

From Comparison to Classification: A Cortical Tool for Boosting Perception

Mor Nahum,^{1*} Luba Daikhin,^{2,3*} Yedida Lubin,^{4*} Yamit Cohen,² and Merav Ahissar^{1,3}

¹Interdisciplinary Center for Neural Computation, ²Department of Neurobiology, The Institute of Medical Sciences, Hadassah Medical School, and Departments of ³Psychology and ⁴Cognitive Science, Hebrew University of Jerusalem, Jerusalem, Israel 91905

Humans are much better in relative than in absolute judgments. This common assertion is based on findings that discrimination thresholds are much lower when measured with methods that allow interstimuli comparisons than when measured with methods that require classification of one stimulus at a time and are hence sensitive to memory load. We now challenged this notion by measuring discrimination thresholds and evoked potentials while listeners performed a two-tone frequency discrimination task. We tested various protocols that differed in the pattern of cross-trial tone repetition. We found that best performance was achieved only when listeners effectively used cross-trial repetition to avoid interstimulus comparisons with the repeated reference tone. Instead, they classified one tone, the nonreference tone, as either high or low by comparing it with a recently formed internal reference. Listeners were not aware of the switch from interstimulus comparison to classification. Its successful use was revealed by the conjunction of improved behavioral performance and an event-related potential component (P3), indicating an implicit perceptual decision, which followed the nonreference tone in each trial. Interestingly, tone repetition itself did not suffice for the switch, implying that the bottleneck to discrimination does not reside at the lower, sensory stage. Rather, the temporal consistency of repetition was important, suggesting the involvement of higher-level mechanisms with longer time constants. These findings suggest that classification is based on more automatic and accurate mechanisms than interstimulus comparisons and that the ability to effectively use them depends on a dynamic interplay between higher- and lower-level cortical mechanisms.

Introduction

One of the hallmarks of human perception is our remarkable discrimination abilities along simple physical dimensions. When performance is assessed with protocols that provide two stimuli between which participants are asked to compare (e.g., “which of the two stimuli is larger?”), measured thresholds are impressively low. This observation was reported for many dimensions, including frequency, intensity, and duration of sounds in the auditory modality (Wright et al., 1997; Karmarkar and Buonomano, 2003; Ulrich et al., 2006) and orientation, position, and contrast of stimuli in the visual modality (Regan, 1985; Regan and Beverley, 1985; Lages and Triesman, 1998; Ulrich et al., 2006). In contrast to this remarkable ability in relative judgment tasks, absolute judgments, in which participants are asked to classify a single stimulus with respect to the same dimensions, are typically quite crude, particularly when assessed with a broad range of stimuli (Garner, 1953; Pollack, 1953; Miller, 1956; Laming, 2004; Stewart et al., 2005).

In an attempt to assess the best thresholds that human performance can achieve, most studies use relative judgment protocols, asking participants to discriminate between sequentially presented stimuli. A relatively common protocol in psychophysical studies is therefore the two-interval two-alternative forced-choice paradigm (2AFC) (Green and Swets, 1966; Leek, 2001; Macmillan and Creelman, 2005), in which participants are asked “which of the two stimuli is longer (or higher, denser, etc.)?”. The common use of this method is based on two assumptions. The first is that our brains can implement interstimulus comparisons. This assumption was directly assessed and supported by electrophysiological studies in monkeys, showing that comparisons can be implemented by a distributed network that includes the prefrontal cortex and the premotor cortex, which can actively retain the information regarding the first stimulus during the temporal interval between the two stimuli (“delayed activity”) (Romo et al., 1999, 2004; Hernández et al., 2002; Zaksas and Pasternak, 2006). The second assumption is that these comparisons can be performed accurately with no loss of information, namely, that the comparison process itself does not impair thresholds.

In parallel to this main dogma, several studies noted that relative judgment procedures do not always yield better performance (Nachmias, 2006; Lapid et al., 2008; Yeshurun et al., 2008). In fact, 60 years ago, Harris (1948) showed that the sequence of stimulus pairs in a block affects performance. Thus, performance is substantially better when a consistent reference stimulus is one of the stimuli in every trial than when the two stimuli in each trial are randomly chosen (Harris, 1948; Moore, 1997). This improvement was attributed to a reduction in the

Received April 14, 2009; revised Oct. 4, 2009; accepted Oct. 26, 2009.

The research was supported by the Israel Science Foundation, a grant from the Israeli Institute of Psychobiology, and a subcontract from the National Institutes of Health (NIH Grant 2R01DC004855). We thank Ehud Ahissar, Laurent Demany, Shaul Hochstein, and Yonatan Loewenstein for insightful comments and discussions and Mor Levi and Nitzan Tal for assistance in collecting the data.

*M.N., L.D., and Y.L. contributed equally to this work.

Correspondence should be addressed to Prof. Merav Ahissar, Department of Psychology and Interdisciplinary Center for Neural Computation, Hebrew University of Jerusalem, Jerusalem, Israel 91905. E-mail: msmerava@mscc.huji.ac.il.

DOI:10.1523/JNEUROSCI.1781-09.2010

Copyright © 2010 the authors 0270-6474/10/301128-09\$15.00/0

Table 1. Participant groups in the behavioral experiments

Participant group	<i>n</i>	Protocol	Number of musically trained
1	21	Reference first	8
2	21	Reference second	8
3	21	Random	10
4	21	Implicit reference	7
5	25	Reference interleaved	9
6	9	Reference first, reference second, random	All were laboratory members with substantial experience in all task protocols

internal noise of the sensory representation of the repeated stimulus (Viemeister, 1970; Lages and Treisman, 1998; Morgan et al., 2000; Nachmias, 2006). Namely, the averaged representation of the reference becomes gradually more accurate with repeated presentations, and this average may replace the reference stimulus when compared with the nonreference stimulus. Recent studies further showed that other cross-trial manipulations, such as the temporal position of the reference within the pair, may significantly affect performance (Morgan et al., 2000; Oxenham and Buus, 2000; Nachmias, 2006; Hairston and Nagarajan, 2007; Georgeson et al., 2008; Lapid et al., 2008; Yeshurun et al., 2008).

However, despite the accumulated findings pointing to the importance of the assessment protocol, the basic concept of a “retain-and-compare” process that can be implemented with no loss of information remained unchallenged. We now asked whether a retain-and-compare mechanism can indeed account for the impressively low discrimination thresholds reported in the literature. Applying a two-tone frequency discrimination paradigm, we found that the perceptual system actively attempts to bypass comparisons between two recently presented stimuli and replace them with a task-related classification (“high” or “low”) based on an internal reference. When the experimental protocol does not allow such classification, discrimination thresholds invariably increase even when a repeated stimulus is presented in each trial and its sensory representation is expected to be accurate. These findings suggest that the retain-and-compare process itself poses a severe bottleneck to perception.

Materials and Methods

Behavioral experiments

Participants

One hundred eighteen participants (mean age, 24 ± 4 years) were engaged in the behavioral part of the study, as detailed in Table 1. Each group included participants with and without previous musical training, i.e., at least 1 year of experience with a tonal instrument (Table 1). Groups 1–5 were composed of naive participants, with no previous experience with the task. Each participant performed one protocol. Group 6 was composed of highly trained laboratory members, each of whom performed the task in all three protocols (“reference first,” “reference second,” and “random”; see below).

