Journal of Experimental Psychology: General
2012, Vol. 141, No. 2, 363–381


0096-3445/12/$12.00 DOI: 10.1037/a0025014

# Implicit Training of Nonnative Speech Stimuli

Eleni L. Vlahou and Athanassios Protopapas
Institute for Language & Speech Processing

Aaron R. Seitz
University of California, Riverside

Learning nonnative speech contrasts in adulthood has proven difficult. Standard training methods have achieved moderate effects using explicit instructions and performance feedback. In this study, the authors question preexisting assumptions by demonstrating a superiority of implicit training procedures. They trained 3 groups of Greek adults on a difficult Hindi contrast (a) explicitly, with feedback (Experiment 1), or (b) implicitly, unaware of the phoneme distinctions, with (Experiment 2) or without (Experiment 3) feedback. Stimuli were natural recordings of consonant-vowel syllables with retroflex and dental unvoiced stops by a native Hindi speaker. On each trial, participants heard pairs of tokens from both categories and had to identify the retroflex sounds (explicit condition) or the sounds differing in intensity (implicit condition). Unbeknownst to participants, in the implicit conditions, target sounds were always retroflex, and distractor sounds were always dental. Post-training identification and discrimination tests showed improved performance of all groups, compared with a baseline of untrained Greek listeners. Learning was most robust for implicit training without feedback. It remains to be investigated whether implicitly trained skills can generalize to linguistically relevant phonetic categories when appropriate variability is introduced. These findings challenge traditional accounts on the role of feedback in phonetic training and highlight the importance of implicit, reward-based mechanisms.

*Keywords:* nonnative speech perception, phonetic training, implicit learning, perceptual learning

Adults fail to differentiate certain distinctions not found in the phonetic inventory of their native language (Flege, 2003). Well-studied cases include perception of the English /r/–/l/ contrast by Japanese listeners (Hattori & Iverson, 2009; Miyawaki et al., 1975) and of the Hindi dental-retroflex contrast by English listeners (Werker & Tees, 1984b). Major theoretical approaches to cross-language speech perception (such as the "perceptual assimilation model"; Best, 1995) and phonetic learning (the "speech learning model"; Flege, 1995) have focused on differences in the difficulty to perceive and learn nonnative contrasts, aiming to predict which contrasts will be least discriminable or most resistant to learning in adulthood (see recent review in Best & Tyler, 2007). Assimilation and perceptual interference, resulting from native language phonetic learning and interacting with second-language experience, are postulated to account for these differences (Iverson, Ekanayake, Hamann, Sennema, & Evans, 2008; Iverson et al., 2003). Beyond accounting for relative difficulty, researchers have examined the effectiveness of different phonetic training methods aiming to improve identification and discrimination of difficult nonnative contrasts. These efforts have typically focused on the selection of training stimuli and presentation schedules (e.g., Iverson, Hazan, & Bannister, 2005; Kondaurova & Francis, 2010; Lively, Logan, & Pisoni, 1993), investigating the role of variability along acoustic dimensions, discriminability of stimuli, and salience of critical acoustic features.

Relatively less emphasis has been given to learning components other than stimulus selection and presentation, such as the nature of the training task (e.g., identification vs. discrimination training; Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; earlier review in Logan & Pruitt, 1995), participant intention and attention (e.g., explicitly focusing on sounds vs. meanings: Guion & Pederson, 2007; or vowels vs. consonants: Pederson & Guion-Anderson, 2010), and the role of feedback (McCandliss, Conway, Protopapas, & McClelland, 2002; earlier review in Logan & Pruitt, 1995). These aspects are important for elucidating the nature of learning that takes place during phonetic training and for understanding the barriers to nonnative phonetic perception and how they arise from previous learning experience. For example, McCandliss et al. (2002) suggested that the difficulties adult listeners face in the perception of some nonnative contrasts may result from Hebbian learning mechanisms that tend to reinforce "whatever response a neural system makes to an incoming stimulus" (p. 91).

In the present study, we are concerned with the conditions and mechanisms of learning. In particular, we approach the problem of learning phonetic contrasts from the perspective of adult perceptual learning (e.g., Ahissar & Hochstein, 1993; Seitz & Dinse, 2007; Seitz & Watanabe, 2005), which has provided fundamental insights into mechanisms guiding plasticity in the adult brain. As such, we address less-investigated aspects of the training procedures concerning listener awareness and intention and the role of feedback. In particular, we apply an incidental learning paradigm (Seitz & Watanabe,

This article was published Online First September 12, 2011.

Eleni L. Vlahou and Athanassios Protopapas, Department of Educational Technology, Institute for Language & Speech Processing, Maroussi, Greece; Aaron R. Seitz, Department of Psychology, University of California Riverside.

We are grateful to Lori L. Holt, Travis Wade, and Radhika Aravamudhan for the recordings of the Hindi stimuli.

Correspondence concerning this article should be addressed to Athanassios Protopapas, who is now at the Department of History and Philosophy of Science (MITHE), University of Athens, Ano Ilissia Campus, GR-157 71 Athens, Greece. E-mail: aprotopapas@phs.uoa.gr

2009) to a difficult nonnative speech contrast, compared with a standard procedure of explicit training with trial-by-trial feedback.

## Statistical Learning and Phonetic Perception

An extensive literature on "implicit learning" suggests that learning can occur after exposure to a stimulus domain, in an unsupervised manner, through powerful mechanisms that extract critical environmental regularities (e.g., Chun & Jiang, 1998; Cleeremans, Destrebecqz, & Boyer, 1998; Kim, Seitz, Feenstra, & Shams, 2009; Pacton, Perruchet, Fayol, & Cleeremans, 2001; Reber, 1989; Seitz, Kim, van Wassenhove, & Shams, 2007). In the context of developmental language acquisition, infants briefly exposed to a continuous stream of artificial syllables, without any acoustic or prosodic cues to word boundaries, were able to extract word-like units based on the transitional probabilities of the syllable pairs between and within nonwords (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). Moreover, infants can exploit prosodic, phonotactic, phonological, and stress patterns from their linguistic environment (Saffran, Werker, & Werner, 2006), and children show increasing sensitivity to several graphotactic and morphological regularities on which they have not received formal instruction (Deacon, Conrad, & Pacton, 2008; Pacton et al., 2001; Steffler, 2001).

Similar statistical mechanisms are often assumed to be operating during the acquisition of phonetic categories. Phonetic learning has been described as a bottom-up process, in which perceptual categories are shaped and modified by the distributional characteristics of phonetic exemplars in the speech input (Guenther & Gjaja, 1996; Maye & Gerken, 2000; Pierrehumbert, 2001, 2003). Maye and colleagues (Maye & Gerken, 2000; Maye, Werker, & Gerken, 2002) have tested these ideas by exposing infants and adults to sounds from a synthetic continuum. All listeners were exposed to the entire continuum, but some were exposed more often to sounds in the midrange of the continuum (unimodal groups) and others more often to sounds near the endpoints (bimodal groups). In discrimination tests after training, unimodal groups treated sounds from the entire continuum as if belonging to a single category, whereas bimodal groups discriminated the sounds, as if assigning them to different categories. Thus, whereas exposure to the unimodal distribution resulted in reduced ability to differentiate sounds that were initially discriminable, Maye, Weiss, and Aslin (2008) showed that exposure to a bimodal distribution resulted in improved discrimination of a contrast that was difficult to discriminate prior to training.

## Adult Phonetic Training and Performance Feedback

Theoretical accounts of learning based on purely unsupervised mechanisms, such as statistical learning, have provided important demonstrations regarding how people can pick up statistical regularities of the environment. However, they cannot fully encompass learning of difficult nonnative phonetic contrasts by adults, as numerous studies have documented that prolonged exposure to nonnative phonetic contrasts is not sufficient for the development of native-like phonetic categories (Munro & Bohn, 2007; Strange, 1995). Additional components or restrictions of learning processes must be posited to account for this apparent insensitivity to shifting environmental statistics. An issue raised by McCandliss et al. (2002) is that when a range of acoustically distinct speech stimuli elicits a common response in the activation of a single perceptual representation, Hebbian learning reinforces the tendency of the sounds to elicit this common percept, thereby further impeding their discrimination. Although this mechanism helps infants ignore sound differences that are not contrastive in their native language, it can pose serious difficulties to adults exposed to a novel linguistic environment. For example, if a Japanese listener has formed a common representation for sounds encompassing the range of both English /r/ and /l/, then Hebbian learning will further strengthen this tendency. Paradoxically, exposure to English has the effect of strengthening Japanese representations, contra statistical learning predictions.

To test the relative importance of supervised versus unsupervised learning mechanisms, McCandliss et al. (2002) trained Japanese listeners to distinguish English /r/-/l/ stimuli, crossing discriminability with feedback. Two groups of listeners were trained with stimuli that were acoustically exaggerated so as to be discriminable, gradually reducing the exaggeration toward natural stimuli (adaptive conditions), whereas two other groups were trained with stimuli at natural levels throughout (fixed condition). In each stimulus condition, one group received trial-by-trial performance feedback, whereas the other group did not. All four groups received the same amount of training, but they did not show the same learning effects. Specifically, performance in the fixed/no feedback condition did not differ from that of a no-training control group. This result is consistent with predictions arising from the Hebbian framework described above (see also Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006, for a replication of this finding) but not with simple statistical learning. However, in the fixed-feedback condition, where the exact same stimuli were heard for the same training period, significant learning was seen, apparently attributable to the feedback provided. Learning was also found in both adaptive conditions, even when feedback was absent. Adaptive learning without feedback is consistent with the Hebbian framework, as sounds rendered discriminable via exaggeration elicit distinct representations throughout training, thereby gradually extending learning to less exaggerated stimuli toward the middle of the continuum. These findings indicate that, in adults, exposure to a stimulus domain may not be sufficient for learning, especially when there is resistance to learning, presumably because of prior learning.

Adult phonetic training studies differ dramatically in their methods (see Bradlow, 2008, for a review of different training regimes), but they typically adopt an *explicit* approach to training. That is, participants are informed about the phonetic distinctions and the number of categories and are required to focus their attention on the phonetic differences. Importantly, listeners are typically provided with some form of feedback on each trial (e.g., Golestani & Zatorre, 2004; Iverson et al., 2005; Lively et al., 1993; Logan, Lively, & Pisoni, 1991; Protopapas & Calhoun, 2000; Pruitt, Jenkins, & Strange, 2006). It has been suggested that feedback may promote learning by directing attention to the critical stimulus characteristics that need to be differentiated (Logan et al., 1991). However, despite its ubiquitous use in phonetic training studies, the exact mechanisms by which feedback contributes to the observed plasticity remain uncertain (Logan & Pruitt, 1995; Tricomi et al., 2006).

Tricomi et al. (2006) used functional magnetic resonance imaging while Japanese listeners were trained on the English /r/–/l/ distinction in blocks with feedback alternating with blocks without feedback. They found that the caudate nucleus was more strongly activated in the feedback condition, in an activation pattern similar to that obtained with a card-guessing task including monetary rewards. According to Tricomi et al., these findings suggest that one way in which feedback may facilitate learning involves recruitment of brain structures and neuromodulatory signals that enhance perceptual plasticity. Although not specific to language learning, involvement of such mechanisms suggests an important role for reward-based learning in first and second language acquisition (Goldstein, King, & West, 2003; Gros-Louis, West, Goldstein, & King, 2006; Tricomi et al., 2006).

## Task-Irrelevant Perceptual Learning

Recent research on perceptual learning has helped clarify how reward-based learning mechanisms can contribute to both explicit and implicit learning situations. Seitz and Watanabe have proposed a model in which perceptual learning is gated by diffuse reinforcement and learning signals that are elicited upon processing of important stimuli (Seitz & Watanabe, 2003, 2005, 2009). According to this model, learning may take place for stimulus features, whether or not they are relevant to the task, as long as they are systematically paired with successfully processed task targets, or rewards, within a critical time window. This model of task-irrelevant perceptual learning (TIPL) has been applied to learning in a variety of tasks in the visual modality (Seitz, Nánez, Holloway, & Watanabe, 2005, 2006; Watanabe, Nánez, & Sasaki, 2001). Recently, Seitz et al. (2010) extended TIPL to the auditory modality, showing that participants could implicitly learn to detect single formant transitions paired at subthreshold levels with attended targets in an unrelated auditory identification task.

