Vision Research 50 (2010) 391–401

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

# Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning

Mor Nahum [a,*], Israel Nelken [a,b], Merav Ahissar [a,c]

[a] Interdisciplinary Center for Neural Computation (ICNC), Hebrew University, Jerusalem, Israel
[b] Department of Neurobiology, Hebrew University, Jerusalem, Israel
[c] Department of Psychology, Hebrew University, Jerusalem, Israel

ARTICLE INFO

*Article history:*
Received 6 May 2009
Received in revised form 8 September 2009

*Keywords:*
Perceptual learning
Auditory
Reverse hierarchy theory
Protocol
Uncertainty
Binaural benefit
Temporal patterning

ABSTRACT

We examined the impact of variability in speech stimuli on improvement of general performance and on accessibility to low-level information as a function of practice. Listeners had to discriminate between two similar words in noise in two configurations that differed only in their low-level binaural information, which was either null or maximal. The difference in performance quantifies the use of binaural low-level information. These configurations were presented in three training protocols: in separate blocks; in a consistently interleaved manner; and in a randomly mixed manner. The first protocol enabled optimal use of the low-level binaural cues already at the first training session. The second, consistently interleaved protocol required more than one training session to reach the same performance. The final, mixed protocol did not enable optimal use of the low-level cues even after multi-session training. Interestingly, training with the first two protocols transferred to the mixed one. These results are in line with recent findings in the visual modality. In both modalities, the effects of variability on learning can be explained by the introduction of obstructions to a search mechanism going down along the sensory processing hierarchy, as suggested by the Reverse Hierarchy Theory.

## 1. Introduction

One of the exciting recent findings in the field of perceptual learning is the specificity of improvement due to training on discrimination of low-level features. For example, learning visual orientation discrimination is specific to both the trained position and the trained orientation (Fiorentini & Berardi, 1980; Schoups, Vogels, & Orban, 1995; Shiu & Pashler, 1992). Similarly, the improvement obtained in motion or direction discrimination is specific to the trained direction of movement (Ball & Sekuler, 1987; Saffell & Matthews, 2003). This specificity suggests that learning involves representations which are narrowly tuned along physical dimensions at the peripheral organ such as retinal position, bar orientation and movement direction. Such specificity characterizes response properties of neurons at low-level visual cortices (Karni & Sagi, 1991, 1993; Schoups, Vogels, Qian, & Orban, 2001; see review in Fahle, 2005), suggesting that learning involves these early, primary and secondary, visual cortices (Fahle, Poggio, & Edelman, 1995; Furmanski, Schluppeck, & Engel, 2004; Poggio, Fahle, & Edelman, 1992; Schoups et al., 2001). However, other studies found very limited stimulus specificity along the same dimensions, despite using similar tasks and stimuli (e.g. Schoups & Orban, 1996).

This discrepancy was explained by the Reverse Hierarchy Theory (RHT), developed by Ahissar and Hochstein (1997, 2004; See also Hochstein & Ahissar, 2002 and Ahissar, Nahum, Nelken, & Hochstein, 2009). RHT postulates that learning always begins at high-level cortices and may gradually progress to lower levels, provided that successful performance requires finer resolution than that allowed by higher levels and that an appropriate low-level neuronal population can be allocated via a top-down guided search (Ahissar & Hochstein, 1997, 2004; Ahissar et al., 2009). This conceptualization implies that although learning may reach low-level populations, and induce a large and specific improvement, such a scenario is not guaranteed, since learning is largely determined by the training protocol. For example, when the task is sufficiently coarse with respect to physical attributes of the stimuli so that successful performance does not require the fine resolution of low-level sensory representations, improvement at high levels of the sensory representation may suffice to achieve successful performance. In that case, access to low levels may not be pursued. Consequently, learning will be general with respects to the physical attributes of the stimulus since it will be based on higher-level neuronal representations. Indeed, when coarse and fine resolutions were trained (typically linked with easy and difficult conditions,

---

[*] Corresponding author. Address: Interdisciplinary Center for Neural Computation (ICNC), The Hebrew University of Jerusalem, Israel 91904, Israel.
*E-mail address:* mornahum@gmail.com (M. Nahum).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.09.004

respectively; see discussion in Nahum, Daikhin, Lubin, Cohen, & Ahissar, submitted for publication), learning of coarse and easy conditions was found to generalize better (Ahissar & Hochstein, 1997; Liu, 1999; reviewed in Ahissar & Hochstein, 2004).

Improvement is similarly expected to be based on high representation levels when successful performance requires low-level resolution, but the training protocol hinders this access. Allocation of an appropriate low-level population requires that it will be detected as informative by a top-down search mechanism. This backward search begins at a task-relevant high-level population, which serves as an initial pointer, and its inputs from lower-level populations are considered. An input population is identified as informative when its activity across trials is best correlated with the internal hypothesis regarding the external stimulus. To estimate the degree of correlation, the same input population should be informative across several consecutive trials. This condition is often satisfied when training is conducted with a very restricted set of stimuli. Indeed, studies that used massive training with the same physical stimuli are precisely those that found substantial improvement in conjunction with a high degree of stimulus specificity (e.g. Ahissar & Hochstein, 2000; Fahle & Morgan, 1996; Karni & Sagi, 1991; Schoups et al., 1995).

However, there is a complementary prediction: when the physical characteristics of the stimuli randomly vary across trials, there will be no consistently informative low-level population, and the backward search will fail, or at least be dramatically slowed (e.g. Parkosadze, Otto, Malania, Kezeli, & Herzog, 2008). Indeed, studies of contrast discrimination which used several reference contrasts found little or no learning when the reference contrasts were randomly interleaved across consecutive trials (Adini, Wilkonsky, Haspel, Tsodyks, & Sagi, 2004; Yu, Klein, & Levi, 2004). Yet, significant learning was obtained when each reference contrast was presented in a separate block (Adini et al., 2004; Yu et al., 2004).

An interesting addition to these observations was recently provided by Kuai, Zhang, Klein, Levi, and Yu (2005) and Zhang et al. (2008). These studies found that presenting the different reference contrasts in repeated temporal sequences enables learning, which in turn transfers to the randomized reference condition. Zhang et al. (2008) further showed that learning under random reference conditions can also occur in the presence of top-down segregating cues (‘tagging’) or by increasing the distance between the reference contrasts so that they become clearly distinct. According to RHT, these conditions enable learning since they eliminate the interference that cross-trial stimulus variability poses to the backward search. High-level segregation presumably enables separate backward search processes, each linked to a different high-level pointer. This segregation eliminates the interference since within each process a consistent input population is informative across trials.

Although most human perceptual learning studies were conducted in the visual modality, similar principles may apply in the auditory modality. Indeed, auditory learning was found using both simple tasks and stimuli (Amitay, Hawkey, & Moore, 2005; Amitay, Irwin, & Moore, 2006; Delhommeau, Micheyl, Jouvent, & Collet, 2002; Demany, 1985; Irvine, Martin, Klimkeit, & Smith, 2000; Karmarkar & Buonomano, 2003; van Wassenhove & Nagarajan, 2007; Wright, Buonomano, Mahncke, & Merzenich, 1997) and using complex stimuli, e.g. in the context of speech perception (e.g. Nygaard & Pisoni, 1998). We now asked whether auditory learning of complex speech stimuli is similarly affected by stimulus variability during training. Our conceptual framework was based on RHT, which we recently mapped onto the auditory modality (Nahum et al., submitted for publication; See also Nelken & Ahissar, 2006). While the auditory processing hierarchy is less understood than the visual one, it has few well-understood characteristics. One relatively well-defined low-level cue is inter-aural phase differences. This cue is calculated early in the auditory pathways, at the level of the Superior Olivary Complex (Batra, Kuwada, & Fitzpatrick, 1997a, 1997b; Yin & Chan, 1990), where its binaural resolution is maximal, and is only partially retained at higher stages (see Nahum et al., submitted for publication for further details).

