LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Distributional Learning of Speech Sounds: An Exploratory Study Into the Effects of Prior Language Experience

Kateřina Chládková [a,b] and Šárka Šimáčková [c]

[a] Institute of Psychology, Czech Academy of Sciences, [b] Amsterdam Center for Language and Communication, University of Amsterdam, and [c] Department of English and American Studies, Palacký University Olomouc

Distributional learning is typically understood as (unattended) tracking of stimulus probabilities. Distributional training with speech yields mixed results and the influencing factors have not yet been fully investigated. This study explored whether prior linguistic experience could have an effect on distributional learning outcomes. Czech and Greek adults, whose native languages contain and lack abstract length categories, respectively, were exposed to novel vowels falling into unimodal or bimodal distributions along the durational dimension. A trending interaction suggested that the Czechs and the Greeks might have been affected differently by the distributional exposure. Improved discrimination of the "trained" contrast was observed in bimodally exposed Czechs (whose prior expectations about length categories could guide learning) and, rather surprisingly, in unimodally exposed Greeks (who, lacking any expectations, might have listened in a noncategorical, auditory mode). Prior linguistic experience could thus affect whether and how experienced language users exploit new distributional speech statistics. This proposal needs to be assessed in future studies.

KC is also affiliated to the Faculty of Arts, Charles University. This work was funded by the Czech Science Foundation, grant 18–01799S. KC was also supported by Charles University grant Primus/17/HUM/19 and by the Netherlands Organisation for Scientific Research (NWO) grant 446-14-012. We would like to thank Barbora Chládková and Anastasia Chionidou for testing the Greek participants, and Václav Jonáš Podlipský for valuable comments on the manuscript.

Correspondence concerning this article should be addressed to Kateřina Chládková, Institute of Psychology, Czech Academy of Sciences, Hybernská 8, 110 00 Praha, Czech Republic. E-mail: k.chladkova@uva.nl

The handling editor for this article was Emma Marsden.

DOI: 10.1111/lang.12432

Keywords distributional learning; prior experience; vowel length; speech sound learning

Introduction

In order to comprehend and produce utterances, language learners first have to acquire the sound inventory of a given language, that is, learn how many and which speech sounds are functionally distinct units. One of the mechanisms they use to do so is distributional learning, which can in this instance be described as tracking without conscious attention of the probability distributions of speech sounds in one's environment. Infants at an age as young as 2 months, that is, even before they have any lexical knowledge, are sensitive to the sounds' distributional information (Maye, Werker, & Gerken, 2002; Wanrooij, Boersma, & van Zuijen, 2014b), indicating that distributional learning can operate as an unsupervised bottom-up mechanism.

Some literature suggests that adults, too, can track the probability distributions of sounds in the ambient language (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Escudero, Benders, & Wanrooij, 2011; Goudbeek, Cutler, & Smits, 2008; Hayes-Harb, 2007; Maye & Gerken, 2001). For instance, Maye and Gerken (2001) showed that listeners exposed to a bimodal distribution (i.e., two distinguishable clusters of sounds) along the voice-onset time continuum between [d] and [d̥] (i.e., prevoiced and voiceless unaspirated alveolar plosives) subsequently discriminated this nonnative contrast better than listeners exposed to a unimodal distribution (i.e., a single cluster of sounds) on the same continuum. However, a number of recently published studies failed to find the expected distributional training effects (Wanrooij, Boersma, & van Zuijen, 2014a; Wanrooij, De Vos, & Boersma, 2015; and, for passively exposed listeners, Ong, Burnham, & Escudero, 2015).

In principle, adults are sensitive to distributional statistics in various modalities (visual: Love, 2003; auditory: Garrido, Teng, Taylor, Rowe, & Mattingley, 2016). However, whether or not this sensitivity results in the formation of new categories or contrasts (i.e., the traditionally expected effects of distributional training) might be subject to factors other than the input statistics. One such factor, for instance, might be prior experience. Specifically for speech, adults, unlike infants, already have at least one speech-sound system in place: that of their native language, which undeniably modulates the extent to which they are able to acquire the sound system of a second language. Thus, the question to be addressed is whether prior phonological knowledge affects—either facilitates or weakens—adults' ability to employ a distributional learning mechanism.

Background Literature

Influences of prior linguistic knowledge on adults' speech and language processing are inevitable and have been widely documented (see the review by Sebastián-Gallés, 2005). The make-up of the native phoneme inventory affects the perception and production of nonnative phonemes. Learners assimilate foreign speech sounds to their native categories, which may result in their failure to distinguish second-language phoneme contrasts (Best & Tyler, 2007). A well-known example is the Japanese difficulty with the English /l/–/ɹ/ distinction (see the review by Holt & Lotto, 2010). Also, learners may associate a second-language contrast with an acoustic dimension other than that used by native speakers of the target language: Finnish learners of English initially rely on duration to differentiate /i/ and /ɪ/ because duration serves as a cue to a similar contrast in their native phonology, but can learn to redefine the contrast as a spectral contrast with sufficient training (Ylinen et al., 2010). It is of interest that Spanish learners of English have also been reported to rely initially on duration to distinguish English /i/ and /ɪ/, even though their native phonology (unlike that of Finnish) does not use duration contrastively (Escudero & Boersma, 2004). According to Escudero and Boersma's (2004) interpretation, adults are most likely to employ distributional learning on a dimension with which they have *no* prior experience from their native language.

