Dimension-specific attention directs learning and listening on auditory training tasks

Lorna F. Halliday · David R. Moore · Jenny L. Taylor · Sygal Amitay

Abstract The relative contributions of bottom-up versus top-down sensory inputs to auditory learning are not well established. In our experiment, listeners were instructed to perform either a frequency discrimination (FD) task ("FD-train group") or an intensity discrimination (ID) task ("ID-train group") during training on a set of physically identical tones that were impossible to discriminate consistently above chance, allowing us to vary top-down attention whilst keeping bottom-up inputs fixed. A third, control group did not receive any training. Only the FD-train group improved on a FD probe following training, whereas all groups improved on ID following training. However, only the ID-train group also showed changes in performance accuracy as a function of interval with training on the ID task. These findings suggest that top-down, dimension-specific attention can direct auditory learning, even when this learning is not reflected in conventional performance measures of threshold change.

Keywords Auditory processing · Training · Learning · Attention · Frequency discrimination · Intensity discrimination · Order effects

L. F. Halliday · D. R. Moore · J. L. Taylor · S. Amitay
MRC Institute of Hearing Research,
University Park,
Nottingham NG7 2RD, UK

L. F. Halliday (✉)
Division of Psychology and Language Sciences,
University College London,
Chandler House, 2 Wakefield Street,
London WC1N 1PF, UK
e-mail: l.halliday@ucl.ac.uk

It is widely established that performance on many auditory perceptual tasks improves following training (Wright & Zhang, 2009). However, the mechanisms underlying this auditory learning are not well understood. A particular point of contention is the relative involvement of "top-down" versus "bottom-up" processes. In the vision literature, psychophysical studies demonstrating that learning is not only specific to the trained stimuli (Ahissar & Hochstein, 1997; Karni & Sagi, 1991), but also to the features attended to during training (Ahissar & Hochstein, 1993), have been interpreted as evidence for an interaction between bottom-up and top-down mechanisms during learning (Ahissar & Hochstein, 2004). However, in the field of auditory learning, evidence for the involvement of top-down processes is less compelling (e.g., Polley, Steinberg, & Merzenich, 2006; Roth, Refael-Taub, Sharvit, & Kishon-Rabin, 2006).[1]

Some evidence that top-down processes may contribute to auditory learning comes from a study by Amitay, Irwin, and Moore (2006). They trained participants on an auditory discrimination task in which there were no physical differences between the stimuli. On each trial, listeners were

---

[1] Polley et al. (2006) found that rats trained to attend to the frequency dimension of tones of various frequencies and intensities that were then examined electrophysiologically showed changes only in cortical frequency mapping, while rats trained to attend to intensity using the same set of stimuli changed only in cortical intensity mapping. However, as it was not possible to ascertain the behavioural transfer from frequency to intensity (and vice versa), it is unclear whether the learning was task specific. In humans, listeners instructed to select the tone with the higher pitch during training using an impossible discrimination task showed improved thresholds on a conventional frequency discrimination task given after training, whereas those told to select the louder tone during training did not (Roth et al., 2006). However, since Roth et al. failed to test the reciprocal task (intensity discrimination), the possibility that their findings were due to preexisting differences between the groups cannot be ruled out.

presented with three identical tones and told to "select the sound that is the odd one out." Despite the fact that this task was impossible, the listeners nonetheless showed significant learning, in the form of improved pre- to posttraining thresholds on a conventional frequency discrimination (FD) probe task in which there were differences between the standard and the target (odd-one-out) sounds. Moreover, the degree of improvement was comparable to that seen in listeners who had received training on a conventional FD task.

Amitay et al. (2006) proposed two explanations for these findings. The first was that listeners were attending to the task-relevant stimulus dimension (frequency) during training, and therefore enhancing their ability to access this dimension and make it available for further processing. Although there were no explicit instructions for the listeners to attend to frequency during training, they had performed a brief FD task immediately prior to the training phase. Listeners may therefore have assumed that they were listening for frequency differences during training. This interpretation assumes that learning was top down— dependent on task instructions and the consequence of attention to a task-specific stimulus dimension. However, the second explanation was that learning was due to listeners creating and refining a global low-level representation of the trained stimulus. At posttraining, listeners could then have used this representation to compare against the heard tones and to identify the target. This interpretation asserts that learning on the impossible discrimination task was bottom up—that it was due to sensitisation/adaptation to the standard and not specific to a particular stimulus dimension. Both of these interpretations assumed that no discriminable differences were present in the stimuli. Amitay, Irwin, and Moore therefore concluded that the learning did not involve fine-tuning of a stimulus comparison mechanism.

