Vision Research 61 (2012) 83–88

Contents lists available at SciVerse ScienceDirect

Vision Research

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

Stimulus uncertainty in auditory perceptual learning

Karen Banai [a,*], Sygal Amitay [b]

[a] University of Haifa, Haifa, Israel
[b] MRC Institute of Hearing Research, Nottingham, UK

ARTICLE INFO

Article history:
Received 24 March 2011
Received in revised form 19 November 2011
Available online 25 January 2012

Keywords:
Auditory perceptual learning
Stimulus uncertainty
Stimulus variability
Roving

ABSTRACT

Stimulus uncertainty produced by variations in a target stimulus to be detected or discriminated, impedes perceptual learning under some, but not all experimental conditions. To account for those discrepancies, it has been proposed that uncertainty is detrimental to learning when the interleaved stimuli or tasks are similar to each other but not when they are sufficiently distinct, or when it obstructs the downstream search required to gain access to fine-grained sensory information, as suggested by the Reverse Hierarchy Theory (RHT). The focus of the current review is on the effects of uncertainty on the perceptual learning of speech and non-speech auditory signals. Taken together, the findings from the auditory modality suggest that in addition to the accounts already described, uncertainty may contribute to learning when categorization of stimuli to phonological or acoustic categories is involved. Therefore, it appears that the differences reported between the learning of non-speech and speech-related parameters are not an outcome of inherent differences between those two domains, but rather due to the nature of the tasks often associated with those different stimuli.

1. Introduction

A large body of research in the visual modality has looked at the effects of stimulus uncertainty on perceptual learning (Aberg & Herzog, 2009; Adini et al., 2004; Kuai et al., 2005; Otto et al., 2006; Tartaglia, Aberg, & Herzog, 2009; Yu, Klein, & Levi, 2004; Zhang et al., 2008). Some of these studies suggest that visual perceptual learning is disrupted by stimulus uncertainty induced by *randomly* interleaving training trials involving different stimuli or tasks (also known as roving) (Adini et al., 2004; Kuai et al., 2005; Yu, Klein, & Levi, 2004), while others (Tartaglia, Aberg, & Herzog, 2009; Zhang et al., 2008) show that this is not always the case. Different accounts were proposed for those discrepant findings (see below), but the effects of stimulus uncertainty on learning are not fully understood. The goal of the current review is therefore to discuss a parallel group of studies, conducted in the auditory modality, with the hope that they might contribute to a better understanding of the effects of stimulus uncertainty in visual perceptual learning. We limit our discussion to cases where two or more stimulus conditions are *randomly* interleaved across trials, such that uncertainty at the level of each trial persists throughout the experiment.

Two accounts were recently proposed for the effects of uncertainty on learning. On the one hand, based on a series of experiments in which line bisection and vernier acuity tasks were practiced, it has been proposed that stimulus uncertainty introduced by randomly interleaving trials of two different types (e.g., vernier acuity with bisection trials or bisection trials with two different bisection stimuli) disrupts perceptual learning when the stimuli used are distinct but share overlapping (though not identical) neural representations (Tartaglia, Aberg, & Herzog, 2009). It has also been shown that practice on two consecutive blocks of a hyperacuity task lead to no learning if stimuli in the two blocks were presented at the same orientation and at the same retinal location, but that learning did occur if the stimuli were presented in different locations or at different orientations (Seitz et al., 2005). While these findings were interpreted in the context of learning consolidation, they also suggest that the degree of overlap between the neural representations of the trained stimuli influences learning. Another account of the effects of stimulus uncertainty on perceptual learning is offered by the Reverse Hierarchy Theory (RHT) (Ahissar & Hochstein, 1997). According to the RHT, conscious perception under ecological conditions is based on stimulus-relevant information that is represented in high-level neural populations. When the sensory resolution provided by a high-level population is not sufficient, as in the case of fine grained discrimination tasks, an attention driven, top-down search process is initiated to locate the neural populations in which sensory representations retain sufficient level of detail. Learning is the process by which those lower-level representations become more accessible to conscious perception (Ahissar & Hochestein, 2004). By this account, stimulus uncertainty should disrupt learning because it obstructs the top-down search process required to access the

* Corresponding author.
E-mail address: kbanai@research.haifa.ac.il (K. Banai).

