Neural correlates of auditory streaming in an objective behavioral task

Naoya Itatani¹ and Georg M. Klump

Cluster of Excellence Hearing4all, Animal Physiology and Behavior Group, Department for Neuroscience, School for Medicine and Health Sciences, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany

Edited by Tony Movshon, New York University, New York, NY, and approved June 16, 2014 (received for review November 18, 2013)

Segregating streams of sounds from sources in complex acoustic scenes is crucial for perception in real world situations. We analyzed an objective psychophysical measure of stream segregation obtained while simultaneously recording forebrain neurons in the European starlings to investigate neural correlates of segregating a stream of A tones from a stream of B tones presented at one-half the rate. The objective measure, sensitivity for time shift detection of the B tone, was higher when the A and B tones were of the same frequency (one stream) compared with when there was a 6- or 12-semitone difference between them (two streams). The sensitivity for representing time shifts in spiking patterns was correlated with the behavioral sensitivity. The spiking patterns reflected the stimulus characteristics but not the behavioral response, indicating that the birds' primary cortical field represents the segregated streams, but not the decision process.

auditory scene analysis | multiunit recording | behavior | songbird

In the natural environment, sounds produced by multiple sources must be segregated to retrieve the information that they convey. At the same time, sequential sound signals from an individual source must be integrated to perform the analysis. The auditory system of humans, as well as that of many vertebrate species, is able to parse signals from multiple sources into separate sets of “auditory streams” based on the signals’ spectral and/or temporal profiles (1–5). Attention affects auditory streaming, and imaging studies have shown that it modulates brain activity during the analysis of segregated streams of signals (6–13). Although to date most psychophysical and imaging studies on auditory streaming have investigated attentive, actively listening human subjects, animal studies have only investigated passively listening subjects not involved in a specific task requiring stream segregation (14–18). Better insight into the cellular mechanisms underlying auditory streaming can be gained by investigating neuronal responses in an actively listening animal.

Here we present neuronal response data obtained from an actively listening bird, the European starling (*Sturnus vulgaris*), performing a task that has been shown in human subjects to be more easily mastered if signals are integrated into a single auditory stream rather than separated into two different streams (19–21). Behavioral performance in such tasks can serve as an objective measure for stream segregation, and correlated neuronal response measures can serve to elucidate the mechanisms underlying the perceptual effects. Behavioral responses indicating auditory stream segregation of pure tone sequences in the European starling have been reported previously (22). Because larger frequency differences lead to better stream segregation both in the physiological response and in perception in a similar way as in humans (16, 22), the European starling is a suitable animal model for studying the central processes involved in auditory scene analysis.

We investigated the detection of irregular timing in a sequence of signals for which the auditory system is more sensitive if the signals are perceived as forming a single stream rather than being perceptually segregated into different streams (20, 21, 23, 24). We recorded multiunit activity in the songbird forebrain in an area (field L complex) that is homologous to the mammalian

primary auditory cortex while presenting sequences of pure tone ABA- triplet stimuli (i.e., ABA-ABA-..., ref. 1). Neural responses were recorded while the bird behaviorally reported a shift in the temporal position of the B tone in the triplet. Based on psychophysical studies in humans, we expected shift detection thresholds to be smaller if the A and B tones were perceived as belonging to one stream rather than to two separate streams. Parallel measurements of neuronal responses and the bird’s perception, as reflected in behavioral responses, allowed us to correlate the sensitivity of the neuronal representation with behavioral sensitivity on a trial-by-trial basis. We applied a similar metric to perception and neuronal representation using signal detection theory and receiver operating characteristic (ROC) analysis. By comparing neuronal responses when the animal detected versus failed to detect a given time shift, we evaluated whether neural activity in the starling’s primary cortical area reflects the stimulus characteristics or whether it also indicates the bird’s perception, as evidenced by its behavioral decisions.

Results

Behavioral Experiment. We determined the threshold for detecting a time shift of the B tone in an ABA- triplet as an objective measure of stream segregation. Detecting the time shift should become more difficult as the frequency difference (Δf) between the A and B tones is increased, thereby increasing the likelihood that they will be perceived as two separate isochronous streams. Time shift detection was determined in experimental sessions comprising 88 trials each, in which the A and B tones had a constant frequency (ranging from 0.4 to 7.0 kHz) and the B tone frequency was 0, 6, or 12 semitones higher than the A tone frequency.

Based on a previous study (22), we expected the percept within a session would be biased to being a one-stream percept ($\Delta f = 0$ semitone), a two-stream percept ($\Delta f = 12$ semitones), or an

Significance

When we listen in the real world, we often segregate sounds into different streams based on their acoustical features. This “auditory streaming” allows us to individuate and separately process sounds from different sources, like different voices at a cocktail party. In the present study we investigated the neuronal mechanism of streaming in a songbird while it performed a sound discrimination task that gave us an objective measure of the streamed percept. The study revealed that the songbirds’ primary auditory forebrain area represents stimulus features with a high sensitivity that reflects the birds’ ability for streaming. However, the bird’s perceptual decisions about streamed sounds are not reflected in the neuronal responses in this auditory area, a result consistent with evidence obtained in earlier non-invasive studies on auditory streaming in human listeners.