The specific benefits of low-level cues for performance (binaural benefits) were measured by calculating the difference in discrimination thresholds under a *diotic* configuration, which contains no binaural information (as both signal and noise are identical in the two ears; see illustration in Fig. 1A) and under a *dichotic* configuration, which maximizes this information. In the *dichotic* configuration the noise presented to the two ears is the same, but the signal is phase reversed to one ear (see Fig. 1B). Consequently, the phase differences (corresponding to time shifts in the scale of less than 1 ms) between the signal reaching the two ears are maximal for each frequency, and a maximal segregation is formed between the origin of the signal and that of the noise. Nevertheless, reversing the phase of the signal does not modify its energy contour (as shown in Fig. 1B), which is considered the important cue for speech perception (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Therefore, the perceptual contents of the *diotic* and *dichotic* configurations do not differ, whereas the thresholds for performance are significantly lower under the *dichotic* configuration. We should note that at the low signal-to-noise ratios (SNR) we used, naïve listeners are unaware of the nature of this inter-aural manipulation, in spite of its impact on performance. Still, highly experienced listeners report that the diotic stimulus forms a focused auditory image at the center of the head, whereas the dichotic stimulus results in a blurred image.

Thresholds for performance under each of these two binaural configurations are affected by many low and high-level factors. However, the difference between these thresholds purely reflects access to low-level resolution. Large binaural benefits indicate that lower level processing stages have been accessed in the performance of the task (e.g. Blauert, 1997; Blauert & Cobben, 1978; Hirsch, 1948; Johansson & Arlinger, 2002; Levitt & Rabiner, 1967; Licklider, 1948).

The speech discrimination paradigm we used was already applied in a previous study where we assessed thresholds under binaurally random and consistent conditions, respectively (Nahum et al., submitted for publication). We found that *diotic* thresholds were not sensitive to the binaural configuration, as expected. However, *dichotic* performance (and hence the benefit from low-level cues) was significantly better in the consistent compared with the randomly mixed condition when using phonologically-similar words. We concluded that uncertainty in the binaural configuration inhibited access to low-level populations.

We now asked whether the limitations on low-level accessibility could be removed with intensive training. We tested the impact of three training protocols applied to different groups of listeners: a *consistent* protocol, in which diotic and dichotic configurations were presented separately; a *mixed* protocol, in which the two binaural configurations were randomly interleaved, and a patterned 1–1 protocol, in which odd trials were diotic and even trials were dichotic. We should note that the cross-trial uncertainty was not explicitly related to the required task, nor was it apparent to the listeners. We figured that it would be relevant because of the structure of the processing hierarchy and the limits it may induce on the backward search.

We found that even massive training did not allow access to low-level binaural information when binaural configurations were presented in a mixed manner. However, patterning the two binaural configurations resulted in a gradual increase in binaural benefit, suggesting that access to low-level representations was achieved. Following training with either the consistent or the temporally-patterned binaural protocols, listeners could fully utilize binaural

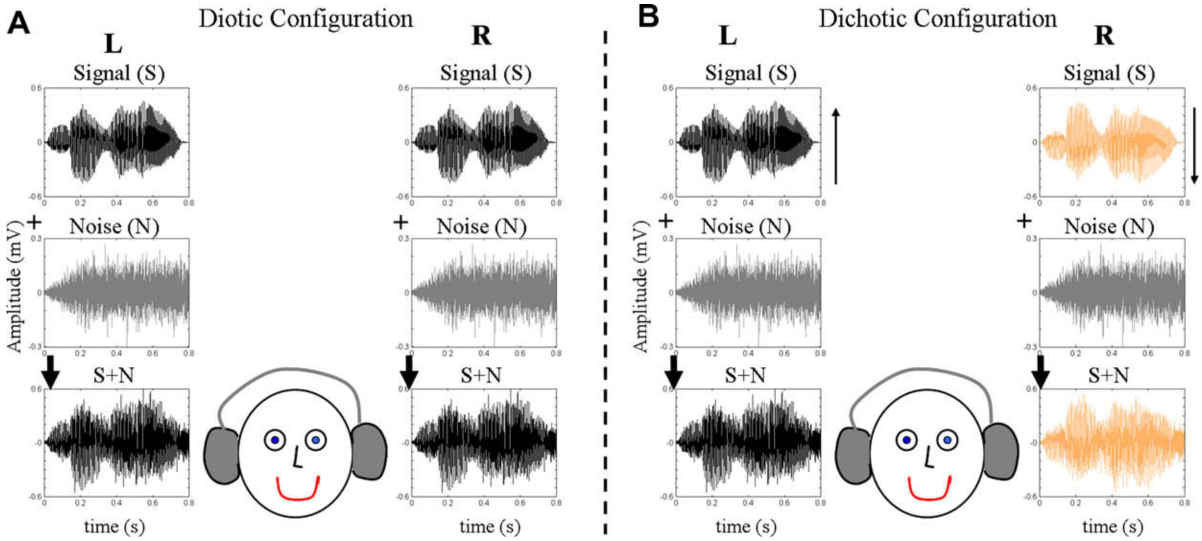

**Fig. 1.** Illustration of the stimuli used in the study. The combined stimuli used in each of the binaural configurations. Stimuli are presented as amplitude (in mV) as a function of time (in seconds). Amplitude values used are for illustration purposes, and are not the actual values used in the study. (A) Diotic configuration. In this configuration, both ears (L – left ear; R – right ear) receive exactly the same signal (S, top) and exactly the same speech noise (N, below). As a result, the signal masked by noise (S + N) is exactly the same to both ears and inter-aural comparison is not useful for segregating signal from noise. (B) Dichotic configuration. In this configuration, the left ear (L) is given S + N, similarly to those in the diotic configuration. However, in the right ear (R) the signal is phase-inverted (inversion on the *y*-axis; compare the L and R signals; arrows point direction of inversion), while the noise is unchanged. In the resulting binaural stimulus the noise is equal in the two ears whereas the signal is opposite at the each point in time, and inter-aural comparison is useful for retrieving this signal from noise. Note that the overall energy contour (stimulus envelope in the left and right dichotic and diotic signals) is equal in these two configurations, and hence their speech content sounds the same.

cues even under the mixed protocol. These findings are in line with the observations in the visual modality described above.

## 2. Materials and methods

### 2.1. Participants

In Experiment I we tested a total of 90 participants (mean age: 24 ± 3 y), 30 in each protocol group (*consistent*, 1–1 and *mixed*). In Experiment II, 42 participants (a subset of Experiment I participants), 14 in each protocol group (*consistent,* 1–1 and *mixed*) were trained. For the control groups, 30 additional participants (mean age: 25 ± 4 y) were used as the naïve control group for the /dilen/ – /tilen/ contrast (10 in each protocol). Ten of these participants (that performed the task in the *consistent* protocol, see below) were also trained on the /dilen/ – /tilen/ discrimination.

All participants were undergraduate students of the Hebrew University of Jerusalem, and were native Hebrew speakers with normal hearing. All participants gave their informed consent for participation in the study.

### 2.2. Stimuli

The stimuli were disyllabic pseudo-word pairs in Hebrew, recorded by the same female speaker. Each word had two different instances. Overall root mean square (RMS) and duration were equated for all words. In Experiment I and most parts of Experiment II, the phonologically-similar pair /barul/ and /parul/ was used. In Experiment II we also used the pair /dilen/ and /tilen/ as control.

The masking noise in all studies was speech noise (Dreschler, Verschuure, Ludvigsen, & Westermann, 2001), played at a constant level of 66 dB SPL (sound pressure level) to both ears. The noise was always identical in both ears. Pseudo-words were played in two different configurations: diotic (also termed $N_0S_0$), in which the word was added to the noise in-phase at both ears (therefore containing no binaural information), and dichotic (termed $N_0S_\pi$), in which the word was inverted in one of the ears before it was added to the noise (Fig. 1). The duration of the noise was 1.4 s, whereas the duration of the word was 0.8 s. Thus, the noise began 0.3 s before and ended 0.3 s after the word. All stimuli were digitally played by a TDT system III signal generator (Tucker Davis Technologies), and presented to listeners through HD-256 Sennheiser headphones.