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2012.01.009

lower-level neural populations in which fine-grained stimulus information is represented (Nahum, Nelken, & Ahissar, 2010). Nevertheless, it may be that if the different to-be-learned stimuli are sufficiently distinct, it is possible to more flexibly shift attention between the relevant perceptual features and thus initiate multiple backward search processes, resulting in learning (Zhang et al., 2008).

In the auditory modality, stimulus uncertainty appears to contribute to the perceptual learning of degraded speech (Davis et al., 2005; Hervais-Adelman et al., 2008; Loebach & Pisoni, 2008; Stacey & Summerfield, 2007) and novel phonetic categories (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Tremblay et al., 1997, 2001). In many of the speech studies, different stimuli (words, sentences) are used in each trial, or multiple stimuli are repeated a few times each. Therefore, it is often hard to disentangle learning and generalization. Nevertheless, the few cases in which such a distinction is possible suggest that stimulus uncertainty does not facilitate the learning of the trained exemplars, but rather it augments the transfer of learning to untrained materials (e.g., Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Lively, Logan, & Pisoni, 1993). For example, Lively, Logan, and Pisoni (1993) showed that native speakers of Japanese trained to identify the English phonemes /r/ and /l/, became more accurate during the training phase whether the training set was comprised of tokens presented by a single-talker (no stimulus uncertainty) or five different talkers (uncertainty about talker identity exists on each trial). Nevertheless, only listeners in the 5-talkers condition generalized their learning to words that were not included in the training set (see Section 2 for further examples). There is also evidence that practice with two or more *randomly* interleaved stimuli either promotes (Amitay, Hawkey, & Moore, 2005), or has no effect (Karmarkar & Buonomano, 2003) on the learning of auditory frequency and temporal interval discrimination as compared to training with each of the stimuli consistently presented on its own. In the case of auditory temporal interval discrimination, stimulus uncertainty induced by randomly mixing stimuli drawn from around two distinct temporal intervals had little effect on the learning rate of the trained intervals compared to a no-uncertainty training regimen. Training with either regimen resulted in no generalization to any untrained temporal interval (Karmarkar & Buonomano, 2003). Similarly, following training on auditory frequency discrimination generalization of learning to untrained frequencies was similar for listeners trained with protocols incorporating either a single or five randomly-interleaved reference frequencies (Amitay, Hawkey, & Moore, 2005).

In the following sections, we will thus focus on the effects of stimulus uncertainty on the perceptual learning of speech and non-speech acoustic elements. This reviewed body of work suggests that stimulus uncertainty plays a similar role in the learning of speech and non-speech auditory elements. In particular, it appears that when classification (based on either phonetic categories or perceptual anchors – consistently repeating reference stimuli) can be used to solve a given task, as is often the case in speech learning, learning is indifferent to stimulus uncertainty. For the purpose of the current review, the term classification describes all the cases in which a response to an experimental task requires assigning a given stimulus to a particular category. Therefore, deciding whether a stimulus is longer or shorter than a given reference (Karmarkar & Buonomano, 2003) is a classification task because the listener determines whether it falls under the category of 'long' or 'short' stimuli. Likewise, determining that a talker speaks with the dialect of a particular region (Clopper & Pisoni, 2004), or identifying that a word one heard was 'bleed' rather than 'breed' also represent instances of classification. By this definition, the number of potential categories along an acoustic dimension (e.g., tone pitch or duration) is determined in a given experiment by the number of randomly interleaved 'base' stimuli and the question posed to the listeners. For example, in the Karmarkar and Buonomano (2003) study mentioned above, tones from two different interval ranges were used and listeners had to decide whether each tone was shorter or longer than a given reference creating a total of four different categories. Because we view categories as ad hoc constructs, defined experimentally, it seems that learning to categorize involves learning to tag (or label) the different categories rather than remapping the acoustic space to discrete units. This means that the classification process does not have to result in reduced within category sensitivity because it does not necessarily involve changes in acoustic representation per se.

