTube data set): 8 subjects; Set 7 (Tamarin stimuli): 8 subjects.

Nonhuman Subject Video Analysis: Preprocessing

Tracking Subjects' Location: To analyze the animals' movement we first performed a manual frame-by-frame coding of the subjects' head locations (using Matlab with Psychophysics Toolbox [10,11]). All videos were coded without audio, and thus blind to the tempo of the musical beat.

Due to limitations on video quality and low contrast between the bird and its backdrop (especially in videos from the online database data set), it was necessary to code the point on the subjects' head that was most easily visible throughout the video. Coding location was kept within a small area (the top of the head, the eye, the top of the beak). For example, if the bird was white (e.g. a cockatoo) against a dark background, the point used was the top of the head or the eye. If the bird was white against a white background, the top-center of the beak or the eye was used. Since all of sections of the head tend to move together the vast majority of the time, slight differences in coding location would not have affected the movement data (see Supplemental Videos). Automated tracking of subject 1 and 2 and videos from online database was not possible due to camera movement and visual complexity of the background scene.

To evaluate the reliability of manual coding, the first 1000 frames of two randomly chosen videos were re-coded. The correlations between the first and second coding were extremely high (dancing+cockatoo4, X: $r=.9685$, Y $r=.9961$; dancing+african grey1, X: $r=.9860$; Y: $r=.9852$). Error in pixels was low (mean difference = 4 pixels; median difference = 2 pixels), especially in comparison to the magnitude of the subjects' movements, which were on the order of 50 to 100 pixels.

Automated tracking was made possible for the tamarin subjects by taking still-camera shots on a plain contrastive background. To ensure that the automated tracking was equivalent to the manual coding, two randomly-selected tamarin subjects were coded both manually and automatically. The automated methods correlated highly with the manual codings (all r values > 0.92), and so the remaining tamarin subjects were coded automatically. Automated coding allowed us to maximize our ability to find evidence of entrainment in the tamarins, since we were able to track more tamarins and larger segments of video, and to track them using two different measures.

Two automated methods were used in order to maximize our ability to find rhythmic movement in tamarin subjects. The first, simplest method was to subtract subsequent frames from each other and compute the mean-squared difference between them. Since the background was completely stable, this gave an estimate of the gross amount of motion of the tamarin. For the second method we took each frame of the video and subtracted out the (known) static background. We then thresholded the remaining image, in order to isolate the tamarin. We then took the thresholded image and fit a maximum likelihood 2D Gaussian, and used the mean of the Gaussian as an index of the tamarin's current location. This allowed us to track the movement of the tamarin's body location specifically, as opposed to simply the total amount of movement.

Correction for 2-D Camera Movement: For sessions in which camera movement may have occurred, a location in the background was chosen to serve as a landmark, and coded for X-Y coordinates in each frame. This value was subtracted from the location of the subjects' head in each frame in order to correct for 2-D camera movement.

Criteria for Exclusion of Video Segments: To isolate the relevant segments of video for analysis, we a-priori excluded certain areas of video from analysis. In order to identify segments of video that fit these criteria, three independent raters watched the videos and identified segments of video that fit each criterion. These ratings were found to agree on 93.3% of frames.

When at least two out of three raters agreed that a frame fit the criteria for exclusion, that frame was excluded from analysis. Areas of video were excluded for the following reasons: the music had not started yet, or had already ended; the subject was engaged in unrelated behaviors, limited to preening feathers, sharpening the beak on the wooden perch, or physically interacting with another individual; the subject's movement was generated by human propulsion; the camera moved in a 3-D manner or zoomed in or out, thus conflating camera movement with the subject's movement. We also excluded segments of the video that did not contain movement which appeared periodic. This was done by manual inspection of the movement function, and confirmed via reliability testing (a second coder agreed with the main coder's judgements on 94.5% of frames, in a randomly selected 10% of the trials).

Nonhuman Subject Video Analysis: Frequency Analysis

Checking for Periodic Movement: After frame-by-frame coding of location, we took the derivative of this function (using the standard approximation to the derivative with discrete samples, the difference between adjacent samples) to obtain the speed of the subjects' movement. We then took the autocorrelation of this function in order to determine if the subject was moving periodically. We defined evidence of periodic movement as the following characteristics in the autocorrelation: 1) a single frame offset should result in a significant and positive autocorrelation (except in the case of rhythmic movement over 4 Hz, which is at a rate where a single frame offset is not expected to be positive) 2) offsets of additional frames after the end of the positive correlations should show significant negative correlations (at least one frame offset should be negative and significant); 3) offsets of additional frames after the end of the negative correlations should show significant positive correlations once again (at least one frame offset should be significant and positive). Sessions in which the autocorrelation did not conform to this rule (and thus did not show periodic movement) were not subjected to further analysis (see Table S2 for list of all videos included for further analysis). Including only the videos that showed quantitative evidence of periodic movement lowered the possibility of false positives, i.e. that a video without true entrainment would appear to show evidence of entrainment by chance, via a chance matching of modal frequency. This procedure resulted in the exclusion of 21 videos.

Did the Periodic Movement Occur by Chance? If the autocorrelation showed periodic movement, we then performed a Fourier transform to identify the extent to which the subject was moving at each frequency. Periodicity can be defined simply as movement at a consistent frequency. If the subject was moving at a consistent frequency, we saw a peak at the frequency range most present in the movement. We used a Monte Carlo simulation to determine whether the consistency of the subjects' movement frequency – and thus any peak in the Fourier spectrum -- was unlikely to have occurred by chance ($p < .05$).

