Sensorimotor integration on a rapid time scale

Jinhong Luo,a,1 Ninad B. Kothari,a and Cynthia F. Mossa

*Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218

Sensing is fundamental to the control of movement. From grasping objects to speech production, sensing guides action. So far, most of our knowledge about sensorimotor integration comes from visually guided reaching and oculomotor integration, in which the time course and trajectories of movements can be measured at a high temporal resolution. By contrast, production of vocalizations by humans and animals involves complex and variable actions, and each syllable often lasts a few hundreds of milliseconds, making it difficult to infer underlying neural processes. Here, we measured and modeled the transfer of sensory information into motor commands for vocal amplitude control in response to background noise, also known as the Lombard effect. We exploited the brief vocalizations of echolocating bats to trace the time course of the Lombard effect on a millisecond time scale. Empirical studies revealed that the Lombard effect features a response latency of a mere 30 ms and provided the foundation for the quantitative audiomotor model of the Lombard effect. We show that the Lombard effect operates by continuously integrating the sound pressure level of background noise through temporal summation to guide the extremely rapid vocal-motor adjustments. These findings can now be extended to models and measures of audiomotor integration in other animals, including humans.

Sensing plays a critical role in the control of movement. In humans, natural behaviors, from grasping a coffee mug to producing intelligible speech, all rely on the guidance of sensing. Similarly, sensory signals lie at the core of most animal behaviors. However, the brain mechanisms underlying sensorimotor integration are not well understood. This is particularly true regarding the control of vocalizations, which are used by a wide range of animal species for communication. At present, a dominant model for motor control is derived from the state feedback control (SFC) theory, which successfully accounts for a wide range of motor behaviors, such as visually guided arm movement (1) and speech production (2). SFC models posit that motor control is based on a comparison of sensory prediction generated from an internal forward model with actual sensory feedback, and sensory feedback is used to train and update the internal forward model. According to this SFC model, sensory feedback is not directly used to guide motor commands, due to its noisy and delayed characteristics, and this notion has been supported by a large body of experimental and theoretical work (1–7). These stand in contrast to motor reflexes, which are movements in direct response to sensory signals. Note, motor reflexes can be graded in response magnitude and can be modulated by cognitive processes (8, 9). One well-known example is the pupillary light reflex, an adjustment in pupil diameter in response to light intensity (10).

The Lombard effect refers to an animal’s increase in vocal signal amplitude in response to an increase in background noise (11). Evidence of the Lombard effect comes from studies of mammals (including humans), birds, frogs, and fish (12–16). Despite over a century of research, the brain mechanisms of the Lombard effect remain elusive. At present, the dominant hypothesis is that the Lombard effect shares the same brain circuits underlying the Fletcher effect (13, 17), an increase in sound-level production in response to a reduction in auditory feedback amplitude (17). Under this hypothesis, the SFC model should account for the Lombard effect. In other words, the Lombard effect should operate through the comparison of predicted vocal amplitude from an internal forward model with the auditory feedback amplitude. Because background noise interferes with the feedback amplitude of an animal’s vocalizations, the magnitude of the Lombard effect grows stronger with increasing background noise. Here, to the contrary, we show that the Lombard effect is a direct response to background noise and features an extremely short latency of about 30 ms. Thus, the Lombard effect can be better described as a reflex.

Results

The echolocating big brown bat, Eptesicus fuscus, produces vocalizations of a few milliseconds in duration, which offers the opportunity to investigate the Lombard effect on a millisecond time scale and to generate data for a quantitative sensorimotor model that differentiates between SFC and reflex hypotheses. This modeling effort depends on detailed analysis of the time course of the Lombard effect. In this study, we trained individual adult bats (five individuals) to rest on a platform and track an approaching tethered insect by echolocation and recorded the bat’s echolocation signals with an array of 14 microphones (Fig. 1A). For each trial, the tethered insect was delivered to the bat via a motorized pulley system and traveled a distance of 3 m in about 2.5 s. The distance between the bat and the tethered insect over the time course of a trial is shown in Fig. 1B. One example of the bat’s sonar tracking behavior is shown in Fig. 1C and D. In the middle of a trial, around 1 s after the tethered insect began its trajectory toward the bat, either a silent sound file (with all samples set to zero) or a 5- to 100-kHz broadband white noise of varying temporal structure was randomly broadcast from a loudspeaker positioned 3.5 m away in front of the bat (Fig. 1A).