Author contributions: N.I. and G.M.K. designed research; N.I. performed research; N.I. and G.M.K. analyzed data; and N.I. and G.M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. Email: naoya.itatani@uni-oldenburg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1321487111/-DCSupplemental.

ambiguous percept, with both one-stream and two-stream percepts ($\Delta f = 6$ semitones). ABA- triplet sounds with a regular rhythm (i.e., no time shift of the B tone, background triplet) were presented continuously as a reference during the experimental session. The bird's task in the operant Go-NoGo test was to jump off a waiting perch to another perch when it perceived an onset time shift of the B tone (target triplet). Data from a set of six sessions (two sessions at each B tone frequency) were analyzed for a given A tone frequency. The A tone and B tone frequencies were chosen to differ by an equal number of semitones from the characteristic frequency (CF) of the recording site for that set. Five different time shifts, with 10 trials per time shift, ranging from 10% to 90% of the 40-ms tone duration (Fig. 1) were presented in each session, and 30 sham trials with no B tone time shift served to measure the false-alarm rate during a session. The first eight trials in each session served as warm-up trials and were excluded from the analysis.

Here we present behavioral data from 64 sets of experiments involving four subjects in which neurophysiological recordings were made simultaneously with measurements of behavioral responses for the three different Δf conditions. There were no significant differences in response probability or sensitivity among subjects [subject random effect, generalized linear mixed model (GLMM) ANOVA]. Thus, data from all subjects were pooled for the subsequent analysis. In general, the reporting probabilities for a time shift of the B tone increased with the amount of shift, as did the associated measure of sensitivity, d' [main effect of time shift on d' : $F(5, 1,091) = 1,260.3$; $P < 0.001$] (Fig. 1A). On average, the value of d' decreased with increasing Δf [$F(2, 1,091) = 115.2$; $P < 0.001$]. A significant interaction between effects of Δf and the magnitude of the time shift [$F(10, 1,091) = 12.5$; $P < 0.001$] was evident in the shallower curves at higher Δf . Thus, sensitivity in reporting the time shift was reduced with increasing Δf .

The behavioral time shift detection thresholds (Fig. 1B) represent the 50% response probability in the psychometric functions fitted to the hit rate data, plotted as a function of the amount of B tone onset shift. The time shift detection threshold increased significantly as Δf increased [main effect of Δf on threshold, $F(2, 126) = 46.8$, $P < 0.001$; $P < 0.001$ for all pairwise comparisons, t test]. The increased threshold with increasing Δf indicates that time shift detection becomes more difficult as the frequency difference between A and B tones increases.

Physiological Experiment. Example peristimulus time histograms (PSTHs) from a recording site with a CF of 1.9 kHz show a clear rise in activity with the onset of each tone in a triplet in the 0 or 6

Fig. 1. Average psychometric functions and time shift detection thresholds. (A) Average psychometric function in relation to Δf . The onset time shift of the B tone is represented as a percentage of the 40-ms tone duration. Sixty-four experiments were conducted in four birds for each Δf condition. Error bars represent $\pm$ SEM. (B) Average behavioral threshold for B tone time shift detection. Threshold values (50% reaction point) were calculated by fitting the psychometric functions relating percent response to the time shift with a single logistic function to the data for each Δf condition in the set of 64 experiments. *Represents a significant difference ($P < 0.001$). Error bars represent $\pm$ SEM. Data were pooled across subjects that showed no significant differences in response probability or sensitivity.

Fig. 2. An example of PSTHs of target triplet responses for different time shifts (rows) and in different semitone conditions (columns) observed from a single multiunit recording site. The vertical axis corresponds to the rate of the number of spikes per second. Dashed lines indicate the onset of each tone of the standard ABA- triplet without a time shift. The PSTHs were constructed from 60 responses for sham trials (0% shift) and 20 responses for targets (10–90% shift). The recording site was tuned to 1.9 kHz, and A and B tone frequencies were set accordingly [at the CF if Δf was 0 semitones and equally distant from the CF (in semitones) if Δf was 6 or 12 semitones].

semitone Δf condition (Fig. 2, *Left* and *Middle*, respectively), with a less pronounced rise in the 12 semitone Δf condition (Fig. 2, *Right*). The time shift of the response parallels the time shift of the B tone.

To correlate the neuronal responses for each Δf condition in relation to the time shift with the behavioral results for the identical stimuli, we constructed neurometric functions using two types of response measures. We first determined spike rates in response to target B tones. An analysis time window of 15 or 40 ms was adjusted to the onset of the B tone in the triplet (i.e., moving with the time shift of the B tone) and delayed by 14 ms to account for the average response latency. By applying ROC analysis to the frequency histograms of the rate response, we obtained a sensitivity measure d_a for comparing the B tone response in a target triplet with that in the triplet preceding the target triplet (Fig. 3). These d_a values, which represent the sensitivity for detecting a time-shifted B tone based on the response rate during B tone presentation, are calculated by transforming the area under the ROC curve to d_a using an inverse normal cumulative distribution function (25). The 0% time shift data represent sham trials with no B tone shift.