In the Monte Carlo simulation, we modelled the null hypothesis that the subjects' periodicity (consistency of frequency) was due to chance. Specifically, we modelled the null hypothesis that each simulated subject 1) was moving with the same pattern of motion as the real subject (e.g., head-bobs), and 2) completed the same total number of units of motion over the course of the simulated session as the subject did in the real session, but 3) that the period of each unit could be any length (length was constrained to a minimum of 5 frames, and to the maximum length that would allow all of the needed head-bob units to fit into the simulated session). We generated data of this type for 5000 simulated subjects, took the Fourier transform of each one, and analyzed the height of the highest peak. We restricted our threshold for

statistical significance to those peak magnitudes attained from the simulated null-hypothesis data less than 5% of the time ($p < .05$). Note that by constraining our simulation to produce the same number of head-bob units, we were in effect running a highly conservative significance test, since performing the same number of head bobs as the bird in the same time severely constrained the possible frequencies of movement.

Does the Animals' Frequency of Movement Match the Musical Beat? We compared the peak frequency in the animals' data for each trial to the range of modal frequencies of human tapping to the same set of auditory stimuli. We checked whether the peak frequency of the animals' movement fell within the range of frequencies identified as the musical beat by human subjects (i.e. within 2 SD of the mean in the maximum likelihood human Gaussian). Anything within this range was counted as matching in frequency.

Musical meters consist of multiple hierarchical levels and, as expected, human subjects often differed from one another in the hierarchical level at which they chose to tap. If the animals were producing human-like entrainment behavior, they could potentially move at a variety of different multiples-of-two of a tempo (e.g. double speed, half speed), any of which are a correct musical beat (Note that this does not imply that the animals have a sense of the metrical hierarchy; for this they must move at multiple metrical levels). To detect entrainment across multiple metrical levels in this analysis, we needed to multiply the human data by the factor of two that placed each subject (a) at the closest metrical level to the animals' peak frequency, and (b) at the same metrical level as the other human subjects.

For this analysis, each individual human subject's data was multiplied or divided by the factor of two that put it closest to the value of the animals' peak frequency. If there was a bimodal distribution of human subjects after this procedure, those subjects at the less-common multiplier were shifted to the more common multiplier, in order to place all subjects' data at the same hierarchical level, thereby avoiding artificial expansion of the range of acceptable frequencies. Outliers (human subjects whose modal inter tap intervals were greater than 2 SD from the mean) were then excluded. In one video, human subjects failed to agree on the music's tempo (dancing+macaw50 from the online database; the final Gaussian spanned a range of over 1 Hz); this video was excluded from further analysis.

Both vertical and horizontal movements were analyzed. Because subjects appeared to be intentionally moving in one dimension (making the other an unintentional artifact), we selected one dimension of movement in each video as its primary dimension. This was done using properties of the Fourier transform, based on the following rule: If one dimension showed significant rhythmic movement, and the other did not, the significant dimension was selected. If both were significant or both were not significant, the dimension closest in peak frequency to the human data was selected.

Nonhuman Subject Video Analysis: Phase Analysis

In addition to the frequency analyses, we asked whether the movements maintained a consistent phase relationship with the musical beat as perceived by human subjects or by an automated beat tracker [12]. Maintaining a consistent phase over a long period of time is extremely unlikely without the real-time error correction of entrainment: If subjects do not actively realign their motor movements with the external beat, they continuously slip out of phase with the musical beat as motor error accumulates.

For the human observers' and the autotracker's data, we divided each inter-tap-interval into bins ranging from $-\pi$ to π (measured in radians), and recorded the bin into which each of the

bird's stopping points fell (i.e., the point at the top or bottom of the rhythmic motion; see 'Designation of Zero-Phase Locations' below). We subjected these data to the Rayleigh test with unspecified mean direction [13] to compare the locations of subjects' turning points to the distribution of turning point locations expected if they were occurring at a random phase relative to the musical beat.

We then asked whether the animals' movements were actually phase-matched (synchronized) with the musical beats. To detect phase matching, we used a modified version of the Rayleigh test [13], this time specifying a mean direction of zero (identical phase). This version of the Rayleigh test asks whether the temporal distribution of the birds' turning points is not only more consistent than expected by chance, but also more aligned with the phase of the music than expected by chance.

Consistent with previous literature, we found that even intentionally entrained human subjects contained slight phase differences from one another and from the autotracked beats [14]. These differences reflected the range of phases produced in human entrainment. As such, consistency of the nonhuman subject with any of the human subjects or autotracker counted as consistent phase, and synchrony with any of the human subjects or autotracker counted as synchronized.

Designation of Zero-Phase Locations

For human tapping data, the key-press time clearly designates the location of the perceived musical beat; this temporal location is zero-phase. The nonhuman subjects' data is continuous. However, there is an a priori reason to expect nonhuman subjects' zero-phase to be at specific locations: human data on freely performed movements.

When humans align free movements to a musical beat, they tend to align the locations of their changes in direction of movement with the beats (points with a speed of zero, either the maximum or minimum location). For example, in a conductor's movements the beat locations are always aligned with changes in direction of movement, never at the maximum speed of movement or other locations in the movement. We thus assigned as a possible zero phase these points of change in direction, what we call "stopping points" or "turning points" in the subjects' motion. As humans seem equally likely to align either the maximum or minimum location, we analyzed both of these possibilities in the nonhuman subjects, asking if either the location minima or the location maxima maintained a consistent and zero-phase relationship to the musical beat (as noted above).