Recent evidence suggests that the relative importance (i.e., weighting) of individual phonetic cues in adults' native phonology may indeed affect the extent to which they learn from distributional statistics, albeit in a different way than proposed by Escudero and Boersma (2004). Schertz, Cho, Lotto, and Warner (2016) showed that after exposure to novel distributions of plosive categories where two cues, fundamental frequency (F0) and voice-onset time, provide conflicting information, adult listeners downweigh the importance of the cue that serves as secondary in their native phonology. Adults thus initially extract a cue conflict, which means they are to some extent sensitive to information provided by both cues; and they continue to trace the distributional information provided by the native-language primary cue but stop relying on the information provided by the native-language secondary cue, which means they keep listening categorically only for the cue that is strongest in their native system.

Given that "listener biases interact in a complex way with changes in distributional information" (Schertz et al., 2016, p. 365), it is quite likely that listeners' prior biases will modulate the extent to which, and/or the way in which, novel categories are learned in distributional training paradigms. The role of previous experience seems to feature also in some recent models of talker and accent adaptation: For instance, in the study by Kleinschmidt and Jaeger (2016), the adaptation of phonetic categories by tracking distributional acoustic information was more successful when the exposure probability distributions did not deviate too greatly from the listener's established phonetic category knowledge. Similarly, prior language knowledge is seen as an interfering factor in statistical learning of new language categories in Pajak, Fine, Kleinschmidt, and Jaeger (2016) experience-based account of second language development.

To summarize, some theories propose that statistical distributional learning is more plausible on phonetic dimensions with which listeners have no prior experience (Escudero & Boersma, 2004; see also Bohn, 1995). In contrast, the results of recent studies indicate that distributional learning might be easier if some prior knowledge exists, for example, in the form of linguistic experience with a particular phonetic dimension (Kleinschmidt & Jaeger, 2016) or as a preference for one particular phonetic dimension over another (Schertz et al., 2016). Whether and how listeners' prior linguistic knowledge affects distributional learning—supposedly one of the main mechanisms responsible for category formation—is yet to be thoroughly investigated.

Initial research in this area reports effects of prior language knowledge on statistical learning of novel words segmented from a stream of speech (Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018) and on learning of lexical tones (Ong, Burnham, Escudero, & Stevens, 2017). Motivated by a lack of interrelation between performance in auditory and in visual statistical learning tasks, Siegelman et al. (2018) formulated the "entrenchment hypothesis," according to which learning of the statistical properties from new language input is modulated by prior linguistic knowledge, but statistical learning from nonverbal auditory input and visual input is unbiased. They tested the hypothesis by investigating native Hebrew speakers' ability to engage statistical learning in (a) extracting words of an artificial language from continuous auditory input, (b) extracting target sequences of nonverbal sounds, and (c) learning novel abstract shapes from visual input. Participants' performance on the latter two tasks was aligned and differed from the performance on the linguistic auditory task, which was subject to entrenchment. In further support of the entrenchment hypothesis, Siegelman et al. found superior statistical learning performance on those artificial verbal stimuli that resembled Hebrew words: That is, prior linguistic knowledge had a facilitative effect.

Given that prior experience modulates statistical learning of syllable sequences, it is plausible that it also modulates statistical learning of isolated speech categories. In a relevant study, Ong et al. (2017) tested distributional learning of lexical tones in two groups of listeners (in separate experiments); they found the expected effects of distributional training in the group who had prior experience with lexical tone from their first language (Mandarin-speaking nonmusicians) and did not find such effects in the other group, who lacked prior experience with lexical tone (Australian English-speaking musicians). Though supporting the facilitative role of prior linguistic knowledge, the finding is not unambiguously conclusive because the two groups in Ong et al. (2017) study were analyzed separately. Moreover, a parallel group of listeners with a nontone first language (Australian English-speaking nonmusicians), reported on in an earlier study (Ong et al., 2015), were able to learn lexical tones distributionally with particular attentional modulations. In light of the inconclusive outcomes to date, it is not possible to infer whether and in what way language experience affects distributional learning of individual speech-sound categories.

The Present Study

As summarized above, the literature to date has not provided conclusive results as to whether and in what ways adults' language background affects their ability to learn new speech-sound categories from distributional input. The aim of the experiment reported in this article was thus to explore whether prior phonological knowledge alone modulates (i.e., facilitates or hampers) the distributional learning of new vowels. We attempted to do this by using a novel experimental design that tests the outcomes of distributional training with multiple stimulus pairs sampled across the entire training dimension. We believe such a design can help to measure changes in perceptual sensitivity more accurately than the previously used type of test that employs only two contrasting stimuli, typically separated by a large acoustic distance (see Figure 1). Testing perceptual discrimination of a single stimulus pair representing a large difference, such as the maximally distinct tone pair in Ong et al. (2017), might not reveal whether bimodally exposed listeners learn from the novel stimulus structures more than unimodally exposed listeners (or vice versa) because everyone may improve in their discrimination of the distinctive stimulus pair merely due to greater familiarity with the task and the particular stimulus difference involved. The present design, with a more varied stimulus set, may be more sensitive