We next compared the spiking pattern in an 80-ms time window starting at the position of the B tone for a 0% time shift between a target triplet and the triplet before the target triplet using the van Rossum (vR) distance (26). The vR distance for the target response is a measure of temporal response similarity between the spikes during the 80-ms time window in the target triplet and the same time period in the reference triplet. The reference distribution for the ROC analysis was obtained by determining the vR distance between spikes during the 80-ms time window in the triplet before the target triplet and its preceding triplet (i.e., comparing two sequential triplets in which no time shift of the B tone occurred). The ROC analyses of the frequency histograms of the vR distance values were used to compute the sensitivity measure d_a for the temporal structure of the neurons' responses. Neurometric functions relating the d_a values to the time shift of the B tone can be directly compared with the behavioral psychometric functions with the corresponding sensitivity d' .

We constructed the average neurometric functions for analyzing multiunit activity from 64 recording sites for four different response measures: a 40-ms or 15-ms rate response and vR distances with a τ value of 3.16 or 31.6 ms (see Fig. 3). A three-way GLMM ANOVA with repeated measurements from the same cells revealed significant effects of response measure, Δf , and time shift [response measure: $F(3, 4,393) = 577.6$, $P < 0.001$; Δf : $F(2, 4,393) = 47.6$, $P < 0.001$; time shift: $F(5, 4,398) = 253.8$, $P < 0.001$] on the sensitivity measure d_a . In general, the d_a values obtained for the comparisons of the vR distance response measures were significantly larger than those obtained for the

Fig. 3. Average neurometric functions based on response rate (Left) and vR distance (Right) representing the response to the time shift of the B tone. A total of 64 experiments were conducted in four birds for each semitone condition. Rate analysis was performed using two different time-shifted windows, 40 ms (Upper) and 15 ms (Lower). The time shift of the window was always matched to the B tone onset shift. The reference distribution was the same as for previous triplet responses. The vR distance was determined using two different time constants, 3.16 ms (Upper) and 31.6 ms (Lower). Error bars represent $\pm$ SEM.

rate response measures (all $P < 0.001$ comparing vR distances with τ of 3.16 or 31.6 ms with rates obtained in time windows of 15 or 40 ms, t test). The da values of the vR distances for the smaller τ were significantly larger than those for the larger τ ($P < 0.001$, t test). Across all measurements, the average sensitivity da value for a Δf of 12 semitones was significantly smaller than the values for Δf 's of 0 and 6 semitones (all $P < 0.001$, t test). The average sensitivity da values for Δf 's of 0 and 6 semitones were quite similar (although also significantly different; $P = 0.003$).

When the slope of the vR distance neurometric functions around the behavioral threshold (between the 30% and 50% time shift conditions) was calculated with a τ of 3.16 ms, the 12 semitone Δf condition had a shallower slope (da change of 0.10 per 10% time shift) compared with the 0 and 6 semitone Δf conditions (0.17 and 0.20 per 10% time shift, respectively). A similar effect of Δf on slope was also observed in the vR distance neurometric functions for $\tau = 31.6$ ms (0.10, 0.21, and 0.24 per 10% time shift in the 12, 6, and 0 semitone Δf conditions, respectively). For the rate measures, the slopes of neurometric functions between the 30% and 50% time shift conditions were all lower.

Because the da values of the four response measures (i.e., 40 ms rate, 15 ms rate, $\tau = 3.16$ ms vR distance, and $\tau = 31.6$ ms vR distance) differed, we analyzed the data obtained for the different response measures in separate GLMM ANOVAs with time shift and frequency difference as fixed main effects. As indicated by da values, spike rates in the 40-ms time window showed a small increase when the B tone time shift was introduced [$F(5, 1,054) = 9.5$, $P < 0.001$] (Fig. 3, Left). Pairwise comparisons of the effect of time shift revealed that the responses for the 0% (sham) and 10% conditions were significantly smaller than the responses for the 70% and 90% conditions ($P < 0.003$, t test). In the rate measure, da values were not significantly affected by Δf condition, and the interaction of the effects of Δf and time shift on da was not significant. The magnitude of rate change and the associated da measured in the 15-ms time window showed patterns similar to those measured in the 40-ms time window, with only a significant main effect of time shift [$F(5, 1,054) = 6.3$, $P < 0.001$]. However, the da values related to the 15-ms onset response were generally smaller than those for the 40-ms response ($P = 0.005$, t test), indicating that the onset spikes provide for a lower sensitivity to detect the time shift compared with the total spikes in the full response time window.

For the vR distance measure, the significance of the effects of time shift and Δf on da (Fig. 3, Right) was more prominent than that observed for the rate response for both the small τ [time shift: $F(5, 1,053) = 146.7$, $P < 0.001$; Δf : $F(2, 1,051) = 33.5$, $P < 0.001$; two-way interaction: $F(10, 1,051) = 3.4$, $P < 0.001$] and the large τ [time shift: $F(5, 1,053) = 149.9$, $P < 0.001$; Δf : $F(2, 1,052) = 37.3$, $P < 0.001$; two-way interaction: $F(10, 1,052) = 4.9$, $P < 0.001$].