Several aspects of this perceptual learning framework make it attractive for reinterpretation of existing findings and extension to novel methodologies regarding learning of nonnative phonetic contrasts. The TIPL model posits that refinement of perceptual representations will occur when two conditions are simultaneously met: First, an incoming stimulus has been neurally registered, regardless of attention, intention, or awareness. Stimulus representation may be minimal or imprecise. Second, a global nonspecific reinforcement signal is produced on the basis of successful completion of processing or other significant eliciting occasion. The cause of this signal need not relate to the stimulus in question, because its effects are nonspecific in that any active representation, however minimal, can be strengthened. In the usual phonetic training situations, the learning signal may be elicited on the basis of successful stimulus discrimination, as in the adaptive/no feedback condition of McCandliss et al. (2002), or on the basis of external feedback provided in the training procedure, as in their fixed conditions and in the comparable conditions of other studies. It may also be related to performance contingencies on an unrelated explicit task, such as success in a video game (Lim & Holt, 2011). In ecologically realistic situations, processing success is generally related to the stimuli that led to the success, so that effects on other incidentally available stimuli tend to cancel out when unsystematic, as in operant conditioning. However, when systematic stimulus-stimulus pairings occur, learning signals reinforce unrelated representations as well, as in classical conditioning.

## Rationale and Overview of the Present Study

The nonspecificity of the purported TIPL learning signal affords the opportunity to train difficult phonetic contrasts in a novel way. The TIPL framework suggests that a number of conditions typically considered necessary for phonetic learning might in fact be superfluous. These include attention to the critical (i.e., distinctive) dimensions of the stimuli to be learned, awareness of the learning situation, intention to learn, a manageable level of initial discriminability, and trial-by-trial performance feedback. In contrast, all that would be necessary for TIPL is an internally generated success signal temporally coinciding with the presence of the critical stimulus attribute(s) to be learned. In the present study, we explore this possibility by removing the usual scaffolding of phonetic training studies and replacing it with an unrelated implicit manipulation.

More specifically, our primary research interest was to test two key predictions that arise from the model of Seitz and Watanabe (2009): First, that processing of speech stimuli with an appropriate reinforcement schedule is sufficient for learning, in the absence of awareness of what needs to be learned, knowledge of the number of categories to be formed or of the fact that there are distinct categories in the training set at all. Second, that plasticity can be enhanced by internally generated learning signals associated with target processing, in the absence of external reinforcements (such as the visual feedback typically provided in phonetic training studies). To test these ideas, we trained different groups of adult Greek listeners on a difficult nonnative phonetic contrast either explicitly or implicitly, with or without performance feedback.

The stimuli were tokens of the Hindi dental /t/ – retroflex ʈ phonetic contrast. Retroflex sounds are only used in 11% of the world's languages (Golestani & Zatorre, 2004), and their nonnative perception declines after the age of 8 months (Werker & Tees, 1984a). Learning of this contrast has proven very difficult for English speakers, but positive results have been obtained after extensive natural experience with Hindi or laboratory training (Golestani & Zatorre, 2004; Werker & Tees, 1984b). Greek listeners have no familiarity with any retroflex sounds, and preliminary testing revealed that they tend to perceive both phonemes as nonrepresentative tokens of the Greek /t/.

We examine the mechanisms of learning in a series of three experiments employing these stimuli. In the first experiment, the standard approach was taken, to produce a baseline reference. Training was conducted with explicit instructions, explicit category labels, and trial-by-trial feedback. Before and after training, listeners' performance was assessed with identification and discrimination tests. In the next two experiments, we employed implicit learning with a similar exposure schedule and stimuli as in Experiment 1, using an intensity difference detection task. Trial-by-trial performance feedback was provided in Experiment 2 and withheld in Experiment 3. After the last training session in these experiments, participants were informed about the phonetic distinctions and were tested on explicit identification and discrimination in tasks identical to those used in Experiment 1. Participants' performance was then compared with the pre-test performance of naive listeners in Experiment 1. It was not possible to obtain pre-training performance for listeners in the last two experiments, because that would provide them with information about the critical task features and would expose them to the stimuli, invalidating the intended implicit nature of the task.

## Experiment 1: Explicit Training With Feedback

The purpose of the first experiment was threefold: first, to examine the initial performance of adult Greek listeners with no prior experience with Hindi sounds and verify the selection of phonetic contrast as difficult but not impossible to learn; second, to use participants' pre-training scores as a baseline reference to assess the effects of explicit and implicit training in this and the following experiments; third, to quantify the extent of phonetic learning under standard laboratory explicit training conditions with performance feedback, for direct comparison with the effects of implicit training in the following experiments.

### Method

**Participants.** Fifteen adult Greek speakers (10 women, 5 men; 22–35 years old) participated in the experiment. Most were students who either volunteered or were given course credit for their participation. No participant reported any hearing or speech impairments or any previous experience with Hindi sounds.

**Stimuli.** Stimuli were natural recordings of Hindi syllables spoken by two male native Hindi speakers. Recordings were made in a sound isolation booth at Carnegie Mellon University. Each speaker pronounced all possible consonant-vowel (CV) and vowel-consonant (VC) combinations involving a dental-retroflex stop contrast, that is, voiced and unvoiced aspirated and unaspirated stop consonants, each combined with five short and four long vowels. The stimuli were read out from a sheet of syllables written in the Devanagari script, each repeated 10 times by each speaker. The CV syllables with an unaspirated unvoiced stop followed by the (long) [a:] were selected as most similar to a corresponding Greek /ta/ syllable. The 10 selected syllables from each speaker were excised from the recording, downsampled to 22,050 Hz (16-bit mono), RMS-amplitude normalized, equated in duration at 170 ms using a 15-ms square-cosine off-ramp when needed, and finally zero-padded to 350-ms total audio duration. There were thus a total of 10 [ta:] and 10 [ʈa:] tokens in each of two voices.

**Procedure.** The experiment consisted of a pre-training test phase, to assess initial performance of naive Greek listeners; a training phase; and a post-training test phase, to assess the effects of training. Stimulus presentation was controlled by DMDX scripts (Forster & Forster, 2003). Participants carried out the training phase at home on laptop computers, but both testing phases were conducted under direct experimenter supervision. The sounds were presented binaurally over headphones provided by the experimenters at an individually adjusted comfortable level. Participant compliance and progress was controlled with daily communication.

***Training.*** During training, recordings from only one of the two Hindi speakers were used, the same one for all participants. Training included six sessions conducted over a period of 3 days, usually two sessions per day. In each trial, participants heard two pairs of Hindi syllables, one pair with two identical tokens beginning with a dental consonant and one pair with two identical tokens beginning with a retroflex consonant. In half of the trials the dental pair preceded the retroflex pair. A 250-ms interval separated the members of each pair and a 500-ms interval separated the two pairs. For example, on a typical trial, participants might hear a syllable beginning with a retroflex sound ([ʈa:]), then after 250 ms the exact same syllable ([ʈa:]), then, after 500 ms, a syllable beginning with a dental sound ([ta:]) and, finally, after 250 ms, the same syllable ([ta:]). The unusual structure of this task was designed to precisely match the exposure schedule to that of the subsequent experiments while retaining critical properties of standard identification training.

Participants were required to listen carefully to both pairs and identify the retroflex pair by pressing the corresponding button on the keyboard. If no response was registered within 3 s after the last syllable was presented, the program marked an incorrect response and moved on to the next trial. At the end of each session, the participant's overall score (percent correct responses in that session) appeared on the screen, as a motivational feature. Each session lasted approximately 10 min and consisted of 100 "normal" and 10 "probe" trials.

*Normal trials.* In the normal trials, participants received immediate feedback after each response: a row of green stars for correct responses or red X's for incorrect responses. Five different tokens (from the same speaker) were presented in these trials.

*Probe trials.* In the second half of each session, two probe trials were randomly interspersed within each block of 10 normal trials. Probe trials differed from normal trials in two ways: (a) No performance feedback was given, and (b) five different tokens were used (from the same speaker as normal trials), which were not heard during normal trials. Therefore, participants never received feedback for their performance on these particular tokens.

***Testing.*** Testing was performed before and after training. Prior to training, testing was preceded by a brief familiarization phase, in which (a) participants were informed that in Hindi there are two different groups of sounds that both sound like the Greek /t/ and an arbitrary label was assigned to each category ("T1" for retroflexes and "T2" for dentals); and (b) they were briefly exposed to 10 tokens from each category. Three types of tasks were used to examine phonetic perception of the target contrast:

*Pair identification.* This test was devised to match the structure of the training trials, maximizing sensitivity to detect learning effects by retaining a familiar testing context. In each trial, participants heard two pairs of Hindi syllables, one with dental and one with retroflex consonants, in random order, and had to indicate the retroflex pair by pressing the corresponding button. There were 80 trials in this task. In one half of the trials, trained tokens were presented, that is, tokens used in normal training trials, for which participants had received feedback during training. In the other half of the trials, tokens presented in probe trials were used, for which no feedback had been given during training.

*Singleton identification.* This was a standard identification task. In each trial, one Hindi syllable was presented and the participant had to categorize it as dental or a retroflex by pressing the corresponding button. There were 100 trials, half with a dental and half with a retroflex sound. As in the preceding test, in one half of the trials, tokens from normal trials were presented, whereas in the other half, tokens from probe trials were presented.

*Discrimination.* This was a standard categorical AX discrimination task. In each trial, two different tokens were presented, half of the time belonging to the same category (dental or retroflex) and half to different categories. Participants had to report whether the two tokens belonged to the same or different categories. There were 80 trials, half of which contained tokens from normal trials and half from probe trials.

Singleton identification and discrimination were also administered using tokens from an untrained voice (recordings of the second Hindi speaker), in otherwise identical structure and procedure. All three tests with the trained voice were completed prior to administering the two tests with the untrained voice. No feedback was provided in any of the testing tasks. Trial presentation order was individually randomized for each participant. In all tasks, if no response was registered within 3 s after the last syllable was presented, the program marked an incorrect response and moved on to the next trial. The entire testing session lasted approximately 20 min.

**Data analysis.** All analyses reported below employed generalized mixed-effects logistic regression models for binomial distributions (Dixon, 2008), via a logit transformation (Jaeger, 2008), with participants and tokens as random factors (Baayen, Davidson, & Bates, 2003; Quené & van den Bergh, 2008), fitted with restricted maximum-likelihood estimation using package lme4 (Bates & Sarkar, 2007) in R (R Development Core Team, 2007). Although standard *t* tests as well as nonparametric tests (Mann-Whitney *U* for group comparisons and Wilcoxon signed ranks for repeated measures) of accuracy were conducted and produced similar patterns of significance, we report GLMM results as they are (a) based on modeling of unbounded log odds for individual response types, rather than aggregate proportions of accuracy bounded within [0,1]; (b) robust to deviations from normality; and (c) insensitive to potential bias, as response odds are contingent on trial type, obviating the need to take explicit account of false alarms by means of signal detection methods such as $d'$. Effect sizes ($\hat{\beta}$) are estimated log odds regression coefficients (zero corresponding to no effect), reported below as absolute values.

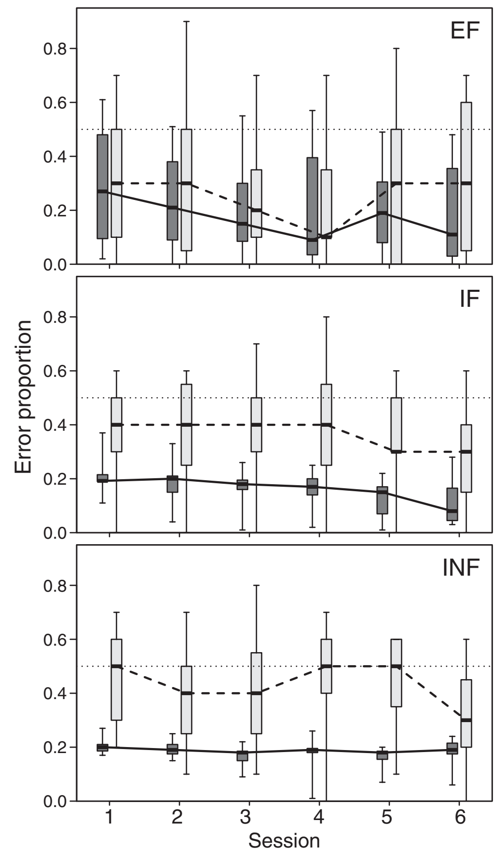


*Figure 1.* Group performance (error proportion) in each training session (1–6) for normal trials (dark gray bars; connected by solid lines) and probe trials (light gray bars; connected by dashed lines). Bars enclose the middle 50% of individual participant error proportions. The median is marked with a thick line. Error bars extend to the full range of the data. The dotted line indicates chance performance (50%). *Top* (EF): Explicit training with feedback (Experiment 1). Displayed proportions represent error in phonetic identification. *Middle* (IF): Implicit training with feedback (Experiment 2). *Bottom* (INF): Implicit training with no feedback (Experiment 3). In the middle and bottom panels, displayed proportions represent error in intensity difference identification for the normal trials, and proportion of retroflex responses for the (equal-intensity) probe trials.