### 2.3. Protocols

#### 2.3.1. Protocol for measuring thresholds

Thresholds for correct identification were measured in all studies using a three down – one up adaptive staircase procedure, converging at 79.4% correct (Levitt, 1971). The level of the masking noise was kept constant while the presentation level of the word was adaptively varied. The step sizes used were: 2, 1, 0.5, 0.3 and 0.1 dB. Each measurement block was composed of 75 trials for each binaural configuration (see below). Diotic and dichotic thresholds were calculated as the arithmetic mean of signal amplitude in the last five reversals. The binaural benefit was calculated as the difference (in dB) between the measured diotic and dichotic thresholds.

#### 2.3.2. Binaural protocols

Three different protocols of presentation of the binaural configurations (diotic and dichotic) were used in the experiments (illustrated in Fig. 2).

○ *Consistent protocol:* diotic and dichotic configurations were measured in different experimental blocks of 75 trials each, administered in immediate succession. The order of the sessions was counterbalanced between-participants. In this protocol the binaural configuration was consistent throughout the block, i.e. it was either diotic or dichotic (Fig. 2A).

○ *1–1 protocol:* the diotic configuration was presented on odd trials, and the dichotic configuration was presented on even trials. Each block consisted of 150 trials, 75 of each configuration.

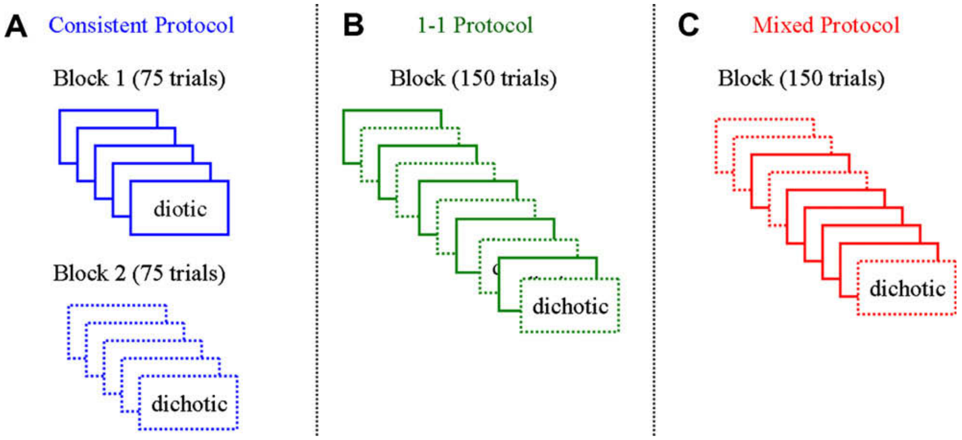

**Fig. 2.** The binaural protocols used in the study. Both experiments tested performance on three different protocols, which varied in the order of diotic and dichotic trials throughout the block. (A) *Consistent* protocol: each block is composed of 75 trials of either diotic or dichotic configuration. Separate staircases were derived for each block. (B) 1–1 protocol: each block is composed of 150 trials, 75 diotic trials and 75 dichotic trials, consistently ordered as diotic–dichotic–diotic–dichotic etc. Separate staircase was calculated for each configuration within the block. (C) *Mixed* protocol: each block is consisted of 150 trials, 75 diotic trials and 75 dichotic trials, randomly interleaved throughout the block. Separate staircase was calculated for each configuration within the block.

Adaptive thresholds for diotic and dichotic configurations were tracked separately throughout the assessment (Fig. 2B).

○ *Mixed protocol:* diotic and dichotic configurations were randomly interleaved across the block, such that on each trial one of these binaural configurations was chosen uniformly in a pseudo-random manner. Each experimental block therefore contained both diotic and dichotic configurations and consisted of 150 trials, 75 for each configuration. Adaptive thresholds for diotic and dichotic configurations were tracked separately throughout the assessment (Fig. 2C).

### 2.4. Experimental procedure

The two experiments were performed in a sound-attenuated booth. Participants' task was a two-word identification task. On each trial, one of two possible words (/barul/ and /parul/ for most phases; /dilen/ and /tilen/ for part of the testing phase of Experiment II) was presented over headphones, masked by noise, and participants were asked to press the left/right patch on the computer screen whose label matched the played word. Visual feedback was given after every button press: a positive feedback for correct responses (happy face) and a negative feedback for incorrect responses (sad face). Participants were instructed to respond as accurately and as quickly as possible.

In Experiment I, three groups of participants performed one block of the task, each using a single binaural protocol: *consistent*, *mixed* or 1–1.

In Experiment II, 14 of the 30 participants in each group continued practicing on the task, using the same initial binaural protocol, for a total of seven practice days. Training was administered with 1–3 day interval between subsequent sessions. Ten of the 14 participants in each group participated in three additional testing sessions (sessions 8–10). On the 8th session, each group performed the task with a different binaural protocol: the *consistent* and 1–1 groups were tested with the *mixed*, whereas the *mixed* group was tested with the 1–1 protocol. On day nine participants were retested with the original training protocol (baseline). On day 10, participants were tested on a new phonologically-similar pair of pseudo-words: /dilen/ and /tilen/ under the originally trained binaural protocol. In each of training and testing sessions, each participant performed three blocks of 150 trials per day (or six blocks of 75 trials each in case of the consistent protocol). Each session took ~25 min.

As a control, a different group of participants (30 participants, 10 per protocol) performed separately the task of discrimination between /dilen/ and /tilen/. The 10 participants that performed the task in the *consistent* protocol underwent additional training of six more days on discrimination between /dilen/ and /tilen/, similarly to the /barul/ – /parul/ training. On day 8, participants were tested on discrimination between /barul/ and /parul/. These data were used for comparison with the /dilen/ – /tilen/ data collected on day 10 of testing.

### 2.5. Data analysis

In Experiment I we used multivariate analysis with dependent factors of diotic threshold and binaural benefit, and an independent variable of protocol (3 levels: consistent, 1–1 and mixed). A post hoc Scheffe contrast was performed to compare the three binaural protocols.

In Experiment II we used a separate analysis for each phase. For the training phase we used separate ANOVAs for diotic thresholds and binaural benefits, using within-subjects factor of days (7 levels) and between-subjects factor of protocol (3 levels). Post hoc Scheffe contrast was used to discriminate between the three protocols. For the transfer across protocol phase we conducted a separate multivariate analysis for each protocol (two dependent variables: diotic thresholds and binaural benefits), with an independent factor of state (3 levels for the variable protocol: naïve, after 1–1 training and after consistent training; 2 levels for the 1–1 protocol: naïve and after mixed training). For the mixed protocol, we additionally conducted a post hoc Scheffe analysis for state. In addition, we used 2-tailed Student *t*-tests to compare performance between the last day or training (day 7) and the testing day (day 8). For the transfer across pairs phase we performed a multivariate analysis (two dependent variables: diotic thresholds and binaural benefits) with between-subjects factors of protocol (3 levels) and state (2 levels: naïve and trained). We conducted a post hoc Scheffe contrast for the protocol variable. For the /dilen/ – /tilen/ learning (a separate group of participants) we performed an ANOVA with factors of aurality (2 levels: diotic thresholds and binaural benefits) and of days (7 levels, day 1 through day 7).