Two-tone frequency discrimination task

All stimuli were presented binaurally through Sennheiser HD-265 linear headphones using a TDT System III signal generator (Tucker Davis Technologies) controlled by in-house software in a sound attenuated room in the laboratory. Tone intensity was 65 dB.

In all protocols but the “implicit reference” protocol (detailed description follows), thresholds for frequency discrimination were measured using a 2AFC. The frequency difference between the two tones in each trial was changed in a three-down, one-up staircase procedure (Levitt, 1971) converging to 79.4% correct responses. Step size decreased every four reversals, from 4.5 to 2 to 1 to 0.5 to 0.1%.

Each trial contained two 50 ms tones, presented with a 950 ms inter-stimulus interval (ISI). Participants were asked to indicate, by pressing a

Table 2. The two participant groups who participated in the ERP experiments

ERP experiment #	<i>n</i>	ERP protocol	Frequency difference (%)
1	10	Reference first, reference second	1 and 2%
2	10	Reference first, reference second, reference interleaved	4 and 8%

button, which of the two tones was higher. A pleasant visual feedback (a happy face) was provided for correct responses, whereas an unpleasant feedback (a sad face) was presented after incorrect responses. The next trial began 1 s after the participant’s response. Each assessment block contained 80 trials (except for the “reference interleaved” protocol; see below). Discrimination thresholds were determined as the mean of the frequency differences in the last seven reversals. Each group of participants (groups 1–5) (Table 1) was assigned to one of the following five protocols and performed three consecutive blocks in that protocol.

Reference first protocol. A fixed reference tone (1000 Hz) was presented on the first interval of every trial. The tone on the second interval (non-reference) was randomly selected to be either higher or lower in frequency. The (absolute) frequency of the nonreference (second) tone was adapted during the assessment as described above.

Reference second protocol. A fixed reference tone (1000 Hz) was presented on the second interval of every trial, whereas the nonreference tone (either higher or lower in frequency) was always presented first. The (absolute) frequency of the nonreference (first) tone was adapted as described above.

Random protocol. There was no fixed reference. On every interval, one tone was selected from the range of 600–1400 Hz, and the other tone (randomly chosen as either higher or lower) differed in frequency according to the same adaptive procedure described above.

Implicit reference protocol. A fixed reference tone (1000 Hz) was presented five times at the beginning of the session. After that, a single tone (the nonreference), which was randomly selected to be either higher or lower than the reference frequency, was presented on every trial. Its temporal position (with respect to the time of response in the previous trial) was equivalent to that of the reference first protocol. Participants were required to reply whether the tone was higher or lower than the initial reference. The (absolute) frequency of the tone was adapted during the assessment as described above.

Reference interleaved protocol. Blocks were of double length, i.e., composed of 160 trials. Odd trials were of type reference first, whereas even trials were of type reference second. Separate staircases and thresholds were calculated for reference first and for reference second trials, each following the same adaptive procedure described above.

Data analysis

Comparison of thresholds under reference first, reference second, random, and implicit reference protocols was conducted using a univariate analysis with between-subjects factor of protocol (four levels). *Post hoc* Scheffé’s contrast was performed to compare thresholds under the four protocols. Comparison of thresholds under reference first protocol when measured separately versus interleaved was done using an unpaired two-tailed Student’s *t* test. A similar comparison was done for the reference second protocol. In addition, we compared thresholds obtained for the reference second trials of the reference interleaved protocol with those of the random protocol, using a similar *t* test.

Event-related potential measurements

Participants

A total of 16 participants (mean age, 25 ± 2 years; five males) took part in the event-related potential (ERP) measurements (Table 2). Four of them (laboratory members) participated in both ERP experiments and had previous experience with the task. These four participants include the experienced subject YC (laboratory member), who initially performed at chance level and was subsequently administered an intensive training protocol with the adaptive reference first protocol for 2 weeks. After

training, this subject participated in the two ERP sessions of experiments 1 and 2, respectively.

Stimuli

The same protocols used for the behavioral experiment were used for the ERP measurements. However, the ERP measurements were conducted using constant, rather than adaptively modified stimuli to collect more data under exactly the same stimulation conditions. ERP measures were taken in two separate experiments. In the first experiment, the frequency differences between the reference and nonreference tone were 1 or 2% (with equal probabilities), whereas in the second experiment, they were 4 and 8% (Table 2).

Protocols

In the first ERP experiment, two 2-tone discrimination protocols were used.

Reference first. The first tone was always 1000 Hz, whereas the second tone was chosen to be 990, 980, 1010, or 1020 Hz with equal probability.

Reference second. Same as reference first protocol, but the order of the tones within the trial was reversed, as described above.

Both protocols were assessed within a single session, composed of four blocks, each composed of 300 trials, in the following sequence: reference first, reference second, reference first, and reference second. The ERP waves reported for each protocol are an average of the trials in the corresponding two blocks.

In the second ERP experiment, the procedure was essentially the same. The frequency differences used were, however, larger, 4 and 8%. Three protocols were assessed in three separate blocks composed of 300 trials each, in the following sequence: reference first, reference second, and reference interleaved. In the reference first block, the first tone was always 1000 Hz, whereas the second tone was chosen from 920, 960, 1040, or 1080 Hz. In the reference second block, the same stimuli were used just in opposite order. In the reference interleaved block odd trials were reference first trials, whereas even trials were reference second trials.

Experimental procedure

Electrophysiological activity was recorded in a sound-attenuated room while participants performed a two-tone frequency discrimination task. Two tones of 50 ms length and with an ISI of 600 ms between them were presented in each trial. Participants were asked to determine which tone was higher (first or second), by pressing either 1 or 2 on the computer keyboard, respectively. Participants were requested to respond immediately after hearing both tones and before the next trial, which began after 1.4 s. No feedback was presented to avoid ERP components related to the feedback stimulus. Based on our pilot studies in which we compared performance in these tasks with and without feedback, we concluded that, although participants prefer having feedback, their discrimination thresholds are not affected by its absence.

The recording session consisted of four blocks (three in the second ERP experiment), each containing 300 trials, with short breaks between them.

EEG recording and averaging

EEG was recorded from 32 active Ag-AgCl electrodes mounted on an elastic cap using the BioSemi ActiveTwo tools. Recordings were referenced to the tip of the nose. Electrode sites were based on the 10–20 system (American Electroencephalographic Society, 1991). Two additional electrodes were placed over the left and right mastoids. Horizontal electro-oculogram (EOG) was recorded from two electrodes placed at the outer canthi of both eyes. Vertical EOG was recorded from electrodes on the infraorbital and supraorbital regions of the right eye in line with the pupil.

EEG and EOG signals were sampled at 256 Hz, amplified, and filtered with an analog bandpass filter of 0.16–100 Hz. Offline analysis was performed using BrainVision Analyzer software. EEG was referenced to the averaged mastoids and was digitally filtered using bandpass of 1–30 Hz. Artifact rejection was applied to the unsegmented data according to the following criteria: any data point with EOG or EEG greater than $\pm 100 \mu\text{V}$ was rejected along with the data ± 300 ms around it. In addition, if the difference between the maximum and the minimum amplitudes of two data points within an interval of 50 ms exceeded $100 \mu\text{V}$, data ± 200

ms around it were rejected. Finally, if the difference between the maximum and the minimum amplitudes of two data points within an interval of 100 ms was below $0.5 \mu\text{V}$, the data point along with the data ± 300 ms around it were rejected. Trials containing rejected data points were omitted from additional analysis, as well as the first three trials of each block. For ERP averaging, the EEG was parsed to 1300 ms epochs starting 100 ms before the first interval and then averaged separately for each condition. Then the baseline was adjusted by subtracting the mean amplitude of the prestimulus period of the averaged ERP epoch from each data point in the epoch.