Details on Interpretation of Analyses

Matched Modal Frequency versus Consistency of Phase versus Phase Matched

Using different analytic techniques, we were able to provide convergent evidence of entrainment and increased power to detect evidence of entrainment. However, it is possible for these measures to provide seemingly divergent results, with one measure showing evidence of entrainment while the other does not. This apparent disagreement is not a conflict but occurs because these measures describe slightly different facets of the animals' behavior. When an individual shows evidence of entrainment by one measure, but not on another, this is mostly due to (1) differences in statistical power between tests; (2) very short total video length, and thus a small amount of available data; or (3) the length of the bouts of entrainment within the video relative to the video's total length.

1) Differences in statistical power can lead to detection of consistency of phase in the Rayleigh test with specified mean direction, but not in the Rayleigh test with unspecified mean direction. The former version of the test detects both consistency and synchrony of phase, and has more statistical power than the unspecified-direction version of the Rayleigh test. It thus may detect consistency even when the unspecified-direction Rayleigh test failed to detect it.

2) Very short total video length, and thus a small amount of available data, can lead to detection of entrainment via matched modal frequency, but failure to detect entrainment by consistency of phase.

Detection of significant phase consistency serves as strong evidence of entrainment for reasons detailed in the main text. However, when trials or videos are extremely short, this measure may fail to detect entrainment in trials where it truly exists. This is because when only a few musical beats are available to analyze, it becomes more difficult to rule out alignment by chance. For some videos of the YouTube data set, the video data available were brief. In these cases, this measure becomes extremely stringent and it may become impossible to reach significance on this measure without more data. Therefore, a failure to maintain a significantly consistent phase relationship does not entirely rule out entrainment for short video clips such as some of those present in the YouTube data set.

3) A short length of the bouts of entrainment within the video relative to the video's total length can lead to detection of entrainment by the consistency of phase analysis, but lack of detection via the matched modal frequency analysis.

When *modal* frequency of movement throughout the entire session matches the beat, this implies that *most of the time* that the animal was moving, it was moving at the correct frequency. Thus, when this measure detects a match, it suggests entrainment. However, an animal could be entrained and not show a modal frequency match if it is entrained for only a subset of time in the trial. For instance, Patel and colleagues' additional data on S2 suggests that entrainment occurs in "bouts" -- the subject will synchronize with the beat for a number of seconds, then fall out of synchrony for a few seconds before re-establishing alignment for another bout [2]. We analyzed the *most common* frequency throughout the whole session; the relative length of each synchronized and nonsynchronized bout will affect this measure. If the majority of nonentrained movements occurred at a specific, nonmatching frequency, and this frequency was more common overall during the session, then overall modal frequency would not match, even if statistically significant entrained 'bouts' occurred during some portion of that trial.

Thus, a short video in which the animal moves at the correct frequency may show evidence of entrainment based on modal frequency while consistency of phase does not reach significance due to lack of sufficient length. In addition, a video in which entrainment occurred in short bouts that formed only a subset of the total movements in the video may show significant consistency of phase, but not matched modal frequency. Therefore what seems to be a discrepancy between the two measures is actually a description of two different facets of the data, each of which can provide evidence of entrainment but not detect it in every case. Together, however, these measures provide sufficient power to reliably find evidence of entrainment should it exist in a particular video.

The Effect of Noise

Some of the videos in our sample have relatively low frame rates and have undergone compression. These factors will add noise to our estimate of the animals' modal frequency, creating the possibility of false detections of entrainment via the matched modal frequency

measure. However, this concern does not arise for the statistical tests we employ, which are specifically designed to be robust against this kind of noise in the data. For example, the value of the Rayleigh test depends explicitly on the amount of noise in the data in order to prevent such false alarms (as does the Monte Carlo simulation for assessing consistency of frequency). The more noise in the data, the more variable our estimate will be of the subjects' stopping points. This variance will decrease the value of the Rayleigh test statistic, and thus we will not find significant entrainment in these cases. In order to test this directly, we ran a simulation of the effects of such noise on the significance of the Rayleigh test.

We generated 1,000 time series of length 10 seconds with a frame rate (sampling rate) of 60 frames/sec, and an underlying periodic movement at a random frequency between 1Hz and 3Hz. We then used the Rayleigh test to examine their rate of matching (falsely) with a periodic signal at 2 Hz. We then downsampled these signals to a rate of only 15 frames/sec and added independent noise to each time point, and once again examined the number of false alarms. We then re-ran this entire simulation 100 times, keeping a count of the number of false alarms before and after downsampling (where we considered a false alarm any signal that generated a significant Rayleigh test but whose underlying frequency was in fact >2.02 or <1.98 Hz). The mean number of false alarms was 0.057 for the original nonnoisy high frame rate signals and 0.053 for the down-sampled and noisy signals. The simulation thus indicates that noise of this type does not create false alarms when using the Rayleigh test.

Supplemental Results

Human Tapping Data: Results

Please see Table S2, column 3 for descriptive statistics regarding the human tapping data.