We also tested the effects for a larger range of time constants τ (Fig. S1), and found similarly large effects for $\tau = 3.16$ ms and $\tau = 10$ ms, with lower sensitivity for $\tau = 1$ ms. The general effects on sensitivity were well represented by $\tau = 3.16$ ms, which may reflect effects of the fine structure, and $\tau = 31.6$ ms, which may include the effects of a large integration time in the analysis of temporal responses.

A decrease in da values based on the vR distance measure with increasing Δf can be expected if there is a decrease in spike activity and less precise onset spiking due to shifting of the A and B tone frequencies farther away from the CF of the recording sites. The average rate change in response to each tone in the sham target triplet in relation to Δf is shown in Fig. S2.

Correlation between Decision and Neural Representation. In the behavioral experiments, the birds exhibited almost no Go responses in the 10% onset shift condition and nearly 100% hit rates in the 90% onset shift condition, along with a low false-alarm rate (only 4.1% on average). The responses in the intermediate onset time shift conditions (~30%) were more variable, however, showing both hits and misses elicited by identical target stimuli.

To correlate the behavioral decisions and neuronal representations, we compared the spike responses when the birds reacted to the time shift with those when they missed the time shift (with equal stimuli eliciting hit or miss responses). For this comparison, we quantified response differences for each recording site first by the area under the ROC curve comparing neuronal responses from trials with a hit versus trials with a miss and converted these to da values. We could not obtain data from experimental conditions in which the subjects either produced no miss (e.g., at a 90% time shift) or no hit (e.g., at a 10% time shift) or no false alarm. Fig. 4 shows the sample size for each experimental condition. In general, for all four response measures (i.e., 40-ms rate, 15-ms rate, $\tau = 3.16$ ms vR distance, $\tau = 31.6$ ms vR distance), the neurometric da values averaged across recording sites were not significantly different from 0 (non-significant intercept in GLMM ANOVA), and the average

Fig. 4. Average neurometric functions based on response rate (Left) and vR distance (Right) representing the decision-related response. Rate analysis was performed using two different time-shifted windows, 40 ms (Upper) and 15 ms (Lower). vR distance was determined using two different time constants, 3.16 ms (Upper) and 31.6 ms (Lower). The sensitivity measure, da , results from a comparison of neuronal responses in trials with a hit and trials with a miss presenting the same stimuli. The blue, red, and green lines represent Δf 's of 0, 6, and 12 semitones, respectively. Each colored number represents the number of recording sites for the corresponding data point. Error bars represent $\pm$ SEM.

values for the different time shifts were not outside the range of -0.5 to 0.5 (Fig. 4). GLMM ANOVA revealed no main effect of onset time shift and Δf on da values comparing the responses to hits and misses. If we consider only false alarms and correct rejections, there also was no significant deviation of the da values from 0. Similarly, there was no significant deviation of the da values from 0 when we analyzed only the data for the 30% onset shift condition, which is closest to the behavioral threshold and should be the most sensitive condition for revealing the effects of decisions on responses.

Discussion

By recording the responses of bird forebrain neurons while a decision was made in a behavioral discrimination task, we can compare the sensitivity observed in the neuronal response with the sensitivity demonstrated in perception. This allows us to relate an objective psychophysical measure of stream segregation to its neuronal basis using equivalent measures of performance. By comparing the neuronal response in trials presenting identical stimuli but differing in behavioral responses, we can discern whether the activity in primary cortical areas reflects the decision itself or only the representation of the stimulus being related to the percept in the one-stream and two-stream situations.

Sensitivity for Time Shift Detection as an Objective Measure of Stream Segregation. One common approach to measuring stream segregation objectively instead of relying on subjects' directly reported streaming percept involves determining a listener's sensitivity to detect changes in the relative timing of sequential sounds (1, 17, 20, 21, 24, 27). Micheyl and Oxenham (20) reported a significant correlation between the shift-detection threshold (objective measure) and the subjective percept of stream segregation in human subjects. The shift-detection threshold was generally higher if the subjective percept was a two-stream percept rather than a one-stream percept. European starlings have been shown to perceive segregated streams in a subjective behavioral task at a Δf of 9 semitones and above (22). Using the objective time shift detection paradigm, we found an elevated time shift detection threshold as Δf was increased to 12 semitones, consistent with the starling subjective stream segregation measurement. At a Δf of 0 semitones, we obtained the smallest time shift detection threshold, which is consistent with a previous study reporting that starlings have a one-stream percept at a Δf of a half semitone (22). Thus, the starling behavioral results indicate that the deterioration of the ability to detect an onset time shift of the middle tone of a triplet reflects the propensity to perceive segregated streams in a manner similar to that observed in humans.

The general reduction in d' values in the psychometric function for time shift detection with increasing Δf suggests that the representation of the temporal structure of the triplet in the auditory pathway deteriorates with increasing Δf . Using a neuronal measure of sensitivity, da , that corresponds to the behavioral d' , we tested whether this deterioration occurs in the stimulus-driven response of the neurons reflecting the input to the auditory forebrain.