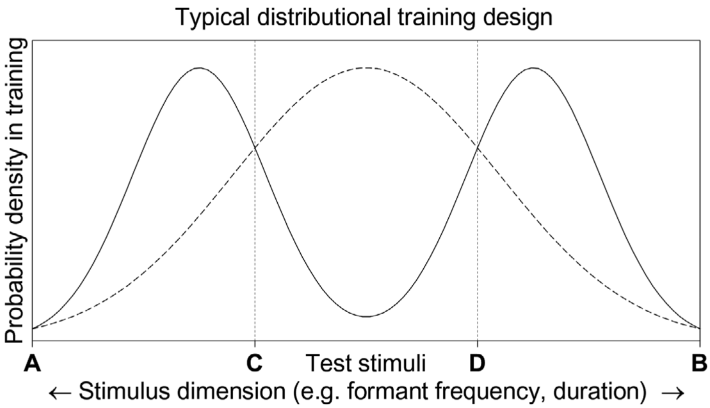

Figure 1 The typical design of distributional training experiments. Two groups of listeners are each exposed to either a unimodal or a bimodal distribution of stimuli. The effects of training are subsequently tested with one or two pairs of stimuli, A versus B or/and C versus D. Improved posttest discrimination of the two tokens within the test stimulus pair in bimodally but not in unimodally trained listeners is taken to reflect distributional learning.

to slight and gradient changes in perception. Furthermore, with its nonrepetitive multiple-stimulus paradigm, the design represents a more ecologically valid method of assessing speech perception than a task with a limited number of repeatedly presented stimulus pairs (as suggested by Rogers & Davis, 2009).

Thus, instead of assessing pre- and postexposure discrimination of a single, repeatedly presented pair of stimuli (whose members either span a category boundary or fall within a single category), we tested discrimination in the entire region whose categorical status differs depending on whether it belongs to a unimodal or to a bimodal training distribution. Assessing perceptual discrimination with such a varied, naturalistic stimulus set can eventually reveal even slight differences between processing mechanisms in different groups of learners (which might be obscured in a design with frequent stimulus repetition, due to, for example, entrainment to test items). If distributional exposure results in listeners' recovery of the underlying categorical structure, perceptual discrimination of the stimuli in such a region should improve in listeners who are trained to perceive this region as a category boundary (i.e., bimodally exposed listeners) but not in listeners who are trained to perceive the region as a category center (i.e., unimodally exposed listeners).

To test whether distributional learning is affected by prior linguistic experience, we investigated learning of novel vowel-length categories in Czech and Greek listeners whose native languages, respectively, contain and lack contrastive vowel length. Both Czech and Greek have five-vowel phoneme inventories, which are similar in terms of vowel quality but differ in the use of vowel duration: Czech employs duration to contrast its vowels (i.e., each of the five vowel qualities occurs as short and long) whereas Greek does not (Arvaniti, 1999; Šimáčková, Podlipský, & Chládková, 2012). The presence versus absence of contrastive duration is the main feature that distinguishes the two vowel systems and defines whether or not listeners enter the distributional training phase with or without prior expectations about the categorical structure of the stimulus space.

Czech and Greek listeners were exposed to either a unimodal or a bimodal distribution of vowels along the duration dimension. Their sensitivity to durational differences was assessed before and after training to reveal whether listeners recovered the categorical structure underlying the distributions of the exposure stimuli. Pre- and postexposure discrimination were measured over a region that spans the valley area in the bimodal distribution and extends over the peak area in the unimodal distribution. As noted above, a valley in a probability distribution corresponds to a boundary between two categories, whereas a peak corresponds to a category center. Hence, two random sounds drawn from the boundary area (found in the bimodal distribution) are more likely to have come from two different categories than are two random sounds drawn from a category center area (found in the unimodal distribution). Perceptual categorization then predicts high discrimination at category boundaries and low discrimination at category centers. The potential outcomes of training, and the potential outcomes in listeners with different language backgrounds, can be summarized as follows:

- If distributional training leads to the formation of new categories, discrimination of the critical test region will improve in participants exposed to the bimodal distribution, for whom the region coincided with a category boundary during exposure.
- If distributional learning of new speech-sound categories is facilitated by the existence of a relevant contrast in the listeners' native language, a bimodal-over-unimodal advantage will be larger in the Czech listeners.
- If distributional learning of new speech-sound categories is weakened by the existence of a relevant contrast in the listeners' native language, the bimodal-over-unimodal advantage will be larger in the Greek listeners.

Method

Participants

The participants were 28 Greeks (14 unimodally trained: age range = 22–35, $M = 27$, 8 female; 14 bimodally trained: age range = 22–34, $M = 27$, 8 female) and 25 Czechs (12 unimodally trained: age range 19–24, $M = 21$, 11 female; 13 unimodally trained: age range = 19–22, $M = 20$, 5 female). They were healthy individuals with normal hearing, and were university students or young professionals. None of the Czechs (tested in Olomouc) or the Greeks (tested in Thessaloniki) had spent an uninterrupted period of time longer than a month outside their country of origin. All participants had been raised monolingually although most reported to have a moderate knowledge of English.

Procedure

At pretest, listeners performed a same–different AX task in which the stimuli were female-voice vowels, naturally produced and subsequently edited (see the subsection Pretest and Posttest Stimuli and Task). On each trial, listeners heard two stimuli, A and X, and had to indicate whether they were the same or different. There was no feedback and no option to replay. There was a total of 35 trials; the task took about 3 minutes to complete.

The pretest was followed by training, during which participants listened passively to a set of 210 randomized vowel tokens for about 4.5 minutes. The stimuli were synthetic male-voice vowels (see the subsection Training Stimuli). The training was followed by a posttest, which was the same AX task as in the pretest, but with a different randomization of the 35 trials.