Case Studies: Results

Subjects 1 and 2

Data for Subject 1's 120 bpm trials are detailed in the main text, as are all data for Subject 2. As noted in the main text, Subject 1 did not show periodic movement in response to the 150 bpm stimuli (as evidenced in the autocorrelation functions, which did not contain the characteristics of periodic movement detailed above). Importantly, this demonstrates that Subject 1 does not simply respond to all musical stimuli with the same behavior (although it may still be the case that Subject 1 may have a preferred tempo of movement of 120 bpm). The lack of response at 150 bpm may further indicate that this tempo was too fast to evoke spontaneous periodic movement in this subject.

Tamarins

The tamarins did not appear to react to the musical beat, and showed no periodic movement visible to human observers. Fourier transforms showed that most frequencies present in tamarin subjects' movements were very low, with peak frequencies never near the frequency of the musical beat as identified by human subjects (peak frequency in the Fourier transform was below 0.55 Hz in all but one minute-long session of one subject). Because our analyses are designed to

test whether discrete periodic movements are entrained to a musical beat, they require the presence of discrete periodic movements (such as head-bobs). The tamarins did not display discrete periodic movements; as such the Monte Carlo simulation and the phase analyses (Rayleigh tests) could not be performed on the tamarin data. Because no possibility of entrainment exists without periodic movement, the tamarins thus failed to show evidence of entrainment.

Online Database Analyses: Results

For information on all search results, see Table S1 and Supplemental Spreadsheet. For motion analysis of all analyzed videos, see Table S2. These data are summarized in the main text. The 49 analyzed videos are available at <http://www.wjh.harvard.edu/~amschach/dataset/>.

The average length of the 49 motion-analyzed videos was 96.73 s, with a SD of 84.14 s and range of 21.76 to 391.6 s (see Figure S1). The mean frame rate for these videos was 22.95 frames per second (fps), with a SD of 7.70 fps, and a range of 10 to 30.3 fps; the most common frame rate was 30 fps (see Figure S2).

Supplemental Discussion

Precise Definition of Entrainment

The core property of entrainment is real-time error correction of movements to realign with the external stimulus. However, entrainment as a behavioral phenomenon appears under multiple definitions depending on the field of inquiry [15]. The current study is concerned with the question of whether a phenomenon similar to auditory/motor entrainment in humans exists in other animals, and thus the definition of entrainment used describes the phenomenon of auditory entrainment in humans [14]. For animals to entrain in a qualitatively similar way to humans, their behavior must have the following characteristics:

- 1) The period of the movement must match the period of the external auditory beat. If the period of the external beat changes, the period of animal's movement must change to match. This change must be due to real-time processing of the external auditory signal. Thus, two birds producing synchronized but temporally inflexible, innately fixed motor/vocal patterns do not demonstrate entrainment. Note that this definition is inclusive of the case of mutual entrainment (e.g. two drummers entraining to one another), in which individuals may intermittently switch between serving as follower and leader (the source of the external auditory beat to which the other entrains).

- 2) The subjects must model their movement off of an auditory stimulus, not simply follow a correlated rhythmic visual stimulus, such as human movement.

- 3) The animals must be able to generalize entrainment to novel auditory stimuli. Simply responding to specific acoustic elements in an auditory signal as a result of behavioral conditioning is not sufficient, as it is not qualitatively similar to human musical entrainment.

Additional Factors Necessary for Entrainment

Additional factors may be necessary in conjunction with vocal mimicry to enable entrainment. Other capacities that may play a role include complex social structure, social attachment to humans, movement imitation, and/or open-ended vocal learning [16]. Our data support the idea

that vocal mimicry is necessary but not sufficient, as we did not see evidence of entrainment in all vocal mimicking individuals or all species, and the majority of entrained species were parrots. While this is at least partially an effect of lower representation of other vocal mimics on the database, the distribution of entrainment in the data set may serve to guide future work investigating other capacities or experiences necessary for entrainment.

Multiple theories of the evolution of entrainment have been proposed, many hypothesizing direct selection for entrainment for enhancing group bonding, affiliation or cooperation [17-21]. Entrainment seems to confer social benefits in humans [21]; however this idea is not necessarily in conflict with the idea that vocal mimicry is necessary for entrainment, nor does not necessarily lead to the conclusion that entrainment was directly selected for in humans: it is still possible that part of the necessary mechanisms for entrainment emerged as a byproduct of selection for vocal mimicry.