Neuronal Sensitivity to the Time Shift. Time shift and Δf affected neuronal sensitivity in a similar way as behavioral sensitivity. For the rate measure reflecting activity during the ongoing B tone, da values were smaller than those representing the vR distance measure, reflecting the change in timing of the B tone response occurring in parallel with its time shift. The frequency separation Δf did not affect the da values obtained from the rate measure, which does not correspond to the observed relationship between sensitivity and Δf in behavior. In general, neurometric functions based on the temporal pattern measure were more sensitive and reflected the behavioral sensitivity better than neurometric functions based on the rate measure. Our findings are consistent with the observations of ferret A1 neurons reported by Walker et al. (28), who found that the human psychometric function for discriminating normal twitter calls of marmosets with calls in

which the temporal structure was locally flipped was well correlated with the neurometric function constructed using temporal spike discharge patterns rather than spike counts (i.e., rate) alone. Despite the notion that a transition from a time code to rate code occurs in the auditory pathway to the forebrain (29, 30), the results of our vR analysis in the present study suggest that at the level of the bird auditory forebrain, sound discrimination still can be reliant on temporal spike discharge patterns. The similarity of the effects of time shift and Δf on the sensitivity in the vR-based neurometric functions and the behavioral psychometric functions suggests that the neuronal responses reflect a bottom-up representation of the stimulus forming the basis for the behavioral decision.

Given that the average temporal neurometric functions in the present study based on the vR response measure showed shallower slopes and lower da values with increasing Δf , the sensitivity for detecting an onset shift is likely related to the strength of stream segregation owing to the differing frequencies of the two tones. Neural correlates of stream segregation were already observed in the auditory periphery and could be derived from preattentive processing of the different tones in separate auditory filters (18, 31–33).

Is the Decision Process Represented in the Primary Auditory Forebrain Area?

A main aim of the present study was to investigate whether the neuronal response at the level of the auditory forebrain is purely stimulus-driven or whether it also relates to an animal's decisions. To date, the relationship between conscious perception of segregated streams and behavioral response has been investigated exclusively with noninvasive methods. In those experiments, human brain activity was recorded while the subject reported a one- or two-stream percept or reported the detection of targets in an informational masking paradigm related to the segregation of streams of targets and maskers (7–11, 34, 35). Although those studies shed light on the involvement of specific brain areas in perceptual stream segregation, neuronal mechanisms at the cellular level remain unclear. By comparing the responses to physically identical stimuli when they are perceived or not perceived, it is possible to conclude whether the activity in a specific brain area is related to the decision process. The comparison of the evoked spike rate or temporal pattern of response obtained for hits with those response measures obtained for misses showed no relationship between animals' behavior and physiological responses.

In comparing neuronal responses associated with hits and misses or associated with false alarms and correct rejections, we found no evidence of a difference in these responses as reflected by the da values. This observation was consistent across all Δf and time shift conditions, indicating that the birds' decision making is not reflected by the neuronal responses in the field L complex of the avian auditory system. Even when confining our analysis to sham trials reflecting the activity associated with false alarms and to trials presenting 30% time shifts that are close to the perceptual threshold and for which small changes of the neuronal activity could tip the decision in one direction, we found no correlation between neuronal response and behavioral decisions.

Previous studies comparing the neuronal responses of hits and misses in an auditory discrimination task have found small but significant differences between the responses to hits and misses in cortical neurons. For example, analyzing local field potentials in the ferret primary auditory cortex, Bizley et al. (36) observed an average area under the ROC curve of up to 0.66 (equivalent to a da value of 0.58) when pooling data from recording sites that significantly represented the information about the stimulus. These authors also compared the area under the choice-derived ROC curves (aROCchoice, indicating the ferrets' decisions) with the area under the stimulus derived ROC curves (aROCstimulus) by a choice index [CI = (aROCchoice – aROCstimulus)/(aROCchoice + aROCstimulus)] and observed a significant positive CI (average, 4.2%) for multiunit activity,

indicating that the neurons were slightly more prone to representing the decision rather than the fundamental frequency of the stimulus. Although our data from the starling forebrain show similarly large maximum *da* values for the decision-related analysis, the fact that these *da* values do not deviate significantly from 0, and that the stimulus-driven *da* values are generally much larger than the choice-driven *da* values, indicates that, unlike in the ferret forebrain, the starling primary cortical area represents only the physical stimulus and not the decision. Niwa et al. (37) reported an area under the ROC curve of 0.512 (equivalent to a *da* value of 0.04) comparing responses from the population of single units in the monkey primary auditory cortex associated with hits and misses in the detection of amplitude modulation. Only when the population of neurons was limited to those showing a significantly greater firing rate for responded trials compared with nonresponded trials did the area under the ROC curve reach average values of 0.608 and 0.611 for single-unit and multiunit responses, respectively (equivalent to a *da* value of ~0.40) (37). The small *da* values obtained when the analysis was limited to a selection of the most sensitive neurons indicates, similar to our results with starlings, only a weak relationship between the choice and the response in the monkey's primary sensory cortical area (see also the review in ref. 38).