Training Stimuli

Training stimuli were isolated vowels synthesized with KlattGrid in Praat (Boersma & Weenink, 1992–2020). The vowels had the quality of /ɑ/, non-native for both Czech and Greek listeners, which lies between both groups' native /a/ and /o/ categories and is more closely assimilable to /a/ in both languages (see Nenonen, Shestakova, Huotilainen, & Näätänen, 2005, for evidence that the nativeness of vowel quality may affect listeners' duration processing; see Fourakis, Botinis, & Katsaiti, 1999, and Skarnitzl & Volín, 2012, for the acoustics of Greek and Czech vowels, respectively). The stimuli modeled a male voice, with the first three formants being 617, 1,179, and 2,231 Hz, respectively, and 17 higher formants being included for a flatter spectrum. The fundamental frequency contour fell linearly from 145 to 131 Hz.

Each vowel was synthesized with a unique duration value drawn from either a bimodal or a unimodal distribution, as shown in Figure 2. From each

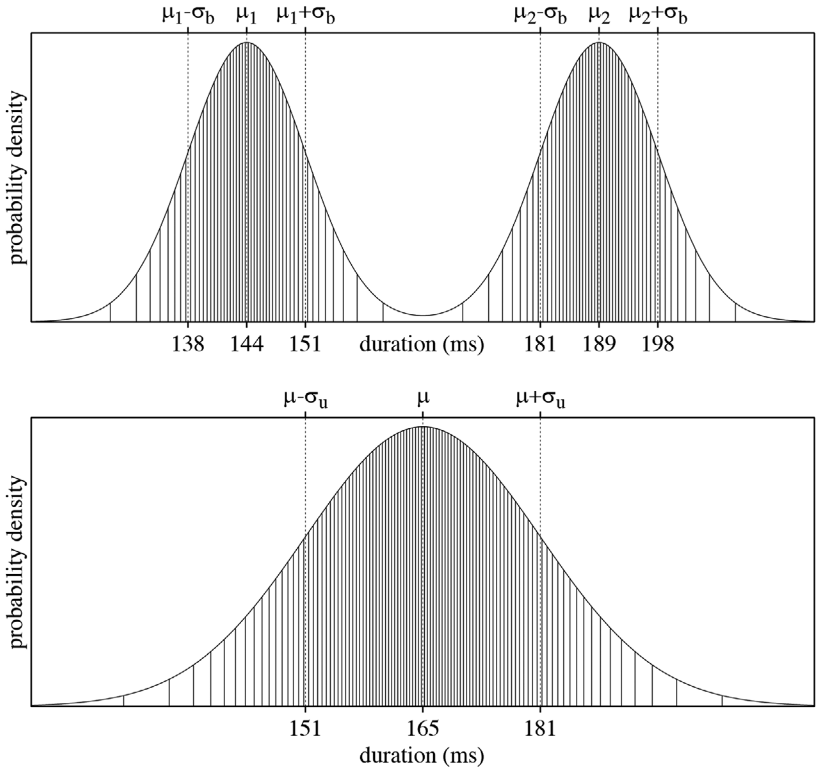

Figure 2 The bimodal (top) and the unimodal (bottom) training distribution. The duration axis (ranging from 122 to 223 milliseconds) was scaled logarithmically; the means (μ for the unimodal distribution, and μ_1 and μ_2 for the bimodal distribution) and standard deviations (σ_u and σ_b for the unimodal and the bimodal distribution, respectively) are shown in milliseconds. Vertical lines illustrate how the training stimuli were sampled from each distribution; for better readability, the figure shows only 100 vertical lines per distribution, whereas a total of 210 stimuli were sampled in the actual experiment. The sampling was done with the equal-area method reported by Wanrooij and Boersma (2013).

distribution, we sampled a total of 210 duration values with the equal-area method described by Wanrooij and Boersma (2013). It can be seen in Figure 2 that most duration values (68.3%) were sampled from the 1-standard-deviation bands around the peaks. For the training exposure, the 210 vowels were randomized and interspersed by an interstimulus interval jittering between 1.03 and 1.15 seconds. Half of the participants from each language group were

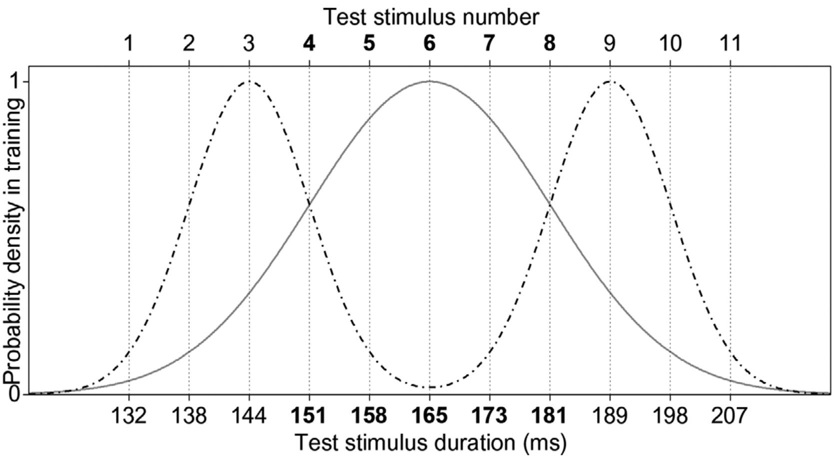

Figure 3 The 11 stimuli used in the test phase, with an overlay of the two training distributions. Unimodal distribution = grey solid curve; bimodal distribution = black dash-dotted curve. The duration axis is scaled logarithmically but the values shown are in milliseconds. The five test stimulus locations shown in bold (i.e., test stimuli 4–8) served as the reference points (i.e., centers) of the 35 AX discrimination pairs presented at test.

exposed to the bimodal distribution and the other half to the unimodal distribution.