Supplemental References

1. McDermott, J., Hauser, M. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition* 94, B11-B21.
2. Patel, A.D., Iversen, J.R., Bregman, M.R., Schulz, I., Schulz, C. (2008). Investigating the human-specificity of synchronization to music. In Proceedings of the 10th Intl. Conf. on Music Perception and Cognition, M. Adachi et al., eds., August 2008, Sapporo, Japan. Adelaide: Causal Productions.
3. Janik, V.M., & Slater, P.J.B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1-11.
4. Poole, J.H., Tyack, P.L., Stoeger-Horwath, A.S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature* 434, 455-456.
5. Sohn, E. Elephant mimics. <http://www.sciencenewsforkids.org/articles/20050330/Note3.asp> (2005).
6. Elephant Said to Speak. http://www.livescience.com/strangenews/060908_ap_elephant_speak.html (2006).
7. Amusement park boasts talking elephant: Experts examining prolix pachyderm. http://english.hani.co.kr/arti/english_edition/e_entertainment/155313.html (2006).
8. Wemmer, C., & Mishra, H. (1982). "Observational learning by an Asiatic elephant of an unusual sound production method." *Mammalia* 46, 556-557.
9. Wemmer, C., Mishra, H., & Dinerstein, E. (1985). "Unusual use of the trunk for sound production in a captive Asian elephant: a second case." *Journal of the Bombay Natural History Society*, 82, 187.
10. Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision* 10, 433-436.
11. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10, 437-442.
12. Ellis, D. (2007). Beat tracking by dynamic programming. *J. New Music Research*, 36, 51-60.
13. Fisher, N. I. (1983). *Statistical analysis of circular data*. (Cambridge, U.K.: Cambridge Univ Press).
15. Clayton, M., Sager, R., & Will, U. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. *ESEM Counterpoint 1*, 3-75.
14. Repp, B.H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969-92.
16. Patel, A.D., Iversen, J., & Bregman, Micah. (in press). Studying synchronization to a beat in nonhuman animals. *Ann. N.Y. Acad. Sci.*
17. Huron, D. (2001). Is music an evolutionary adaptation? *Ann. N.Y. Acad. Sci.* 930, 43-61.
18. Bispham, J. (2006). Rhythm in music: What is it? Who has it? And why? *Music Perception*, 24, 125-134.
19. Merker, B. (2000). Synchronous chorusing and human origins. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 315-327). Cambridge, MA: MIT Press.
20. Cross, I. (2001). Music, cognition, culture and evolution. *Ann. NY. Acad. Sci.* 930, 28-42.
21. Wiltermuth, S.S. & Heath, C. (2008). Synchrony and Cooperation. *Psychol. Sci.* 20, 1-5.

Table S1. List of All Search Terms by Search Step

Search Step	Search Terms	Reason for Search	Number of Results (Videos)
1	animal+dancing	Human contact	7
1	bear+dancing	Human contact	3
1	camel+dancing	Human contact	9
1	cat+dancing	Human contact	23
1	chicken+dancing	Human contact	3
1	cow+dancing	Human contact	0
1	dancing+bird	Human contact	34
1	dog+dancing	Human contact	38
1	duck+dancing	Human contact	6
1	ferret+dancing	Human contact	42
1	fish+dancing	Human contact	18
1	goat+dancing	Human contact	12
1	goose+dancing	Human contact	5
1	hamster+dancing	Human contact	3
1	horse+dancing	Human contact	37
1	lizard+dancing	Human contact	13
1	mouse+dancing	Human contact	3
1	pig+dancing	Human contact	1
1	rabbit+dancing	Human contact	11
1	rat+dancing	Human contact	8
1	sheep+dancing	Human contact	5
1	snake+dancing	Human contact	11
1	squirrel+dancing	Human contact	18
1	turtle+dancing	Human contact	6
2	dancing+owlet	Order of Neoaves (Aegothelidae)	0
2	dancing+hummingbird	Order of Neoaves (Apodiformes)	12
2	dancing+nightbird	Order of Neoaves (Caprimulgiformes)	0
2	dancing+gull	Order of Neoaves (Charadriiformes)	18
2	dancing+stork	Order of Neoaves (Ciconiiformes)	8
2	dancing+mousebird	Order of Neoaves (Coliiformes)	0
2	dancing+pigeon	Order of Neoaves (Columbiformes)	25
2	dancing+kingfisher	Order of Neoaves (Coraciiformes)	1
2	dancing+cuckoo	Order of Neoaves (Cuculiformes)	0
2	dancing+turaco	Order of Neoaves (Cuculiformes)	0
2	dancing+hawk	Order of Neoaves (Falconiformes)	0
2	dancing+grouse	Order of Neoaves (Galliformes)	15
2	dancing+loon	Order of Neoaves (Gaviiformes)	2
2	dancing+crane	Order of Neoaves (Gruiformes)	7
2	dancing+hoatzin	Order of Neoaves (Opisthocomiformes)	0
2	dancing+sparrow	Order of Neoaves (Passeriformes)	2
2	dancing+booby	Order of Neoaves (Pelecaniformes)	4
2	dancing+flamingo	Order of Neoaves (Phoenicopteriformes)	13
2	dancing+woodpecker	Order of Neoaves (Piciformes)	3
2	dancing+grebe	Order of Neoaves (Podicipediformes)	2

2	dancing+albatross	Order of Neoaves (Procellariiformes)	28
2	dancing+parrot	Order of Neoaves (Psittaciformes)	37
2	dancing+penguin	Order of Neoaves (Sphenisciformes)	3
2	dancing+owl	Order of Neoaves (Strigiformes)	9
2	dancing+trogon	Order of Neoaves (Trogoniformes)	0
3	dancing+african grey	Species seen entrained in Step 2	42
3	dancing+budgerigar	Species seen entrained in Step 2	30
3	dancing+cockatoo	Species seen entrained in Step 2	30
3	dancing+conure	Species seen entrained in Step 2	44
3	dancing+lovebird	Species seen entrained in Step 2	37
3	dancing+macaw	Species seen entrained in Step 2	46
3	dancing+parakeet	Species seen entrained in Step 2	38
4	dancing+accentor	Oscine	0
4	dancing+apostlebird	Oscine	0
4	dancing+babbler	Oscine	1
4	dancing+batise	Oscine	0
4	dancing+berrypecker	Oscine	0
4	dancing+bird of paradise	Oscine	5
4	dancing+blackbird	Oscine	2
4	dancing+bluebird	Oscine	0
4	dancing+bowerbird	Oscine	1
4	dancing+bristlebird	Oscine	0
4	dancing+bulbul	Oscine	1
4	dancing+bunting	Oscine	0
4	dancing+butcherbird	Oscine	0
4	dancing+chickadee	Oscine	0
4	dancing+chough	Oscine	0
4	dancing+cisticola	Oscine	0
4	dancing+crow	Oscine	4
4	dancing+currawong	Oscine	0
4	dancing+dipper	Oscine	0
4	dancing+donacobius	Oscine	0
4	dancing+drongo	Oscine	0
4	dancing+fantail	Oscine	4
4	dancing+figbird	Oscine	0
4	dancing+finch	Oscine	7
4	dancing+flowerpecker	Oscine	0
4	dancing+flycatcher	Oscine (European flycatchers)	0
4	dancing+gnatcatcher	Oscine	0
4	dancing+grassbird	Oscine	0
4	dancing+grosbeak	Oscine	0
4	dancing+honeyeater	Oscine	0
4	dancing+hylia	Oscine	0
4	dancing+hyliota	Oscine	0
4	dancing+hypocolius	Oscine	0
4	dancing+jay	Oscine	0
4	dancing+kinglet	Oscine	0