2. Perceptual learning of speech under conditions of stimulus uncertainty

With training, human listeners can gradually learn to interpret degraded speech materials that are incomprehensible to naïve listeners (e.g., Davis et al., 2005; Logan, Lively, & Pisoni, 1991; Pallier et al., 1998; Peelle & Wingfield, 2005; Song et al., 2011; Stacey & Summerfield, 2007). Speech degradation can be achieved by embedding the speech signal in background noise (not discussed here), by using speech materials produced by listeners with foreign accents or uncommon dialects, or by artificially distorting the acoustic properties of the signal. In particular, both the spectral content and the intensity of naturally produced speech change over time in a characteristic way known as the speech envelope, and these fluctuations carry information that is relevant for speech perception (Rosen, 1992). Therefore, it is possible to experimentally degrade the speech envelope or its content (spectrum and fine structure) independently. A popular form of manipulation is vocoding. In vocoded speech, the temporal-envelopes are extracted from natural signals. Those envelopes are then imposed on tone (single frequency) or band-limited noise (range of frequencies) carriers. The result is a signal that maintains the low-level temporal structure of the original signal, but is devoid of natural spectral content, and is thought to simulate the speech signal that is available for cochlear implant users. While natural speech is hard to use in learning studies because it is so well learned, vocoded speech is often initially unintelligible, especially when using a limited number of tones or frequency bands.

In a typical experiment, listeners are exposed to lists of sentences or words during the training phase and are subsequently tested on a *different* list to demonstrate learning. In those experiments, there is no consistent across trial repetition of a fixed or standard comparison stimulus. Furthermore, even though the vocoding manipulation stays the same across trials, a different stimulus is vocoded on each trial therefore randomly changing the acoustics of the trained stimuli on a trial-by-trial basis. Nonetheless, robust learning is often observed (Davis et al., 2005; Hervais-Adelman et al., 2008; Loebach & Pisoni, 2008; Stacey & Summerfield, 2007). Furthermore, because learning generalizes across words, non-words and environmental sounds, it appears that learning occurs at a sub-lexical acoustic–phonetic level (Hervais-Adelman et al., 2008; Loebach & Pisoni, 2008; Loebach, Pisoni, & Svirsky, 2009). Because learning occurred at a level in which stimulus representation is acoustic and not conceptual, this type of learning is considered perceptual. The modification of acoustic representations with training further makes it likely that learning did not result only due to the use of meaningful language materials. No comparison to consistent (non-variable) training protocols was conducted within this line of work, but it appears that the pattern of generalization in the degraded speech studies (e.g., from words to non-words and environmental sounds or across different carrier stimuli used in the vocoding process) (Hervais-Adelman et al., 2011; Loebach, Pisoni, & Svirsky, 2009) is wider than that obtained in non-speech auditory training studies. Although experimental tests of generalization in the non-speech literature are often limited to different values of the trained acoustic parameters, there is evidence that there is little generalization between acoustic dimensions. For example, listeners trained on an impossible frequency discrimination task, in which they were asked to discriminate between identical tones based on their frequency, improved on a subsequent test of frequency discrimination but not on a subsequent test of intensity discrimination (Halliday et al., 2011). Similarly, learning on frequency discrimination did not generalize to duration discrimination even though the tones used in both conditions were similar in frequency and duration (Banai & Ahissar, 2009).

Direct comparisons between conditions varying in the degree of training set uncertainty in the domains of accent and dialect learning suggest that the perceptual learning of both regional dialects and the ability to comprehend foreign accented English were enhanced when the training sets included several, as compared to a single, talker (Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Wade, Jongman, & Sereno, 2007). In those studies, uncertainty is introduced into the training set by varying the number of talkers representing each accent or dialect during training and randomly mixing the utterances produced by the different talkers. Thus, when asked to transcribe English sentences produced by novel non-native English speakers with Chinese or Slovakian accents, listeners who received practice with a training set of five Chinese-accented speakers showed more generalization to untrained sentences and an untrained talker than those who were exposed to a single talker during training, even though the accuracy of listeners exposed to multiple talkers during the training phase itself was lower. In both cases learning did not generalize to an untrained (Slovakian) accent (Bradlow & Bent, 2008). Similarly, after training on a dialect categorization task, in which listeners were asked to categorize utterances based on talkers' dialects, listeners who practiced with a variable training set in which each dialect was represented by three different talkers generalized their learning to novel sentences spoken by unfamiliar talkers. This is in contrast to listeners who practiced with only a single speaker of each dialect who showed no generalization to untrained talkers despite of their higher accuracy during the training phase (Clopper & Pisoni, 2004). It could be that randomly mixing talkers that share an accent or a dialect allows listeners in the multi-talker conditions more opportunity to extract talker-independent accent-specific information compared to listeners who were exposed to a single talker. It therefore appears that stimulus uncertainty during learning facilitates the creation of abstract representations of the properties unique to each accent/dialect, at the cost of less learning of the talker-specific information during the training phase.