Pretest and Posttest Stimuli and Task

For pretest and posttest stimuli, we used naturally produced /ɑː/, recorded by a female Estonian speaker with duration of 374 milliseconds and sustained spectral quality. The stable 216-millisecond-long middle portion of the vowel was extracted and its duration manipulated with the time-domain pitch-synchronous overlap-and-add method in Praat (Boersma & Weenink, 1992–2020). We created a total of 11 vowels with duration values equidistant on a logarithmic scale, as shown in Figure 3. The stimuli are publicly available at https://osf.io/xdq79/ and on IRIS at http://www.iris-database.org.

The 11 vowels were used to create stimuli for a same–different AX task. The AX task was targeted on the central part of the continuum that stretched between locations 4 through 8. As shown in Figure 3, the locations 4–8 fall in the valley area of the bimodal training distribution (i.e., they coincide with the category boundary) and in the peak area of the unimodal distribution (i.e., they coincide with the category center). Each of the five locations 4–8 served as a reference point for creating same–different AX pairs with four varying step sizes: an identical pair (no acoustic difference), two intermediate step sizes (representing, on average, durational differences of 15 milliseconds and of a twice-as-large 30 milliseconds), and one large step size (representing a durational difference of ∼45 milliseconds, i.e., three times that for the smallest acoustically distinct pair). The differences were computed on the log scale, such that, for instance, the absolute size of the largest step ranged from 41 to 49 milliseconds, for the shortest and longest stimulus pair, respectively. The various AX stimulus pairs were created for each of the five reference points as follows: For instance, for the middle reference point 6, we tested discrimination of Stimulus 6 versus Stimulus 6 (an identical trial), Stimuli 5 versus 7 (a 15-millisecond difference), Stimuli 4 versus 8 (a 30-millisecond difference), and Stimuli 3 versus 9 (a 45-millisecond difference). The two intermediate step sizes of 15 and 30 milliseconds represent a change of 9% and 17% in relative terms, and both fall within the range of previously reported magnitudes of the just-noticeable difference for duration in vowels and vowel-like stimuli (i.e., between 5% and 25%; Carlson & Granström, 1975; Smits, Sereno, & Jongman, 2006).

All the nonidentical, physically different, pairs were presented in both orders, that is, AX as well as XA. In total there were thus 35 AX trials, resulting from {5 reference points × [1 identical pair + (3 nonidentical pairs × 2 orders)]}. The 35 stimulus pairs were randomized and presented in a same–different AX task. The interstimulus interval between the two stimuli within a trial was 1 second, and the trial-initial silence after registering a response on a previous trial was 400 milliseconds.

Statistical Models

We analyzed the data using generalized linear mixed-effects models (the *lmerTest* package; Kuznetsova, Brockhoff, & Christensen, 2017) in R (R core team, 2016). We inferred and compared estimated means using the package *emmeans* (Lenth, Singmann, Love, Buerkner, & Herve, 2018).

We first fitted four models differing in their complexity (number of predictors) and compared them using the *anova*() function. In the models, the dependent, binomial, variable was the response "different." The simplest model (Model 1) included the main and interaction effects of language, training, and test. The two more complex models added step size (continuous predictor, Model 2a) or reference point (categorical predictor, Model 2b), as another, fourth, predictor respectively. The most complex model (Model 3) included both step size and reference point and their interactions with one another as well as with the other three predictors. For each categorical factor we specified orthogonal contrasts as follows: one contrast for language, with Czech coded as −0.5 and Greek as +0.5; one contrast for training, with unimodal coded as −0.5 and bimodal as +0.5; and one contrast for test, with pretest coded as −0.5 and posttest as +0.5. The contrasts for reference point compared the lateral reference points to the central point, 4 and 8 versus 6 and 5 and 7 versus 6 (with reference points 4 and 8, as well as 5 and 7, each coded as −0.25, and reference point 6 as +0.5), as well as comparing the short-duration to the long-duration reference points, 4 versus 8 and 5 versus 7 (with 4 as well as 5 coded as −0.5, and 8 as well as 7 coded as +0.5). In Model 1, participant and reference point were entered as random effects with per-test random slopes. In Model 2a, the random structure additionally included random slopes for step size. In Model 2b, the random effect was participant with per-test and per-reference-point random slopes, in addition to which Model 3 included per-step-size random slopes.

The data and the analysis script (including the full versions of the models) are available at https://osf.io/xdq79/ and on IRIS at iris-database.org.

Results

Model comparisons showed improvement of fit (indicated by a lower Akaike information criterion, AIC) from Model 1 to Model 2a (*AIC* 5,105 vs. 4,041, *df* 14 vs. 28, $p < .001$) but not from Model 1 to Model 2b (*AIC* 5,105 vs. 5,154, *df* 14 vs. 61, $p = .545$); Model 2a also fitted better than the more complex Model 3 (*AIC* 4,041 vs. 4,078, *df* 28 vs. 108, $p = .001$). Below we report results from the best-fitting model, 2a, with language, training, test, and step size as fixed effects, and participant and reference point as random effects with per-test and per-step-size random slopes.