## 3. Results

Listeners were asked to discriminate between two phonologically-similar Hebrew pseudo-words, /barul/ and /parul/, embedded

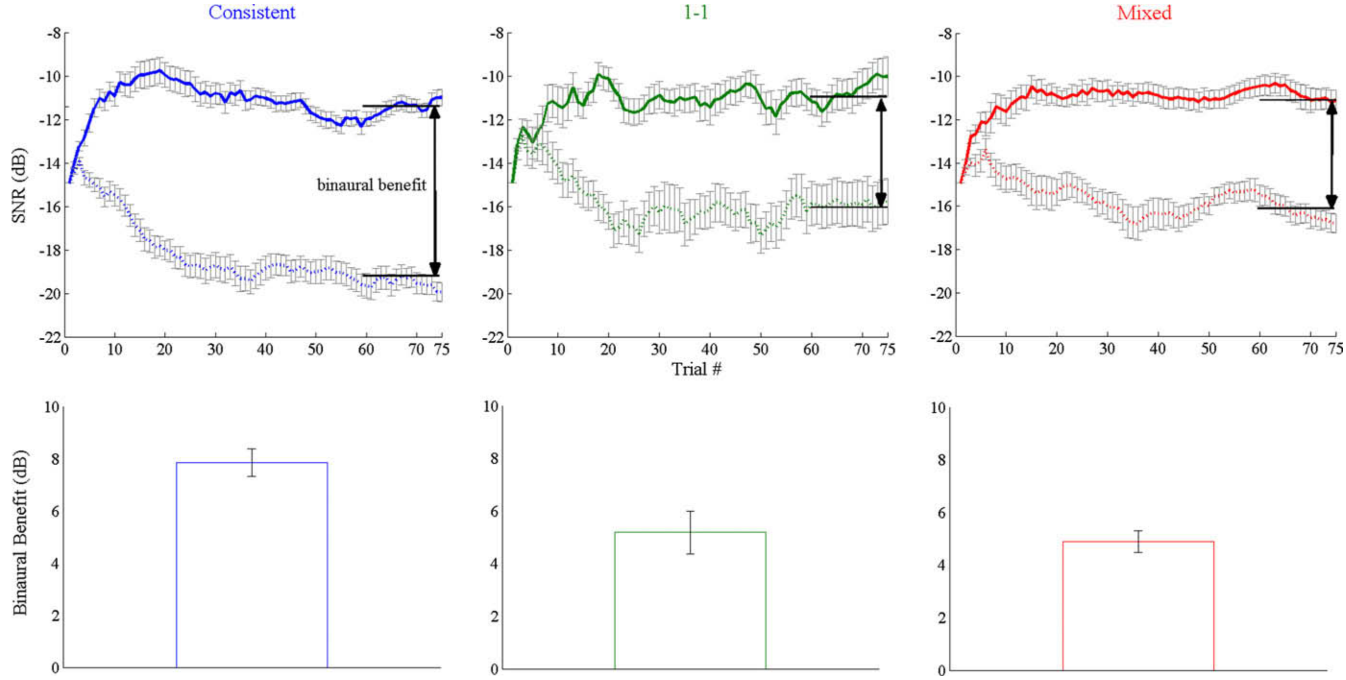

**Fig. 3.** The effect of protocol for naïve subjects. Top: the dynamics of the adaptive threshold assessment as a function of trial number under each of the three binaural protocols (averaged across participants ± SEM, $N = 30$ in each protocol group): *consistent* (blue, left), 1–1 (green, middle) and *mixed* (red, right). The level of the signal was modified adaptively, following a three-down-one-up staircase procedure. Threshold measurements are denoted in dB signal-to-noise ratio (SNR). Diotic threshold measurements are denoted by solid lines, dichotic by dashed lines. The solid black lines on each panel mark the actual averaged diotic and dichotic thresholds, calculated as the means of last five reversals (see Section 2). Double arrow in each panel shows the calculated binaural benefit for that protocol. Bottom: binaural benefits, calculated as the difference between the diotic and dichotic thresholds (averaged across participants ± SEM). Binaural benefits were significantly larger under the consistent protocol although task difficulty (diotic thresholds) did not differ between protocols.

in speech noise (Dreschler et al., 2001). Experiment I was aimed to assess the effect of protocol on the ability of naïve participants to utilize low-level binaural information for this discrimination. Experiment II tested the effects of multi-session training on the ability of participants to use these cues.

### 3.1. Experiment I: the effect of protocol on the use of low-level information

In Experiment I, three groups of participants were tested on the speech discrimination task under three protocols (see Fig. 2): *consistent*, in which diotic and dichotic configurations were measured in separate blocks, *mixed*, in which the diotic and dichotic configurations were pseudo-randomly interleaved in the block; And 1–1, in which the diotic configuration was presented on odd trials, and the dichotic configuration was used on even trials. For this group, binaural information was presented in a variable, but fully predictable manner. Each subject performed the task using a single protocol, and participants were unaware, based on their introspective evidence, of the manipulation of low-level variability. The first two protocols (*consistent* and *mixed*) have been previously tested by us (Nahum et al., submitted for publication), whereas the 1–1 protocol is novel.

Fig. 3 (top) depicts the signal-to-noise ratio (SNR, the difference between stimulus and noise levels) in dB as a function of trial number during the first assessment block of each of the three protocols. Using a three-down-one-up staircase procedure (Levitt, 1971), the initial SNR was the experimenters' pre-determined one, but subsequently varied according to listeners' performance, converging at about 80% correct. A steady state level of performance was typically reached by the 40th trial. As seen in Fig. 3 (top), diotic thresholds did not significantly differ between the three protocols ($F(2, 87) = 0.36$, $p = 0.7$, n.s.). On the other hand, dichotic thresholds (dashed lines in Fig. 3) differed. As a result, binaural benefits (the difference between diotic and dichotic thresholds; bottom panels and double arrows on top panels, Fig. 3), significantly differed between protocols ($F(2, 87) = 7.2$, $p < 0.002$; Fig. 3, bottom) being largest for the *consistent* protocol (7.8 ± 0.5 dB, blue bar[1]), and significantly smaller for the other two protocols (5.2 ± 0.8 and 4.9 ± 0.4 dB for 1–1 and *mixed* protocols, respectively). A post hoc Scheffe contrast between the three protocols showed that the *consistent* protocol is significantly different from both the 1–1 protocol ($p < 0.01$) and the *mixed* protocol ($p < 0.005$), while the two latter conditions are not significantly different from each other ($p = 0.95$, n.s.).

The results of Experiment I show that the implicit manipulation of binaural configuration did not change overall task difficulty, as measured by the diotic thresholds. Yet, it affected the degree of binaural benefits. Optimal benefits were obtained only when dichotic thresholds were measured separately (i.e. the binaural configuration within the block was fixed; see also Nahum et al., submitted for publication). When binaural configuration varied between consecutive trials (1–1 protocol), utilization was again sub-optimal, as in the *mixed* protocol. Thus, performance was hampered by cross-trial variability even when it had a consistent structure across the block.

### 3.2. Experiment II: the effect of training on the use of low-level information

In Experiment II we asked whether the use of low-level binaural cues, which was sub-optimal when performed in the first sessions of the *mixed* and 1–1 protocols, improves following multi-session training.

---

[1] For interpretation of color in Figs. 3, 4 and 6 the reader is referred to the web version of this article.

**Table 1**
Experimental procedure and stimuli, Experiment II.

| | Training | Testing | | |
|---|---|---|---|---|
| | | Another Protocol | Baseline | Another Pair |
| Day | 1–7 | 8 | 9 | 10 |
| *N* | 14 | 10 | | |
| Stimuli | /barul/ vs. /parul/ | | | /dilen/ vs. /tilen/ |
| | Protocol | | | |
| *Consistent Group* | Consistent | Mixed | | Consistent |
| 1–1 *Group* | 1–1 | Mixed | | 1–1 |
| Mixed *Group* | Mixed | 1–1 | | Mixed |

Fourteen participants of each assessment group continued to train on the same speech discrimination task (between /barul/ and /parul/) using the same protocol they were initially assigned to. The three groups (‘*consistent*’, ‘1–1’, and ‘*mixed*’) trained for six additional sessions. Following training, 10 participants in each group participated in three more sessions: a test session using a different binaural protocol (day 8), another baseline session (day 9), and an additional session with a different phonological contrast (day 10). The experimental procedure and stimuli used are detailed in Table 1.

### 3.2.1. Improvement in utilization of low-level binaural information following training

Fig. 4 (upper panels) depicts the average diotic and dichotic thresholds on each training day. Diotic thresholds of all three groups improved to a similar extent (~4–5 dB; effect of days: $F(6, 216) = 49.6$, $p < 0.00001$; no significant effect of protocol: $F(2, 36) = 1.2$, $p = 0.3$, no significant interaction: $F(12, 216) = 1.7$, $p = 0.09$).