Further evidence that learning under conditions of stimulus uncertainty is not restricted to complex linguistic materials and can result in modifications to the acoustic representations of speech comes from studies on learning to discriminate consonant–vowel (CV) syllables (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Tremblay et al., 1997, 2001). Learning phonetic distinctions that are not present in one's native language (e.g., distinguishing [mba] which does not exist in English from the native [ba]) is obtained with a roving training protocol in which different variants of the native and non-native stimuli were presented on each trial, and furthermore, learning transferred to an untrained contrast ([ba] vs. [nba]) (Tremblay et al., 1997, 2001). Perhaps the most compelling evidence for the claim that stimulus uncertainty facilitates learning by contributing to greater generalization to untrained tokens comes from studies in which native speakers of Japanese were trained to discriminate the English phonemes /r/ and /l/, a phonetic distinction that does not occur in Japanese (Bradlow et al., 1999; Iverson, Hazan, & Bannister, 2005; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991). In those studies a minimal pair identification task was used. On each trial listeners heard a word containing either /r/ or /l/ and were asked to select the word they heard from two alternatives presented in written form (e.g., breed/bleed). After several weeks of training, small but significant improvements in both the accuracy and speed of identification were observed among listeners who trained on either single- or multi-talker sets. Generalization to untrained tokens on the other hand was observed only after multi-speaker training, in which uncertainty about talker identity persisted throughout the experiment (Lively, Logan, & Pisoni, 1993). /r/–/l/ identification learning did not depend on the lexical status of the token (that is whether the token was a real word or a pseudo-word), but was sensitive to phonetic environment (e.g., the position of the critical phoneme within the word). These outcomes suggest that learning occurred at an acoustic–phonetic rather than at a lexical level, a conclusion similar to that reached for the learning of vocoded speech as discussed above.

Considering the studies discussed above in the context of the theoretical accounts from the visual modality (see Section 1), the speech learning findings appear consistent with the proposal that uncertainty does not interfere with learning if the roved stimuli are sufficiently distinct along the dimension that is relevant to the task at hand (Tartaglia, Aberg, & Herzog, 2009; Zhang et al., 2008). They further suggest that this may be the case because uncertainty allows learners to acquire the invariant, task relevant, information, and thus allow attention switching to the appropriate perceptual features, as proposed by Zhang and colleagues (2008). In the cases of auditory learning described above this relevant, invariant information would be the information about the dialect or the phonetic category that is common across speakers, as opposed to the information about the identity of each individual talker or the semantic content of each utterance.

Taken together, the studies reviewed in this section suggest that stimulus uncertainty during auditory training contributes to the generalization of the perceptual learning of speech. Uncertainty might promote generalization by allowing listeners to sharpen the abstract representations of the acoustic–phonetic properties that underlie the identification of individual tokens based on the category to which they belong. Although in the auditory system categorization might play a more critical role in the perception of speech (Holt & Lotto, 2010) than in the perception of non-speech acoustic properties of sound, it seems plausible that stimulus uncertainty could similarly contribute to learning on non-speech tasks when the training scenario emphasizes (or at-least affords) categorization rather than within category discrimination. It seems that a similar distinction can be drawn in the visual modality between category and shape learning in which stimulus uncertainty was proven helpful (e.g., Eimas & Quinn, 1994; Posner & Keele, 1968) and visual feature discrimination learning in which uncertainty often impedes learning (see Section 1, above), although a detailed comparison between the two modalities is beyond the scope of the present review.