Figure 4 shows the data per the variables language, training, and test, as well as the pretest to posttest improvement for each group. Table 1 gives the summary statistics of the best-fitting model, 2a. Converting the intercept estimate from the logit to the response scale as $e^{-0.176}/(1 + e^{-0.176})$ shows that the overall probability of responding “different” was .456, 95% CI [.411, .502], which is effectively at chance, aligning well with the fact that over half of the trials (i.e., including the “same” trials) contained stimuli distinguished by approximately the just-noticeable difference. The main effect of test shows that listeners generally improved in discriminating the test stimuli as a result of exposure. The effect of step size indicates that, as expected, stimuli differing by a larger acoustic distance were more often perceived as different than stimuli differing by a smaller distance. Step size also interacted with language,

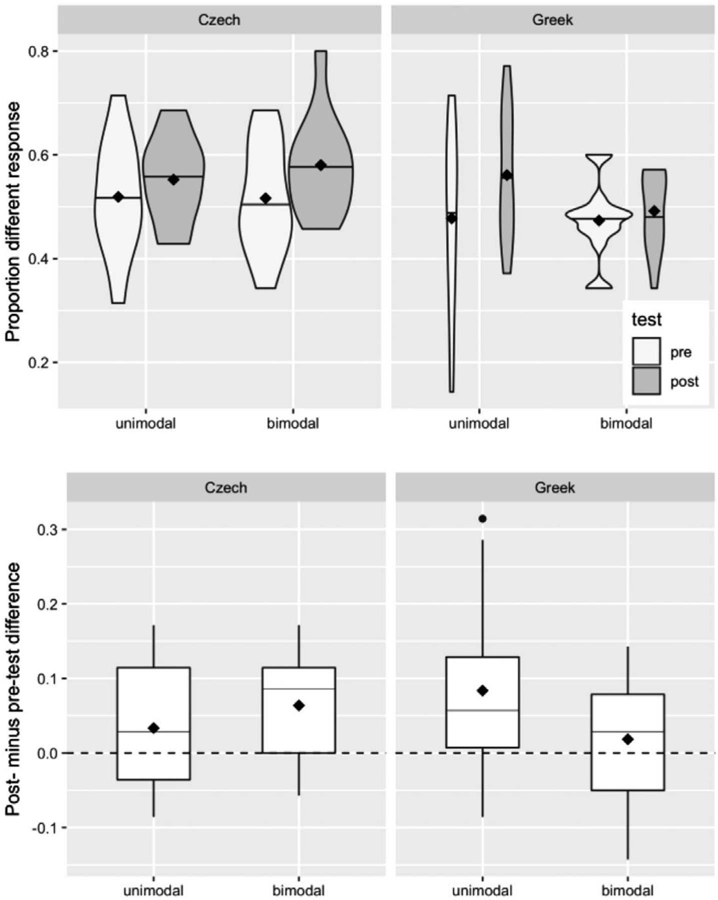

Figure 4 Data per language, training, and test, and the pre- to posttest improvement for each group. *Upper graph*: proportion of "different" responses per language, training, and test, collapsed across step sizes. Violins represent vertically arranged density plots and are trimmed to the range of the data. Horizontal lines are medians; black diamonds are means. *Bottom graph*: posttest minus pretest difference per Language and Training. Boxes range from the 1st to the 3rd quartile; whiskers cover 1.5 times the interquartile range. Horizontal bars show medians; diamonds are means. The dashed line runs through 0, which marks the level of no pretest-to-posttest improvement. Note that unimodally trained Greeks seem to yield a more dispersed response distribution at pretest than bimodally trained Greeks; this pattern can only be coincidental because participant assignment to training conditions was random. The box-plot visualization in the bottom graph shows that the pretest to posttest *improvement* (i.e., the measure of interest) had comparable dispersion across the groups.

Table 1 Summary of the best-fitting model, 2a

	Estimate			
Parameter	*M*	95% CI	*z*	*p*
Intercept	**−0.176**	**[−0.360, 0.008]**	**−1.879**	**.060**
Language (−Cz +Gr)	−0.172	[−0.480, 0.136]	−1.097	.273
Training (−uni +bi)	−0.081	[−0.389, 0.227]	−0.519	.604
Test (−pre +post)	**0.274**	**[0.084, 0.464]**	**2.813**	**.005**
Step Size	**1.312**	**[1.138, 1.486]**	**14.822**	**<.001**
Language × Training	−0.223	[−0.838, 0.392]	−0.709	.478
Language × Test	−0.003	[−0.328, 0.322]	−0.017	.987
Training × Test	−0.032	[−0.357, 0.293]	−0.192	.848
Language (−Cz +Gr) × Step Size	**−0.366**	**[−0.693, −0.039]**	**−2.188**	**.029**
Training × Step Size	0.103	[−0.224, 0.430]	0.618	.537
Test × Step Size	0.107	[−0.067, 0.281]	1.200	.230
Language × Training × Test	**−0.610**	**[−1.259, 0.039]**	**−1.842**	**.065**
Language × Training × Step Size	−0.091	[−0.748, 0.566]	−0.273	.785
Language × Test × Step Size	−0.108	[−0.457, 0.241]	−0.606	.544
Training × Test × Step Size	−0.263	[−0.612, 0.086]	−1.472	.141
Language × Training × Test × Step Size	0.277	[−0.423, 0.977]	0.777	.437

Note. Significant and trending effects (below alpha 0.05 and below 0.10, respectively) are in bold. Values modeled and shown are on the logit scale. CI = confidence interval; Cz = Czech; Gr = Greek; uni = unimodal; bi = bimodal; +/– before the levels of language, training, and test indicates contrast coding.

suggesting that the step size effect was larger for Czech (the +abstract length L1) than for Greek (the −abstract length L1) listeners. Finally, and most directly addressing the question of whether language background affects training outcomes, the triple interaction involving language, training, and test was just above the statistical significance threshold of .05 (estimated as −0.610, 95% CI [−1.259, 0.039], $p = .065$).