Thus, in the field L complex of the European starling, the neurons' response better predicted the physical time shift of the B tone presented rather than the behavioral decision leading to a response. Similar to the present study, noninvasive studies of humans have failed to find differences in stimulus-driven responses in the primary auditory cortex as a function of whether or not the same stimulus was detected in different trials (8, 9). This similarity suggests a common perceptual mechanism in humans and songbirds, in that primary auditory cortical areas of humans (especially in the auditory cortex core region) and of the starling (field L complex) represent crucial information for perception, but do not represent the decision itself. Further studies of secondary cortical areas are needed to identify the fields in the bird forebrain that correspond to the human brain areas in which the response is modulated by attention and reflects the behavioral decision (8, 10).

Materials and Methods

Experimental Setup. Behavioral experiments with simultaneous recording of the neuronal responses from the forebrain were carried out in unrestrained and freely moving European starlings (*S. vulgaris*) in a test cage (56 cm L × 36 cm W × 33 cm H) located inside a radio-shielded sound-insulated acoustic booth (IAC 402A; Industrial Acoustics). The cage was equipped on one side with an upper perch that the bird preferred to use as a resting position and on the other side with a lower perch located in front of a feeder. Stimuli were presented from a loudspeaker (KEF Audio SP3253; level variation in transfer function $\leq \pm 3$ dB over the range of 0.1–10 kHz) positioned approximately 70 cm above the bird's head, which minimized the effects of the bird's head turns and movements on the sound level. The multiunit signals were transmitted via radio telemetry by a small FM radio transmitter (FHC 40-71-1; Frederick Haer and Co.), received by a bipolar antenna placed around the test cage, and demodulated by an FM tuner (TX-970; Pioneer). The amplified (40–50 dB, custom-built amplifier) and bandpass-filtered (500–5,000 Hz) multiunit signal was digitally recorded onto a hard disk (16 bit, 44.1 kHz; Hammerfall DSP Multiface II; RME) for later off-line analysis.

Auditory Streaming Stimuli. We presented sequential pure-tone ABA- triplets (1, 16) known to elicit auditory streaming (22). By introducing a time shift of the onset of the middle B tone and measuring the sensitivity for detecting the shift, we obtained an objective measure for the auditory streaming percept. It has been demonstrated that the sensitivity for detecting the shift is lower during a two-stream percept than during a one-stream percept (20, 21, 24). The frequencies of A and B tones were chosen according to the tuning characteristics of the neurons recorded during the behavioral task. The frequency of A and B tones was chosen to be equally distant from the characteristic frequency of the recording site, so that similar spike responses to both tones would be evoked. The frequency difference between A and B tones (Δf) was 0, 6, or 12 semitones, and the B tone frequency was always higher than the A tone frequency. Tone duration was 40 ms gated with 5-ms Hanning ramps. A 40-ms silent interval separated the tones of a triplet, and a silent interval of 120 ms separated the triplets; total triplet duration was

320 ms. The level of each triplet of 65 dB sound pressure level (SPL) was roved by ± 3 dB during continuous triplet presentation, counteracting the possibility of perceived level changes related to temporal integration when the silent gap between A and B tones was changed. Triplets without a time shift were presented as a background during the entire experimental session. In a target triplet, the onset time of B tone was delayed compared with that in the background triplet. The time shift of the B tone was either 0% (sham trial; 37.5% of the targets) or 10%, 30%, 50%, 70%, or 90% of the tone duration (test trials; 12.5% for each of the targets) while the temporal positions of the A tones and the total triplet duration were kept constant.

Behavioral Testing Procedure. Four wild-caught adult European starlings (1 male and 3 females) were trained before implanting a microdrive and starting physiological recordings in parallel with the behavioral experiments. The behavioral experiments were performed with an operant Go-NoGo paradigm. For presentation of target triplets, the bird had to sit on the upper perch and targets were presented after a random waiting time of 2–7 s. For a Go response on a target, the bird had to leave the upper perch within 2 s after the end of the target triplet. Within each session, target triplets were presented in randomized order. The total number of target triplets in one session was 88, with a session composed of 10 blocks of eight trials each presenting five targets (each with a different time shift; test) and three targets without a time shift (sham). An additional block of eight trials at the beginning of each session provided warm-up trials using target stimuli with a time shift of 90%; these warm-up trials were eliminated from the analysis.

If the bird responded to a test within 2 s, then a "hit" was scored, and food reward (meal worm) was given. If the bird did not respond to a test within 2 s, then a "miss" was scored. A response in a sham trial was scored as a "false alarm." Regular testing together with recording of forebrain multiunit activity started after the bird's detection probability for the largest time shift condition reached 80%. Data from a total set of six sessions (three Δf conditions × two repeats) were collected for each isolated forebrain site.