Analysis of P3 peaks

Quantitative analysis was specifically performed for the P3 component obtained from the Cz electrode (this electrode was chosen because the magnitude of the P3 component measured with this electrode was largest, as expected from the literature for the P3b component) (Squires et al., 1975; Polich, 2007). P3 peaks were defined as the maximal values between 350 and 500 ms after stimulus onset for each of the two stimuli in a trial. The amplitude and the latency of the P3 peaks were measured for each subject and for each protocol (or trial type) separately. An ANOVA with within-subjects factors of measurement (two levels: separate vs interleaved) and protocol (reference first vs reference second) was conducted separately for the first and second peak values.

Results

Behavioral experiments

Five groups of listeners were tested with five protocols, respectively (see Materials and Methods and Table 1). All groups performed a two-tone frequency discrimination task (with the exception of the implicit reference protocol; see below), in which participants were asked, “which of the two sequentially presented tones is higher?”

Three groups were tested with three different protocols in which one interval, whose position was fixed throughout the assessment blocks (either first or second of the two stimuli in a trial), contained sufficient information for task performance. Discrimination thresholds in these three groups were similarly low. Figure 1*a* shows the dynamics of the first threshold assessment as a function of trial number in these groups, performing the reference first (reference tone presented in the first interval on each trial; solid blue line), reference second (reference tone presented in the second interval; solid red line), and implicit reference (a single tone, above or below the implicit 1000 Hz reference, presented in every trial; solid green line) protocols, respectively. Thresholds did not significantly differ between the groups (3.8 ± 1.4 , 4.5 ± 1.1 , and $4.9 \pm 3\%$ difference between the two tones in a trial, respectively), even when the reference was never explicitly presented (implicit reference).

Two additional groups were tested with protocols in which there was no consistent interval that contained sufficient task-related information (Fig. 1*b*). One group (group 4) was tested with the random protocol (no repeated reference), which requires interstimulus comparisons between the two tones in a trial. As shown in Figure 1*b* (black line), thresholds measured for the random protocol were substantially higher ($14 \pm 2\%$) than those measured in the three protocols illustrated in Figure 1*a* ($F_{(3,80)} = 5.6$; $p < 0.003$; *post hoc* Scheffé’s contrast: $p < 0.03$ for reference first, reference second, and implicit reference compared with random; $p = 0.99$, NS, for comparing the first three protocols).

The other group (group 5) was tested with the reference interleaved protocol, in which the position of the reference tone was systematically switched between consecutive trials. Here successful performance requires either an interstimulus comparison or a trial-by-trial switch of the classified interval, which contains the

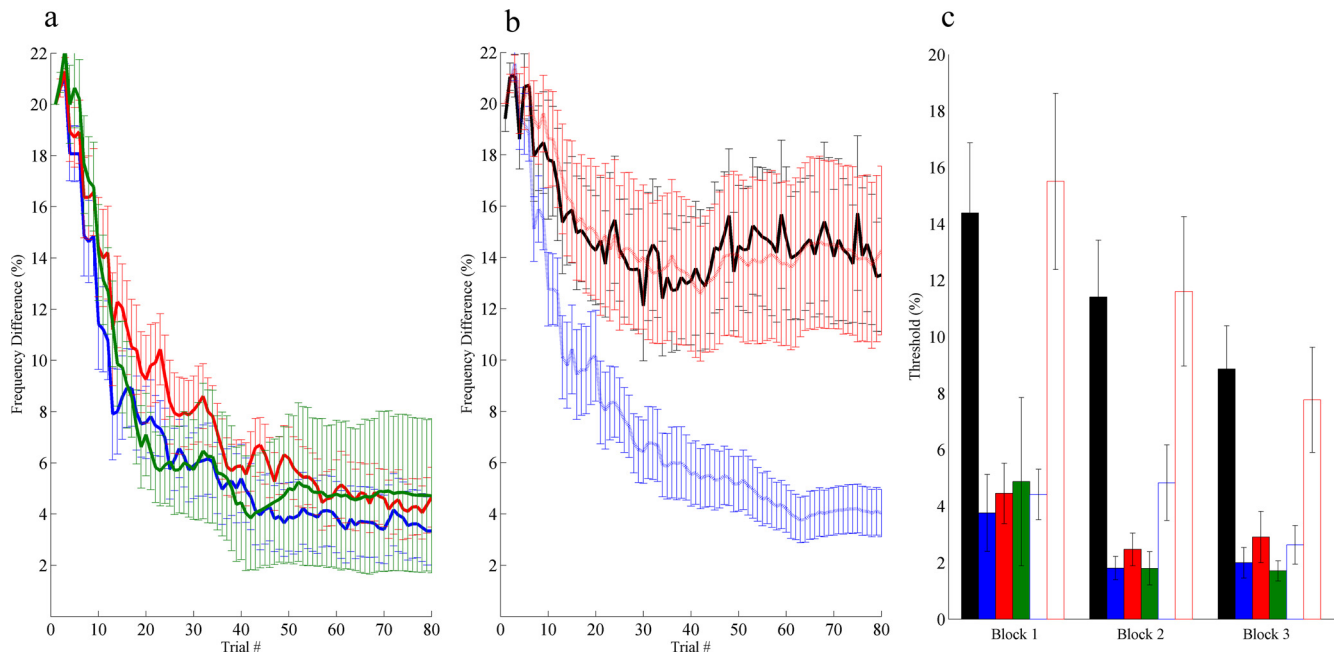


Figure 1. Behavioral performance in two-tone frequency discrimination under four different protocols. **a**, Percentage frequency difference between tones as a function of trial number in the first assessment block under three protocols: reference first (blue curve), reference second (red curve), and implicit reference (green curve). Performance was measured using an adaptive staircase procedure (Levitt, 1971), which began with a 20% frequency difference under all protocols and was adaptively modified according to listener's performance. All three protocols yielded similarly low thresholds. **b**, Percentage frequency difference along the first assessment for two additional protocols: random (no repeated reference) and reference interleaved (reference first on odd trials and reference second on even trials) measured with the same adaptive paradigm. Note that the random and reference second trials yielded similarly high thresholds, whereas the reference first trials yielded low thresholds, almost as low as those measured separately. **c**, Thresholds (percentage frequency difference) in each of the three blocks performed consecutively on each of the five protocols: random (black bars), reference first (filled blue bars), reference second (filled red bars), implicit reference (green bars), and reference interleaved (open blue and red bars for reference first and reference second thresholds, respectively). Although all thresholds showed improvement, their ranks were retained across assessments. Cross-subject averages and SEMs are shown; $n = 21$ in each group, except for the reference interleaved group in which $n = 25$.

nonreference tone. In this protocol, the thresholds obtained in reference first (odd trials (Fig. 1*b*, dotted blue line) greatly differed from those obtained in reference second (even) trials (Fig. 1*b*, dotted red line). In the interleaved reference first trials, performance was just as good as in the group that performed the reference first protocol separately ($4.4 \pm 0.9\%$ compared with $3.8 \pm 1.4\%$ for the separate measurement; $F_{(1,43)} = 1.5$, $p = 0.23$) (compare with blue line in Fig. 1*a*). However, in the interleaved reference second trials, performance dramatically degraded compared with that of the group that performed the reference second protocol separately ($15.5 \pm 3\%$ compared with $4.5 \pm 1\%$ for the separate measurement; $F_{(1,43)} = 8.8$, $p < 0.005$) (compare with red line in Fig. 1*a*) and was as poor as in the random protocol (comparing reference second interleaved with random: $F_{(1,44)} = 0.75$, $p = 0.79$). The asymmetry that was found between the odd and the even trials within the reference interleaved protocol cannot be explained in terms of the accuracy of the internal representation, because both trial types contained a repeated reference tone in a fixed position.