4	dancing+lark	Oscine	0
4	dancing+logrunner	Oscine	0
4	dancing+longspur	Oscine	0
4	dancing+lora	Oscine	0
4	dancing+lyrebird	Oscine	3
4	dancing+magpie	Oscine	2
4	dancing+melampitta	Oscine	0
4	dancing+nicator	Oscine	0
4	dancing+nuthatch	Oscine	1
4	dancing+oriole	Oscine	0
4	dancing+oxpecker	Oscine	0
4	dancing+palmchat	Oscine	0
4	dancing+pardalote	Oscine	0
4	dancing+pipit	Oscine	0
4	dancing+pitohius	Oscine	0
4	dancing+puffback	Oscine	0
4	dancing+rhabdornis	Oscine	0
4	dancing+robin	Oscine	0
4	dancing+rockfowl	Oscine	0
4	dancing+scrub bird	Oscine	3
4	dancing+shrike	Oscine	0
4	dancing+sittella	Oscine	0
4	dancing+starling	Oscine	14
4	dancing+sugarbird	Oscine	0
4	dancing+sunbird	Oscine	0
4	dancing+swallow	Oscine	0
4	dancing+tanager	Oscine	0
4	dancing+thornbill	Oscine	0
4	dancing+thrush	Oscine	0
4	dancing+tit	Oscine	0
4	dancing+titmouse	Oscine	0
4	dancing+trecreeper	Oscine	0
4	dancing+vanga	Oscine	0
4	dancing+vireo	Oscine	0
4	dancing+warbler	Oscine	0
4	dancing+wattlebird	Oscine	0
4	dancing+waxwing	Oscine	0
4	dancing+weaver	Oscine	0
4	dancing+whistler	Oscine	0
4	dancing+wren	Oscine	0
4	dancing+antbird	Suboscine	7
4	dancing+asitie	Suboscine	0
4	dancing+bellbird	Suboscine	5
4	dancing+broadbill	Suboscine	12
4	dancing+cotinga	Suboscine	0
4	dancing+crescentchest	Suboscine	0
4	dancing+gnateater	Suboscine	0

4	dancing+miner	Suboscine	0
4	dancing+ovenbird	Suboscine	0
4	dancing+pitta	Suboscine	0
4	dancing+sapayoa	Suboscine	0
4	dancing+tapaculo	Suboscine	0
4	dancing+tityra	Suboscine	0
4	dancing+woodcreeper	Suboscine	0
4	dancing+manakin	Suboscine	8
4	dancing+sharpbill	Suboscine	0
4	bat+dancing	Theory (Vocal Mimic)	2
4	dolphin+dancing	Theory (Vocal Mimic)	14
4	elephant+dancing	Theory (Vocal Mimic)	27
4	seal+dancing	Theory (Vocal Mimic)	8
4	whale+dancing	Theory (Vocal Mimic)	12
4	chimpanzee+dancing	Theory (Vocal Nonmimic)	8
4	dancing+ape	Theory (Vocal Nonmimic)	2
4	dancing+orangutan	Theory (Vocal Nonmimic)	3
4	dancing+siamang	Theory (Vocal Nonmimic)	1
4	gibbon+dancing	Theory (Vocal Nonmimic)	0
4	gorilla+dancing	Theory (Vocal Nonmimic)	2
4	monkey+dancing	Theory (Vocal Nonmimic)	2
4	Sea lion+dancing	Theory (Vocal Nonmimic)	30
4	walrus+dancing	Theory (Vocal Nonmimic)	4

Column 4 shows number of search results (out of top 50) which included a nonhuman animal of the species specified by the search. Note: these searches also resulted in a total of 52 additional videos of animals, of a species other than that specified by the search. These videos are not listed here but were included and analyzed with the rest of the data set.

Table S2.