Even though the interaction of test, language, and training was not significant at alpha .05, we cannot exclude the possibility that the effect of test is further modified by language and training. Therefore, we compared the estimated discrimination probabilities at pretest versus posttest across the two languages, and the two training types; the estimated means and confidence intervals of the discrimination probabilities are listed in Table 2 and visualized in Figure 5. The pairwise comparisons of pretest and posttest data show

Table 2 Pairwise comparisons for the three-way Language × Training × Test interaction, showing means and 95% confidence intervals (CIs) for the proportion responding “different” in each condition

		Pretest		Posttest		Effect size[a]			
Language	Training	*M*	95% CI	*M*	95% CI	*M*	95% CI	*z*	*p*
Czech	Unimodal	0.523	[0.427, 0.617]	0.576	[0.488, 0.660]	0.216	[−0.137, 0.569]	−1.204	.229
	Bimodal	0.515	[0.421, 0.608]	0.614	[0.530, 0.692]	0.404	[0.053, 0.755]	−2.261	.024
Greek	Unimodal	0.459	[0.373, 0.548]	0.575	[0.495, 0.651]	0.466	[0.148, 0.784]	−2.881	.004
	Bimodal	0.460	[0.374, 0.548]	0.485	[0.407, 0.564]	0.102	[−0.201, 0.414]	−0.642	.521

Note. Statistical comparisons and effect size calculations were done on the log odds ratio scale.
[a]Simple effect sizes were calculated as posttest minus pretest log odds score in each group.

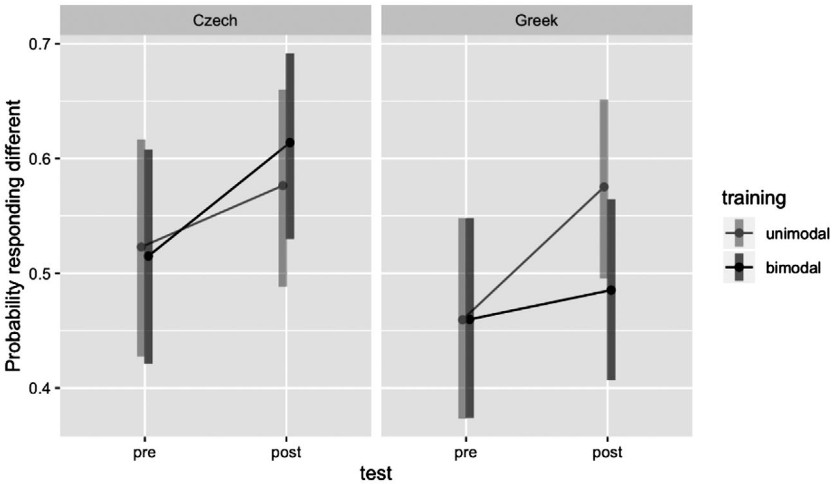

Figure 5 Estimated probabilities of responding “different” per language (separate panels), training (grey = unimodal; black = bimodal), and test. Filled circles correspond to the estimated mean proportion responding “different”; bars represent 95% confidence intervals. Pretest-to-posttest change is depicted by lines connecting the means for each condition.

that discrimination accuracy improved significantly from pretest to posttest in bimodally exposed Czechs (from mean 0.515 to 0.614, $p = .024$) and in unimodally exposed Greeks (from mean 0.459 to 0.575, $p = .004$; see Table 2).

Discussion

Present Questions and Findings

Our experiment tested whether prior language experience affects the outcomes of distributional training, that is, whether native-language phonology can help or hinder listeners’ ability to learn novel speech-sound categories from the sounds’ probability distributions. It has been widely acknowledged that language background affects the degree of difficulty that adults experience when processing and learning nonnative speech sounds. Yet, until very recently the recurring assumption in distributional training experiments seems to have been that the learning mechanism (ultimately leading to various degrees of attainment) does not change when learning subsequent languages (i.e., after the first). That is, it has been taken for granted that adults, whenever exposed to statistical distributions of stimuli, trace distributional statistics and, with various degrees of success, recover the underlying categories (e.g., Gulian, Escudero,

& Boersma, 2007; Hayes-Harb, 2007; Ong et al., 2015). This implicit assumption is beginning to be questioned, and we have explored its plausibility in this article.

Listeners were exposed to bimodal and to unimodal distributions of a nonnative categorical contrast, and before and after exposure performed a discrimination task on 20 unique stimulus pairs that were sampled across the critical stimulus region covering the valley area of the bimodal distribution (i.e., the boundary region between two novel categories) and the peak area of the unimodal distribution (i.e., the category center of a single novel category). In order to assess the (potentially gradient) impact of distributional training on learning speech-sound categories, the pre- and postexposure tests did not target a single stimulus pair (such as the two crossing points of the unimodal and bimodal distributions, or the two maximally distinct points on the stimulus scale) but covered the critical stimulus region with a more fine-grained approach. This region has different functions in the unimodal and bimodal training distributions, and different performance is therefore predicted for these training conditions. In line with the reasoning behind the definition of categorical perception (Repp, 1984), if listeners trace the statistics and attempt to recover the underlying categories, our bimodal exposure to durational differences would lead to improved sensitivity, whereas our unimodal exposure will lead to reduced sensitivity to acoustic durational differences in the critical area (representing the boundary and the category center, respectively, in the two exposure types).