Administering these five protocols for three consecutive assessments induced some improvement in all of them but did not affect the magnitude of the benefit of the reference-containing protocols over the others (Fig. 1*c*). Thus, by the third block, thresholds in the random and reference second trials in the reference interleaved protocol were more than three times larger than the thresholds in the other protocols.

This more than threefold difference in thresholds, in favor of the protocols containing a reference at a fixed temporal position in the trial, was consistent across groups with various degrees of previous musical background and training with the specific task.

Thus, when the group of naive participants (shown in Fig. 1) was subsequently divided according to individuals' previous musical experience, similar patterns showing large benefits for reference use were found among both subgroups: those with no musical background (Fig. 2*a*) and those with musical background (Fig. 2*b*).

We further assessed this effect among individuals with lots of previous experience with the specific task (laboratory members, group 6) (Table 1). As expected from previous results in the perceptual learning literature (Tremblay et al., 1997; Irvine et al., 2000; Delhommeau et al., 2002; Demany and Semal, 2002; Hawkey et al., 2004; Amitay et al., 2005; Wright and Fitzgerald, 2005; van Wassenhove and Nagarajan, 2007; Halliday et al., 2008), thresholds for these highly experienced individuals were substantially lower. However, they showed a similarly large impact of protocol as inexperienced participants (Fig. 2*c*), indicating that the benefit of the reference (explicit or implicit) does not reflect the first stage of a faster long-term perceptual learning mechanism.

Despite the dramatic effect of the protocol on discrimination performance, listeners were unaware of any cross-trial repetition and naturally unaware of its temporal consistency. When asked immediately after assessments, all participants reported that they listened to the first tone, retained it in memory, listened to the second tone, and then compared the two. Trained laboratory members (group 6), who all performed reference first, reference second, and random protocols, were also unaware of the protocol while performing the task and had a similar introspection regarding their strategy for task performance.

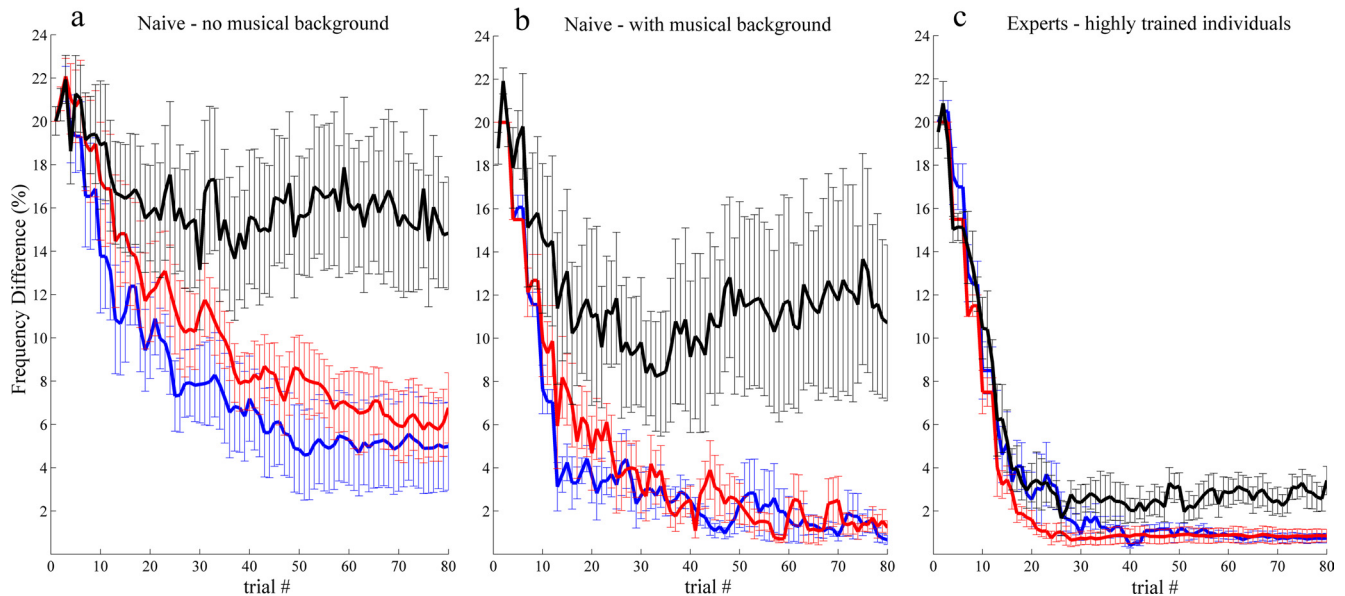


Figure 2. The advantage of the reference-containing conditions over the random protocol in different participant groups. Performance on reference first (blue), reference second (red), and random (black) protocols is shown for three participant groups: **a**, naive to task with no musical training ($n = 8$ for reference first and reference second protocols; $n = 10$ for random protocol); **b**, naive to task with at least 1 year of musical training ($n = 13$ for reference first and reference second; $n = 11$ for random); and **c**, laboratory members who are highly trained with the task ($n = 9$; each subject did all three protocols). The data for the total group of participants shown in **a** and **b** was shown in Figure 1. Graphs show frequency difference as a function of trial number throughout the first assessment for the three protocols; overall performance is best for the expert group (**c**) and is worst for the naive with no musical training group (**a**). However, all groups show a similarly large benefit in thresholds for the reference-containing protocols compared with the no-reference (random) protocol (2.8, 6, and 3.7 for the threshold ratio between random and reference first protocols for the 3 subject groups, respectively).

ERP experiments

Two ERP experiments were conducted. In the first experiment, measurements were taken while participants performed the reference first and reference second protocols in separate blocks. The protocols were similar to those used in the behavior-only experiments described above, except that constant (rather than adaptive) frequency differences (1–2%, in the range of the thresholds reached by most participants after a few hundred trials using the adaptive procedure) were used. The average ERP responses (at central electrode Cz) obtained under these two protocols are similar in terms of the N1, P2, and N2 components (Fig. 3, blue and red lines, for reference first and reference second, respectively). Each of these components was induced in response to each of the two tones in each trial; the N1 response to the second tone was somewhat smaller, in line with previous reports for successive tones with a temporal interval smaller than 1 s (Hari et al., 1982; Nahum et al., 2009).