Significance

Humans and other animals continuously monitor sensory information to inform the selection of motor commands for adaptive behaviors. Acoustic communication, for example, utilizes auditory feedback to fine-tune vocal production parameters. Because most animal species produce vocalizations that last several hundred milliseconds, it is difficult to dissect the temporal dynamics of audio-vocal feedback control. Here we took advantage of the brief echolocation signals of bats and mapped out the time course of vocal adjustments to background noise on a millisecond time scale. The high temporal resolution data provided the foundation for the model of audio-vocal volume control. We discovered that temporal summation, a shared auditory process across the animal kingdom, lies at the core of adaptive vocal volume control.

Author contributions: J.L. and N.B.K. designed research; J.L. and N.B.K. performed research; J.L. analyzed data; and J.L. and C.F.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. Email: jlu18@jhu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702671114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1702671114

PNAS | June 20, 2017 | vol. 114 | no. 25 | 6605–6610
illustrated by the red bar at the bottom of each figure panel. These examples show that the peak amplitude of the calls in both 40-dB and 50-dB noise conditions exhibited a gradual increase after the noise started, and a gradual decrease after the noise ended. In contrast, there is little variation in the peak amplitude of calls in the silence control. The same phenomenon of noise-related changes in call amplitude can be seen in the scatter plots of the root-mean-square (RMS) call amplitude from all five bats per condition, after aligning the sonar calls in time relative to the noise onset (Fig. 2 E–H). The gray line in each panel represents the spline fitting, showing the average trend. To compare differences in the call amplitude over the time course statistically, we combined data points from each neighboring 50-ms time window. Fig. 2 I and J shows the mean ± SEM for each time window and each noise condition and the normalized mean ± SEM, respectively. The normalized means were calculated by taking the differences in means between the noise conditions and the silence control to account for the inherent time-related changes in call amplitude when bats were tracking the approaching insect. Before the noise was broadcast, there was little variation in call amplitude across conditions, with a maximum difference in the averaged call amplitude of 1.1 dB. After the noise started, bats produced gradually more intense calls until reaching a plateau. After the noise ended, the call amplitude decreased gradually and returned to the baseline level of silence control. The rising and the plateau (maximum) of the call amplitude depended on the noise level: the higher the noise level, the steeper the initial rise and the greater the maximum increase. Once the noise started again, the animals increased call amplitude again before dropping to the baseline amplitude.

**Time Course of the Lombard Effect on a Millisecond Scale.** In the first experiment, we presented the bats with 500-ms broadband noise at 30 dB, 40 dB, and 50 dB SPL (sound pressure level relative to 20 μPa), along with a silence control. One example of the echolocation behavior from each condition is shown in Fig. 2 A–D. The time course of the noise, relative to the sonar behavior of the bat, is indicated by a red bar at the bottom of each figure panel. These examples show that the peak amplitude of the calls in both 40-dB and 50-dB noise conditions exhibited a gradual increase after the noise started, and a gradual decrease after the noise ended. In contrast, there is little variation in the peak amplitude of calls in the silence control. The same phenomenon of noise-related changes in call amplitude can be seen in the scatter plots of the root-mean-square (RMS) call amplitude from all five bats per condition, after aligning the sonar calls in time relative to the noise onset (Fig. 2 E–H). The gray line in each panel represents the spline fitting, showing the average trend. To compare differences in the call amplitude over the time course statistically, we combined data points from each neighboring 50-ms time window. Fig. 2 I and J shows the mean ± SEM for each time window and each noise condition and the normalized mean ± SEM, respectively. The normalized means were calculated by taking the differences in means between the noise conditions and the silence control to account for the inherent time-related changes in call amplitude when bats were tracking the approaching insect. Before the noise was broadcast, there was little variation in call amplitude across conditions, with a maximum difference in the averaged call amplitude of 1.1 dB. After the noise started, bats produced gradually more intense calls until reaching a plateau. After the noise ended, the call amplitude decreased gradually and returned to the baseline level of silence control. The rising and the plateau (maximum) of the call amplitude depended on the noise level: the higher the noise level, the steeper the initial rise and the greater the maximum increase.

These data allowed us to estimate the response latency of call amplitude adjustment. For the averaged data in the time window between 50 and 100 ms after noise onset (referenced at 75 ms, Fig. 2 J), the bats produced 2.2 ± 0.7 dB higher-amplitude calls in the 50-dB noise condition (T test, P < 0.001, df = 162); whereas there was no change in call amplitude at 25 ms after noise onset (–0.35 ± 0.47 dB; T test, P = 0.62, df = 150), which suggests that the response latency lies between 25 and 75 ms.