## Results and Discussion

**Training.** Figure 1 (top) shows performance during training for normal and probe trials separately. Training performance was analyzed with a model including fixed effects of session and trial type as well as their interaction, in addition to a random effect of participants. In R notation, this was specified as

```
response~session * trialtype +
                    (1 + session|subject),
```

with two types of responses ("correct" or "incorrect") regressed onto six sessions (specified numerically, 1. . .6) and two trial types ("normal" and "probe"). A random slope of session was included (as shown in the parenthesized factor) to model variation in learning rates among participants, as this led to improved model fit (Baayen, 2008). A significant interaction between session and trial type ($\hat{\beta}$ = .11, $z$ = 2.23, $p$ = .026) led to separate analyses for the two trial types: For normal trials, the linear effect of session was significant ($\hat{\beta}$ = .18, $z$ = 3.26, $p$ = .001), whereas for probe trials it was not ($\hat{\beta}$ = .06, $z$ = 1.07, $p$ = .287). Therefore, it seems that participants improved during training, but only on tokens for which they received feedback.

**Testing.** In this and subsequent experiments, identification response polarity was adjusted as follows: In each identification task, if a participant's total error rate exceeded 50% (including normal and probe tokens considered together), then each response was reversed, leading to a total score less than 50%. The rationale for this transformation was that error rates exceeding 50% might reflect adequate ability of a participant to identify the sounds correctly while systematically applying the categorization labels incorrectly. The transformation was highly conservative in that it applied most often in the pre-training tests, where participants were highly uncertain about the novel labels (and indeed about the task itself), so that estimated learning effect sizes became smaller as a result of this transformation. This issue is taken up further in the General Discussion.

Singleton identification prior to training was examined with a model of the form

```
response~stimtype * speaker + (1|subject)
                              + (1|token),
```

with two types of responses ("retroflex" and "dental") regressed onto two types of stimuli ("retroflex" and "dental") and two speakers ("trained" and "novel"), also including random effects of both participants and tokens. Performance, as a main effect of stimulus type, was well above chance, as the odds of responding differentially to tokens of the two types were significant ($\hat{\beta} = 1.11$, $z = 10.17$, $p < .0005$), regardless of speaker (no interaction, $\hat{\beta} = .24$, $z = 1.57$, $p = .117$), suggesting that (a) Greek listeners were able to differentiate the sounds to some extent prior to training and (b) this ability was not dependent on some peculiarity of the (subsequently) trained speaker. Still, error rates were high (above 30% in all measures), so there was plenty of room for improvement with training.

Effects of training on singleton identification were examined with models of the form

$$\mathtt{response} \sim \mathtt{stimtype} * \mathtt{time} * \mathtt{trialtype} + (\mathtt{1}|\mathtt{subject}) + (\mathtt{1}|\mathtt{token}),$$

as an interaction of stimulus type by testing time ("naive" vs. "trained"), separately for the trained and novel speaker. The trial type factor ("normal" and "probe") was relevant for the trained speaker only; when interacting, analysis was broken down into separate tests. Similar models were used for the other tests: In pair identification, stimulus type and response type were "retroflex first" or "dental first"; in discrimination, stimulus type and response type were "same" or "different" instead of "retroflex" and "dental."

Table 1 and Figure 2 show participants' performance (error rates) before and after training in each test. For the trained speaker, significant improvement from pre- to post-test was found in pair identification ($\hat{\beta} = 1.01$, $z = 3.97$, $p < .0005$), singleton identification ($\hat{\beta} = .66$, $z = 3.00$, $p = .003$), and discrimination ($\hat{\beta} = .94$, $z = 3.82$, $p < .0005$). There was no (triple) interaction of this improvement with trial type for identification (pair or singleton; both $p$s $> .5$), but there was an interaction for discrimination ($\hat{\beta} = .77$, $z = 2.22$, $p = .026$), owing to a significant pre- to post-test improvement for normal trials ($\hat{\beta} = .95$, $z = 3.84$, $p < .0005$) but not for probe trials ($\hat{\beta} = .17$, $z = .72$, $p = .471$). There was no evidence for generalization of learning effects to the untrained voice, as there was no significant stimulus type by testing time interaction in either identification ($\hat{\beta} = .02$, $z = .11$, $p = .91$) or discrimination ($\hat{\beta} = .04$, $z = .24$, $p = .81$). (Data from one participant in untrained voice tasks were not available because of technical problems.)

In sum, the results indicate that (a) participants improved in identification and discrimination of tokens on which they were trained with feedback; (b) the improvement in identification extended to tokens to which participants were exposed but did not receive feedback, whereas discrimination for those tokens did not improve; and (c) there was no improvement in identification or discrimination of tokens spoken by the untrained voice.

At 600 trials total, the training procedure was rather brief, by the standards of nonnative phonetic training studies, which typically provide several thousand trials to each participant (e.g., 3,000 in Iverson et al., 2005; about 4,000 in Lively et al., 1993; upward of 8,000 in Pruitt et al., 2006). Still, this relatively minimal amount of training did result in reduction of mean singleton identification error rate by 15%–30% (compared with the typical findings of about 50% in long-term training studies; cf. e.g., Iverson et al., 2005, Figure 3—Initial position, trained talkers and words—and Figure 4). The aim of this experiment was not to achieve maximum asymptotic performance but to provide a baseline learning effect, especially in the most rapidly progressing initial stages, for comparison against different procedures. The achieved improvement was quite substantial, allowing ample opportunity to discern both smaller and larger significant learning effects in subsequent experiments.

Table 1
*Percentage of Incorrect Responses (Group Means and Standard Deviations) and a Signal-Detection Measure of Sensitivity (d′ Group Mean) in the Identification (ID) and Discrimination Tests in Experiment 1*

| | | Trained voice | | | Untrained voice | | |
|---|---|---|---|---|---|---|---|
| Task | Tokens | *M* | *SD* | *d′* | *M* | *SD* | *d′* |
| Before training | | | | | | | |
| Pair ID | trained | 35.16 | 18.77 | .76 | | | |
| | probes | 36.00 | 13.15 | .61 | | | |
| Singleton ID | trained | 39.33 | 12.02 | .73 | 42.50 | 6.22 | .40 |
| | probes | 34.67 | 10.02 | .88 | | | |
| Discrimination | trained | 44.83 | 9.70 | 1.14 | 48.13 | 6.68 | .73 |
| | probes | 42.50 | 9.64 | 1.01 | | | |
| After training | | | | | | | |
| Pair ID | trained | 24.67 | 20.76 | 1.33 | | | |
| | probes | 23.83 | 16.55 | 1.21 | | | |
| Singleton ID | trained | 32.13 | 19.75 | 1.43 | 42.29 | 7.66 | .43 |
| | probes | 29.87 | 11.43 | 1.25 | | | |
| Discrimination | trained | 34.67 | 17.39 | 1.66 | 48.13 | 8.49 | .88 |
| | probes | 40.33 | 12.99 | 1.24 | | | |

*Note.* The distinction between trained and probe tokens applies to the trained voice only; none of the tokens from the untrained voice were trained.

## Experiment 2: Implicit Training With Feedback

The results of Experiment 1 confirmed the effectiveness of explicit training with feedback. The purpose of the second experiment was to test whether learning can be achieved in the absence of awareness or intention related to the nonnative phonetic contrast. To achieve this, we employed a similar exposure and reward schedule as in Experiment 1, but the external feedback was provided on the basis of performance on an unrelated explicit task. Our manipulation consisted in systematically pairing the phonetic contrast with an intensity difference as the explicit task target. As in Experiment 1, in each trial, participants were presented with tokens from both categories and reinforcement was associated with the retroflexes. The key difference from Experiment 1 was that the reinforcement of the retroflexes now took place implicitly, in parallel with the explicit intensity-difference detection task. In this way, reinforcement signals that contribute to plasticity (Seitz & Watanabe, 2005) were indirectly but consistently associated with tokens from a fixed speech category.

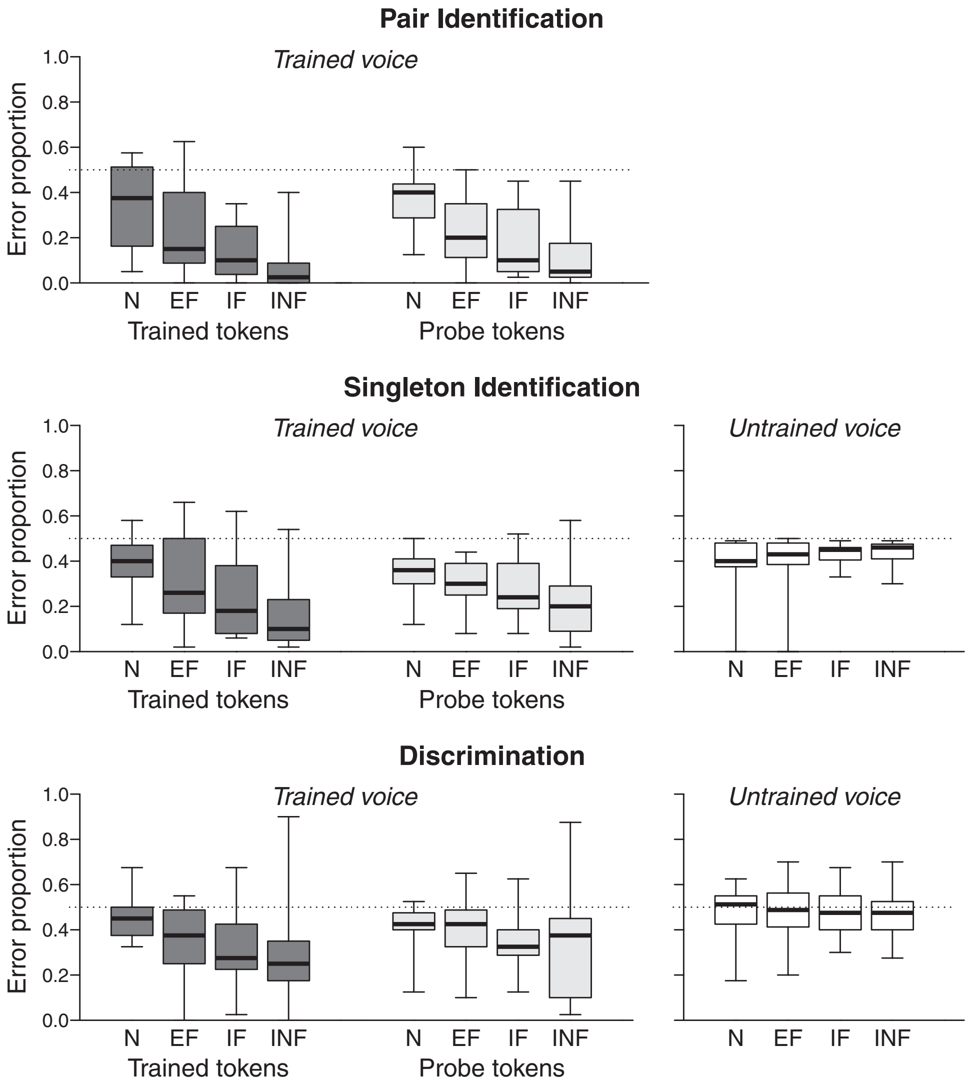


*Figure 2.* Group performance (error proportion) in each test for trained and probe tokens from the trained and untrained voice. Bars enclose the middle 50% of individual participant error proportions. The median is marked with a thick line. Error bars extend to the full range of the data. The dotted line indicates chance performance (50%). N = naive listeners (Experiment 1, before training); EF = explicit training with feedback (Experiment 1, after training); IF = implicit training with feedback (Experiment 2); INF = implicit training with no feedback (Experiment 3).