	Species	Frame Rate (fps)	Mean modal human tempo (stdev of modes) in bpm	Phase		Frequency	
				Consistent phase relationship (* = p<.05; with humans, autotracker)	Phase-matched (* = p<.05; with humans, autotracker)	Matched modal frequency (Z-score; * = Z-score < 2)	Consistent frequency (* = p < .05 under Monte Carlo simulation)
1	Asian Elephant	25	49.0 (2.3)	3/8 *	1/8 *	0.39 *	p>.1
2	Asian Elephant	30.3	65.1 (1.7)	1/8 *	0/8	-0.42 *	p>.1
3	Asian Elephant	10	20.1 (0.5)	3/8 *	3/8 *	2.29	p>.1
4	Asian Elephant	15	70.1 (1.5)	0/8	0/8	-1.99 *	p<.01 *
5	African Grey Parrot	30	127.4 (4.6)	0/9	1/9 *	0.68 *	.05<p<.1
6	African Grey Parrot	30	116.6 (3.8)	7/8, A *	0/8	1.81 *	p<.01 *
7	African Grey Parrot	30	120.3 (4.3)	1/6 *	1/6 *	4.96	p>.1
8	African Grey Parrot	13.04	119.1 (3.2)	0/8	0/8	1.56 *	p<.01 *
9	Blue and Gold Macaw	10	65.8 (1.4)	1/7 *	0/7	-9.61	p<.01 *
10	Blue and Gold Macaw	30	140.2 (4.8)	1/7 *	0/7	3.38	p<.01 *
11	Blue-Crowned Conure	30	110.9 (3.7)	2/4 *	1/4 *	-0.43 *	p>.1
12	Green Conure	30	126.3 (7.0)	1/9, A *	1/9 *	0.18 *	p>.1
13	Green-Winged Macaw	25	98.9 (3.7)	0/8	5/8 *	-6.18	p>.1
14	Green-Winged Macaw	15	79.1 (9.8)	0/10	0/10	-0.84 *	p>.1
15	Green-Winged Macaw	15	80.4 (3.4)	1/8 *	0/8	-1.77 *	p<.01 *
16	Hyacinth Macaw	30	121.0 (3.8)	0/7, A *	0/7	-6.77	p<.01 *
17	Indian Ringneck Parakeet	15	22.6 (0.4)	1/6 *	0/6	-0.85 *	p>.1
18	Moluccan Cockatoo	30	64.3 (2.7)	1/8, A *	1/8 *	-16.28	p>.1
19	Nanday Conure	14.29	122.6 (7.0)	1/10 *	0/10	6.28	p>.1
20	Nanday Conure	15	147.0 (6.1)	0/7	0/7	-0.09 *	p>.1

21	Peachface Lovebird	15	315.4 (15.9)	0/1	0/1	0.92 *	p>.1
22	Peachface Lovebird	15	28.9 (6.0)	1/10 *	0/10	0.72 *	p<.01 *
23	Quaker Parrot	30	175.4 (4.16)	0/1	1/1 *	2.80	p>.1
24	Sulphur-Crested Cockatoo	28.57	106.9 (1.3)	7/7, A *	1/7 *	0.89 *	p<.01 *
25	Sulphur-Crested Cockatoo	25	176.4 (5.1)	1/2, A *	0/2, A *	0.02 *	p>.1
26	Sulphur-Crested Cockatoo	12	34.5 (1.1)	1/9 *	1/9 *	0.73 *	p>.1
27	Sulphur-Crested Cockatoo	10	64.5 (2.1)	0/9, A *	0/9	-0.38 *	p>.1
28	Sulphur-Crested Cockatoo	25	180.6 (2.9)	0/2	0/2	-1.64 *	p<.01 *
29	Sun Conure	30	137.2 (2.6)	2/7 *	0/7	4.43	p<.01 *
30	Umbrella Cockatoo	11.54	103.7 (1.8)	0/6	3/6 *	1.39 *	p>.1
31	Umbrella Cockatoo	30	140.2 (3.0)	0/8	0/8	-0.83 *	p<.05 *
32	Umbrella Cockatoo	30	62.79 (1.5)	0/8	0/8	1.56 *	.05<p<.1
33	Umbrella Cockatoo	30	111.5 (2.4)	0/8	0/8	1.16 *	p>.1
34	African Grey Parrot	30	104.8 (2.8)	0/5	0/5	-3.43	p<.01 *
35	African Grey Parrot	15	130.5 (3.2)	0/5	0/5	-4.88	p<.01 *
36	African Grey Parrot	15	113.9 (3.9)	0/8	0/8	9.00	p>.1
37	Alexandrine Parakeet	30	128.8 (1.3)	0/8	0/8	5.20	.05<p<.1
38	Blue and Gold Macaw	25	112.5 (3.3)	0/5	0/5	6.08	p<.01 *
39	Catalina Macaw	30	166.1 (5.3)	0/5	0/5	-2.03	p>.1
40	Dog	30	180.5 (2.9)	0/3	0/3	-6.32	p>.1
41	Green Cheek Conure	15	13.5 (0.2)	0/10	0/10	8.12	p<.01 *
42	Hyacinth Macaw	30	122.8 (2.0)	0/5	0/5	-6.71	p>.1
43	Masked Lovebird	15	128.5 (2.2)	0/7	0/7	-2.22	p>.1

44	Orangutan	15	147.1 (3.1)	0/4	0/4	-2.87	p<.01 *
45	Orangutan	30	247.7 (4.6)	0/0	0/0	2.06	p>.1
46	Quaker Parrot	25	161.3 (5.3)	0/2	0/2	-3.24	p>.1
47	Sea Lion	30	34.2 (0.5)	0/7	0/7	2.55	p>.1
48	Sun Conure	30	106.0 (1.6)	0/9	0/9	-3.62	p>.1
49	Umbrella Cockatoo	25	165.2 (4.6)	0/3	0/3	-7.47	p>.1