3. Perceptual learning of non-speech auditory discriminations under conditions of stimulus uncertainty

Whereas the measurement of discrimination thresholds for frequency, intensity and other acoustic parameters under roving or uncertain conditions is relatively common, only few published accounts on learning under such conditions are available. In one group of studies, listeners learned to discriminate tonal sequences that differed on a single (adaptively changing) component (Leek & Watson, 1984; Watson, Kelly, & Wroton, 1976) and found that any variability in the pitch, timing or order of the different components was detrimental to learning. Similar observations have been made in the visual modality (Kuai et al., 2005). The common interpretation of these findings has been that detecting a target in a multi-component sequence relies on attention to the temporal position (or timing) of the target within the stimulus sequence, and that stimulus uncertainty disrupts this attentional process.

More recently, learning under conditions of stimulus uncertainty was demonstrated in the auditory modality using tasks that are more similar in nature to those used in the visual literature in that sensory discriminations between simple tones (rather than tonal sequences) were used. For example, auditory temporal-interval discrimination thresholds are known to improve with practice if a single temporal-interval is consistently practiced over multiple training sessions using either a discrimination task in which listeners should discriminate between two tones on each trial (Banai et al., 2010; Lapid, Ulrich, & Rammsayer, 2009; Wright et al., 1997), or a single interval classification task in which listeners are presented with a reference tone at the beginning of the experiment and are subsequently required to determine whether subsequently presented tones are longer or shorter than that reference (Karmarkar & Buonomano, 2003). A typical finding is that the discrimination of the trained temporal-interval improves and that learning transfers across frequency (that is across a task-irrelevant dimension, to the discrimination of the practiced temporal-interval marked with tones of untrained frequencies) but not to any untrained temporal-intervals. When two distinct temporal intervals are randomly interleaved, both intervals are learned and the pattern of generalization is similar to that observed under a single interval training regimen (Karmarkar & Buonomano, 2003). These findings suggest that stimulus uncertainty does not necessarily disrupt auditory perceptual learning of fine-grained acoustic discriminations. Similar findings were reported for auditory frequency discrimination (Amitay, Hawkey, & Moore, 2005).

Several factors probably contributed to learning and generalization under conditions of uncertainty. Amitay, Hawkey, and Moore (2005) showed that the patterns of learning and generalization depended on initial discrimination thresholds and on the frequency difference between the different roved tracks. Thus, compared to a single condition (non-roving) training regimen with a fixed standard frequency, learning of listeners with good starting thresholds (2/3 of the listeners) was slower when the randomly interleaved base frequencies were 50 Hz apart but not when the differences between the base frequencies were larger than 200 Hz. Listeners with poorer initial thresholds, on the other hand, learned more slowly on both roving conditions remained higher than those of the good listeners. As for generalization, no transfer to untrained conditions was observed among poor listeners. Among good listeners, an asymmetric pattern was observed. Listeners who trained on a roving training regimen transferred their learning fully to a fixed, untrained condition. Even though they received only one fifth of the exposure to stimuli drawn from the fixed condition during training, roving-trained listeners performed it as well as listeners who practiced on this specific condition throughout the practice phase. In contrast, listeners who practiced with a fixed regimen did not transfer their learning to the untrained roving condition, and performed it equivalently to naïve listeners. This pattern of findings suggests that learning under conditions of stimulus uncertainty can, under some conditions, be superior to practice with consistent presentation regimens in that it generalizes more broadly, in keeping with the speech studies discussed in Section 2 above (for a similar conclusion reached based on visual orientation and contrast discrimination training see Xiao et al. (2008)).

In contrast to the finding that auditory temporal interval discrimination can be learned under conditions of uncertainty (Karmarkar & Buonomano, 2003), we found that, whereas the discrimination of two temporal intervals (100 and 350 ms) improved with training when the two intervals were practiced sequentially during each training session (Banai et al., 2010), no learning on either interval was observed when the two intervals were randomly interleaved (Banai et al., 2007). Two differences could account for the discrepancy between these two studies: First, whereas Karmarker and Buonamano used a single interval classification task in which listeners had to determine whether a stimulus was longer or shorter from reference stimuli that were presented at the onset of each block of trials (similar to the categorization tasks used in the speech literature), we have used a 2-alternatives forced-choice (2AFC) discrimination task in which listeners heard two tones on each trial (a fixed reference tone of 100 or 350 ms and a longer test tone) and were asked to determine which of the two stimuli was longer. This interpretation is consistent with the findings of the perceptual learning of speech studies discussed in Section 2, suggesting that identifying stimuli based on either acoustic or phonetic categories may be learned using similar mechanisms. Second, to enable performance on the single interval task, Karmarkar and Buonomano (2003) had to mark each of the intervals with marker tones of different frequencies, whereas in our study the marker frequency was the same for both intervals. While the use of different marker frequencies could potentially result in less overlap between the representations of the two intervals, therefore enabling learning (Tartaglia, Aberg, & Herzog, 2009), we have preliminary data to the contrary. Thus, using two randomly interleaved trained intervals (100 and 350 ms) and a 2AFC task, we find *no* learning even when the two intervals were marked with different frequencies (1 and 4 kHz).