The crucial difference between the ERP responses in the two protocols was in the position of the later P3 component. In the reference first trials (blue plot), in which the nonreference stimuli were presented in the second interval, a clear P3 was elicited at the end of the trial (average peak amplitude, $3.4 \pm 0.5 \mu\text{V}$), and no P3 was elicited after the first stimulus (average amplitude, $1.4 \pm 0.2 \mu\text{V}$). Conversely, in the reference second protocol (red plot), in which the informative (nonrefer-

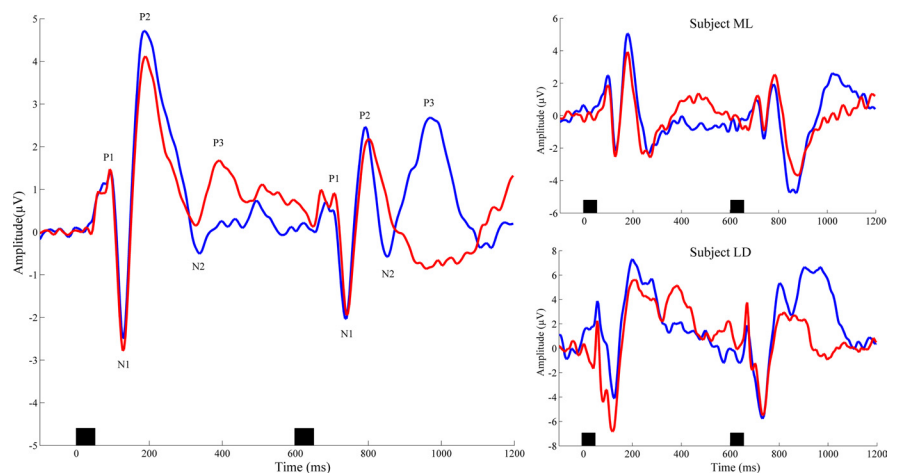


Figure 3. ERP waves recorded while participants performed the two-tone frequency discrimination task with a repeated reference in a fixed position across trials. Participants performed the reference first (blue) and reference second (red) protocols in separate blocks, with 1–2% difference between the reference and nonreference tones. Most participants (9 of 10) were good listeners and reached $\sim 95\%$ correct responses (averaged performance on reference first, $95 \pm 4\%$; reference second, $93 \pm 6\%$). One participant had chance-level performance on the first session, and her data were analyzed separately (see Fig. 4). The plotted waves are taken from the central electrode Cz and were averaged across participants ($n = 9$). The temporal location of the tones in a trial is marked by the black rectangles at the bottom of each plot. The relevant components are marked on the averaged waveforms. A clear P3 component can be seen after the second tone in the reference first protocol and after the first (nonreference) tone in the reference second protocol. The insets on the right show examples of two participants: ML had no previous experience with the task and no musical background, and LD had previous experience with the task and musical background (see Materials and Methods). Participants typically responded ~ 400 ms after the second tone in both protocols.

ence) stimulus was presented in the first interval and the second (reference) tone was not necessary for determining which tone is higher, no P3 was elicited after the second tone (average amplitude, $0.6 \pm 0.2 \mu\text{V}$); instead, a clear P3 response followed the first, nonreference tone ($2.4 \pm 0.5 \mu\text{V}$). These differences were consistent across individuals (comparing the first P3 between proto-

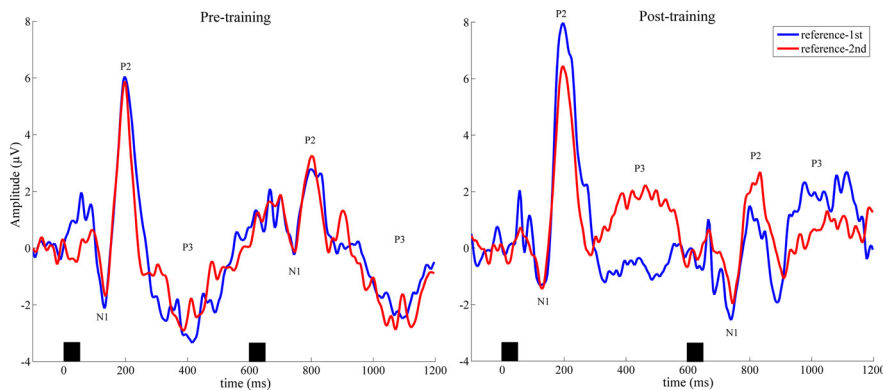


Figure 4. ERP waves recorded for the reference first and reference second protocols before (left) and after (right) 2 weeks of training. Participant YC performed around chance level in the initial ERP assessment of the reference-containing protocols (1–2% frequency difference; reference first, 59.5% correct; reference second, 41.5% correct). The initial ERP waves reflect this poor performance, showing no P3 component for reference first (blue curve) or reference second (red curve) trials. After training with the adaptive protocol, another ERP measurement session was administered. Performance reached 91 and 94% success for reference first and reference second blocks, respectively. Posttraining ERP measures (right) reflect this behavioral change. A clear P3 component is seen after the second (nonreference) tone in the reference first protocol (blue curve) and after the first but not the second tone for the reference second protocol (red curve).

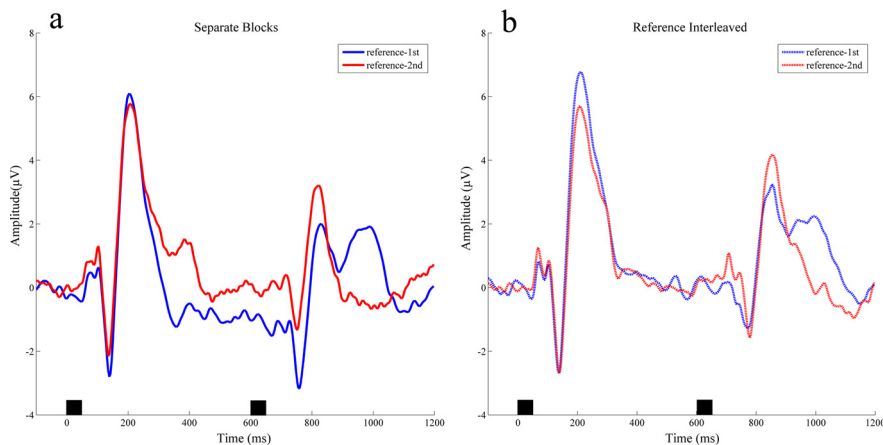


Figure 5. ERP measurements of the reference first and reference second performed separately (**a**) and in (reference) interleaved blocks (**b**). **a**, Averaged response traces recorded from participants ($n = 10$) performing reference first (blue trace; $96 \pm 2\%$ correct) and reference second (red trace; $87 \pm 6\%$ correct) protocols in separate blocks, using 4 and 8% frequency difference. A clear P3 component is evident after the second tone for reference first protocol and after the first tone for reference second protocol. **b**, Averaged response traces recorded from the same participants while performing the reference interleaved protocol (dotted blue trace, reference first trials, $81 \pm 5\%$ correct; dotted red trace, reference second trials, $84 \pm 4\%$ correct). The same frequency difference of 4 and 8% between the reference and nonreference tones was used. A clear P3 component was elicited at the end of the trial for the reference first trials but not after the first tone for the reference second trials.

cols, $p < 0.02$; second P3, $p < 0.0001$). Thus, the P3 response was produced in each of the two protocols ~ 350 ms after the presentation of the nonreference stimulus. Because this component is reliably, and even implicitly, produced when a perceptual decision is made (Donchin and Coles, 1988; Verleger, 1988; Nieuwenhuis et al., 2005), the shift of its position suggests that an implicit decision was made immediately after the presentation of the category determining stimulus, even when this stimulus was presented first (as in reference second). Thus, an implicit decision was made with no online interstimulus comparison, in contrast to listeners' subjective report and in contrast to the prevailing comparison assumption.