Behavioral Data Analysis. For the behavioral analysis, we constructed psychometric functions by combining data from two sessions with identical stimulus parameters to calculate the response probability for each shift condition. The function was fitted with a single logistic function, $\alpha + (\beta - \alpha)/(1 + e^{-(X-M)/S})$, where α , β , X , M , and S represent false-alarm probability, maximum function, stimulus level, midpoint of the psychometric function, and spread parameter, respectively (39, 40). The 50% response probability of the fitted function designated the behavioral threshold. Psychometric functions relating the sensitivity d' to the time shift value were calculated using signal detection theory based on the response probabilities for test and sham targets.

Neuronal Recording Procedure. The care and treatment of the animals in the behavioral experiments and with respect to all procedures related to neural recordings were in accordance with the permit for animal experimentation issued by the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit.

The details of the surgical procedures are described elsewhere (41); only a brief description is provided herein. All surgical procedures were performed using isoflurane anesthesia. A custom-built small microdrive with four custom-built Teflon-insulated platinum-iridium wire electrodes (wire diameter, 25 μ m) with etched tips (A-M Systems) (42), with impedance ranging from 0.6 to 7.0 M Ω , was mounted on the bird's skull. The array of four electrodes was stereotactically implanted into the field L complex of the right forebrain hemisphere, with blood vessels on the brain surface serving as landmarks (16). For recording, one electrode from the array was connected to the transmitter mounted next to the microdrive. Recording started after the bird had completely recovered.

Before each set of six behavioral sessions, the bird was restrained briefly, and electrodes were positioned by adjusting the microdrive while providing stimulation with 200-ms pure tones (~75 dB SPL) (41). Adjustment was completed when the sound-evoked multiunit activity had an RMS amplitude at least 1.3-fold larger than that of the background activity. Then the bird was released into the test cage. Without the bird performing any behavioral task, a tuning curve was recorded using 200-ms pure-tone stimuli with frequencies presented in 0.25-octave steps ranging from 1.5 octaves below to 1.5 octaves above the estimated best frequency (41). From the tuning curve, the final CF of the recording site was determined, which served to set the frequency of the A and B tones in the behavioral experiments.

Throughout the behavioral experiments, the neuronal responses toward all presented background and target triplets were recorded and stored. After data collection, the birds were perfused and their brains cut into 50- μ m

parasagittal sections and Nissl-stained to verify the position of the electrode tracks using established histological landmarks for the field L complex.

Neuronal Data Analysis. B tone-driven spike rates were calculated within a 14-ms latency and time shift-adjusted temporal window of 15 ms or 40 ms. The similarity of B tone-driven temporal response patterns relative to that observed in previous background triplets was evaluated using the vR distance (26). The vR distance is a measure for evaluating the similarity of two spike trains, in which each spike train is convolved with a single exponential function, and the similarity of the resulting trains is then computed as the Euclidean distance between the pair of convolved spike trains. The convolution can be interpreted as representing a postsynaptic potential (26). By varying the decay time constant, τ , of the exponential, the weighting of the precision of spiking coincidence on this measure is altered (43). A small τ value reflects a comparison based on coincident spikes, whereas a large τ value reflects a comparison based on the ongoing spiking rate. The 80-ms temporal window for that analysis always started at 80 ms from the onset of a triplet. In the present analysis, we used two τ values, 3.16 and 31.6 ms, representing short and long integration times, respectively.

For determining neurometric da values for the ROC analysis, a reference distribution of rate responses or vR distances was obtained for the triplet with no time shift presented before the triplet with a target. This reference distribution was compared with the distribution of rate responses or vR distances for the triplet with the target. The vR distances were calculated by comparing spike trains in response to B tones in two subsequent triplets. For

the distribution of vR distances in triplets with the target, the responses in the target triplet were compared with those obtained in the triplet before the target triplet. For the reference distribution of vR distances, the responses in the triplet before the target triplet were compared with those in its predecessor. To relate behavioral and neuronal sensitivity measures, the neurometric da values were compared with the d' values obtained for the behavioral data.

To evaluate the joint effects of the different factors on the spike responses, we used GLMM ANOVA with da value as the dependent variable; amount of B tone time shift (0%, 10%, 30%, 50%, 70%, or 90% of tone duration), Δf (0, 6, or 12 semitones), and type of measurement (spike rate within a 15- or 40-ms time window or vR distance with a 3.16- or 31.6-ms time constant) as the fixed effects; and recording site ($n = 64$) as the random effect. The same fixed effects were used for the statistical comparisons of the hit responses and miss responses, which also were quantified as da values. For a more sensitive analysis of the effects of B tone time shift and Δf , we also analyzed the responses from each type of measurement separately. Subsequent pairwise comparisons within each effect were carried out using Bonferroni-corrected t tests. The criterion for statistical significance in all tests was $\alpha = 0.05$. All analyses were performed using SPSS 20.0.

ACKNOWLEDGMENTS. We thank Mark Bee for his comments on a previous version of the manuscript. This research was supported by Deutsche Forschungsgemeinschaft Grant SFB TRR 31.