4. Categorization vs. the use of low level acoustic information

We have suggested above that stimulus uncertainty prompts 'categorization learning' of both speech and non-speech auditory stimuli by facilitating the abstract representation of category specific information. This proposal is consistent with predictions derived from the RHT (Ahissar & Hochstein, 1997; Ahissar et al., 2009). According to the RHT the detailed (non-categorical) acoustic information required to make fine-grained discriminations (such as temporal-interval discrimination) becomes available to the listener only if a top-down initiated backward search of that information can be carried out. Stimulus uncertainty impedes this backward search by changing the search parameters on a trial by trial basis, and therefore the learning outcomes will depend upon the necessity of low level acoustic information for learning (Ahissar et al., 2009). The findings of the speech learning studies described above are generally in line with the suggestions of the RHT. Successful learning of novel phonetic categories (like in the /r/–/l/ studies) or of category identification under non-ideal listening conditions (like in the degraded speech studies) does not require the use of the low level acoustic information that makes the different tokens taken from each category distinct, but rather the higher level phonetic information that makes the different instances categorically similar.

In support of the dissociation of categorical and low level acoustic information, Nahum, Nelken, and Ahissar (2010) demonstrated that listeners could learn to categorize phonetically similar words without a corresponding improvement in the use of low level (binaural) acoustic information. In this study, listeners were required to identify pseudo-words embedded in background noise under conditions differing in how low level binaural information was presented to the two ears. One condition was diotic, meaning that the stimuli presented to the two ears were identical. In the other, dichotic, condition the phase of the speech signal presented to one of the ears was inverted thus creating a disparity between the ears. Training followed either a consistent protocol in which diotic and dichotic stimuli were presented in separate blocks of trials, or a random protocol in which diotic and dichotic stimuli were randomly interleaved on a trial-by-trial basis. Diotic thresholds improved with training on either protocol thereby suggesting that categorization can be learned under conditions of stimulus uncertainty. On the other hand, the magnitude of the binaural advantage (defined as the threshold difference between the diotic and dichotic conditions) which provides an estimate of the use of low-level acoustic information, did not reach optimal level even after prolonged training with the random protocol (Nahum, Nelken, & Ahissar, 2010). This is particular noteworthy because under consistent presentation conditions, the low-level acoustic information was fully available even to naïve listeners (Nahum, Nelken, & Ahissar, 2008).

Learning whether stimuli are phonetically similar (as described in the previous paragraph) or whether they belong to the same phonetic categories (as in the /r/–/l/ studies that were described in Section 2) may not require the use of fine-grained low level acoustic information and can thus proceed even under conditions of stimulus uncertainty. This doesn't necessarily mean that the ability to discriminate between stimuli from the same category (based on acoustic differences) is lost, but rather that it may depend on whether experimental conditions make it possible for listeners to reach a decision based on acoustics rather than on category labels. Indeed, within-category discrimination also appears to improve with practice under roving conditions (Amitay, Hawkey, & Moore, 2005; Karmarkar & Buonomano, 2003). In these studies listeners were asked to perform frequency discriminations (Amitay, Hawkey, & Moore, 2005) or decide whether a presented temporal-interval was longer or shorter than an exemplar presented in the beginning of the experimental block (Karmarkar & Buonomano, 2003). Categorization was therefore not sufficient; an additional, within-category, decision was also necessary. That this further decision was possible is inconsistent with the idea that task similarity between the randomly interleaved conditions blocks learning (Tartaglia, Aberg, & Herzog, 2009) because in both studies listeners were required to perform the same task with each of the different (category) stimuli.