One participant was a poor listener and performed the task (the nonadaptive procedure used during ERP recordings) at chance level. This participant showed no P3 responses in either

the reference first or the reference second blocks (Fig. 4, left, blue and red lines). However, after 2 weeks of intensive practice, performance improved, and ERPs measured during task performance were similar to those of good listeners (Fig. 4, right). These findings suggest that P3 is not produced when the task is not performed successfully. However, once successful performance is obtained, an implicit perceptual decision is made immediately after the informative, nonreference tone.

The data presented thus far are consistent with a quick, implicit strategy change from an online comparison between the two tones to task-related classification of the tone presented in the informative, nonreference interval. To further test this hypothesis, we measured ERP responses under the reference interleaved protocol. Based on our previous behavioral results with the reference interleaved protocol (Fig. 1b), we chose to use larger frequency differences between the tones (4–8%) so that performance in both reference first and reference second trials will be adequate. To ensure that the use of a larger frequency difference does not impact the pattern of ERP responses in the separate protocols, we administered them as well. As shown in Figure 5a, ERP response patterns in the separate protocols were similar to those obtained with the smaller frequency differences (compare the right side of Fig. 5 with the plots in Fig. 3): a clear P3 was formed after the second tone in reference first and after the first tone in reference second blocks. However, the potentials induced in the reference interleaved protocol (Fig. 5b, dotted lines) were different and showed the same type of asymmetry between the reference first and reference second trials that was found in the behavioral experiment. A relatively clear P3 was induced at the end of the odd, reference first trials ($3.1 \pm 0.9 \mu\text{V}$, similar in magnitude to the $3.5 \pm 0.8 \mu\text{V}$ measured separately; effect of measurement: $F_{(1,9)} = 0.001$, $p = 0.98$, NS). However, in the reference second trials, no clear P3 was induced after either the first ($1.2 \pm 0.4 \mu\text{V}$) or the second ($2.1 \pm 0.4 \mu\text{V}$; significant effect of protocol, $F_{(1,9)} = 6.6$, $p < 0.05$; significant interaction of protocol \times measurement, $F_{(1,9)} = 7.03$, $p < 0.03$) interval. Only a marginal P3 was induced after the second interval (Fig. 5b, dotted red curve). The lack of a clear P3 in reference second trials suggests that, in these trials, participants could not reach a clear perceptual decision at any consistent point in time with respect to stimulus presentation.

Summary of results

We measured behavioral performance and ERP waves under several stimulation protocols of a two-tone frequency discrimination task that differed in the presence and location of a consistent

reference tone. Behaviorally, we found that best performance (lowest thresholds) was achieved for reference first, reference second, and implicit reference protocols, as well as for reference first trials in the reference interleaved protocol. Thresholds were more than threefold higher (i.e., worse) in the random protocol (no repeated reference) and in the reference second trials of the reference interleaved protocol. The advantage of the reference-containing protocols was consistent across the degree of musical training and familiarity with the task.

ERP measurements, recorded while participants performed the reference first and reference second protocols, mainly differed in the location of the P3 component. P3 followed the second tone in reference first blocks and the first tone in reference second blocks. Thus, in both cases, it followed the informative, nonreference tone. In the reference interleaved protocol, it followed the second tone in reference first trials and was almost absent in reference second trials. This asymmetry is in line with the asymmetrical performance in these trial types when they are interleaved. Reference first performance is not affected, whereas reference second performance does not benefit from the repeated reference.

Discussion

The main conclusion of this study is that the classical approach, which uses fixed thresholds for characterizing human discrimination ability, should be modified. We found that seemingly minute changes in the assessment protocols induce dramatic changes to the measured thresholds. Specifically, discrimination thresholds are systematically dictated by the temporal cross-trial structure of the assessment protocol. Thus, when characterizing discrimination thresholds, the protocol used for assessment should be part of the description.

Some effects of the measurement protocol on performance have been reported previously (Morgan et al., 2000; Nachmias, 2006; Hairston and Nagarjan, 2007; Lapid et al., 2008). However, previous studies attributed the improvement in threshold to increased accuracy of the internal representation of the repeated reference (Durlach and Braida, 1969, 1988). We now find that discrimination thresholds may be quite poor, even when the reference is repeatedly presented on every trial and its temporal position is fully predicted, as in the reference second trials of the reference interleaved protocol. Thus, decreasing internal noise and providing sufficient information regarding the informative interval is not sufficient for attaining low thresholds. Moreover, impressively low thresholds were obtained even when the reference stimulus was never explicitly presented, as in the implicit reference protocol. Thus, decreasing internal noise by using a repeated reference is not necessary for attaining low thresholds. We propose that the protocols that afford impressively low discrimination thresholds are those that allow our perceptual system to switch the underlying mechanism from interstimulus comparison to classification, which is based on comparison of a single stimulus, typically in a fixed interval, to an internal reference. The dramatically improved performance stems from the greater accuracy of the classification mechanism.

Classification versus comparison mechanisms

The three protocols that consistently yielded lowest discrimination thresholds share two characteristics. First, successful performance (except for the first few trials) could be based on a single, nonreference tone (lower or higher than 1000 Hz). Second, this informative tone is presented at a fixed interval throughout the assessment block. These characteristics afford classification based on a single, fixed interval.

In these protocols, successful performance was correlated with the formation of a P3 response immediately (300–350 ms) after the presentation of the informative (nonreference) tone. A P3 response followed the first tone in the reference second protocol and the second tone in the reference first protocol. Several investigators suggested that the P3 component is produced when a clear task-related categorization (“high” or “low” in our case) is made (Donchin and Coles, 1988; Verleger, 1988; Mecklinger and Ullsperger, 1993; Nieuwenhuis et al., 2005). In our study, we could therefore use it as a marker for the point in time in which a decision had been made, even when listeners are unaware of such a decision. Our findings indicate that the implicit decision in the two-tone discrimination task had been made based on the single informative stimulus.

The difference in performance between the successful protocols and those inducing higher thresholds was apparent already early within the first assessment. We therefore propose that the switch to classification of stimuli that are evaluated as informative is quick. Moreover, it did not result from participants’ explicit intention or even their awareness and may therefore be automatic. However, under some conditions, as in the random protocol, in which classification of a single stimulus is not sufficient for successful performance, a comparison between the representations of two recently presented stimuli is required. When classification of a single stimulus is sufficient but its position is not fixed, a mixed strategy may be used. Crucially, both protocols involve interstimulus comparisons, which require explicit decision and activation of working memory mechanisms to retain and compare the two recently presented stimuli. We suggest that the accuracy of the implicit classification processes is inherently better than the accuracy obtained under explicit comparison mechanisms.