- van Noorden LPAS (1975) *Temporal Coherence in the Perception of Tone Sequences* (Eindhoven Univ of Technology, Eindhoven, The Netherlands).
- Bee MA, Micheyl C (2008) The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol* 122(3):235–251.
- Moore BCJ, Gockel HE (2012) Properties of auditory stream formation. *Philos Trans R Soc Lond B Biol Sci* 367(1591):919–931.
- Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
- Micheyl C, et al. (2007) The role of auditory cortex in the formation of auditory streams. *Hear Res* 229(1–2):116–131.
- Zion Golumbic EM, et al. (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron* 77(5):980–991.
- Gutschalk A, et al. (2005) Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci* 25(22):5382–5388.
- Gutschalk A, Micheyl C, Oxenham AJ, Wilson EC, Melcher JR (2008) Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol* 6(6):e138.
- Dykstra AR, et al. (2011) Widespread brain areas engaged during a classical auditory streaming task revealed by intracranial EEG. *Front Hum Neurosci* 5:74.
- Wiegand K, Gutschalk A (2012) Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. *Neuroimage* 61(1):62–69.
- Kondo HM, Kashino M (2009) Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J Neurosci* 29(40):12695–12701.
- Kashino M, Kondo HM (2012) Functional brain networks underlying perceptual switching: Auditory streaming and verbal transformations. *Philos Trans R Soc Lond B Biol Sci* 367(1591):977–987.
- Hill KT, Bishop CW, Miller LM (2012) Auditory grouping mechanisms reflect a sound’s relative position in a sequence. *Front Hum Neurosci* 6:158.
- Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 151(1–2):167–187.
- Fishman YI, Arezzo JC, Steinschneider M (2004) Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am* 116(3):1656–1670.
- Bee MA, Klump GM (2004) Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *J Neurophysiol* 92(2):1088–1104.
- Micheyl C, Tian B, Carlyon RP, Rauschecker JP (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48(1):139–148.
- Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. *Curr Biol* 18(15):1124–1128.
- Vliegen J, Oxenham AJ (1999) Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am* 105(1):339–346.
- Micheyl C, Oxenham AJ (2010) Objective and subjective psychophysical measures of auditory stream integration and segregation. *J Assoc Res Otolaryngol* 11(4):709–724.
- Thompson SK, Carlyon RP, Cusack R (2011) An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J Exp Psychol Hum Percept Perform* 37(4):1253–1262.
- MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W (1998) Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *J Acoust Soc Am* 103(6):3581–3587.
- Neff DL, Jesteadt W, Brown EL (1982) The relation between gap discrimination and auditory stream segregation. *Percept Psychophys* 31(5):493–501.
- Vliegen J, Moore BCJ, Oxenham AJ (1999) The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J Acoust Soc Am* 106(2):938–945.
- Macmillan NA, Creelman CD (2005) *Detection Theory: A User’s Guide* (Lawrence Erlbaum Associates, Mahwah, NJ).
- van Rossum MCW (2001) A novel spike distance. *Neural Comput* 13(4):751–763.
- Roberts B, Glasberg BR, Moore BCJ (2008) Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *J Exp Psychol Hum Percept Perform* 34(4):992–1006.
- Walker KM, Ahmed B, Schnupp JW (2008) Linking cortical spike pattern codes to auditory perception. *J Cogn Neurosci* 20(1):135–152.
- Dolžel LV, Itatani N, Günther S, Klump GM (2012) Auditory streaming by phase relations between components of harmonic complexes: A comparative study of human subjects and bird forebrain neurons. *Behav Neurosci* 126(6):797–808.
- Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84(2):541–577.
- Beauvois MW, Meddis R (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Am* 99(4 Pt 1):2270–2280.
- McCabe SL, Denham MJ (1997) A model of auditory streaming. *J Acoust Soc Am* 101(3):1611–1621.
- Hartmann WM, Johnson D (1991) Stream segregation and peripheral channeling. *Musik Percept* 9(2):155–184.
- Cusack R (2005) The intraparietal sulcus and perceptual organization. *J Cogn Neurosci* 17(4):641–651.
- Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci* 18(1):1–13.
- Bizley JK, Walker KMM, Nodal FR, King AJ, Schnupp JWH (2013) Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr Biol* 23(7):620–625.
- Niwa M, Johnson JS, O’Connor KN, Sutter ML (2012) Activity related to perceptual judgment and action in primary auditory cortex. *J Neurosci* 32(9):3193–3210.
- Nienborg H, Cohen MR, Cumming BG (2012) Decision-related activity in sensory neurons: Correlations among neurons and with behavior. *Annu Rev Neurosci* 35:463–483.
- Lam CF, Mills JH, Dubno JR (1996) Placement of observations for the efficient estimation of a psychometric function. *J Acoust Soc Am* 99(6):3689–3693.
- Lam CF, Dubno JR, Ahlstrom JB, He NJ, Mills JH (1997) Estimating parameters for psychometric functions using the four-point sampling method. *J Acoust Soc Am* 102(6):3697–3703.
- Itatani N, Klump GM (2009) Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J Neurophysiol* 101(6):3212–3225.
- Hofer SB, Klump GM (2003) Within- and across-channel processing in auditory masking: A physiological study in the songbird forebrain. *J Neurosci* 23(13):5732–5739.
- Narayan R, Graña G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96(1):252–258.