## Method

**Participants.** Seventeen adult Greek speakers (11 women, 6 men; 19–32 years old) participated in the experiment. One was excluded because of failure to understand testing instructions and another because of chance performance in the explicit task; thus, data from 15 participants are reported below. Most participants were students who either volunteered or were given course credit for their participation. None reported any hearing or speech impairments or previous experience with Hindi sounds.

**Stimuli.** The same Hindi syllables were used as in Experiment 1, that is, 10 tokens of [ta:] and 10 tokens of [ʈa:] spoken by each of two native Hindi speakers. The intensity of tokens by Speaker 1 (the one used for training in Experiment 1) was manipulated using Praat (Boersma, 2001) to create additional pairs of identical tokens differing in RMS amplitude by 0.5, 1, 2, 3, and 4 dB.

**Procedure.** As in Experiment 1, there was a training phase and a post-training test phase. Unlike Experiment 1, there was no pre-training test phase, because participants had to remain naive as to the purposes of the experiment. Therefore, there was only one testing phase, after the end of the last training session, in which participants were informed about the existence of two phonetic categories and about the experimental manipulation. Scheduling, equipment, supervision, and performance reporting were identical to those in Experiment 1.

***Training.*** During training, recordings from only one of the two Hindi speakers were used, the same one for all participants, as in Experiment 1. In each trial, participants heard two pairs of Hindi syllables, one pair with two identical tokens beginning with a dental consonant and one pair with two identical tokens beginning with a retroflex consonant. Stimulus timing and counterbalancing were the same as in Experiment 1. The task was to listen carefully to both pairs and identify the pair differing in intensity by pressing the corresponding key. Each session lasted approximately 10 min and again consisted of 100 normal trials and 10 probe trials.

*Normal trials.* In the normal trials, the syllables with dental consonants were always identical tokens, as in Experiment 1. In contrast, the syllables with retroflex consonants always differed in intensity by an adaptively varied amount between 0.5 and 4 dB. Therefore, unbeknownst to the participants, correct identification of the pair differing in intensity was equivalent to correct identification of the retroflex pair. The intensity difference was set at the maximum difference of 4 dB in the beginning of each session and was varied in 1 dB steps (except that a minimum difference of 0.5 dB was presented instead of 0 dB) in a 3-down 1-up schedule, to ensure adequate attention, motivation, and at least 75% correct performance during training. In these trials, participants received immediate feedback after each response, as in Experiment 1 (but here feedback was explicitly related to the intensity discrimination task).

*Probe trials.* In the second half of each session, two probe trials were randomly interspersed within each block of 10 normal trials. There was no intensity difference in probe trials; therefore, there was no correct response with respect to the explicit task. As in Experiment 1, no performance feedback was given in probe trials, and different tokens were used, not heard during normal trials. Therefore, participants never received feedback for their performance on these particular tokens.

***Testing.*** The testing procedure was identical to that in Experiment 1.

## Results and Discussion

**Training.** Figure 1 (middle) shows performance during training. Error proportion for normal trials concerns misidentification of the intensity difference. Here, the linear effect of session was significant ($\hat{\beta}$ = .18, $z$ = 4.18, $p$ < .0005). As there was no intensity difference within pairs in the probe trials, there were no correct responses in terms of the explicit task, so the two types of pairs should have been selected equally often. However, if participants were influenced (whether implicitly or explicitly) by the consistent pairing of retroflex with an intensity difference, then a learning effect during training might be evident as an increase in tendency to select the retroflex pair in probe trials. Indeed, the linear effect of session was significant for these trials as well ($\hat{\beta}$ = .11, $z$ = 2.14, $p$ = .032). Therefore, participants evidently used the phonetic distinction in performing the task: (a) For probe trials, the phonetic difference was the only available cue. (b) Even for normal trials, the improving performance despite the adaptive nature of the intensity task indicates that recorded judgments were not based on intensity alone. Apparently, the systematic phonetic pairing allowed participants to extend their success on the explicit task into regions of subthreshold intensity differences.

Analysis of the intensity difference level achieved by participants throughout training, shown in Figure 3 (top), corroborates this interpretation. Intensity difference in dB, as a continuous dependent variable, was linearly regressed onto trial and session via GLMM with participants as a random factor. Concentrating on the second session of each day, to exclude settle-down trajectories from high initial starting levels, and considering normal trials only, there was a significant linear effect of trial ($\hat{\beta}$ = .005, $t$ = 4.63, $p$ < .0005) and of day ($\hat{\beta}$ = .45, $t$ = 14.44, $p$ < .0005), consistent with continuing improvement in explicit task performance within and across training days. The interaction failed to reach significance ($\hat{\beta}$ < .001, $t$ = 1.74, $p$ = .082), indicating relatively stable improvement rate over the 3-day training period.

**Testing.** Identification responses were transformed as in Experiment 1. Table 2 and Figure 2 show listeners' performance (error rates) in Experiment 2. Training effects were assessed in comparison of participants' performance to that of untrained listeners, that is, of Experiment 1 participants tested prior to training, using the same statistical model formulation, such that the testing time factor was now between participants.

Post-training performance on the trained voice in Experiment 2 was significantly better than pre-training performance in Experiment 1 (naive listeners) in pair identification ($\hat{\beta}$ = 2.37, $z$ = 8.12, $p$ < .0005), singleton identification ($\hat{\beta}$ = 1.53, $z$ = 6.57, $p$ < .0005), and discrimination ($\hat{\beta}$ = 1.41, $z$ = 5.56, $p$ < .0005). There was no interaction with trial type in pair identification ($\hat{\beta}$ = .35, $z$ = .88, $p$ = .381). However, there was an interaction in singleton identification ($\hat{\beta}$ = 1.05, $z$ = 3.24, $p$ = .001) and discrimination ($\hat{\beta}$ = .71, $z$ = 2.01, $p$ = .044), because the performance difference was much larger for normal tokens (singleton identification $\hat{\beta}$ = 1.56, $z$ = 6.57, $p$ < .0005; discrimination $\hat{\beta}$ = 1.41, $z$ = 5.56, $p$ < .0005) than for probe tokens (singleton identification $\hat{\beta}$ = .48, $z$ = 2.15, $p$ = .032; discrimination $\hat{\beta}$ = .69, $z$ = 2.77, $p$ = .006), even though the difference was significant in both trial types. With the untrained voice, there was a marginal difference in performance between participants in Experiment 2 and naive listeners in singleton identification ($\hat{\beta}$ = .30, $z$ = 1.91, $p$ = .056), which was in the opposite direction (i.e., Experiment 2 post-test performance was slightly worse than Experiment 1 pre-test performance). There was no difference in discrimination ($\hat{\beta}$ = .02, $z$ = .13, $p$ = .90).

In sum, it appears that implicit training resulted in incidental learning of the specific tokens that were associated with feedback during training and, to a lesser extent, of tokens presented during training for which no feedback was provided. There was no evidence of transfer of learning to the untrained voice.

Debriefing after training revealed that most participants made correct, partially correct, or incorrect hypotheses about the experimental manipulation after the second training session or later. For example, some participants reported that there were "many different sounds" and some of them always differed in intensity (a partially correct hypothesis); one participant reported that the sounds that differed in intensity were different tokens (an incorrect hypothesis). A few participants reported that they relied on their hypotheses rather than on perceived intensity for responding. However, most subjects claimed that they did not attempt to test their hypotheses deliberately during training.

Because of the explicit hypotheses, relating the sound tokens and the intensity difference task, which were made by most participants in this experiment, it is difficult to draw a clear distinction between explicit and implicit processes in learning through this paradigm. It is unclear whether participants were sensitive to some

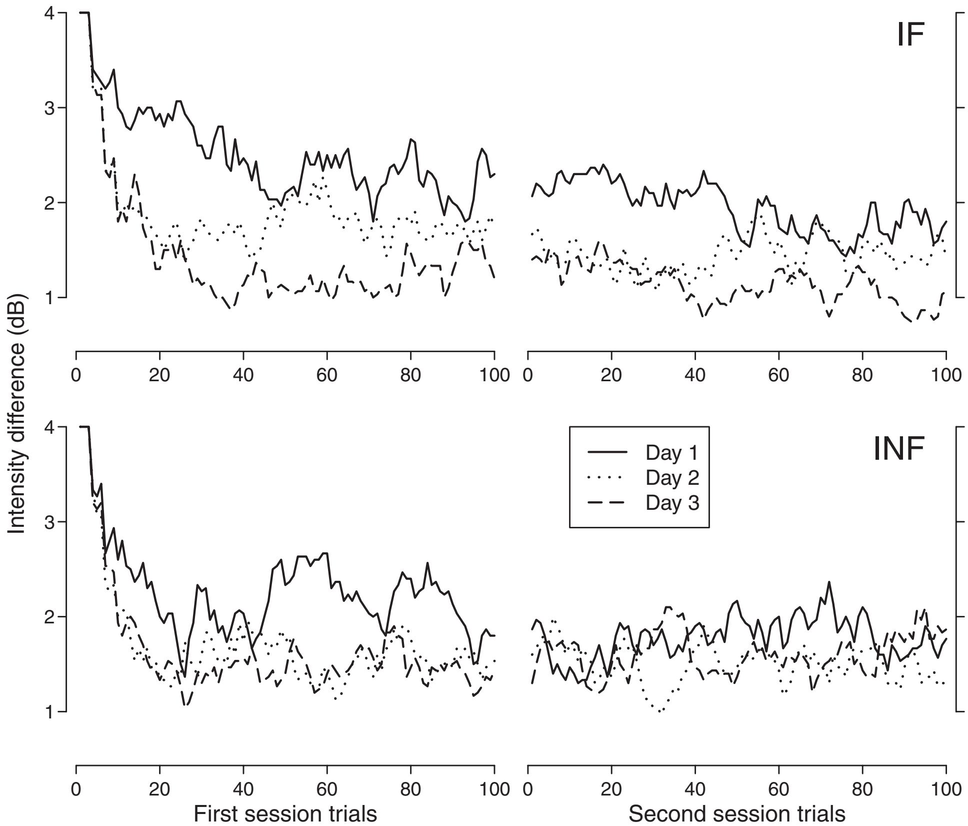


*Figure 3.* Group performance in the adaptive explicit task in Experiments 2 (IF, top) and 3 (INF, bottom). Lines depict mean intensity difference per trial in each training session, grouped by day (normal trials only). IF = implicit training with feedback; INF = implicit training with no feedback.

phonetic differences from the beginning of the training and then realized the pairing of some tokens with the explicit target, or whether they became increasingly more sensitive because of implicit learning, in which case implicit and explicit processes interacted. On the other hand, attention of participants to their explicit hypotheses might have hindered further learning, especially in case of incorrect or partially correct hypotheses. In any case, any observed learning for this group could be characterized as incidental to the extent that the exploitation of some perceived phonetic differences took place in the course of training without explicit instructions to do so.

We suspected that one possible reason for the explicit, hypothesis-driven attitude of the participants was the presence of feedback with regard to the explicit task. Feedback could have helped subjects confirm, reject, or modify their hypotheses, especially when the intensity difference was very low (1 dB or less) and they could not rely entirely on the explicit task to achieve optimal performance. Therefore, in the next experiment, we used an identical design dispensing with feedback altogether. We hypothesized that this manipulation would not impact the degree of task-irrelevant perceptual learning, because successful processing of a target is thought to be sufficient to trigger learning signals in the absence of external rewards (Seitz & Watanabe, 2005, 2009). At the same time, participants might be less likely to notice and form explicit hypotheses relating different sound tokens to differences in intensity.

## Experiment 3: Implicit Training Without Feedback

Experiment 3 was identical to Experiment 2, that is, participants were unaware of the phoneme distinctions and the retroflex-intensity difference manipulation, but this time they did not receive any external feedback during training. Probe trials did not differ from previous experiments but in the context of this experiment they did not stand out any more, because there was no feedback in normal trials either.