There is ample evidence that these two types of comparisons are indeed implemented in the brain by different mechanisms. For example, Hernández et al. (1997) trained monkeys on a tactile discrimination task using a reference first protocol, until a low threshold was reached. After training, the monkeys were tested on the implicit reference protocol, in which they retained their trained level of accuracy, and on the random protocol, which they completely failed. This result suggests that, in the testing phase, the monkeys kept classifying the second stimulus with respect to the learned internal reference and ignored the first stimulus. Attaining successful performance in the random protocol, which requires the monkeys to retain-and-compare sequentially presented stimuli, needed intensive additional training. The authors concluded that monkeys (and probably humans, who also participated, but were not trained, in this study) have difficulties using retain-and-compare mechanisms and use classification when possible. The similarity in the impact of the assessment protocol on the relative thresholds of highly trained monkeys, highly trained humans, and naive human performers suggests that similar mechanisms underlie classification with respect to a recently formed reference and to a reference retrieved from long-term memory.

A similar type of separation between underlying mechanisms using seemingly similar tasks is provided by tasks that require to indicate whether a stimulus had been presented recently compared with those asking to indicate when was it presented. The first task requires only implicit automatic mechanisms, whereas the latter requires explicit working memory mechanisms. For example, when listeners are asked to perform an *n*-back task, i.e., report whether an element is a repetition of the element presented *n* steps earlier in the sequence, they have no difficulty reporting that the element was recently presented, supported by the implicit memory classification mechanism. However, they have

huge difficulties reporting its exact serial position, supported by explicit working memory (Gray et al., 2003). Similarly, when monkeys are asked to retrieve the serial order of stimulus presentation, their default behavior relies on the highly trained rather than on the most recent presentation order (Orlov et al., 2000).

Additional evidence for the separation of these two types of mechanisms comes from studies of dyslexia. Although activating explicit working memory mechanisms is typically more demanding, some individuals have specific difficulties in gaining benefits from the implicit mechanisms using cross-trial repetitions. These individuals have difficulties in switching from comparison to classification even when the protocols afford such a switch for the general population. Recent findings (Ahissar et al., 2006; Banai and Ahissar, 2006; Ahissar, 2007) indicate that such is the case for many individuals with reading disabilities, who perform the random protocols adequately, whereas their performance in the reference-containing protocols is impaired compared with that of controls. These findings suggest that they have adequate inter-stimulus comparison mechanisms but difficulties in the mechanisms underlying the formation of internal references and automatic stimulus classification. Therefore, a broad range of tasks remains effortful for them.

Underlying neuronal mechanisms

The formation of the P3 component that we recorded provides a useful indication that an implicit decision had been made. However, it is not too informative with respect to the exact timing of this decision, its neuronal origin, or its underlying mechanism. The accumulative findings suggest that P3 reflects the response of the norepinephrine system to the outcome of internal decision-making processes and the consequent effects of noradrenergic potentiation of information processing (Nieuwenhuis et al., 2005). As such, it is probably evoked by a spread network of activity, is an outcome of a variety of decision-making mechanisms, and can only provide an upper boundary for when a decision had been made.

However, single-unit recordings in monkeys using similar tasks shed light on the neural origins of the response. A series of seminal studies by Romo and his colleagues directly assessed the neural mechanisms underlying performance in these two protocols. The monkeys were trained to perform a two-vibration tactile discrimination task, in which they were asked which of two sequentially administered vibrations had a higher frequency (Hernández et al., 2002; Romo et al., 2002). Neuronal activity was recorded in a series of cortical areas while monkeys performed an equivalent of our random protocol. Whereas S1 neurons responded only during the presentation of the stimuli, neurons in a series of higher-level areas, including the prefrontal cortex and ventral and medial premotor cortices, were active during the delay interval. This observation suggests that, in retain-and-compare demanding conditions, the magnitude of the first stimulus is actively retained with delayed, poststimulus activity. Delayed activity is not an automatic property of these areas (Miller et al., 1996), and its accurate retention may be difficult (Seung, 1996).

Indeed, when monkeys performed the reference first protocol, they ceased to produce delayed activity but exhibited fantastic discrimination thresholds (Romo et al., 2002). Similar results were recently obtained when monkeys performed an auditory equivalent of this task, in which they were required to compare the rate of two sequentially presented auditory pulse trains. Similarly, no delayed activity was found in primary auditory cortex A1 (Lemus et al., 2009a), whereas neurons in ventral premotor cortex showed clear delayed activity (Lemus et al., 2009b). Thus,

although activity in A1 is related to both the behavioral condition (Fritz et al., 2003; Otazu et al., 2009) and the protocol of assessment (Ulanovsky et al., 2003), current data suggest that A1 (or primary somatosensory cortex S1) does not implement inter-stimulus comparisons (Salinas et al., 2000; Romo and Salinas, 2003; Yoon et al., 2006).

Pinpointing the neural mechanisms underlying classification is difficult perhaps because it is the automatic or the default mode of our perceptual system. We propose that the automatic, implicit mechanisms of classification benefit from the sharpening of the reference by adaptation mechanisms, whose impact is retained through intervening trials (Li et al., 1993; Ulanovsky et al., 2004).

Conclusion

Our findings suggest that our impressively low discrimination thresholds do not reflect refined explicit retain-and-compare mechanisms, as assumed by most behavioral studies, but rather an effective implicit strategy shift from discrimination to classification, based on quickly established internal references. The similarity between the impact of the assessment protocol on naive and trained (both human and monkey) performers suggests that similar mechanisms underlie classification, whether the reference was learned long ago or just recently. The impressive effectiveness of this strategy shift probably stems from the quick ability of higher-level mechanisms to detect stimulus repetitions at lower levels but only those repetitions that follow the specific temporal patterns that these levels are predisposed to detect. This shift may serve as an automatic mechanism for releasing the specialized high-level working memory networks to engage in novel comparisons (for such distinction, see Zatorre et al., 1998), whereas the quickly-learned classifications are “down-sourced” and more efficiently implemented by lower-level cortical networks (Schneider and Shiffrin, 1977; Ohl and Scheich, 2005).

References

- Ahissar M (2007) Dyslexia and the anchoring-deficit hypothesis. *Trends Cogn Sci* 11:458–465.
- Ahissar M, Lubin Y, Putter-Katz H, Banai K (2006) Dyslexia and the failure to form a perceptual anchor. *Nat Neurosci* 9:1558–1564.
- Amitay S, Hawkey DJ, Moore DR (2005) Auditory frequency discrimination learning is affected by stimulus variability. *Percept Psychophys* 67:691–698.
- Banai K, Ahissar M (2006) Auditory processing deficits in dyslexia: task or stimulus related? *Cereb Cortex* 16:1718–1728.
- Delhommeau K, Micheyl C, Jouvent R, Collet L (2002) Transfer of learning across durations and ears in auditory frequency discrimination. *Percept Psychophys* 64:426–436.
- Demany L, Semal C (2002) Learning to perceive pitch differences. *J Acoust Soc Am* 111:1377–1388.
- Donchin E, Coles MG (1988) Is the P300 component a manifestation of context updating? *Behav Brain Sci* 11:357–374.
- Durlach NI, Braida LD (1969) Intensity perception. I. Preliminary theory of intensity resolution. *J Acoust Soc Am* 46:372–383.
- Durlach NL, Braida LD (1988) Peripheral and central factors in intensity perception. In: *Auditory function: neurological basis of hearing* (Edelman G, Cohen M, Gall WE, eds). New York: Wiley.
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223.
- Garner WR (1953) An informational analysis of absolute judgments of loudness. *J Exp Psychol* 46:373–380.
- Georgeson MA, Yates TA, Schofield AJ (2008) Discriminating depth in corrugated stereo surfaces: facilitation by a pedestal is explained by removal of uncertainty. *Vision Res* 48:2321–2328.
- Gray JR, Chabris CF, Braver TS (2003) Neural mechanisms of general fluid intelligence. *Nat Neurosci* 6:316–322.
- Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. New York: Wiley.