**Real-Time Transfer of Background Noise into Vocal-Motor Control.** To test the hypothesis that the noise-induced call amplitude changes are a result of the bat’s real-time processing of the background noise and not simply triggered by the onset of the noise, we collected additional data using two paradigms in three additional experiments. In the first paradigm, we presented bats with noise of different durations: 50, 150, 250, and 800 ms. In the second paradigm, we presented bats with two repeated noise elements of either 150 or 250 ms duration, separated by a silence gap of either 10, 50, or 100 ms. The noise amplitude was 40 dB SPL across all these conditions. We found that the call amplitude is determined by both the duration of the noise and the duration of the silence gap (Fig. 3). Call amplitude adjustments in all noise conditions showed the same pattern of a gradual increase after the noise started, remaining constant if the noise was longer than about 300 ms and gradually decreasing after the noise ended. The gapped noise paradigm offers strong support for the real-time processing hypothesis, as not only the peak of the call adjustment time course was determined by the noise duration (as in paradigm 1), but also the valley of the time course was directly affected by the duration of the silence gap. The longer the silence gap, the greater the decrease in call amplitude. Once the noise started again, the animals increased call amplitude again before dropping to the baseline amplitude.

**A Computational Model for the Lombard Effect.** Here we present a computational model that accounts for the noise-induced changes in call amplitude (Fig. 4 A). The proposed model contains three free parameters: time constant (τ, in ms), response latency (L, in ms), and the maximum increase in call amplitude for a given noise level [A(tin), in dB]. The time constant determines the maximum integration time for the background noise. The response latency determines how fast the noise information is reflected by a change in call amplitude. The maximum increase in call amplitude represents the compensation goal for a given noise level.

This model works as follows. First, after the background noise enters the auditory system, the envelope of the sound pressure level is extracted. Second, the pressure envelope is temporally integrated over the effective noise duration, and this process is leaky. The effective noise duration equals the time elapsed after the noise began minus the response latency. Subsequently, the integrated pressure envelope is continuously transformed into call amplitude compensation if the noise continues. If the noise ends, the integrated pressure envelope follows an exponential decay function that shares the same time constant with the leaky integration process, and then the output is continuously transformed into vocal amplitude compensation (Fig. 4 A).
We used mean vocal-level adjustment data in the 500-ms noise condition of experiment 1 to search for the three free parameters for the model. (Note that median vocal adjustment values differed by less than 0.1 dB from the mean, on average; Table S1). Predictions based on the combinations of the three parameters that resulted in the best fit, i.e., the smallest sum of squared errors, are shown in Fig. 4B. The optimal time constant and response latency were 267 and 30 ms for the 40-dB noise condition; and 252 and 33 ms for the 50-dB noise conditions, respectively. Thus, the differences in the vocal amplitude compensation between the 40- and 50-dB noise conditions were driven nearly entirely by the third parameter, the maximum increase in call amplitude (5.9 dB vs. 9.8 dB). One important parameter returned by the model is the response latency of about 30 ms, which is remarkably short. Fig. 4C–J shows that the model accounts for the data from the other eight experimental conditions quite well. The absolute differences between the model prediction and the data for these conditions were 0.6 ± 0.45 dB. In other words, once the three parameters are established, the fine time course of the Lombard effect can be predicted precisely.

Discussion
Taking advantage of the echolocating bat’s adaptive sonar behavior, we have successfully traced the time course of the Lombard effect. Importantly, we have demonstrated that background noise level is continuously monitored to guide the vocal amplitude control, as revealed by the noise-gap paradigm, in which vocal amplitude adjustments were affected by both the duration of the noise element and the silence gap (Fig. 3). The success of our model in accounting for the fine time course of the Lombard effect offers theoretical support that background noise is directly used to guide vocal amplitude control. Specifically, our model clearly shows that neither internal forward model predictions nor a comparison of these predictions with background noise is required, as would be expected by the SFC models or by the Fletcher effect hypothesis. Furthermore, this study also revealed that transforming sensory inputs into vocal-motor control underlying the Lombard effect is extremely rapid, with a latency of about 30 ms.

How can the Lombard effect operate so rapidly? The Lombard effect response latency of 30 ms is even shorter than the auditory laryngeal reflex, an increase in fundamental frequency in response to the presentation of a click, with a latency of ∼50 ms in humans (18). Nevertheless, the 30-ms response latency is well within the physiological limit of audio-vocal integration, as the acoustic crycothyroid muscle reflex can be as short as 6 ms (19). Our model shows that the short response latency of the Lombard effect is probably due to the very few audio-vocal processes involved: Envelope extraction, leaky integration, and scaling. Because both envelope extraction and leaky integration can be accomplished by the peripheral auditory system (20, 21), it is likely that the initiation of the Lombard effect does not require higher auditory centers. The output from the peripheral auditory system can be directly sent to the vocal-motor system of the
brainstem to guide the vocal amplitude adjustments, as indicated by the finding that decerebrate cats show the Lombard effect (22).