It could be that in those cases of successful learning of within category discrimination, sufficient acoustic information was maintained at the level of categorical representation to allow within category classifications. This suggestion is consistent with the finding that the generalization of learning on a vocoded speech task depended on the specific acoustics properties of the manipulation used to degrade the speech (Hervais-Adelman et al., 2011). Alternatively, if listeners are able to categorize the interleaved stimuli correctly, then within a category they may be able to form and maintain an anchor (based on the repeated reference tone or the base stimulus) against which subsequent stimuli could be classified, rather than discriminated (Nahum, Daikhin, et al., 2010). This latter possibility is consistent with the RHT, as long as the different categories are sufficiently distinct. In this case, a backward search of the neural populations in which the required low level information is represented might be implemented within each category from the higher, categorical level to the lower, acoustical level. Indeed, when listeners trained on frequency discrimination with randomly interleaved stimuli from five different frequency categories that were clearly separable (wide roving, 570, 840, 1170, 1600, 2150 Hz), learning was faster than learning under a similar regimen in which the frequency categories were less separable (narrow roving, 900, 950, 1000, 1050, 1100 Hz). Furthermore, uncertainty precluded learning entirely among listeners whose initial discrimination was poor, and who therefore were less likely to find the different categories separable (Amitay, Hawkey, & Moore, 2005). Note though that listeners in the narrow roving condition never reached an asymptotic level of performance on the trained roving condition. After 3500 trials of training, thresholds on the roving condition were similar to those of naïve listeners on the non-roving conditions, yet, as mentioned above, at the same time, thresholds of the roving-trained listeners on the non-roving condition were equivalent to those of listeners who trained on that condition the entire time. This is inconsistent with the RHT, unless one assumes that once learning was initiated on the roving condition, the low-level information becomes fully accessible, but listeners are unable to fully use it under roving conditions.

5. Summary and conclusion

Uncertainty about the values of the acoustic parameters or in the timing of the presentation of the stimuli within the sequence slows or prevents learning that is dependent on the use of low level acoustic information such as that required for fine-grained auditory discriminations (Banai et al., 2007; Leek & Watson, 1984; Nahum, Nelken, & Ahissar, 2010). This however is not always the case. Learning under conditions of stimulus uncertainty has been observed across 'simple' acoustic parameters (Amitay, Hawkey, & Moore, 2005; Karmarkar & Buonomano, 2003), and tasks involving speech stimuli. In particular, talker variability contributes to the generalization of learning on various speech tasks (phonetic discriminations, dialect identification) (Clopper & Pisoni, 2004; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991), perhaps because in these cases the information critical for learning is categorical rather than purely acoustic in nature. Therefore, we are led to the conclusion that learning of speech and non-speech elements might differ not due to inherent differences between those two domains, but rather due to the different emphasis placed by the tasks typically used in each domain on categories vs. individual exemplars. In the speech domain, identifying talkers or phonetic categories emphasizes the categorical nature of the stimulus, making information that is easily available to conscious perception sufficient for learning. In those cases, stimulus uncertainty might help extracting features that are common to all category members. In the acoustic domain, the emphasis is often on within category discriminations. This acoustic information is typically represented in lower levels of the auditory system, and uncertainty might impede the backward search required to use this information to make perceptual decisions, consistent with the RHT. Nevertheless, some of the findings (e.g., "super generalization" from a variable to a fixed condition, Amitay, Hawkey, & Moore, 2005), are inconsistent with current theories of learning. This conclusion joins a growing body of work arriving at similar conclusions by analyzing the patterns of generalization of auditory perceptual learning following training on non-variable protocols (Wright & Zhang, 2009), as well as by looking at the learning of task-relevant vs. task irrelevant information in auditory tasks (Amitay, 2009).

From a practical standpoint, we suggest that the use of stimulus uncertainty should depend on the goals of training. Whereas fine discrimination of acoustic features might be better learned when there is no uncertainty, the properties of more abstract categories might are better generalized if learned with multiple, randomly presented stimuli. Because the effects of uncertainty might also depend on listener characteristics (Amitay, Hawkey, & Moore, 2005), deciding on an optimal training procedure should take those factors into account as well.