- Hairston IS, Nagarajan SS (2007) Neural mechanisms of the time-order error: an MEG study. *J Cogn Neurosci* 19:1163–1174.
- Halliday LF, Taylor JL, Edmondson-Jones AM, Moore DR (2008) Frequency discrimination learning in children. *J Acoust Soc Am* 123:4393–4402.
- Hari R, Kaila K, Katila T, Tuomisto T, Varpula T (1982) Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalogr Clin Neurophysiol* 54:561–569.
- Harris J (1948) Discrimination of pitch: suggestions toward method and procedure. *Am J Psychol* 61:309–322.
- Hawkey DJ, Amitay S, Moore DR (2004) Early and rapid perceptual learning. *Nat Neurosci* 7:1055–1056.
- Hernández A, Salinas E, García R, Romo R (1997) Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *J Neurosci* 17:6391–6400.
- Hernández A, Zainos A, Romo R (2002) Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33:959–972.
- Irvine DR, Martin RL, Klimkeit E, Smith R (2000) Specificity of perceptual learning in a frequency discrimination task. *J Acoust Soc Am* 108:2964–2968.
- Karmarkar UR, Buonomano DV (2003) Temporal specificity of perceptual learning in an auditory discrimination task. *Learn Mem* 10:141–147.
- Lages M, Treisman M (1998) Spatial frequency discrimination: visual long-term memory or criterion setting? *Vision Res* 38:557–572.
- Laming D (2004) Human judgment: the eye of the beholder. Andover, UK: Cengage Learning EMEA Higher Education.
- Lapid E, Ulrich R, Rammsayer T (2008) On estimating the difference limen in duration discrimination tasks: a comparison of the 2AFC and the reminder task. *Percept Psychophys* 70:291–305.
- Leek MR (2001) Adaptive procedures in psychophysical research. *Percept Psychophys* 63:1279–1292.
- Lemus L, Hernández A, Romo R (2009a) Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *Proc Natl Acad Sci U S A* 106:9471–9476.
- Lemus L, Hernández A, Romo R (2009b) Neural encoding of auditory discrimination in ventral premotor cortex. *Proc Natl Acad Sci U S A* 106:14640–14645.
- Levitt H (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:Suppl 2:467+.
- Li L, Miller EK, Desimone R (1993) The representation of stimulus familiarity in anterior inferior temporal cortex. *J Neurophysiol* 69:1918–1929.
- Macmillan NA, Creelman CD (2005) Detection theory: a user's guide. London: Erlbaum.
- Mecklinger A, Ullsperger P (1993) P3 varies with stimulus categorization rather than probability. *Electroencephalogr Clin Neurophysiol* 86:395–407.
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16:5154–5167.
- Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for information processing. *Psychol Rev* 63:81–97.
- Moore BCJ (1997) An introduction to the psychology of hearing. London: Academic.
- Morgan MJ, Watamaniuk SN, McKee SP (2000) The use of an implicit standard for measuring discrimination thresholds. *Vision Res* 40:2341–2349.
- Nachmias J (2006) The role of virtual standards in visual discrimination. *Vision Res* 46:2456–2464.
- Nahum M, Renvall H, Ahissar M (2009) Dynamics of cortical responses to tone pairs in relation to task difficulty: a MEG study. *Hum Brain Mapp* 30:1592–1604.
- Nieuwenhuis S, Aston-Jones G, Cohen JD (2005) Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol Bull* 131:510–532.
- Ohl FW, Scheich H (2005) Learning-induced plasticity in animal and human auditory cortex. *Curr Opin Neurobiol* 15:470–477.
- Orlov T, Yakovlev V, Hochstein S, Zohary E (2000) Macaque monkeys categorize images by their ordinal number. *Nature* 404:77–80.
- Otazu GH, Tai LH, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* 12:646–654.
- Oxenham AJ, Buus S (2000) Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions. *J Acoust Soc Am* 107:1605–1614.
- Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118:2128–2148.
- Pollack I (1953) The information of elementary auditory displays. II. *J Acoust Soc Am* 25:765–769.
- Regan D (1985) Masking of spatial-frequency discrimination. *J Opt Soc Am A* 2:1153–1159.
- Regan D, Beverley KI (1985) Post adaptation orientation discrimination. *J Opt Soc Am A* 2:147–155.
- Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci* 4:203–218.
- Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399:470–473.
- Romo R, Hernández A, Zainos A, Lemus L, Brody CD (2002) Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat Neurosci* 5:1217–1225.
- Romo R, Hernández A, Zainos A (2004) Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41:165–173.
- Salinas E, Hernandez A, Zainos A, Romo R (2000) Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *J Neurosci* 20:5503–5515.
- Schneider W, Shiffrin RM (1977) Controlled and automatic human information processing. I. Detection, search, and attention. *Psychol Rev* 84:1–66.
- Seung HS (1996) How the brain keeps the eyes still. *Proc Natl Acad Sci U S A* 93:13339–13344.
- Squires KC, Squires NK, Hillyard SA (1975) Decision-related cortical potentials during an auditory signal detection task with cued observation intervals. *J Exp Psychol Hum Percept Perform* 1:268–279.
- Stewart N, Brown GD, Chater N (2005) Absolute identification by relative judgment. *Psychol Rev* 112:881–911.
- Tremblay K, Kraus N, Carrell TD, McGee T (1997) Central auditory system plasticity: generalization to novel stimuli following listening training. *J Acoust Soc Am* 102:3762–3773.
- Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6:391–398.
- Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* 24:10440–10453.
- Ulrich R, Nitschke J, Rammsayer T (2006) Crossmodal temporal discrimination: assessing the predictions of a general pacemaker counter model. *Percept Psychophys* 68:1140–1152.
- van Wassenhove V, Nagarajan SS (2007) Auditory cortical plasticity in learning to discriminate modulation rate. *J Neurosci* 27:2663–2672.
- Verleger R (1988) Event-related potentials and cognition: a critique of the context updating hypothesis and an alternative interpretation of P3. *Behav Brain Sci* 11:343–427.
- Viemeister NF (1970) Intensity discrimination: performance in three paradigms. *Percept Psychophys* 8:417–419.
- Wright BA, Fitzgerald MB (2005) Learning and generalization of five auditory discrimination tasks as assessed by threshold changes. In: *Auditory signal processing: physiology, psychoacoustics, and models* (Pressnitzer D, de Cheveigne A, McAdams S, Collet L, eds). New York: Springer.
- Wright BA, Buonomano DV, Mahncke HW, Merzenich MM (1997) Learning and generalization of auditory temporal-interval discrimination in humans. *J Neurosci* 17:3956–3963.
- Yeshurun Y, Carrasco M, Maloney LT (2008) Bias and sensitivity in two-interval forced choice procedures: tests of the difference model. *Vision Res* 48:1837–1851.
- Yoon JH, Curtis CE, D'Esposito M (2006) Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. *Neuroimage* 29:1117–1126.
- Zaksas D, Pasternak T (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J Neurosci* 26:11726–11742.
- Zatorre RJ, Perry DW, Beckett CA, Westbury CF, Evans AC (1998) Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proc Natl Acad Sci U S A* 95:3172–3177.