Micheyl, McDermott, and Oxenham (2009) offered a third explanation for these findings. Their calculations showed that, because of internal noise (Green & Swets, 1966), physically identical stimuli were likely to have been perceived as physically different, and that those perceived differences were on a par with those induced by stimuli that differed in their actual physical characteristics at just-suprathreshold levels. Micheyl et al. therefore argued that

learning on the impossible discrimination task could have been due, at least in part, to fine-tuning of bottom-up sensory discrimination mechanisms during training. On the other hand, noise could equally have been introduced at the decision variable level or in the decision criterion, and could therefore just as easily represent the top-down effect of attention on the perception of the stimulus.

The present study aimed to assess the effect of dimension-specific attention on learning, by training listeners on an impossible discrimination task (Amitay et al., 2006) and manipulating the task instructions. To do this, one group of listeners were instructed to attend to the frequency dimension during training, a second group were instructed to attend to the intensity dimension during training, and a third group did not receive any training. If learning is dependent on bottom-up processes, it would be expected to occur on the trained stimuli regardless of the task instructions. If, however, learning is dependent on top-down processes, it would be expected to occur only for the dimension to which listeners were instructed to attend. In the latter case, because the training stimuli were identical, differences in auditory performance following training must depend on attention to a specific stimulus dimension, as directed by the instructions for each task.

Method

Participants

A total of 64 adults (23 males, 41 females) 18–40 years of age were recruited via posters from the Nottingham University and Queen's Medical Centre campus. The listeners all had normal hearing (pure tone thresholds ≤20 dB HL across 0.5–4 kHz) and no prior experience of psychophysical testing.

Design

The experiment comprised three phases: pretraining, training, and posttraining (Fig. 1). All phases were completed in a single session, in a sound-attenuated booth. The study
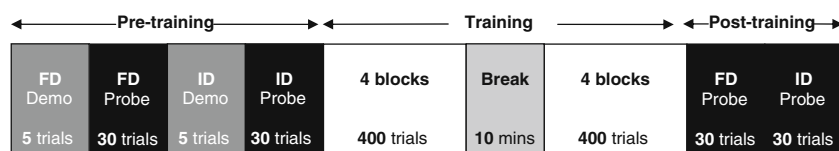

Fig. 1 Experimental design. The frequency discrimination (FD-train) and intensity discrimination (ID-train) groups completed all three phases (pretraining, training, and posttraining). The only difference between the FD-train and ID-train groups was the instructions that

they received during training ("Which tone was higher in pitch?" vs. "Which tone was louder in volume?"). The control group completed the pre- and posttraining phases but did not complete any training trials during the training phase

protocol was approved by the Nottingham NHS Trust Research Ethics Committee.

Stimuli and equipment

The stimuli for the pretraining, training, and posttraining phases consisted of 100-ms pure tones (10-ms raised cosine ramps), with an interstimulus interval of 500 ms. Stimuli were presented diotically using Sennheiser HD-25-1 headphones. The standard tone had a frequency of 1000 Hz and an intensity of 60 dB SPL (reference level 20 μPa). During the training phase, only standard tones were presented. During the pre- and posttraining phases, the target tone was adaptively varied in frequency for the FD task and in intensity for the ID task. Testing was administered using a computer game format in which a visual interface (cartoon characters) cued sound presentation. The FD and ID tasks were represented on screen using different visual interfaces. Responses were recorded via a touch screen.