Binaural benefits (Fig. 4, lower panels) significantly changed during learning as well (effect of days: $F(6, 216) = 6.9$, $p < 0.0001$), but this change mainly resulted from their improvement in the 1–1 group (Fig. 4B; effect of protocol: $F(2, 36) = 7.8$, $p < 0.005$), which increased from 5.4 ± 0.5 dB to 8.4 ± 0.6 dB from 1st to 7th day of training. In the *mixed* group (Fig. 4C) there was a small gradual increase from 5 ± 0.5 on the 1st day to 6.1 ± 0.6 dB on the 7th day. No increase was found for the *consistent* group (Fig. 4A), whose binaural benefits were large and remained similar throughout the training days (7.2 ± 0.3 and 7.6 ± 0.5 dB for 1st and 7th days, respectively). The major increase in binaural benefits for the 1–1 group occurred between the 1st and 2nd days, although some increase continued until the 5th day of training.

By the end of the training, diotic and dichotic thresholds (and hence binaural benefits) were similar in the *consistent* and 1–1 groups. However, in the *mixed* group binaural benefits were significantly smaller than in the *consistent* and 1–1 groups (post hoc Scheffe contrast: $p < 0.02$, $p < 0.005$, respectively; the latter two did not differ, $p = 0.93$, n.s.).

### 3.2.2. Transfer across protocols

Following 7 days of practice, 10 of the 14 participants in each group performed the same discrimination with a different binaural protocol than the one they were trained with. Participants in the *consistent* and 1–1 groups performed the *mixed* protocol, while participants in the *mixed* group performed the 1–1 protocol. In order to assess generalization across protocols, performance was compared to that in the first and last blocks for participants who were trained with the tested protocol.

Learning under the *consistent* and the 1–1 protocols fully transferred to the *mixed* protocol, resulting in improvements both in diotic thresholds (from −11.2 ± 0.5 for performance in first block on the mixed protocol to −15 ± 0.6 and −14 ± 0.7 dB SNR following *consistent* and 1–1 training, respectively; $F(3, 56) = 8$, $p < 0.001$; red bars in Fig. 4, top) and in binaural benefits (from 4.9 ± 0.4 dB in 1st block on the mixed protocol to 7.4 ± 0.5 and 7.9 ± 0.4 dB, respectively; $F(3, 56) = 7.6$; $p < 0.001$; red bars in left and middle panels

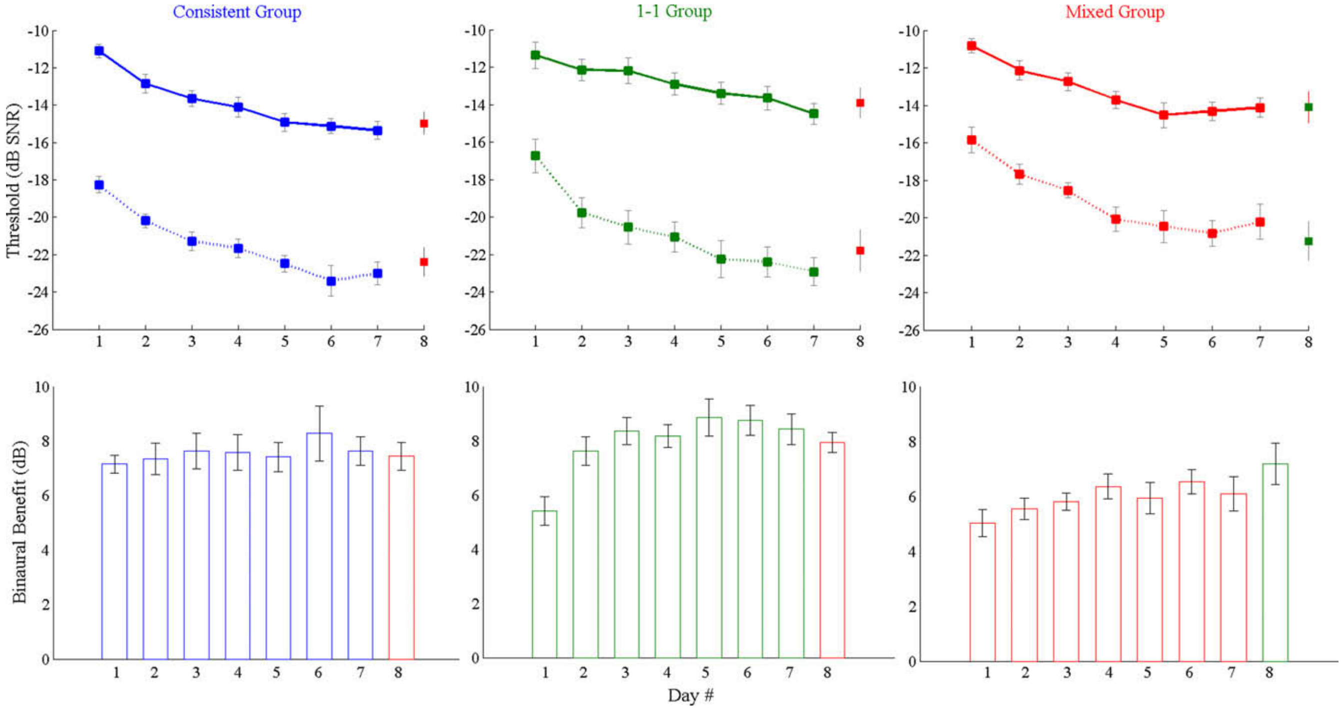

**Fig. 4.** Learning and transfer across protocols. Thresholds (top row) and binaural benefits (bottom row) obtained during training (days 1–7) and protocol testing (day 8) in each of the three training groups: *consistent* (left), 1–1 (middle) and *mixed* (right). Top: diotic (solid line) and dichotic (dashed line) thresholds. For each subject, the threshold in each day is calculated as the average of thresholds obtained in the three blocks performed that day. Thresholds are denoted in dB signal-to-noise ratio (SNR). Bottom: binaural benefits, calculated as the difference between the diotic and dichotic thresholds above. $N = 14$ for each group for days 1–7; $N = 10$ for each group for day 8.

of Fig. 4, bottom). Moreover, diotic thresholds and binaural benefits were similar to those obtained on the last (7th) day of practice on each protocol ($t$-tests: diotic thresholds: $p = 0.97$ and $p = 0.77$ for *consistent* and 1–1 protocols, respectively; binaural benefits: $p = 0.77$ and $p = 0.43$ for *consistent* and 1–1 protocols, respectively).

Learning with the *mixed* protocol showed transfer of diotic thresholds to the 1–1 protocol (from −10.8 + 0.6 for the 1st block of 1–1 to −14.1 ± 0.8 dB SNR following *mixed* training; $F(2, 46) = 8.6$, $p < 0.001$; green bars in left panel of Fig. 4), and a tendency, which approached statistical significance, for improved binaural benefits (7.2 ± 0.8 dB following training ($F(2, 46) = 2.9$, $p = 0.06$, n.s.). Although the improvement was not significant, following training with the *mixed* protocol, binaural benefits on the 1–1 protocol did not significantly differ from those obtained following 1–1 training (Post hoc Scheffe contrast: $p = 0.76$, n.s.), as can be seen by comparing the right and middle plots of Fig. 4.

### 3.2.3. Specificity to the trained phonological contrast

To assess transfer across phonological contrasts we tested the trained participants ($N = 10$ in each group) with a novel pair of phonologically-similar pseudo-words containing a different phonological contrast (/dilen/ – /tilen/). Each group was tested with the protocol they were trained with. Their performance was compared to that of naïve (untrained) participants performing the same speech discrimination task with the same protocol (see Section 2).

As with the originally trained (/barul/ – /parul/) word pair, diotic thresholds of naïve participants did not differ between the three protocols (top plots in Fig. 5; effect of protocol: $F(2, 56) = 0.36$, $p = 0.7$, n.s.), whereas binaural benefits significantly differed ($F(2, 56) = 5.05$, $p < 0.01$; Fig. 5, bottom). They were larger for the *consistent* protocol compared with the 1–1 (post hoc Scheffe contrast: $p < 0.03$) and *mixed* ($p < 0.03$) protocols. Thus, the effect of low-level consistency was similar in this phonological contrast.