Acknowledgments

We thank Beverly Wright for helpful discussions of the issues raised in this review and David Moore for providing helpful comments on the manuscript. K.B. is supported by a Marie Curie International Reintegration Grant 224763.

References

Aberg, K. C., & Herzog, M. H. (2009). Interleaving bisection stimuli – randomly or in sequence – does not disrupt perceptual learning, it just makes it more difficult. *Vision Research, 49*(21), 2591–2598.

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*(12), 993–1005.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*(10), 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. (2009). Forward and reverse hierarchies in auditory perceptual learning. *Learning and Perception, 1*(1), 59–68.

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception and Psychophysics, 67*(4), 691–698.

Banai, K., & Ahissar, M. (2009). Perceptual learning as a tool for boosting working memory among individuals with reading and learning disability. *Learning and Perception, 1*(1), 115–134.

Banai, K., Ortiz, J. A., Oppenheimer, J. D., & Wright, B. A. (2007). Daily training requirements for bypassing the stimulus specificity of perceptual learning (A). *Cognitive Neuroscience Society* (New York).

Banai, K., Ortiz, J. A., Oppenheimer, J. D., & Wright, B. A. (2010). Learning two things at once: Differential constraints on the acquisition and consolidation of perceptual learning. *Neuroscience, 165*(2), 436–444.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*(5), 977–985.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47*(Pt 3), 207–239.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*(2), 222–241.

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development, 65*(3), 903–917.

Halliday, L. F., Moore, D. R., Taylor, J. L., & Amitay, S. (2011). Dimension-specific attention directs learning and listening on auditory training tasks. *Attention, Perception, and Psychophysics, 73*(5), 1329–1335.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 460–474.

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 283–295.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, and Psychophysics, 72*(5), 1218–1227.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/–/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*(5), 3267–3278.

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning and Memory, 10*(2), 141–147.

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*(11), 1497–1499.

Lapid, E., Ulrich, R., & Rammsayer, T. (2009). Perceptual learning in auditory temporal discrimination: No evidence for a cross-modal transfer to the visual modality. *Psychonomic Bulletin and Review, 16*(2), 382–389.

Leek, M. R., & Watson, C. S. (1984). Learning to detect auditory pattern components. *Journal of the Acoustical Society of America, 76*(4), 1037–1044.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*(3 Pt 1), 1242–1255.

Loebach, J. L., & Pisoni, D. B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *Journal of the Acoustical Society of America, 123*(2), 1126–1139.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2009). Transfer of auditory perceptual learning with spectrally reduced speech to speech and nonspeech tasks: Implications for cochlear implants. *Ear and Hearing, 30*(6), 662–674.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*(2), 874–886.

Nahum, M., Daikhin, L., Lubin, Y., Cohen, Y., & Ahissar, M. (2010). From comparison to classification: A cortical tool for boosting perception. *Journal of Neuroscience, 30*(3), 1128–1136.

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology, 6*(5), e126.

Nahum, M., Nelken, I., & Ahissar, M. (2010). Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning. *Vision Research, 50*(4), 391–401.

Otto, T. U., Herzog, M. H., Fahle, M., & Zhaoping, L. (2006). Perceptual learning with spatial uncertainties. *Vision Research, 46*(19), 3223–3233.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition, 26*(4), 844–851.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1315–1330.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3), 353–363.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 336*(1278), 367–373.

Seitz, A. R., Yamagishi, N., Werner, B., Goda, N., Kawato, M., & Watanabe, T. (2005). Task-specific disruption of perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 102*(41), 14895–14900.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex*. doi:10.1093/cercor/bhr196.

Stacey, P. C., & Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *Journal of the Acoustical Society of America, 121*(5 Pt 1), 2923–2935.

Tartaglia, E. M., Aberg, K. C., & Herzog, M. H. (2009). Perceptual learning and roving: Stimulus types and overlapping neural populations. *Vision Research, 49*(11), 1420–1427.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America, 102*(6), 3762–3773.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1–P2 complex after speech-sound training. *Ear and Hearing, 22*(2), 79–90.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica, 64*(2–3), 122–144.

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America, 60*(5), 1176–1186.

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.

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 364*(1515), 301–311.

Xiao, L. Q., Zhang, J. Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology, 18*(24), 1922–1926.

Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision, 4*(3), 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*(8), e197.