Pre- and posttraining phase

Difference limens for frequency (DLFs) and intensity (DLIs) were ascertained for all listeners using both FD and ID "probe" procedures. Each of the probes comprised 30 trials, and each trial followed a three-interval, three-alternative forced choice (3I-3AFC) procedure, with both tasks using the same standard stimulus. Listeners were instructed to select the interval that they believed contained a different pitch (FD task) or volume (ID task). Listeners received trial-by-trial feedback, in which correct responses were marked by a brief animation of the correctly chosen character. To familiarise listeners with the task requirements, a short demonstration that contained a combination of "easy" (target stimulus 1500 Hz or 80 dB SPL) and impossible (target stimulus 1000 Hz or 60 dB SPL) trials ($N = 5$, also 3I-3AFC) was administered before each probe in the pretraining phase. The order of the two probes was counterbalanced across listeners.

A three-down, one-up adaptive staircase procedure using logarithmic steps was used to target 79.4% correct on the psychometric function (Levitt, 1971). ΔF (or ΔI) varied adaptively according to the following rule: Starting with an increase of 50% of the standard for FD (i.e., 500 Hz, giving a starting target value of 1500 Hz) and 33% for ID (20 dB SPL, starting target value 80 dB SPL), this difference was divided by 2 following every correct response until the first incorrect response. Thereafter, ΔF (or ΔI) was divided by $\sqrt{2}$ after three correct responses, and multiplied by $\sqrt{2}$ after one incorrect response. DLFs and DLIs were estimated as the 79.4% correct point on the logistic

psychometric function (the probability of responding correctly as a function of the log-transformed adaptive parameter value ΔF or ΔI; see Eq. 1), fitted to the 30 trials in each probe using the fitting technique described by Wichmann and Hill (2001):

$$\Psi(x) = \gamma + \frac{1 - \gamma - \lambda}{1 + e^{-\left(\frac{x}{\alpha}\right)^{\beta}}} \tag{1}$$

x represents the adaptive parameter value, α denotes the midpoint of the psychometric function (the parameter value corresponding to the halfway point between chance and asymptotic performance), β is the slope parameter, γ is the guessing rate (chance performance level), and λ is the asymptotic performance level, used to estimate the lapse rate. α was constrained to be within "reasonable" psychophysical limits (0.01%–60%). β was unconstrained. λ was also unconstrained, because fitting yielded estimates <5%. γ was fixed at 0.33 (for a 3AFC paradigm). The logistic function was chosen following earlier work that had found it the most stable for FD data, even for as few as 30 data points (Amitay, Nelson, Hawkey, Cowan, & Moore, 2006).

Listeners were allocated to one of three groups based on their pretraining DLFs, so as to match the groups as closely as possible on initial FD ability. A one-way ANOVA confirmed that the training groups were well matched on pretraining DLFs, $F(2, 55) = 0.09$, $p = .91$. It was not possible to simultaneously match participants on ID ability, but the differences in pretraining thresholds did not differ significantly, $F(2, 55) = 1.94$, $p = .15$.

Training phase

During the training phase, two groups received training on an impossible auditory task. Listeners in these groups completed eight training blocks of 100 trials each, with a 10-min rest period following the fourth training block (Fig. 1). During each trial, listeners were presented with three intervals, all of which contained the standard tone. Listeners in the FD-train group ($n = 19$) were instructed to select the interval that they believed was different in pitch. Listeners in the ID-train group ($n = 20$) were instructed to select the interval that they believed was different in volume. Listeners were told that the task was very difficult and that they should guess if they could not tell the difference between the tones. The different tasks were cued on screen using the same visual interfaces as for the FD and ID probes during the pre- and posttraining phases. Listeners received trial-by-trial random feedback signalling "correct" responses (33% of the trials). A third, control group were not given any tasks to complete during the training phase (approximately 90 min).

Participant exclusions

Listeners were excluded from the analysis on the basis of two criteria. First, in order to reduce variance in performance, listeners who obtained DLFs at pretraining that were greater than 5% or less than 0.5% were excluded. This resulted in the exclusion of 2 listeners. Second, listeners whose slope estimates on any of the FD or ID pre- or posttraining probes were less than 0.1 were excluded. This was based on the calculation that these estimates had a greater than 10% measurement error, and resulted in the exclusion of a further 4 listeners. This left a total sample size of 58 (19 males, 39 females).