Following training on the original pair (/barul/ – /parul/) there was only a minor improvement in the diotic thresholds of the untrained pair (/dilen/ – /tilen/; −14.4 ± 0.5 and −15.3 ± 0.5 dB SNR for naïve (untrained) and trained, respectively; effect of state: $F(1, 56) = 4.3$, $p < 0.05$). This minor improvement mainly stemmed from the mildly reduced thresholds for the *consistent* protocol. Binaural benefits when discriminating the untrained pair did not increase ($F(1, 56) = 0.02$, $p = 0.9$, n.s.; Fig. 5). Thus, the use of binaural information in the untrained pair did not improve following training on a different pair. This was the case for all protocols. To a large extent, learning is therefore specific to the phonetic contrast of the trained pair. In particular there was no transfer in the use of low-level binaural cues.

Since diotic thresholds for naïve performance are about 3 dB lower for the /dilen/ – /tilen/ discrimination than for the /barul/ – /parul/ discrimination (see Fig. 6), it was not clear whether the low thresholds for /dilen/ – /tilen/ following /barul/ – /parul/ training are the result of learning or not. Moreover, it was not clear whether improvement in these thresholds, which were initially lower, could even occur following training. In order to verify the specificity of learning to the trained pair, we performed a complementary control study, in which the reversed pattern was tested: participants were trained on /dilen/ – /tilen/ discrimination using

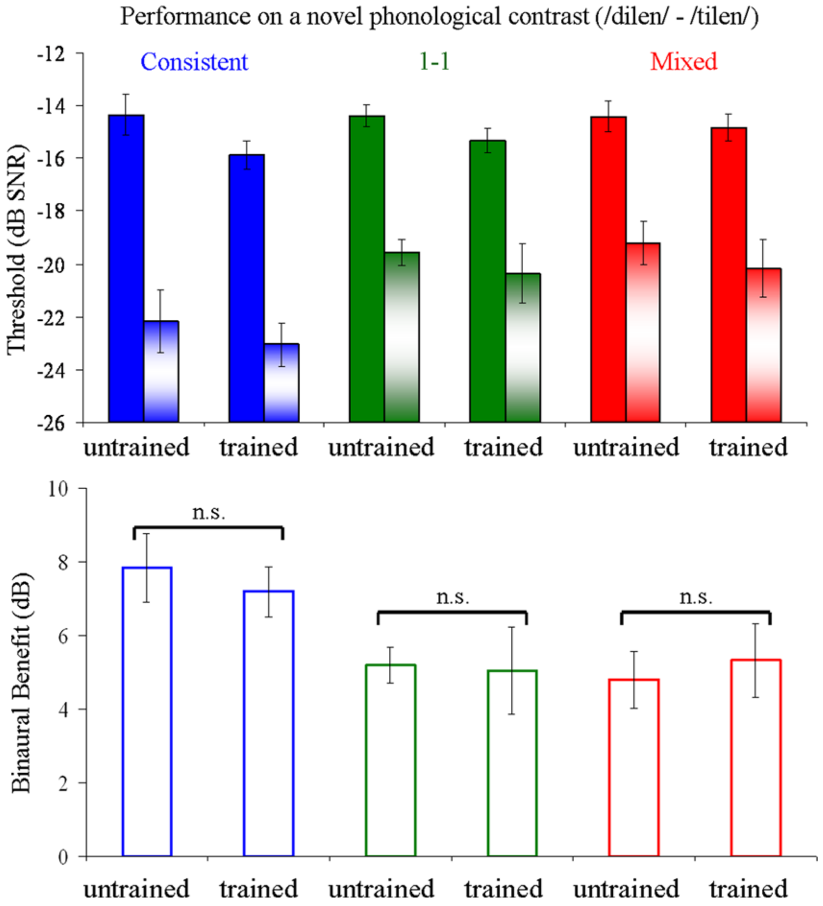

**Fig. 5.** Generalization of learning to a novel phonological contrast. Thresholds (average ± SEM; top; filled bars: diotic thresholds, shaded bars: dichotic thresholds) and binaural benefits (bottom) for discrimination between an untrained word pair (/dilen/ vs. /tilen/). Performance is shown for untrained (naïve, $N = 10$ per protocol) and trained ($N = 10$ per protocol) participants for all three protocols. Trained participants were trained on a different discrimination (/barul/ – /parul/), and were then tested on the /dilen/ – /tilen/ discrimination (see Section 2). Note the difference in thresholds and binaural benefits between consistent and the other two protocols for untrained participants, similarly to what is found in the originally trained pair (Fig. 1).

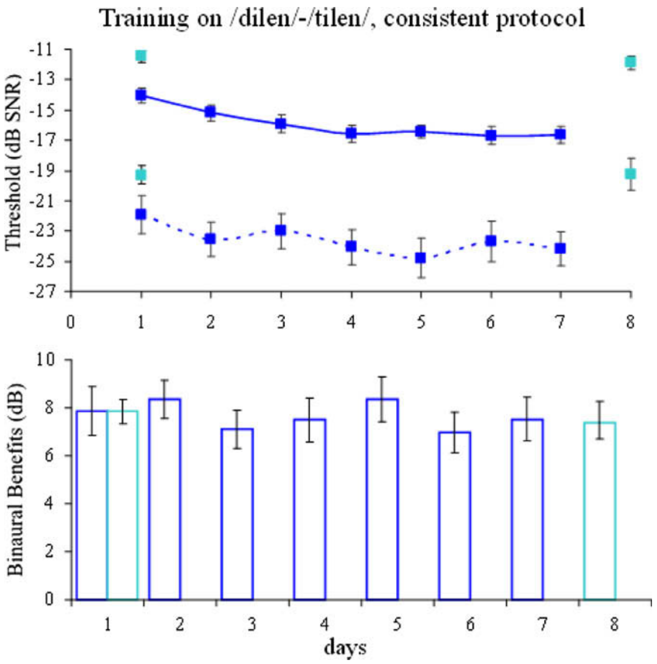

**Fig. 6.** Specificity of training on a different phonological contrast. Ten participants were trained on discrimination between /dilen/ – /tilen/ for 7 days, under the consistent protocol (dark blue). On the 8th day, performance on /barul/ – /parul/ was tested (light blue). Top: diotic (solid line) and dichotic (dashed line) thresholds (average ± SEM; $N = 10$) obtained during training. Bottom: corresponding binaural benefits. Results of the /barul/ – /parul/ discrimination are presented in light blue, for both naïve participants (day 1, taken from Experiment I, here re-plotted in day 1) and following /dilen/ – /tilen/ training (day 8).

the *consistent* protocol, and were then tested on the discrimination between /barul/ and /parul/.

Diotic thresholds for discrimination between the trained words (/dilen/ and /tilen/) improved (from $-14 \pm 0.5$ to $-16.6 \pm 0.6$ dB SNR; $F(6, 48) = 6.4$, $p < 0.005$), whereas binaural benefits remained large and constant (~8 dB) throughout training, as expected for the *consistent* protocol ($F(6, 48) = 1.2$, $p = 0.3$, n.s.; Fig. 6). However, there was no significant generalization to the untrained pair (/barul/ – /parul/). Namely, naïve and post-training thresholds (and binaural benefits) did not differ (compare light blue bars in day 1 and day 8 in Fig. 6; *t*-test: $p = 0.8$ and $p = 0.4$, n.s, for diotic thresholds and binaural benefits, respectively). These results further support the conclusion that learning is specific to the trained phonological contrast.

### 3.3. Summary of results

We tested the effect of variability and uncertainty of a low-level cue (binaural phase differences) on the ability to use it for improved discrimination between phonologically-similar words in noise, initially and following training. Three groups of listeners were tested and trained on the task using three binaural protocols: *consistent* binaural information (the same binaural configuration was presented throughout the block), *mixed* binaural information (randomly interleaved binaural configurations) and consistently interleaved (temporally-ordered binaural configurations in a 1–1 sequence).