Details of analyses from online database data set. 49 of the 70 videos analysed frame-by-frame showed evidence of periodic movement via autocorrelation analyses, and were thus included for further analyses of the phase and frequency of this periodicity (as described in the Supplemental Methods) and in the current table. Phase analyses: The Rayleigh test was used to determine if the animal's movements maintained a consistent phase relationship to any of the human subjects' tapping data or to an automated beat tracker; a different version of the Rayleigh test was used to determine if the animal's movements were significantly synchronized, or phase-matched with human subjects' tapping data or an automated beat tracker. This also implies significant consistent phase, however since the latter is a more powerful statistical test, it is possible for the synchrony test to be significant while the consistency-only test is not. Matching modal frequency: Z-score refers to location of modal frequency within distribution of human subjects' modal frequencies tapping to the same stimuli; < 2 SD was termed correct frequency. Consistent frequency: p-value refers to likelihood under Monte Carlo simulation; significance implies movement at a consistent frequency throughout the trial unlikely to have occurred by chance. Tempo refers to the mean of subjects' modal tap frequencies in beats per minute. Standard deviation, in parentheses, refers to the distribution of the subjects' modal tap frequencies. Frame rates are specified in frames per second (fps).

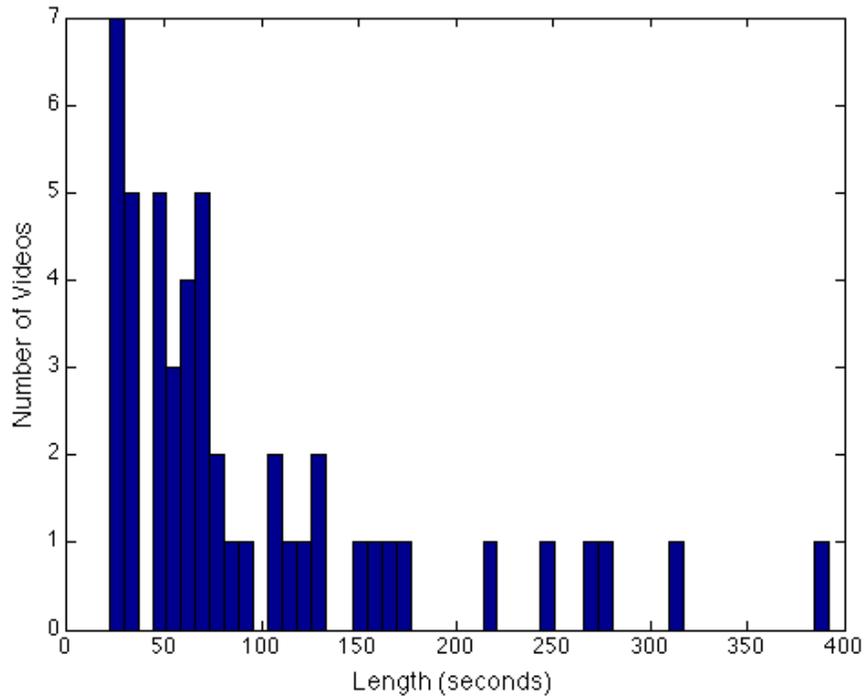


Figure S1. Histogram of the Lengths of the Motion-Analyzed Videos of the Online Database Data Set

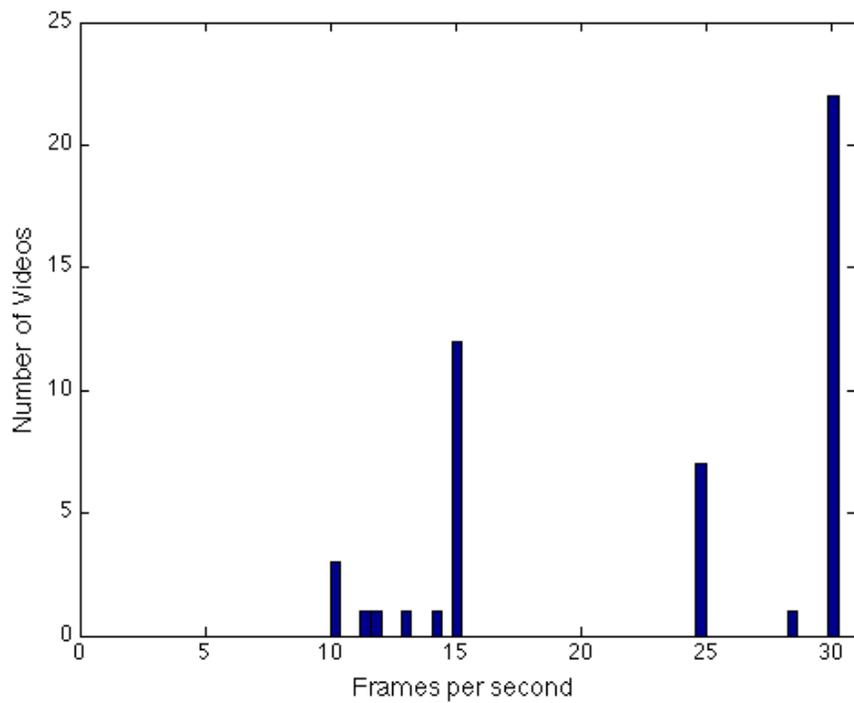


Figure S2. Histogram of the Frame Rates of the Motion-Analyzed Videos of the Online Database Data Set