### Method

**Participants.** Seventeen adult Greek speakers (7 women, 10 men; 20–35 years old) participated in the experiment. One participant was excluded because of failure to understand the testing instructions and another because of technical problems preventing completion of training. Therefore, data from 15 participants are reported below. Most participants were students who either volunteered or were given course credit for their participation. None reported any hearing or speech impairments or previous experience with Hindi sounds.

**Stimuli and procedure.** All aspects of stimuli and procedure were identical to those in Experiment 2, with a single difference: In this experiment, no feedback was provided in any training trials (normal or probe).

Table 2
*Percentage of Incorrect Responses (Group Means and Standard Deviations) and a Signal-Detection Measure of Sensitivity (d′ Group Mean) in the Identification (ID) and Discrimination Tests in Experiment 2*

| | | Trained voice | | | Untrained voice | | |
|---|---|---|---|---|---|---|---|
| Task | Tokens | *M* | *SD* | *d′* | *M* | *SD* | *d′* |
| Pair ID | trained | 14.50 | 12.14 | 1.72 | | | |
| | probes | 17.17 | 14.91 | 1.57 | | | |
| Singleton ID | trained | 24.00 | 17.34 | 1.84 | 43.13 | 4.32 | .38 |
| | probes | 29.33 | 13.85 | 1.29 | | | |
| Discrimination | trained | 30.33 | 16.71 | 2.00 | 48.50 | 5.83 | .76 |
| | probes | 34.83 | 11.74 | 1.63 | | | |

*Note.* "Trained" and "probe" tokens refer to the trained voice only.

## Results and Discussion

Informal debriefing after training revealed that the vast majority of participants did not notice any patterns during training. A few participants reported that they formed a few hypotheses after the end of the first session but rejected them in subsequent sessions.

**Training.** Figure 1 (bottom) shows performance during training. The linear effect of session was very small and did not quite reach significance, either for normal trials ($\hat{\beta} = .04$, $z = 1.81$, $p = .071$) or for probe trials ($\hat{\beta} = .07$, $z = 1.49$, $p = .137$). There was, therefore, no strong evidence for learning on the explicit intensity task (consistent with the adaptive nature of the task) or for use of the dental-retroflex distinction in responding to the explicit task. However, it may be important that the proportion of retroflex pair selection in Session 6 probe trials was significantly higher than in Sessions 1, 4, and 5 ($\hat{\beta} = .12/.32/.55$, $z = 2.20/2.60/2.25$, $p = .028/.009/.024$) and likely reflects the learning of the phonetic contrast that was observed in the post-test performance.

Analysis of the intensity difference levels, shown in Figure 3 (bottom), showed a significant linear effect of day ($\hat{\beta} = .07$, $t = 2.30$, $p < .021$), no effect of trial ($\hat{\beta} = .001$, $t = 1.12$, $p = .261$) and no interaction ($\hat{\beta} < .001$, $t = 0.41$, $p = .680$). The effect of day was six times smaller than in Experiment 2, a large and significant difference (tested as interaction by experiment, $\hat{\beta} = .38$, $t = 8.55$, $p < .0005$), consistent with very limited between-day learning and no within-day learning along the explicit dimension.

**Testing.** Identification responses were transformed as in Experiment 1. Table 3 and Figure 2 show listeners' performance (error rates) in Experiment 3. Training effects were assessed in comparison of participants' performance to that of untrained listeners in Experiment 1, as was done for Experiment 2.

Similar to Experiment 2, post-training performance on the trained voice in Experiment 3 was significantly better than pre-training performance in Experiment 1 (naive listeners) in pair identification ($\hat{\beta} = 3.65$, $z = 10.54$, $p < .0005$), singleton identification ($\hat{\beta} = 2.55$, $z = 9.99$, $p < .0005$), and discrimination ($\hat{\beta} = 1.47$, $z = 5.84$, $p < .0005$). There was no interaction with trial type in pair identification ($\hat{\beta} = .73$, $z = 1.57$, $p = .117$). However, there was an interaction in singleton identification ($\hat{\beta} = 1.09$, $z = 3.10$, $p = .002$) and, marginally, in discrimination ($\hat{\beta} = .69$, $z = 1.95$, $p = .051$), because the performance difference was larger for normal tokens (singleton identification $\hat{\beta} = 2.63$, $z = 9.95$, $p < .0005$; discrimination $\hat{\beta} = 1.47$, $z = 5.85$, $p < .0005$) than for probe tokens (singleton identification $\hat{\beta} = 1.45$, $z = 5.95$, $p < .0005$; discrimination $\hat{\beta} = .78$, $z = 3.18$, $p = .001$), even though the difference was significant in both trial types. With the untrained voice, there was a significant difference in performance between participants in Experiment 2 and naive listeners in singleton identification ($\hat{\beta} = .34$, $z = 2.21$, $p = .027$), again in the opposite direction, and no difference in discrimination ($\hat{\beta} = .07$, $z = .43$, $p = .67$).

Overall, the results of Experiment 3 demonstrated significant improvement in the perception of the nonnative contrast for the trained voice, generalizing to some extent over "untrained" tokens (i.e., tokens not paired with the intensity difference) but not to the untrained voice.

**Cross-experiment comparisons.** Finally, we examined the relative effectiveness of implicit versus explicit training by directly comparing the post-test performance of trained participants between experiments (ignoring pre-training performance in Experiment 1). These comparisons were tested as interactions between experiment (i.e., participant group) and stimulus type, affecting the log odds of response types. Trial type was also included as factor, allowing for separation of effects on tokens from normal and probe trials. Participant and token random effects were included in the models, as above.

For the trained voice, there was a significant difference between explicit (Experiment 1) and implicit training with feedback (Experiment 2) in the post-training performance of participants in pair identification ($\hat{\beta} = 1.35$, $z = 4.46$, $p < .0005$) and singleton identification ($\hat{\beta} = .89$, $z = 3.71$, $p < .0005$) and a trend in discrimination ($\hat{\beta} = .43$, $z = 1.69$, $p = .091$). The difference in singleton identification interacted with trial type ($\hat{\beta} = .93$, $z = 2.78$, $p = .005$), such that it was significant for normal tokens ($\hat{\beta} = .90$, $z = 3.71$, $p < .0005$) but not for probe tokens ($\hat{\beta} = .05$, $z = .21$, $p = .836$). There was no interaction with trial type for pair identification ($\hat{\beta} = .51$, $z = 1.23$, $p = .220$) or discrimination ($\hat{\beta} = .08$, $z = .21$, $p = .831$).

There was also a significant difference between explicit (Experiment 1) and implicit training without feedback (Experiment 3) in pair identification ($\hat{\beta} = 2.64$, $z = 7.38$, $p < .0005$), singleton identification ($\hat{\beta} = 1.92$, $z = 7.33$, $p < .0005$), and discrimination

Table 3
*Percentage of Incorrect Responses (Group Means and Standard Deviations) and a Signal-Detection Measure of Sensitivity (d′ Group Mean) in the Identification (ID) and Discrimination Tests in Experiment 3*

| | | Trained voice | | | Untrained voice | | |
|---|---|---|---|---|---|---|---|
| Task | Tokens | *M* | *SD* | *d′* | *M* | *SD* | *d′* |
| Pair ID | trained | 8.33 | 12.94 | 2.18 | | | |
| | probes | 11.67 | 14.47 | 1.96 | | | |
| Singleton ID | trained | 16.27 | 16.09 | 2.44 | 43.47 | 5.64 | .37 |
| | probes | 21.20 | 15.62 | 2.05 | | | |
| Discrimination | trained | 28.67 | 21.61 | 2.34 | 47.00 | 6.44 | .85 |
| | probes | 33.17 | 22.35 | 2.05 | | | |

*Note.* "Trained" and "probe" tokens refer to the trained voice only.

($\hat{\beta}$ = .52, $z$ = 2.04, $p$ = .042). The difference in singleton identification again interacted with trial type ($\hat{\beta}$ = .97, $z$ = 2.66, $p$ = .008), such that it was larger for normal tokens ($\hat{\beta}$ = 1.97, $z$ = 7.32, $p < .0005$) than for probe tokens ($\hat{\beta}$ = .94, $z$ = 3.68, $p < .0005$) while significant for both. There was no significant interaction with trial type for pair identification (although a substantial trend approached significance, $\hat{\beta}$ = .89, $z$ = 1.86, $p$ = .063) or discrimination ($\hat{\beta}$ = .08, $z$ = .23, $p$ = .820).

Finally, there was a significant difference between implicit training with feedback (Experiment 2) and without feedback (Experiment 3) in post-training performance in pair identification ($\hat{\beta}$ = 1.29, $z$ = 3.33, $p$ = .001) and singleton identification ($\hat{\beta}$ = 1.01, $z$ = 3.71, $p < .0005$) but not in discrimination ($\hat{\beta}$ = .10, $z$ = .38, $p$ = .704). There was no interaction with trial type (all $p$s > .4).

For the untrained voice, there were significant differences between explicit training and both kinds of implicit training in singleton identification (Experiment 1 vs. 2 $\hat{\beta}$ = .31, $z$ = 2.01, $p$ = .045; Experiment 1 vs. 3 $\hat{\beta}$ = .36, $z$ = 2.31, $p$ = .021), reflecting the fact that Experiment 1 post-training performance in untrained speaker identification was very slightly yet reliably *better* than that of the subsequent experiments, a pattern opposite from that observed for the trained speaker. This distinct pattern was also evident as a three-way interaction between experiment, stimulus type, and speaker ("trained" vs. "novel"; pooling across trial types), which was significant in both cases (Experiment 1 vs. 2 $\hat{\beta}$ = .78, $z$ = 3.47, $p$ = .001; Experiment 1 vs. 3 $\hat{\beta}$ = 1.77, $z$ = 7.54, $p < .0005$). There was no significant difference between the two kinds of implicit training in singleton identification (Experiment 2 vs. 3 $\hat{\beta}$ = .05, $z$ = .29, $p$ = .77) and no significant difference in discrimination between any pair of experiment groups (all $p$s > .5).

Thus, the observed differences between experiments were systematic and specific to the trained speaker and to the explicit identification task. This pattern is inconsistent with potential attribution of the findings to nonspecific and uncontrolled differences among participant groups in initial phonetic, perceptual, or learning ability, because if there were such differences we might expect them to have similar consequences for both speakers and tasks. This observation is important, as the implicit nature of the training task in Experiments 2 and 3 precludes testing participants to ensure equivalent pre-training performance.

In sum, the improvement in category identification, which is the standard test of explicit performance, was greater for either type of implicit training (with or without feedback) than for standard explicit training, at least for normal tokens. Moreover, improvement in this critical task was greater for implicit training without feedback than for either type of training with feedback (explicit or implicit), for both normal and probe tokens. Therefore, the implicit experimental manipulation has proved successful in inducing robust learning effects in the absence of external feedback, in accord with the TIPL hypothesis.

## General Discussion

We have found that, following standard explicit training with feedback as well as implicit training with and without feedback, participants substantially and significantly outperformed untrained listeners in all identification and discrimination tasks using trained tokens from the trained voice and in most tasks using unrewarded (probe) tokens from the same voice. Moreover, and most important, implicit training not only produced robust learning but also led to superior learning than explicit training of the same amount. Our training regimen involved a relatively brief schedule, providing five to 10 times fewer training trials than typical studies of nonnative phonetic training and, thus, does not address asymptotic differences in learning between these methods. Nevertheless, this study demonstrates rapid adult learning of natural speech stimuli from a difficult nonnative phonetic contrast without task awareness, explicit instructions, known number of categories, or any trial-by-trial feedback with regard to the phonetic task. In both implicit training conditions, participants were evidently able to extract patterns from the training environment without a clear awareness of what must be learned and without knowing the number or the labels of the categories to which they were exposed. This experimental setting seems more closely related to ecologically realistic situations of phonetic category learning outside the lab. We believe that such implicit training procedures have significant implications for speech perception research and may help us understand the processes of acquisition of novel speech categories under natural settings, where external feedback is absent and implicit and explicit processes interact.