Statistical analyses

DLFs and DLIs (in Hertz and decibels, respectively) were log transformed to normalise the data, and all statistical analyses were carried out on the log-transformed thresholds.

Results

Threshold change

To assess the effects of training, we first calculated pre- and posttraining DLFs and DLIs for the three groups (Figs. 2A and B). Learning indices for each group were then calculated as the difference between pre- and posttraining log-transformed DLFs/DLIs (Figs. 2C and D). Paired-samples t tests were conducted to assess whether any of the groups showed learning on either of the two tasks. For the FD task, only the FD-train group showed significant learning ($p \leq .02$, correcting for three comparisons). For the ID task, all groups showed significant learning, although learning for the FD-train group was not significant after controlling for multiple comparisons ($p = .04$). These findings were supported by analyses showing that, for the FD task, only the FD-train group had a significantly higher proportion of "learners" (i.e., listeners who showed a pre- to posttraining improvement that was $>\sqrt{2}$, the step size in

Fig. 2 (Top) (A) Frequency discrimination (FD) thresholds (mean ± SEM) and (B) intensity discrimination (ID) thresholds during pre- and posttraining for the FD-train, ID-train, and control groups. (Middle) Overall learning (mean ± SEM) for (C) the FD task and (D) the ID task for the three groups. Asterisks mark significant learning ($p < .05$; Bonferroni corrected for three multiple comparisons). (Bottom) Pre- and posttraining (E) FD and (F) ID thresholds for listeners in the three groups. "Learners" (see text) are represented by black lines. "Nonlearners" (see text) are represented by grey lines
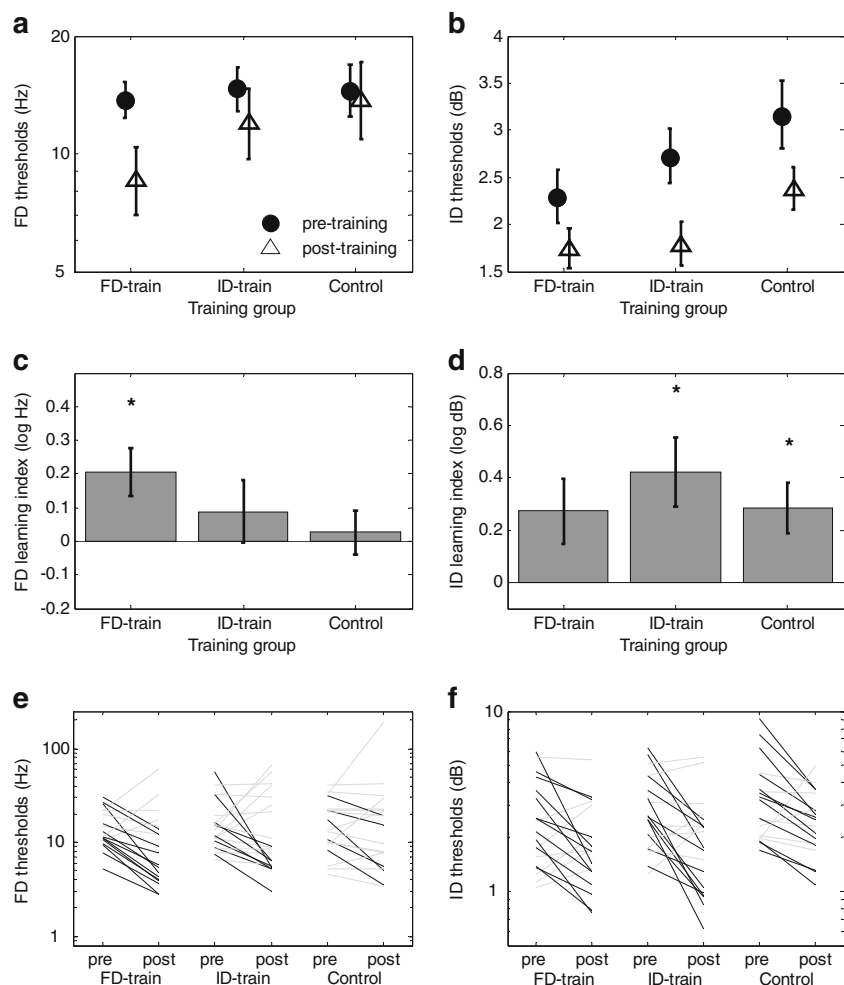
the adaptive staircase) than of "nonlearners" (those who did not) [$\chi^2(2) = 7.10$, $p = .03$]. For the ID task, in contrast, there was no significant difference between the three groups in the proportions of learners and nonlearners [$\chi^2(2) = 0.06$, $p = .97$] (Figs. 2E and F).