We found that in the diotic configuration, where access to low-level binaural information does not improve performance, both initial and trained performance did not differ between protocols. Dichotic thresholds, on the other hand, were sensitive to the binaural protocol. For the *consistent* protocol they were low already in the initial training session, indicating optimal usage of binaural information right from the beginning. For the *mixed* protocol they were initially higher and improved to a similar extent as the diotic thresholds, without increased binaural benefits. For the 1–1 protocol dichotic thresholds in the first session were as high as in the *mixed* protocol. However, by the second session they were already significantly lower and were similar to those of the *consistent* protocol. The binaural benefits measured for the *consistent* and 1–1 protocols were fully transferred to the *mixed* protocol. In all three protocols improvement in both diotic and dichotic configurations was largely specific to the trained phonological contrast.

## 4. Discussion

### 4.1. The impact of low-level uncertainty on initial performance

According to RHT, naïve (untrained) performance is based on higher levels of representations. Access to specific low-level populations requires a backward search, and can therefore be achieved only when such backward search can be successfully implemented (Ahissar & Hochstein, 1997). Its implementation requires top-down guidance and will be disabled when the informative population changes across trials (Ahissar et al., 2009). This is the case when low-level uncertainty is introduced. It therefore follows that the impact of low-level uncertainty (the impact of protocol in our study) on performance depends on whether access to low-level populations is expected to result in performance gains (see Ahissar et al., 2009 for a review).

Indeed, in our study, initial diotic thresholds, which presumably do not benefit from access to low-level binaural cues (see Nahum et al., submitted for publication), were not affected by the protocol manipulation, whereas initial dichotic thresholds were. We should note that our choice of phonological discrimination was intentionally aimed to be based on temporally-localized cues (/barul/ vs. /parul/, or /dilen/ vs. /tilen/), which we assumed would benefit from access to low-level information in the dichotic configuration. RHT further proposes that, when phonetic differences are spread throughout the word (as is the case when discriminating between two very different words, e.g. /dilen/ vs. /barul/), dichotic performance will not be sensitive to low-level uncertainty, since in that case the auditory hierarchy did not yield any convergence and high-level representations contain all the information (for a detailed discussion see Nahum et al., submitted for publication).

Indeed, the impact of cross-trial uncertainty on initial performance varied between studies, both in the visual and in the auditory modalities. By design, the current study illustrates both types of results within the same task. In the auditory modality, performance in a two-tone frequency discrimination task is extremely sensitive to the protocol (Nahum et al., submitted for publication; Amitay et al., 2005). The sensitivity of visual contrast discrimination to protocol varied between studies (Adini et al., 2004; Kuai et al., 2005; Yu et al., 2004), whereas a study which assessed performance on a line bisection task reported no sensitivity (Parkosadze et al., 2008). Detailed analysis of all these results in the context of the principles of RHT and the impact of the combination of conditions under which they were administered is beyond the scope of the current study.

In general, however, recent results in the visual modality are in line with RHT's prediction (Ahissar et al., 2009; Nahum et al., submitted for publication) according to which naïve performance and learning depend on the level which is relevant for task performance. For example, recent findings from perceptual learning of orientation discrimination (Jeter, Dosher, Petrov, & Lu, 2009) suggest that when high precision is required (analogous to "temporally-local" in the auditory domain), and access to low-level information may be beneficial, learning involves low-level representations. However, when coarse evaluation is required in the same task (analogous to "global" contrasts), performance involves only higher levels regardless of the degree of task difficulty. These findings are therefore in line with RHT's prediction that performance and learning in conditions that require high precision will be sensitive to the behavioral protocol (e.g. local cross-trial uncertainty), whereas performance of coarse evaluations will not be sensitive to behavioral protocol.

### 4.2. The impact of low-level uncertainty on learning

Several previous studies tested the impact of low-level variability on learning. Adini et al. (2004) and Yu et al. (2004) assessed the impact of reference variability when observers were asked to compare contrasts between a reference and a test stimulus. In the auditory modality, Amitay and colleges (2005) roved the reference tone frequency in a frequency discrimination task. In both modalities, roving the reference decreased the amount of improvement obtained with practice. In both modalities (Amitay et al., 2005; Zhang et al., 2008) the impact of this uncertainty within the assessment protocol was scaled with listeners' performance. Thus, Amitay et al. (2005) reported that for good listeners, a small degree of roving interfered with learning, but a larger degree of roving did not, whereas for poor listeners even a large degree of roving interfered with learning. These observations are similar to those reported in the visual modality (Zhang et al., 2008): when the cross-trial stimulus variability is sufficiently large to be detected as clearly different stimuli by high-level representations, learning takes place. A consistent finding was reported by Karmarkar and Buonomano (2003): the authors trained listeners on interval discrimination using very different two base intervals presented in a randomly interleaved manner. The two intervals were further marked by two different frequencies, so that there was no high-level confusion, and indeed, learning took place.

In the current study we varied a low-level cue that was not directly relevant to the explicit task. Yet, this variation affected the consistency of the informative low-level population (Nahum et al., submitted for publication). We found a similar effect to that reported by Zhang et al. (2008) and Amitay et al. (2005). Random variability made high-level learning possible, as indicated by the improvement in the absolute thresholds for speech discrimination obtained under both binaural configurations in the *mixed* protocol. However, there was no improved access to low-level populations, as indicated by the fact that the improvement in the two binaural configurations, diotic and dichotic, was similar. Thus, the use of the inter-aural phase differences, measured by the threshold difference (the binaural benefit), did not improve, and binaural benefits did not improve to the level obtained using the *consistent* protocol.

Interesting phenomenology was described by Parkosadze et al. (2008), who tested learning of bisection stimuli under roving. The authors found that when subjects are given massive training (almost 20,000 trials), they can gradually improve even when two different distances are presented in a mixed manner. However, learning only began following almost 4000 trials (which is more than twice the length of our entire training). Whether improvement began when top-down mechanisms detected the two reference distances as distinct or slow bottom-up mechanisms were involved is hard to deduce from that study.

In the *consistent* protocol, which contained no low-level uncertainty, maximal use of binaural information (i.e. ideal listener level of performance, Nahum et al., submitted for publication) was obtained quickly, already by the first session. According to the RHT interpretation, this cross-trial consistency enables a quick backward search, and hence efficient utilization of binaural cues.

Finally, all three protocols resulted in learning that was specific to the trained contrast, in line with previous studies which trained with a very limited stimulus range (e.g. Ahissar & Hochstein, 1997; Fiorentini & Berardi, 1980; Karni & Sagi, 1991; Saffell & Matthews, 2003; Schoups et al., 2001; Shiu & Pashler, 1992).

### 4.3. The impact of low-level uncertainty on performance following training

Once listeners reached an asymptotic level of performance, cross-trial uncertainty was no longer important. Following training with the *consistent* or 1–1 protocols, we found complete transfer to the randomly interleaved condition, despite the fact that performance following the *mixed* (randomly interleaved) protocol itself did not show a corresponding benefit.

Here too, our findings are in line with recent results obtained in the visual domain (Adini et al., 2004; Kuai et al., 2005; Zhang et al., 2008). For example, Zhang et al. (2008) demonstrated that following perceptual learning on a contrast discrimination task using temporal structures, performance was undisrupted by a roving session (if administered at least 4 h after training), nor by extended roving training. Thus, performance transferred to the roved condition. Once perceptual learning has been completed and the stimulus information consolidated, performance could no longer be degraded by extended roved training.

The "immunity" to cross-trial uncertainty obtained through learning, can be interpreted by RHT in more than one way. According to one interpretation, once the backward search had been trained and "marked", it is automatically retrieved with stimulus presentation, when the trained context is re-presented. Alternatively, the backward search strengthens bottom-up inputs that were determined to be informative, and hence, following training, it is no longer necessary. The latter interpretation is in line with concepts suggested by Dosher and Lu (1999) who proposed that learning results from selective strengthening of informative inputs. The selection may be difficult to implement when the relevant input changes across trials. However, once these inputs were selected and their weights were increased, their benefit is expected to be apparent regardless of cross-trial uncertainty. These two interpretations yield different behavioral predictions. The first suggests that the sensitivity to the broad stimulation context will remain following training, whereas the latter suggests that this sensitivity will be eliminated with training. These predictions should be tested in further investigations.

Interestingly, training with the *mixed* protocol showed some transfer to the 1–1 protocol. Thus, diotic thresholds improved on the 1–1 protocol following *mixed* training, consistent with the high-level nature of learning in the case of the *mixed* protocol.