Why have past studies overestimated the response latency for the Lombard effect? The Lombard effect was first described over a century ago and is among the most widely studied audio-vocal integration phenomena. The response latency of the Lombard effect was estimated to be about 150–175 ms for humans (23, 24), 150 ms for birds (25), and less than 150 ms for bats (26). There are at least two possible explanations for the overestimation. First, the duration of the vocalizations under investigation sets the finest latency measurement resolution. For the human and bird experiments (23–25), the vocalizations studied were about 150 ms long, which constrains the lower latency limit of the Lombard effect that can be measured. Second, estimations based on statistical assessment are by nature overestimations, because all experimental data are intrinsically noisy. For example, using the data from the 500-ms continuous noise experiment (experiment 1) we estimated the response latency for the Lombard effect to be between 25 and 75 ms because calls were $2.2 \pm 0.7$ dB more intense for the 75-ms call group statistically, but not for the 25-ms call group.

In addition to explaining the extremely short vocal response latency to noise, our model offers a mechanistic explanation of the Lombard effect that applies to all vertebrates. This stems from the fact that two basic auditory processes, envelope extraction and leaky integration, are fundamental to the vertebrate auditory system. Specifically, both envelope extraction and leaky integration underlie the process of temporal summation, which has been demonstrated for fish (27), frogs (28), birds (29), and mammals (20). Thus, the basic auditory system of vertebrates is wired to show the Lombard effect.
A computational model for the Lombard effect. (A) The model takes the white noise as the input, followed by envelope extraction, leaky integration, and scaling. There are three free parameters: the response latency (ms), time constant (ms), and maximum compensation (dB). (B) Data from the 500-ms continuous noise conditions were used to find the optimal combination of the three model parameters through minimizing the sum of squared errors. The corresponding model predictions are colored in gray. (C–J) A comparison of the model prediction, based on the three parameters optimized from the 40-dB SPL noise condition (267 ms of time constant, 30 ms of response latency, and 5.9 dB of maximum compensation), and the data. The model precisely predicts the time course of the call amplitude for all conditions. For B–J, the red bar at the bottom shows the time course of noise presentation.

30 ms, suggests that the Lombard effect can be described as a reflex. Thus, the Lombard effect is fundamentally different from the Fletcher effect and does not require a comparison of internal forward model predictions with the sensory feedback. Our results are in line with the data from an earlier study on another bat species (Rhinolophus ferrumequinum) that exhibited the Lombard effect in the first call after noise onset (26). However, the term reflex should not be interpreted as that the Lombard effect is a fixed response to background noise. In fact, the Lombard effect in humans is affected by both the communication context (30) and content (31) and can be inhibited through training (32). These studies highlight that cognitive processes can modulate the Lombard effect.

**Materials and Methods**

Five adult big brown bats (Eptesicus fuscus), four males and one female, were trained to rest on a platform and track an approaching tethered insect by echolocation (Fig. 1A). Echolocation calls of the bat were recorded at a sample rate of 250 kHz, and the noise was generated at a sample rate of 2.5 MHz. The noise was presented with a custom-made electrostatic loudspeaker, which has a relative flat frequency response (± 5 dB) between 20 and 90 kHz that is the main frequency range of the bats’ echolocation calls. The presented noise level was directly measured with a free-field measurement microphone (7016, 1/4 inch Condenser microphone, ACO Pacific; with protection grid removed) placed at the position of the bat and directed toward the loudspeaker. Each array microphone was directed to the platform with a laser beamer. We localized the spatial position of the array microphones by playing and recording chirp sequences (5- to 25-kHz FM sweeps) through a loudspeaker placed at different locations. The exact location of the speaker was recorded by a motion tracking system containing three high-speed cameras (MX T40, Vicon Motion Systems). The distance between the platform and each microphone was confirmed by direct measurements from a laser distance meter (GLR 500, Bosch).