Interval order effects

Analyses of threshold changes therefore partially supported the notion that auditory learning is directed by dimension-specific attention, in that only the group who attended to the dimension of frequency during training showed improvements in FD. However, further analyses revealed that one of the assumptions of the 3AFC task—that of equal response accuracy as a function of stimulus presentation order—had been violated in this study. Our results showed that during the pre- and posttraining probes, the likelihood of participants responding correctly to a given trial was dependent not only on the physical difference between the stimuli, but also on the temporal order in which those stimuli were presented. Recent studies have shown that when such order effects (also known as "time order errors"; see, e.g., Fechner, 1860) are present, these can systematically over- or underestimate thresholds

(García-Pérez & Alcalá-Quintana, 2010; Ulrich, 2010; Ulrich & Vorberg, 2009). Moreover, changes in order effects following training have recently been reported to be associated with the M300 (Hairston & Nagarajan, 2007), a component believed to be involved in the allocation of attention (see Soltani & Knight, 2000, for a review). We therefore went on to examine whether order effects might provide an insight into the role of dimension-specific attention in auditory learning.

We found significant order effects for both tasks [FD, $F(2, 330) = 14.74$, $p < .001$; ID, $F(2, 330) = 3.39$, $p = .04$] (Figs. 3A and B, respectively). These effects were driven by the lower response accuracy to Interval 1 relative to Interval 3 in both tasks, and also for Interval 2 relative to Interval 3 in FD (FD, $p < .001$; ID, $p = .01$). For the ID (but not the FD) task, there was also a significant interaction between interval and group [$F(2, 330) = 2.93$, $p = .02$], resulting from a training-induced increase in accuracy for the ID group alone to targets in Interval 1 ($p = .002$) (Fig. 3C). This effect was driven, in turn, by a significant reduction in the false alarm rate (i.e., choosing Interval 1 when the target was in a different interval) between pre- and posttraining [$t(19) = 2.83$, $p = .01$], coupled with a nonsignificant increase in the number of