Our study was not intended to elaborate on the source of the difficulty in perceiving the nonnative contrast (cf. Best, 1995; Flege, 1995) or to consider the effects of learning on the representation of the acoustic-phonetic space (cf. Iverson et al., 2003, 2008). Instead, we have focused on components of training procedures that have been considered critical, or at least very useful, in related theorizing and standard practice. Based on theoretical developments regarding the effects of nonspecific, internally generated learning signals (Seitz & Watanabe, 2003, 2005, 2009), we have hypothesized and experimentally confirmed that it is possible to eliminate common training elements without reducing learning at the token level. If anything, the novel manipulations may have led to *increased* learning, compared with the standard procedure, although this is difficult to maintain with absolute certainty because of stimulus variability in intensity in the implicit procedures (Experiments 2 and 3) that was not present in the explicit procedure (Experiment 1). This variability, although orthogonal to the phonetic distinction, may have somehow enhanced learning, either by allowing sampling over a larger acoustic range or by acting as a redundant orienting cue, thereby limiting the extent to which we can ascribe learning to the implicit nature of the task per se. The nature of our manipulation precludes fully addressing this issue; therefore, it is left to future replications to provide further support for our hypothesis.

### Robustness of Data Analysis

To maximize the reliability of our conclusions we have applied the most conservative approach possible, applying multiple types of analysis subject to distinct conditions and limitations. Thus, we are confident in the reality of the basic finding, specifically, that implicit learning effects are robust and maximum in the no-feedback condition.

Moreover, we have applied a highly conservative approach to the treatment of identification responses, which are arguably the most important indicator of the explicit expression of learning effects. Because several participants exhibited very high error rates in the original baseline condition (above 50%), we hypothesized that they might be performing the desired phonetic distinction successfully while misapplying the novel labels. If that interpretation were correct, treating their high error rates as truly reflecting their identification ability might artificially inflate our training effect estimates, because training might have the effect of learning the labels rather than learning the phonetic distinction. Therefore, in each identification test with an error rate exceeding 50%, response labels were reversed prior to the analysis.

Table 4 shows the number of participants in each group meeting this criterion for one or both testing speakers (voices) in the singleton identification test. The data do not support the incorrect labeling assumption, because (a) the number of participants meeting the criterion for both voices is extremely small, suggesting uncertainty rather than consistent mislabeling, and (b) the number of reversals closely parallels the overall relative identification performance among groups. Specifically, average performance seems to vary around chance performance for untrained listeners in both voices and for all listeners in the untrained voice, with half (seven or eight) of the 15 participants in each group meeting the reversal criterion. In contrast, very few sessions in trained groups met the reversal criterion, fewest (zero) in the implicit-no feedback group (Experiment 3). Therefore, our reversal manipulation has clearly worked against the learning effect in every group (by inflating the average performance of seven untrained listeners) and against the superiority of implicit training without feedback (by inflating the average performance of three explicitly trained listeners). When analyses are conducted without response reversal all learning effects come out larger, exhibiting even stronger and more consistent statistical significance (whereas the inverse effects for the untrained voice disappear). The fact that the critical findings survive this correction for any possibility of mislabeling is further testimony to the robustness of the reported effects.

## External Feedback Versus Internally Generated Signals

The finding of more robust learning after implicit training without external feedback may seem at first blush not only counterintuitive but also inconsistent with previous demonstrations that supervised phonetic learning is superior to unsupervised (Goudbeek, Cutler, & Smits, 2008; McCandliss et al., 2002; Tricomi et al., 2006). Here we define external feedback to be explicit, task-related performance information conveyed to experimental participants via elements of the task procedure. In this sense, external feedback may be rare in natural settings. Nonetheless, in training adults with difficult nonnative phonetic contrasts, external feedback has been shown to be highly beneficial for listeners, allowing them to overcome interference arising from their native phonetic system. Thus, external feedback provided to listeners on a trial-by-trial basis seems to be tacitly acknowledged as a critical component of training, guiding and shaping modification of responses (Logan & Pruitt, 1995) and presumably guiding attention to the relevant acoustic/phonetic properties that consistently differentiate the target phonemes (Logan et al., 1991). Similarly, in the case of visual perceptual learning, external feedback has typically been shown to be helpful and sometimes required (Herzog & Fahle, 1997; Petrov, Lu, & Dosher, 2006; Seitz et al., 2006).

However, it may be incorrect to consider our "no feedback" condition (i.e., Experiment 3) as entirely lacking any form of feedback. Rather, it might best be considered as a situation in which no *external* performance feedback was explicitly provided by the experimenters. The manageability of the explicit task allowed ample opportunity for participants to generate their own *internal* feedback upon successful detection of the intensity differences (Herzog & Fahle, 1998; Seitz & Watanabe, 2005). Thus, internal feedback refers to endogenously provided information that reflects the participant's own judgments of the stimuli and of their performance. The distinction between external and internal feedback may seem somewhat artificial insofar as natural situations may provide relevant information to guide learning that cannot easily be classified as external in the intended sense. For example, word knowledge can shape phonetic perception to contrast-dependent and talker-dependent adjustment of acoustic processing of ambiguous (native-language) speech sounds (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Norris, McQueen, & Cutler, 2003). Although (internal) word knowledge is the guiding internal information, it cannot be considered feedback, because it is taken into account for processing itself; in natural situations of communication it would be the usage context that provides the (external) feedback for this sort of perceptual adjustment.

More generally, from the point of view of availability of guiding reinforcement signals, the question is not whether no feedback is superior to feedback but, rather, which type of performance information is most effective in which type of situation and schedule of reinforcement. In this view, both external and internal feedback were available in Experiment 2 versus only internal feedback in Experiment 3. Results showed that the former situation was not significantly more effective than standard explicit training, whereas the latter was. The question, then, is why was the combination of external and internal feedback not more beneficial than internal feedback alone? One possibility is that external visual feedback (as provided in Experiments 1 and 2) may have led to explicit, hypothesis-driven strategies relating sound tokens with corresponding responses. In contrast, such strategies would not be afforded in the absence of external feedback (as in Experiment 3). Although it is not possible to draw a clear distinction between implicit and explicit processes in any of the three conditions, it seems reasonable to assume that external feedback provides cues lending themselves to hypothesis testing, thus inducing a more explicit attitude during training. For example, participants in Experiment 2 could easily generate and test hypotheses based on the immediate feedback received after each response, without relying on the intensity difference task, especially when it became too difficult because of the adaptive procedure. This may explain their high performance during training: Despite intensity differences progressively dropping to indiscriminable levels (at or below 1 dB), via an adaptive procedure targeting 75% correct performance,

Table 4
*Number of Participants With More Than 50% Incorrect Singleton ID Responses*

| Condition | Trained voice | Untrained voice | Both |
|---|---|---|---|
| Experiment 1, pre | 7 | 8 | 2 |
| Experiment 1, post | 3 | 8 | 1 |
| Experiment 2 | 2 | 8 | 1 |
| Experiment 3 | 0 | 7 | 0 |

significant improvement was observed, reaching a mean performance of 89% correct in the final session.

In contrast, participants in Experiment 3 were forced to rely on the intensity difference task itself because there was no other information to explicitly take into consideration. This interpretation is consistent with their performance on the explicit task, which remained around 80% throughout training for the normal trials, showing no sign of improvement. For these participants, then, learning may have proceeded unhindered by incorrect hypotheses and unbiased by excessive focus on salient but suboptimal cues or on incomplete sets of cues. Internally generated reward signals elicited upon processing of the target stimuli could apply simultaneously on all aspects of the incoming signals. Cumulative strengthening of individual stimulus representations might therefore allow consistent patterns of cues to eventually prevail, canceling out inconsistent and irrelevant cues, however salient. This procedure may thus be optimal, as well as more ecologically valid, for learning in complex domains such as phonetic perception, where external feedback is not normally available and the nature of the stimuli is too complex for explicit, hypothesis-testing approaches.

Previous research of task-irrelevant perceptual learning (TIPL) has typically demonstrated that internal feedback alone can drive perceptual learning (for reviews, see Seitz & Watanabe, 2005, 2009). Although previous studies of TIPL have not directly explored how external feedback interacts with internal feedback to produce learning, some studies in the visual domain have demonstrated that attention can interfere with learning. In one study, more learning was found for weak, parathreshold, stimulus features, which were thought to be at stimulus strengths below the threshold of attention, but not for suprathreshold stimuli (Tsushima, Seitz, & Watanabe, 2008). In a related study, Choi, Seitz, and Watanabe (2009) found that task-irrelevant learning was inhibited when exogenous attention was directed to the trained feature. In the auditory modality, Wade and Holt (2005) found that an implicit procedure resulted in better perceptual learning of difficult nonspeech categories than explicit attempts at categorization, whereas there was no difference for the easier, linearly separable categories. Together, these studies suggest that subjects' attentional strategies can inhibit learning, especially for stimulus features that can be distracting to the subjects' main task. In the context of the present study, it may be the case that subjects' attention to the Hindi sounds may have impaired learning for those sounds.

Clearly, further research will be necessary in order to understand the role of feedback in the context of explicit and implicit training procedures. Even if our interpretation about external versus internal feedback is on the right track, many issues will need to be clarified. For example, the difference between Experiment 2 and Experiment 3 in the pattern of performance during training, if reliably replicated, might indicate that external feedback is indeed a powerful facilitator of learning in explicit situations, because it (might have) helped Experiment 2 participants improve on intensity discrimination itself. This possibility was not tested in our study, because psychoacoustic performance was not within our primary focus. Moreover, the importance of feedback has been clearly demonstrated in studies of explicit training, even though its precise role in the process remains unclear (Tricomi et al., 2006). More comprehensive approaches will need to identify, quantify, and disentangle separable contributions of procedure and stimulus components on distinct learning effects.

## Passive Learning

An alternative account for the learning effects may be developed on the assumption that simple exposure to the stimulus domain is sufficient for learning, regardless of reward contingencies. Such an approach would be consistent with the literature on implicit statistical learning. In this framework, learning can be viewed as a byproduct of powerful statistical mechanisms assumed to be triggered upon exposure to structured stimuli (e.g., Jiang & Chun, 2001; Perruchet & Pacton, 2006). Certain aspects of language acquisition can be accounted for by postulating such mechanisms (e.g., Saffran et al., 2006). Within the context of phonetic learning, in particular, it has been suggested that humans track the frequency with which phonetic exemplars occur in the speech input and organize their perceptual space accordingly (e.g.,Lacerda, 1995; Maye & Gerken, 2000; Maye et al., 2008; Maye et al., 2002; Pierrehumbert, 2003). In support of this hypothesis, studies with infants (Maye et al., 2002) and adults (Maye & Gerken, 2000) have demonstrated that exposure to different distributions (unimodal or bimodal) of the same stimuli affect listeners' subsequent discrimination performance.

A key limitation of the statistical learning approach is that it cannot explain the differences in learning between our three experiments. In each case, the statistics of stimulus presentation were the same. Although it has been shown that attention can shape what is learned through statistical learning (Turk-Browne, Jungé, & Scholl, 2005), such results indicate that more statistical learning will be found for attended stimuli, which seems to be the opposite of what we observed in the present study. However, the statistical learning account is consistent with the learning evident during training in Experiment 2, where participants apparently learned to apply the retroflex-dental distinction to the intensity difference task in the probe trials. Thus, although some of our results may be related to principles of statistical learning, this framework seems insufficient to account for important aspects of our findings.

Another apparent limitation of this account is that statistical computations per se cannot explain how the cognitive system can overcome the perceived similarity between distinct nonnative speech sounds, which may be functional within a native linguistic environment but dysfunctional in a different one. As noted in the introduction, under a Hebbian account, prolonged exposure to stimuli that are initially indiscriminable may further entrench the listeners' tendency to perceive them as belonging to the same category. This view may help explain why listeners sometimes fail to learn to distinguish contrasting phonemes, even after years of natural experience (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004) or intensive laboratory training (Bradlow, 2008; McCandliss et al., 2002). McCandliss et al. (2002) addressed this issue by exaggerating the natural differences of the stimuli so as to render them initially discriminable, gradually reducing the exaggeration in the course of training. Using this manipulation, they found learning after exposure to the target stimuli in the absence of external feedback. In contrast, learning was possible only with trial-by-trial feedback when stimulus differences remained at natural levels (hence indiscriminable) throughout training.

One cannot dismiss the possibility that longer periods of exposure might be sufficient for learning, even without discriminability or feedback. Still, the evidence suggests that feedback is beneficial even when unnecessary, as in comparison between the adaptive conditions of McCandliss et al. (2002), where more rapid improvement was found with than without feedback. Other studies have suggested that feedback-based mechanisms may be active even during first language acquisition. For example, Goldstein et al. (2003) found that mothers' contingent positive responses to infant vocalizations were associated with more mature vocal behavior of the infants. These findings suggest that perceptual systems will take advantage of reinforcements whenever available (McCandliss et al., 2002; Vallabha & McClelland, 2007).