### 4.4. Temporal patterns and learning

In two recent studies, Kuai et al. (2005) and Zhang et al. (2008) have tested the benefit of using fixed temporal patterns when presenting several contrast references. They found that this consistency enabled learning. The learning process was not quick, i.e. it was not obtained within a single session, suggesting that detection of the patterns is a challenging process, requiring several training sessions. Our results are in line with these findings. In the first block of assessment, performance on the 1–1 protocol was similar to that of the *mixed* (randomly interleaved) protocol, showing only partial use of binaural information. However, already by the 2nd session, dichotic performance improved and became similar to that of the *consistent* protocol.

The RHT account asserts that access to low-level information requires a successful, top-down guided backward search (Ahissar & Hochstein, 2004). The finding that patterning of the implicit (or explicit in previous studies) configuration of stimulus presentation enables learning, is interpreted as an indication that such patterning increases the salience of the different stimuli to higher level representations. The repeated temporal pattern may be detected gradually if the stimuli composing this pattern are not sufficiently distinct to be identified as different within very few trials, but are sufficiently distinct so that some increase in saliency uniquely identifies them when they form some global cross-trial structure. Consequently, following some practice, the lower-level populations that represent different stimuli in the pattern are accessed through separate higher-level representations. A further RHT prediction is therefore that when the interleaved stimuli are too similar, patterning will not suffice to induce learning.

Along this line of interpretation, we propose that the two binaural configurations, diotic and dichotic, are not initially perceived as different and hence cannot be learned separately under the *mixed* protocol (unlike learning with two very distinct stimuli, as was found in Zhang et al., 2008, and Parkosadze et al., 2008). However, since the two binaural configurations were associated with a number of differences in the perceived qualities of the sounds, which were amplified by the consistent temporal pattern (1–1 protocol) they were eventually implicitly categorized as different stimuli by higher levels, making it possible to conduct separate search processes for each configuration. For example, the perceptual differences between the two configurations could be due to the different thresholds (with the dichotic configuration having a significantly lower threshold), so that the 1–1 protocol yielded a higher-lower intensity pattern; alternatively, it could be related to the larger size of the sound image in the dichotic configuration. This presumed separation occurred rather quickly, and by the second day of training, dichotic performance in the 1–1 protocol was similar to that of the *consistent* protocol (see Fig. 2). The observation that most dichotic learning occurred already during training is consistent with several previous reports in the perceptual learning literature (e.g. Ahissar & Hochstein, 1997; Karni & Sagi, 1993; Zhang et al., 2008).

This interpretation is consistent with the *stimulus tagging concept* proposed by Zhang et al. (2008). The authors proposed that learning in a mixed stimuli protocol requires conceptually different tagging (similar to the "high-level" identification in RHT) of each stimulus. In line with RHT, this concept emphasizes top-down influence in perceptual learning. It further suggests that conceptual or semantic tagging could be useful. Stimulus temporal patterning with rhythm may enable learning by providing a unique tag for each roving stimulus, which enables the backward search required by RHT. The stimulus tagging concept therefore suggests that the 1–1 protocol employed here enabled the correct tagging of each configuration, which initiated the top-down search. However, a more direct test of the 'tagging' process could be by labeling each configuration, diotic and dichotic, and checking whether such labeling would enable a gradual increase in binaural benefits in the *mixed* protocol.

In summary, our findings suggest that similar principles underlie the dynamics and protocol sensitivity of auditory and visual perceptual learning. In both modalities, the impact of protocol can be largely predicted by basic principles denoted by the Reverse Hierarchy Theory.

## References

Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M., & Sagi, D. (2004). Perceptual learning in contrast discrimination: The effect of contrast uncertainty. *Journal of Vision, 4*, 993–1005.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*, 401–406.

Ahissar, M., & Hochstein, S. (2000). The spread of attention and learning in feature search: effects of target distribution and task difficulty. *Vision Research, 40*(10–12), 1349–1364.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*, 457–464.

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 364*(1515), 285–299.

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception and Psychophysics, 67*, 691–698.

Amitay, S., Irwin, A., & Moore, D. R. (2006). Discrimination learning induced by training with identical stimuli. *Nature Neuroscience, 9*, 1446–1448.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*(6), 953–965.

Batra, R., Kuwada, S., & Fitzpatrick, D. C. (1997a). Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses. *Journal of Neurophysiology, 78*, 1222–1236.

Batra, R., Kuwada, S., & Fitzpatrick, D. C. (1997b). Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. II. Coincidence detection. *Journal of Neurophysiology, 78*, 1237–1247.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. Cambridge, Massachusetts: MIT Press.

Blauert, J., & Cobben, W. (1978). Some consideration of binaural cross correlation analysis. *Acustica, 39*, 96–103.

Delhommeau, K., Micheyl, C., Jouvent, R., & Collet, L. (2002). Transfer of learning across durations and ears in auditory frequency discrimination. *Perception and Psychophysics, 64*, 426–436.

Demany, L. (1985). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America, 78*, 1118–1120.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology, 40*, 148–157.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*(19), 3197–3221.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*, 154–160.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research, 35*, 3003–3013.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology, 6*(3), 292–297.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287*, 43–44.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology, 14*, 573–578.

Hirsch, I. (1948). The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America, 20*, 536–544.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791–804.

Irvine, D. R. F., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *Journal of the Acoustical Society of America, 108*(6), 2964–2968.

Jeter, P. E., Dosher, B. A., Petrov, A., & Lu, Z. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision, 9*(3), 1–13.

Johansson, M. S., & Arlinger, S. D. (2002). Binaural masking level difference for speech signals in noise. *International Journal of Audiology, 41*, 279–284.

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning and Memory, 10*, 141–147.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88*(11), 4966–4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250–252.

Kuai, S. G., Zhang, J. Y., Klein, S. A., Levi, D. M., & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience, 8*, 1497–1499.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477.

Levitt, H., & Rabiner, L. R. (1967). Binaural release from masking for speech and gain in intelligibility. *Journal of the Acoustical Society of America, 42*, 601–608.

Licklider, J. (1948). The influence of interaural phase relation upon the masking of speech by white noise. *Journal of the Acoustical Society of America, 20*, 150–159.

Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. *Proceedings of the National Academy of Sciences of the United States of America, 96*(24), 14085–14087.

Nahum, M., Daikhin, L., Lubin, Y., Cohen, Y., & Ahissar, M. (submitted for publication). From comparison to classification-a cortical tool for boosting perception.

Nelken, I., & Ahissar, M. (2006). High-level and low-level processing in the auditory system: The role of primary auditory cortex. In P. Divenyi, S. Greenberg, & G. Meyer (Eds.), *Dynamics of speech production and perception* (pp. 343–354). Amsterdam: IOS Press.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60*, 355–376.

Parkosadze, K., Otto, T. U., Malania, M., Kezeli, A., & Herzog, M. H. (2008). Perceptual learning of bisection stimuli under roving: Slow and largely specific. *Journal of Vision, 8*(5), 1–8.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science, 256*, 1018–1021.

Saffell, T., & Matthews, N. (2003). Task specific perceptual learning on speed and direction discrimination. *Vision Research, 43*, 1365–1374.

Schoups, A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *Journal of Physiology, 483*(3), 797–810.

Schoups, A. A., & Orban, G. A. (1996). Interocular transfer in perceptual learning of a pop-out discrimination task. *Proceedings of the National Academy of Sciences of the United States of America, 93*(14), 7358–7362.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature, 412*(6846), 549–553.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303–304.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics, 52*, 582–588.

van Wassenhove, V., & Nagarajan, S. S. (2007). Auditory cortical plasticity in learning to discriminate modulation rate. *Journal of Neuroscience, 27*(10), 2663–2672.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience, 17*(10), 3956–3963.

Yin, T. C., & Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology, 64*, 465–488.

Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision, 4*, 169–182.

Zhang, J. Y., Kuai, S. G., Xiao, L. Q., Klein, S. A., Levi, D. M., & Yu, C. (2008). Stimulus coding rules for perceptual learning. *PLoS Biology, 6*, 1651–1660.