Fig. 4. A computational model for the Lombard effect. (A) The model takes the white noise as the input, followed by envelope extraction, leaky integration, and scaling. There are three free parameters: the response latency (ms), time constant (ms), and maximum compensation (dB). (B) Data from the 500-ms continuous noise conditions were used to find the optimal combination of the three model parameters through minimizing the sum of squared errors. The corresponding model predictions are colored in gray. (C–J) A comparison of the model prediction, based on the three parameters optimized from the 40-dB SPL noise condition (267 ms of time constant, 30 ms of response latency, and 5.9 dB of maximum compensation), and the data. The model precisely predicts the time course of the call amplitude for all conditions. For B–J, the red bar at the bottom shows the time course of noise presentation.
We collected sound recordings of individual bats from four experiments in total. In the first experiment, we presented bats with 500-ms-long white noise at 30, 40, 50, and 60 dB SPL level. In the second experiment, we presented bats with white noise of different noise durations: 50, 150, and 250 ms. In the third experiment, we presented bats with two repeated noise elements of 150 ms duration each that were separated by silence gaps of 10, 50, and 100 ms. In the fourth experiment, we presented the bats with either 800 ms noise or two repeated noise elements of 250 ms duration that were separated by a silence gap of 100 ms. Each experiment had its own silence control. For the last three experiments, the noise amplitude was 40 dB SPL. During the data collection, presentations of the treatments were randomized, and the experimenter was not aware of the exact treatment. For each recording day, typically 20–30 trials were run for each bat in about 30 min. A minimum of 4 d was devoted to data collection for each experiment to contain a minimum of 30 trials per condition per bat.

Sound analysis was performed with custom-written programs in MATLAB (R2015a, MathWorks), based on Luo et al. (16, 34, 35). We first checked the waveform of the echolocation call sequence for each trial and excluded the trials in which the bat did not track the prey target, as indicated by no decrease in the pulse-interval over time, which were rare cases (<5% of the total trials). Then, we accounted for the frequency response of the microphones by filtering the recorded calls with each microphone’s compensatory impulse response (5th-order finite-impulse response filter) and high-pass filtered all recordings at 10 kHz. Because of the spectral notch of the microphones at around 75 kHz, we low-pass-filtered the recordings at 70 kHz. Subsequently, we identified echolocation calls from the microphone that was directed toward the bat (central microphone, 0-degree azimuth and elevation), from which each echolocation call was cut with a fixed time window of 5 ms before the call and 5 ms after the call. Then we estimated the call duration (~20 dB relative to the peak amplitude) and the RMS amplitude over the duration. Through cross-correlation, calls in all other channels were readily available for analysis after compensating for the frequency- and distance-specific transmission loss. The quality of the sound analysis program was manually checked by displaying the waveform and spectrogram at varying stages of the analysis and its high quality was confirmed. Each analyzed call represented the call of the maximum RMS amplitude of the 14 recorded calls from the 14 microphones, so that the influence of the changes in calling direction of the bat on signal parameter estimation is minimized.

We used the Global Optimization Toolbox from MATLAB to search for the free parameters of the model. Optimization for each of the three noise conditions (30, 40, and 50 dB SPL level) was run repeatedly until the optimization output did not improve anymore. Specifically, we used the mean vocal-level adjustments, as shown in Fig. 2J, to search for the free parameters of the model. Nevertheless, we wish to point out that we also computed median level adjustments, and there was little difference between the mean and median vocal-level adjustments (Δ0.1 dB on average; Table S1). Mathematical processes of the model and the simulation details are provided as MATLAB scripts in Dataset S1. In short, the computational model takes the background noise (i.e., the sound pressure level of the noise over the entire time course) and the three free parameters as inputs. The outputs are predictions of the amplitude adjustments (Fig. 4A).

There was individual variation in the compensation magnitude of the Lombard effect, and different individuals contributed different number or percentage of calls (i.e., sample size) to the population analysis (ranging from 13 to 25%; Table S2). To assess how much the differential sample size of individual bats affected the population analysis, we compared the mean call amplitude of our population data with the mean call amplitude from bootstrapped data, in which each bat contributed equally to the population analysis [bootstrat = bootstrp(1000, @mean, ind_data); MATLAB R2015a], using the data from experiment 1. We found that there was a negligible effect of the differential sample size on the estimation of the mean call amplitude compensation at the population level across the three noise conditions. The differences in maximum increase in call amplitude (based on data within time window of 250–500 ms) between the empirical and bootstrapped data were 0.11, 0.14, and 0.14 dB for the 30-, 40-, and 50-dB noise conditions, respectively.

ACKNOWLEDGMENTS. We thank Melville Wohlgemuth for suggestions during animal training, Klemen Kosej and Lutz Wiegrebe for suggestions to the model, and James Garmon for engineering supports. Two anonymous reviewers provided valuable comments for improving the manuscript. This project is funded by National Science Foundation (IOS-1010193), the Human Frontiers Science Program (RGPF0040), the National Science Foundation (IOS-1460149), the Office of Naval Research (N00014-12-1-0339), and the Air Force Office of Scientific Research (FA9550-14-1-039) (to C.F.M.), and by a longitudinal postdoctoral fellowship from Human Frontier Science Program (to J.L.) (LT000279-2016-1).