Fig. 3 Mean performance accuracy (± *SEM*) for Intervals 1, 2, and 3 during pre- and posttraining of the three groups combined on (**A**) the frequency discrimination (FD) task and (**B**) the intensity discrimination (ID) task. (**C**) Performance accuracy for Interval 1 during pre- and posttraining for each of the three groups on the ID task. (**D**) Proportions of false alarms versus misses to targets in Interval 1 during pre- and posttraining for the ID-train group. The asterisk marks significant pre- to posttraining change ($p < .05$). (**E** and **F**) Frequencies of responses selected (± *SEM*) during the training phase for Intervals 1, 2, and 3, as a function of training block, for (**E**) the FD-train group and (**F**) the ID-train group
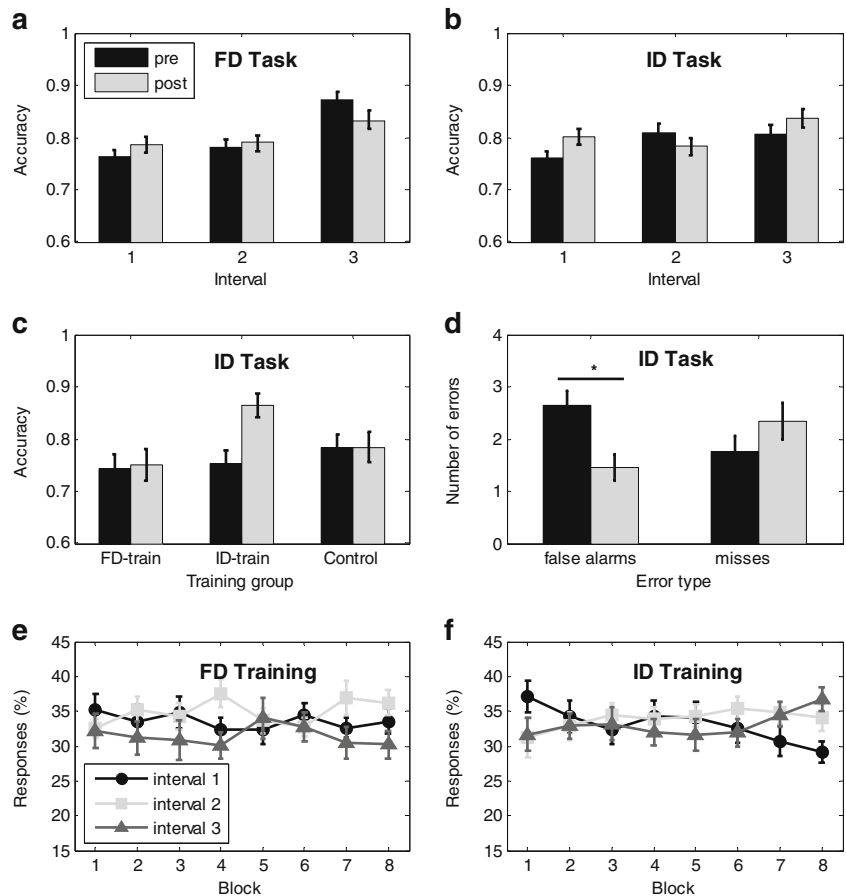
misses [$t(19) = –1.71$, $p = .10$; Fig. 3D], suggesting that the ID-train group was less likely to incorrectly select Interval 1 as the target following training.

Finally, we examined the frequency with which each target interval was selected during training for the FD and ID tasks (Figs. 3E and F, respectively) to establish whether the shift in response bias described above was training related. Whereas interval selection did not change consistently with training in the FD-train group (Fig. 3E), there was a significant change in interval selection in the ID-train group [$\chi^2(14) = 46.11$, $p < .001$] (Fig. 3F). Inspection of the adjusted residuals confirmed that this was due to a reduction in the frequency for which Interval 1 was selected with training and an increase in the frequency for which Interval 3 was selected with training. Taken together, changes to the order effects resulting from ID training demonstrated a dimension-specific learning effect that was not accompanied by a threshold shift. Note that whilst we also observed order effects for the FD task, these did not change with learning. Consequently, whilst order effects may have influenced absolute pre- and posttraining thresholds for the FD task, they are unlikely to have impacted upon the relative difference between them.

Discussion

The implications of our findings are threefold. First, they suggest that dimension-specific attention can direct auditory learning, whether this is reflected in increased performance measures or not, adding to a growing body of evidence from the visual literature that learning is guided by the top-down orienting of attention to task-relevant properties. Our findings do not directly challenge those of Micheyl et al. (2009), in that they do not speak to whether or not listeners perceived these physically identical stimuli as being different. Rather, they suggest that if auditory learning involves the fine-tuning of discrimination mechanisms, this tuning is likely to be directed to the particular dimension to which the listener is attending.

Second, our results provide evidence for the mechanisms underlying attention-driven learning. In a recent review, Amitay (2009) suggested that these might include the selection and enhancement of task-relevant information and the filtering of task-irrelevant information. The finding that only the group who were instructed to attend to the frequency dimension during training showed improved FD thresholds following training appears to provide evidence for selection and enhancement. Moreover, the finding that the ID-train group showed a decreased tendency to select Interval 1 with increasing training, despite the fact that the stimuli were physically identical, may be suggestive of a filtering mechanism. Indeed, a plausible explanation of this result is that the neural response to stimuli in the first interval was stronger, which was interpreted by neurons in primary auditory cortex as "louder" (see also Hairston & Nagarajan, 2007, for comparable MEG data). The random and meaningless feedback during the training phase could then have had the effect of improving listeners' abilities to ignore or "filter out" these neural biases.