## Category Learning

It is possible that some form of feedback is necessary or useful for some kinds of learning but not others. To make further progress on this issue, we will have to carefully analyze what exactly is learned in each condition. In our current studies, for example, it may seem that participants in the implicit training conditions learned a novel phonetic category distinction. However, there are two issues that need to be resolved before such a conclusion can be drawn: one concerning the notion of "category" distinctions and the other regarding "phonetic" distinctions. With respect to the former, it may be noted that there was no specific categorization requirement in the procedure, as the main explicit task concerned detection of an intensity difference. One may argue that categorization of the speech tokens as retroflex, as opposed to dental, could support the desired detection. Independent evidence for increased categorization performance as a result of implicit task relevance has been provided by Wade and Holt (2005), who embedded complex nonspeech sounds in a video game such that correct categorization of the sounds would facilitate higher performance in the game. They found that listeners learned to categorize the various sounds, to some extent, without instructions or awareness of doing so. More recently, Lim and Holt (2011) used the same implicit procedure to improve categorization performance of synthesized /r/–/l/ stimuli by Japanese adults, confirming the applicability of this approach to speech sounds.

On the other hand, it may appear somewhat more difficult in our case to argue that listeners learned to perceive the phonetic contrast categorically as a direct outcome of training. Not only were no labels provided in the implicit tasks but also there was no need for such categorization. Listeners might alternatively become more attuned to acoustic cues signaling retroflex consonants by increasing their perceptual sensitivity. This description would seem more in line with other recent demonstrations of task-irrelevant perceptual learning in the visual (Seitz et al., 2005; Watanabe et al., 2001) and auditory (Seitz et al., 2010) domain. In those studies, participants were presented with subthreshold stimuli along an unattended psychophysical dimension while performing an unrelated explicit task. After prolonged training, perceptual sensitivity to the unattended dimension was demonstrated, even though the stimuli presented during training remained subthreshold throughout the procedure. No categorization was necessary in that case; rather, a refinement of low-level perceptual representations seems to account for the observed learning effects.

Could such a basic representational refinement account for our findings? On the one hand, it seems necessary to postulate an increased sensitivity to the acoustic cues signaling the retroflex-dental distinction in the context of at least some speech syllables, because stimuli that were initially difficult or impossible to discriminate became easier to discriminate after training. Here, the importance of acoustic context must be emphasized, because it is known that sensitivity to acoustic features can vary greatly depending on whether a cue is presented in isolation or as part of a complex stimulus, particularly speech. For example, native Japanese speakers are unimpaired in distinguishing isolated formant transitions that, in the context of additional formants making up a speech syllable, would differentiate an /r/ from an /l/, a distinction these same speakers would be unable to perform (Miyawaki et al., 1975). Our manipulation has clearly enhanced the perception of acoustic cues supporting the distinction between Hindi /t/ and /ʈ/ for our Greek listeners. On the other hand, listeners were unaware of the existence of separate categories and of their labels. Therefore, their performance in the post-training identification task cannot be explained by the training task alone, because the identification test required them to apply specific labels to the two categories, something they could not have learned during implicit training yet all could perform to some extent (especially in Experiment 3).

A low-level perceptual refinement is also consistent with relevant findings from the animal literature. Passive stimulation has been successfully used as a control condition, not leading to expansion of primary cortical representation, contrasted with rewarded (and presumably attended) discrimination, which resulted in expanded and more finely tuned cortical representations accompanying increased perceptual sensitivity (Recanzone, Schreiner, & Merzenich, 1993). Yet passive listening suffices to produce expanded primary cortical representations in the context of temporally coincident stimulation of neuromodulatory nuclei involved in stimulus assessment and reward (Bao, Chan, & Merzenich, 2001; Kilgard & Merzenich, 1998). To the extent that such neuromodulatory signals may function as the alerting-reinforcement signals thought to underlie task-irrelevant perceptual learning (Seitz & Watanabe, 2005), cortical expansion may be a reasonable hypothesis regarding the physiological learning outcome of our implicit training manipulation. Taking into account that expansion of cortical representations is associated with perceptual refinement and discrimination (Recanzone et al., 1993) rather than categorization (Guenther et al., 2004), considered together with previous results of task-irrelevant perceptual learning, it seems more likely that Experiment 3 implicit training increased the sensitivity of our listeners to the critical acoustic cues necessary for the perception of retroflex consonants.

In sum, one possibility is that listeners acquired two acoustic categories during implicit training, to which they immediately applied the two labels provided in the identification task after training. A second possibility is that listeners acquired an increased perceptual sensitivity to the relevant acoustic cues, which they rapidly applied to the required categorization along with the novel labels. Although the latter option appears more in line with previous studies focusing on psychophysical sensitivity, it remains to be determined what exactly is learned during implicit training and how it is subsequently brought to bear on the post-training tests.

### Phonetic or Stimulus-Specific Learning?

The second issue in need of further inquiry concerns the purported phonetic nature of learning and performance in our training and testing tasks. A persistent finding across all three experiments reported here was the lack of generalization of learning to the novel speaker, as trained participants did not differ from naive listeners when tested with the untrained voice. Although this finding was not unexpected, as past research has shown that variability in the training set is crucial for generalization (Bradlow, 2008), it raises further questions as to what was actually learned. Roughly, the issue is this: If participants learned categories, what kind of categories were they, including tokens from one voice but not from another? If, instead, participants refined their acoustic representations, what feature was represented, and why did it fail to support categorization in a different acoustic context? Either way, one may argue that listeners formed perceptual representations of a specific set of stimuli (or a category encompassing them) rather than of a more general phonetic nature.

A straightforward interpretation of the testing results is that listeners were able to develop distinct categories that enabled them to successfully differentiate the two phonemes in the training set. Rather than attributing lack of generalization to the development of *incomplete* representations, perhaps the issue is that they developed *overspecified* ones. As has long been pointed out, listeners encode specific instances (exemplars) and not abstract, context-independent categories, retaining detailed information about talkers' voices (Pisoni & Lively, 1995). In this view, poor performance with the novel speaker may indicate that the newly developed categories included phonetically irrelevant talker-specific details. Therefore, when exposed to a different speaker, listeners were unable to ignore factors related to the speaker's voice and attend to the critical dimensions that differentiate the target phonemes across speakers.

Logan et al. (1991) have argued that in order to learn which acoustic cues are critical for categorization, listeners must be exposed to a broad range of speech tokens, produced by many different speakers, in diverse phonetic environments. In their study they trained Japanese listeners with natural recordings of words containing /r/ and /l/ in various positions (word-initial, word-final, intervocalic, singleton or in a cluster) produced by five different speakers. They found not only successful learning of the training set but also transfer to novel tokens and speakers across phonetic contexts. Lively et al. (1993) further demonstrated that talker variability is necessary for generalization: When listeners were trained with tokens produced by a single speaker, they did not show transfer of learning to new speakers even when trained with a broad range of stimuli in various contexts. A schedule of gradually increasing the number of training voices and vowel contexts throughout training was more recently used by Protopapas and Calhoun (2000) and Pruitt et al. (2006) to support simultaneous generalization over multiple dimensions, emphasizing the point that learning generalizes only over dimensions in which variability is provided during training. It remains to be seen in future research whether implicit training with increased variability leads to generalization in the same way as explicit training.

In our experiments, we used one voice and one phonetic environment in order to simplify the training set and focus on our main goal, which was the implicit nature of learning. This has demonstrably resulted in lack of generalization to a novel voice, as expected. Presumably, had we tested that possibility, we would also have found lack of generalization to novel phonetic contexts, such as syllables with different vowels following the dental/retroflex consonants, or the same consonants in syllabic coda position, even when spoken by the same (trained) voice. Therefore, it would be incorrect to claim that we have demonstrated *phonetic* learning, insofar as phonetic distinction is theoretically assumed to be abstract and context-independent (but see Pisoni & Lively, 1995, and Port, 2007, for further discussion of this assumption). Regardless of the degree of presumed abstraction, at the very least a phonetic distinction must support adequate performance with arbitrary speakers and phonetic contexts, if it is to function distinctively in verbal communication.

If our participants have not learned a phonetic distinction but might have learned a phonetic distinction given more variability in the data set, what is it they have learned? A first approach to this question may consider the results of testing with the probe tokens, which were never rewarded during training, either explicitly (in Experiments 1 and 2) or implicitly (in Experiment 3, because there was no intensity difference in them). There was evidence for transfer of learning to those tokens in each experiment, even though it was weaker than learning of the trained tokens. Because there were fewer probe trials than normal trials in training, we cannot dismiss the possibility that direct learning of probe tokens took place during training, due to passive exposure, and that the differences in testing emerged as a result of lower frequency rather than imperfect post-training transfer. However, if we consider correct identification of probe trials as an indication of generalization over tokens, then we can interpret our pattern of findings as consistent with the phonetic training literature, as follows: There was variability in trained tokens, supporting transfer of learning to novel tokens; there was no variability in trained voices, preventing transfer of learning to novel voices.

This interpretation is consistent with a view in which listeners encode specific details about the training tokens, and not of any purported decontextualized constituent features. Repeated instantiations of these tokens result in strengthened and refined representations of their acoustic properties, supporting subsequent categorization and discrimination of these particular tokens. Alternation of a multitude of tokens introduces variability along the corresponding dimensions. Repeated processing of variable tokens causes their common properties to stand out by eventually canceling out inconsistent differences, as in connectionist learning gradually adjusting weights to account for an entire training set simultaneously. If reinforced tokens happen to exhibit a common property, such as belonging to distinct phonetic categories differentially rewarded, then the outcome of this learning will be a representation of the acoustic properties defining this phonetic distinction. In a related vein, Kraljic and Samuel (2007) found talker-specific perceptual adjustments of categorical boundaries for native fricative consonants but not for stop consonants. Consistent with learning of speech details (and the overspecified representation hypothesis posited above), they speculated that when acoustic cues provide talker identify information, as in fricatives, then perceptual adjustments are talker-specific, whereas general adjustments are observed in the absence of identification cues. In this view, there is no real abstraction; rather, generalization is the result of averaging a multitude of exemplars spanning the functional range of a target category. Use of a perceptual distinction in the context of language communication is what makes the learned perceptual category phonetic, rather than any specific property of the sounds or of the training procedure.

## Task-Irrelevant Perceptual Learning

The present result fits nicely within the framework of TIPL, where reinforcement from an explicit, attended task is thought to interact with temporally coincident stimulus-driven activity to produce enhanced representations of attended and unattended stimuli alike (Seitz & Watanabe, 2005). Seitz and Watanabe (2003) have postulated that these reinforcement signals are released as a result of successful recognition of the task targets, such as the intensity cues of the present study. In both the rapid serial visual presentation tasks used in previous research and the intensity discrimination task used in the present study, participants are cognizant of their own performance and thus become self-reinforcing. In other studies, similar learning effects have been observed through external reinforcers in the absence of any explicit task (Frankó, Seitz, & Vogels, 2010; Seitz, Kim, & Watanabe, 2009), ascribed to stimulus-reward contingencies. In this framework, it is important that the task-irrelevant stimuli are presented in conjunction with high task accuracy, to ensure high stimulus-reward contingencies. However, it is also important that tasks are challenging, to ensure that correct performance remains highly rewarding. This hypothesized trade-off between task difficulty and task accuracy in TIPL has not yet been directly addressed empirically and will require further research for clarification and quantification.

As noted above, the present study differs from previous research using TIPL in that it concerns categorical learning. Most previous research of TIPL has been framed as an enhancement of processing low-level features that were paired with task targets or rewards in an explicit task. Thus, one may reasonably question the extent to which our results reflect category learning per se rather than enhancement of processing for a specific set of stimuli, namely, the retroflex tokens. In principle, the TIPL model posits that stimulus-driven activity can be reinforced at any (and possibly all) levels of perceptual processing. Thus, categorical activation and other higher level features may also be reinforced. Consistent with this idea, TIPL has also been found for visual contours (Rosenthal & Humphreys, 2010) and natural scenes (Lin, Pype, Murray, & Boynton, 2010), in which learning can no longer be explained simply as an enhancement of low-level visual features. In the present study, transfer of learning to untrained tokens suggests learning likely involving more than simply an enhancement of basic auditory features. Further research needs to address the extent to which learning concerns categories and to determine the reinforcement conditions best leading to the formation of categorical boundaries between stimuli. In particular, rewarding both dental and retroflex sounds (albeit differentially) might lead to even more efficient training than the experimental conditions reported here.