Third, the results provide, for the first time, evidence that dimension-specific attention might alter perceptual bias. It is clear that the order effects we observed in this study were not due to participants' idiosyncratic preferences for a particular buttonpress, since the patterns differed between the two tasks. Rather, our findings appear to reflect genuine sensory/perceptual interactions between successive stimuli that (a) change with training (see also Hairston & Nagarajan, 2007; Jamieson & Petrusic, 1975, 1976, 1978), (b) are independent of threshold change, and (c) vary according to the particular stimulus dimension to which the listener is attending. The findings of this study therefore suggest that not only does the way in which listeners perceive successive stimuli alter with training, but the mechanisms of learning may differ according to the particular stimulus dimension to which one is attending. Again, this interpretation is consistent with reports that the M300 might comprise the neural correlate for order effects in audition (Hairston & Nagarajan, 2007).

Finally, it is worth pointing out that where these biases are present, threshold estimates are likely to be more difficult to ascertain (for information about the influence of order effects on thresholds in two-alternative forced choice paradigms, see García-Pérez & Alcalá-Quintana, 2010; Ulrich, 2010; Ulrich & Vorberg, 2009). However, we are confident that our data are robust because, whereas threshold changes for the FD task were observed in the absence of changes in order effects, changes in order effects were seen for the ID task in the absence of differential threshold change. It is therefore evident that these two mechanisms (threshold change and order effects) were acting somewhat independently. This independence suggests that rather than being a nuisance, order effects might instead provide insight in addition to that provided by conventional measures for assessing learning following training.

References

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual-learning. *Proceedings of the National Academy of Sciences, 90,* 5718–5722.

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

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

Amitay, S. (2009). Forward and reverse hierarchies in auditory perceptual learning. *Learning and Perception, 1,* 59–68.

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

Amitay, S., Nelson, A., Hawkey, D. J. C., Cowan, J., & Moore, D. R. (2006). A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners. *Journal of the Acoustical Society of America, 119,* 1616–1625.

Fechner, G. T. (1860). *Elemente der Psychophysik (Erster Teil) [Elements of psychophysics (Vol. 1)].* Leipzig: Breitkopf & Hartel.

García-Pérez, M. A., & Alcalá-Quintana, R. (2010). Reminder and 2AFC tasks provide similar estimates of the difference limen: A reanalysis of data from Lapid, Ulrich, and Rammsayer (2008) and a discussion of Ulrich and Vorberg (2009). *Attention, Perception, & Psychophysics, 72,* 1155–1178.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hairston, I. S., & Nagarajan, S. S. (2007). Neural mechanisms of the time-order error: An MEG study. *Journal of Cognitive Neuroscience, 19,* 1163–1174. doi:10.1162/jocn.2007.19.7.1163

Jamieson, D. G., & Petrusic, W. M. (1975). Presentation order effects in duration discrimination. *Perception & Psychophysics, 17,* 197–202.

Jamieson, D. G., & Petrusic, W. M. (1976). Bias induced by provision of feedback in psychophysical experiments. *Acta Psychologica, 40,* 199–206.

Jamieson, D. G., & Petrusic, W. M. (1978). Feedback versus an illusion in time. *Perception, 7,* 91–96.

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, 88,* 4966–4970.

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

Micheyl, C., McDermott, J. H., & Oxenham, A. J. (2009). Sensory noise explains auditory frequency discrimination learning induced by training with identical stimuli. *Attention, Perception, & Psychophysics, 71,* 5–7.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Reorganization through top-down influences. *Journal of Neuroscience, 26,* 4970–4982.

Roth, D. A., Refael-Taub, R., Sharvit, R., & Kishon-Rabin, L. (2006). Can learning a frequency discrimination task occur without discrimination? *Journal of Basic Clinical Physiology and Pharmacology, 17,* 159–171.

Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology, 14,* 199–224.

Ulrich, R. (2010). DLs in reminder and 2AFC tasks: Data and models. *Attention, Perception, & Psychophysics, 72,* 1179–1198.

Ulrich, R., & Vorberg, D. (2009). Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators. *Attention, Perception, & Psychophysics, 71,* 1219–1227.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313.

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society B, 364,* 301–311.