## Conclusion and Further Directions

In conclusion, we have demonstrated implicit learning of speech stimuli from a difficult nonnative phonetic contrast, over the small extent of variability provided during training (over tokens), eschewing aspects of phonetic training previously considered crucial or at least important: task awareness, focused attention and intention to learn the phonetic distinction, and trial-by-trial performance feedback. Consistent with predictions based on the unified model of task-relevant and task-irrelevant learning (Seitz & Watanabe, 2003, 2005, 2009), we found that implicitly pairing the target phonetic distinction with an irrelevant acoustic distinction that was sufficiently easy for participants to understand and perform successfully resulted in better post-training identification and discrimination performance than after a standard explicit training procedure with feedback using the same stimuli in the same number of trials. It remains to be investigated whether full generalization to voices and phonetic contexts is also possible with implicit training as it has been in explicit procedures.

If our findings can be replicated and confirmed, they raise significant potential implications for training applications. If consistent pairing of a relatively easier task with a more demanding target domain suffices to enable or boost learning for the latter, then it may be possible to greatly increase the efficiency of training programs while at the same time reducing the subjective cognitive effort and sustaining motivation due to perceived success in the explicit task. The implicit procedure has another advantage, which may turn out to be of crucial importance both in applications and in our theoretical understanding of the role of feedback. Specifically, even though the same number of trials can be presented in the explicit and the implicit design, if the target task is very difficult but the explicit task only moderately so (but not too easy, to sustain attention and motivation), then most of the internally generated feedback in the implicit condition will be positive reinforcement. In contrast, in a difficult explicit task much if not most of the externally provided feedback will necessarily be negative, indicating an incorrect guess. Apart from the emotional effects on motivation, the role of negative feedback may be distinct from the role of positive feedback in the eventual strengthening of the target representations.

This view is consistent with the well-known maxim in psychophysical training that, for optimal efficiency, most trials should be presented at a level leading to successful performance (Ahissar & Hochstein, 1997). In the past, this was thought to be due to discriminability, so manipulations such as exaggeration of acoustic differences have been successfully applied to enhance training of difficult contrasts (e.g., McCandliss et al., 2002). If, however, it is not discriminability per se that makes the difference but proportion of successful performance leading to positive reinforcement signals, then a novel approach to training may emerge. This approach would require an explicit task consistently paired with the difficult target task, so as to maximize the frequency of positive reinforcement, leading to perceptual representations of the target domain eventually sufficiently strong and refined to support explicit task performance.

## References

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences, USA, 90,* 5718–5722. doi:10.1073/pnas.90.12.5718

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401–406. doi:10.1038/387401a0

Aoyama, K., Flege, J. E., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32,* 233–250. doi:10.1016/S0095-4470(03)00036-6

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324. doi:10.1111/1467-9280.00063

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, England: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Bao, S., Chan, V. T., & Merzenich, M. M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature, 412,* 79–83. doi:10.1038/35083586

Bates, D., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes* (R package ver. 0.99875–6). Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). Philadelphia, PA: Benjamins.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Bradlow, A. (2008). Training non-native language sound patterns. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). Philadelphia, PA: Benjamins.

Choi, H., Seitz, A. R., & Watanabe, T. (2009). When attention interrupts learning: Inhibitory effects of attention on TIPL. *Vision Research, 49,* 2586–2590. doi:10.1016/j.visres.2009.07.004

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71. doi:10.1006/cogp.1998.0681

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2,* 406–416. doi:10.1016/S1364-6613(98)01232-7

Deacon, H., Conrad, N., & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology, 49,* 118–124. doi:10.1037/0708-5591.49.2.118

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59,* 447–456. doi:10.1016/j.jml.2007.11.004

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67,* 224–238. doi:10.3758/BF03206487

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–358). Berlin, Germany: Mouton de Gruyter. doi:10.1515/9783110895094.319

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35,* 116–124. doi:10.3758/BF03195503

Frankó, E., Seitz, A. R., & Vogels, R. (2010). Dissociable neural effects of long-term stimulus-reward pairing in macaque visual cortex. *Journal of Cognitive Neuroscience, 22,* 1425–1439. doi:10.1162/jocn.2009.21288

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, USA, 100,* 8030–8035. doi:10.1073/pnas.1332441100

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage, 21,* 494–506. doi:10.1016/j.neuroimage.2003.09.071

Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication, 50,* 109–125. doi:10.1016/j.specom.2007.07.003

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30,* 509–516. doi:10.1177/0165025406071914

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America, 100,* 1111–1121. doi:10.1121/1.416296

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research, 47,* 46–57. doi:10.1044/1092-4388(2004/005)

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 57–77). Philadelphia, PA: Benjamins.

Hattori, K., & Iverson, P. (2009). English /r/–/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America, 125,* 469–479. doi:10.1121/1.3021295

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research, 37,* 2133–2141.

Herzog, M. H., & Fahle, M. (1998). Modeling perceptual learning: Difficulties and how they can be overcome. *Biological Cybernetics, 78,* 107–117.

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., & Evans, B. G. (2008). Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1305–1316. doi:10.1037/0096-1523.34.5.1305

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,* 3267–3278. doi:10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87,* B47–B57. doi:10.1016/S0010-0277(02)00198-1

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54*A, 1105–1124. doi:10.1080/02724980042000516

Kilgard, M. P., & Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science, 279,* 1714–1718. doi:10.1126/science.279.5357.1714

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461,* 145–149. doi:10.1016/j.neulet.2009.06.030

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics, 38,* 569–587. doi:10.1016/j.wocn.2010.08.003

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56,* 1–15. doi:10.1016/j.jml.2006.07.010

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of the exemplar-based phonetic memory. In K. Elenius & P. Branderyd (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 140–147). Stockholm, Sweden: KTH and Stockholm University.

Lim, S.-J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*. Advance online publication. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1551-6709.2011.01192.x/abstract

Lin, J. Y., Pype, A. D., Murray, S. O., & Boynton, G. M. (2010). Enhanced memory for scenes presented at behaviorally relevant points in time. *PLoS Biology, 8,* e1000337. doi:10.1371/journal.pbio.1000337

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,* 1242–1255. doi:10.1121/1.408177

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,* 874–886. doi:10.1121/1.1894649

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351–377). Timonium, MD: York Press.

Maye, J., & Gerken, L. (2000). Learning phoneme categories without minimal pairs. In S. C. Howel, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11,* 122–134. doi:10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101–B111. doi:10.1016/S0010-0277(01)00157-3

McCandliss, B., Conway, M., Protopapas, A., & McClelland, J. (2002). Success and failure in teaching the –[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience, 2,* 89–108. doi:10.3758/CABN.2.2.89

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340. doi:10.3758/BF03211209

Munro, M. J., & Bohn, O.-S. (2007). The study of second language speech: A brief overview. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 3–11). Philadelphia, PA: Benjamins.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47,* 204–238. doi:10.1016/S0010-0285(03)00006-9

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401–426. doi:10.1037/0096-3445.130.3.401

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *Journal of the Acoustical Society of America, 127,* EL54–EL59. doi:10.1121/1.3292286

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10,* 233–238. doi:10.1016/j.tics.2006.03.006

Petrov, A., Dosher, B., & Lu, Z.-L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Research, 46,* 3177–3197.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam, the Netherlands: Benjamins.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46,* 115–154. doi:10.1177/00238309030460020501

Pisoni, D. B., & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 433–459). Timonium, MD: York Press.

Port, R. (2007). The graphical basis of phones and phonemes. In O. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 349–365). Philadelphia, PA: Benjamins.

Protopapas, A., & Calhoun, B. (2000, August). Adaptive phonetic training for second language learners. In P. Delcloque (Ed.), *Integrating speech technology in language learning and the assistive interface (Proceedings of InSTIL, 2000)* (pp. 31–38). Dundee, England: InSTIL.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America, 119,* 1684–1696. doi:10.1121/1.2161427

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59,* 413–425. doi:10.1016/j.jml.2008.02.002

R Development Core Team. (2007). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235. doi:10.1037/0096-3445.118.3.219

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *The Journal of Neuroscience, 13,* 87–103.

Rosenthal, O., & Humphreys, G. W. (2010). Perceptual organization without perception: The subliminal learning of global contour. *Psychological Science, 21,* 1751–1758. doi:10.1177/0956797610389188

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928. doi:10.1126/science.274.5294.1926

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infants auditory world: Hearing, speech and the beginnings of language. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception and language* (Vol. 2, pp. 58–108). New York, NY: Wiley.

Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology, 17,* 148–153. doi:10.1016/j.conb.2007.02.004

Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron, 61,* 700–707. doi:10.1016/j.neuron.2009.01.016

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory association. *Perception, 36,* 1445–1453. doi:10.1068/p5843

Seitz, A. R., Nánez, J. E., Holloway, S. R., & Watanabe, T. (2005). Visual experience can substantially alter critical flicker fusion thresholds. *Human Psychopharmacology: Clinical and Experimental, 20,* 55–60. doi:10.1002/hup.661

Seitz, A. R., Nánez, J. E., Holloway, S. R., & Watanabe, T. (2006). Perceptual learning of motion leads to faster flicker perception. *PLoS ONE, 1,* e28. doi:10.1371/journal.pone.0000028

Seitz, A. R., Protopapas, A., Tsushima, Y., Vlahou, E. L., Gori, S., Grossberg, S., & Watanabe, T. (2010). Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. *Cognition, 115,* 435–443. doi:10.1016/j.cognition.2010.03.004

Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature, 422,* 36. doi:10.1038/422036a

Seitz, A. R., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Sciences, 9,* 329–334. doi:10.1016/j.tics.2005.05.010

Seitz, A. R., & Watanabe, T. (2009). The phenomenon of task-irrelevant perceptual learning. *Vision Research, 49,* 2604–2610. doi:10.1016/j.visres.2009.08.003

Steffler, D. J. (2001). Implicit cognition and spelling development. *Developmental Review, 21,* 168–204. doi:10.1006/drev.2000.0517

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 3–45). Timonium, MD: York Press.

Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience, 18,* 1029–1043. doi:10.1162/jocn.2006.18.6.1029

Tsushima, Y., Seitz, A. R., & Watanabe, T. (2008). Task-irrelevant learning occurs only when the irrelevant feature is weak. *Current Biology, 18,* R516–R517. doi:10.1016/j.cub.2008.04.029

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564. doi:10.1037/0096-3445.134.4.552

Vallabha, G. K., & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective & Behavioral Neuroscience, 7,* 53–73. doi:10.3758/CABN.7.1.53

Wade, T., & Holt, L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of the Acoustical Society of America, 118,* 2618–2633. doi:10.1121/1.2011156

Watanabe, T., Nánez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413,* 844–848. doi:10.1038/35101601

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence of perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63. doi:10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866–1878. doi:10.1121/1.390988

Received March 10, 2011
Revision received June 28, 2011
Accepted June 28, 2011 ■

## Correction to Vlahou et al. (2011)

The article “Implicit Training of Nonnative Speech Stimuli,” by Eleni L. Vlahou, Athanassios Protopapas, and Aaron R. Seitz (*Journal of Experimental Psychology: General*, Advance online publication. September 12, 2011. doi:10.1037/a0025014) contained a production-related error.

In the Training section of Experiment 1, the sentence “For example, on a typical trial, participants might hear a syllable beginning with a retroflex sound ([ʈa:]), then after 250 ms the exact same syllable ([ʈ]), then, after 500 ms, a syllable beginning with a dental sound ([ta:]) and, finally, after 250 ms, the same syllable ([ta:])” should read “For example, on a typical trial, participants might hear a syllable beginning with a retroflex sound ([ʈa:]), then after 250 ms the exact same syllable ([ʈa:]), then, after 500 ms, a syllable beginning with a dental sound ([ta:]) and, finally, after 250 ms, the same syllable ([ta:]).” All versions of this article have been corrected.

DOI: 10.1037/a0025992