We reasoned that the makeup of a listener's native phonology, specifically, their prior experience with categorical structures on the durational dimension, could determine whether or not they could employ distributional learning for novel vowel-length categories, /ɑ/ versus /ɑː/. We tested two groups differing in whether or not they have prior expectations about underlying categories: Czech listeners who have prior experience with durationally cued speech-sound contrasts, and Greek listeners who do not. We formulated two competing predictions for the effects of prior experience. If distributional learning is more likely to occur in listeners without a prior bias (i.e., blank-slate learners; Escudero & Boersma, 2004), Greeks could be expected to show larger effects of distributional training (i.e., a boost for discrimination of a new contrast after bimodal exposure and/or lower discrimination after unimodal exposure) than Czechs. If, on the other hand, distributional learning is facilitated by the existence of a prior bias toward perceiving the stimulus dimension categorically (e.g. Schertz et al., 2016; Siegelman et al., 2018), Czechs could be expected to show larger effects of distributional training than Greeks.

Discrimination performance, operationalized as (the probability of) responding “different,” was analyzed using generalized linear-mixed models. The results showed that, unsurprisingly, larger stimulus differences yielded higher discrimination scores than smaller stimulus differences, confirming that our novel stimulus design relatively finely measured the listeners’ perceptual performance and could therefore also be expected to uncover any gradient training-induced changes in perception. The Czech listeners were overall more affected by the size of stimulus difference than the Greek listeners, suggesting that the Czechs traced the stimuli’s physical differences slightly more accurately than the Greeks.

As to exposure-induced effects, posttest discrimination was overall higher than pretest discrimination, which means that exposure to the stimuli and/or to the task in general led to better performance at posttest. It appears that the pretest-to-posttest improvement might have been further modified by language and training. Although a slight numeric improvement from pretest to posttest was apparent across all four groups, only two groups improved significantly, namely, the Czech bimodally exposed group and the Greek unimodally exposed group. This pattern of results might be merely coincidental: The possibility that the pretest-to-posttest improvement was differentially modified by training in Czechs versus Greeks is addressed by the triple interaction of language, training, and test, whose marginal statistical significance ($p = .065$) suggests that the data are equivocal on this point. Alternatively, it might have been brought about by language-specific distributional learning mechanisms.

The observation that bimodal exposure led to improvement in Czech listeners but did not do so, or perhaps did so to a lesser extent, in Greek listeners is compatible with the hypothesis that prior language experience affects the outcomes of distributional learning. Specifically, it aligns with the prediction that an existing prior bias *facilitates* the learning of novel categories from distributional exposure.

At the same time, however, the finding of language-specific effects of unimodal exposure, namely, that unimodal exposure led to improvement in Greek listeners but to no or less improvement in Czech listeners, is suggestive of another learning mechanism employed by the Greeks. We elaborate on a possible interpretation of the improvements seen in the bimodally exposed Czechs and the unimodally exposed Greeks in the next subsection.

To sum up the current subsection, there was a main effect of test, and the four pairwise comparisons suggested a similar direction of change, that is, numeric improvement at posttest across all four groups (which turned out to be largest and statistically significant in unimodally exposed Greeks and bimodally exposed Czechs). This means that above and beyond any language-specific or training-distribution effects, listeners generally improved from pretest to posttest in their discrimination of durational differences in [ɑ(ː)]-like vowels.

Explaining the Outcomes of Training

The typically expected outcome of distributional training is that exposure to distributional probabilities leads to tracking and recovery of the underlying categories, such that at posttest bimodally exposed participants will improve in their ability to discriminate stimuli spanning a category boundary, whereas unimodally exposed participants will do so to a lesser extent or not at all. This is also what we seem to find here for the Czech participants. Supposedly, the bimodally exposed Czechs warped the acoustic space into two distinct categories and did so because they could link the auditory stimulus properties (i.e., short and long instances of novel vowels) to their existing bias toward recovering categorical structures along the durational dimension. Their prior bias thus enabled them to trace and uncover the bimodal short–long contrast on the novel /ɑ/–/ɑː/ continuum. The unimodally exposed Czechs might have, likewise, detected a category structure, this time unimodal, and come to treat any durational differences in the new vowels as within-category variation.

In contrast, the present observation for the Greek listeners cannot be explained by the typical distributional learning mechanism. Categorical perception posits that discrimination *within* category centers should be relatively low (see Holt & Lotto, 2010). Because it was the unimodally trained group who improved their discrimination of the critical region, we are inclined to conclude that the Greeks did not recover the categorical structures underlying the distributional exposure. Our speculation is that the Greeks, who lacked prior biases regarding linguistic categories cued by duration, did not perceive durational variation in the exposure stimuli as *phonologically* relevant and thus did not warp the stimuli into linguistic categories. Instead, they may have listened in a purely auditory mode (Pisoni & Lazarus, 1974; Werker & Logan, 1985) and improved in their auditory listening and discrimination of (subtle) acoustic differences for stimuli that they heard frequently. It happens that most tokens with which the unimodally exposed (Greek) participants were trained came from the critical test region. This is why they improved in their auditory discrimination of duration within that region.

To speculate further, the Greeks' auditory, noncategorical sensitization to phonetic detail could have occurred because the Greek listeners (unlike the Czechs) had no abstract representations in their phonology to link the acoustics to. This proposed explanation aligns well with the work of Kronrod, Feldman, and colleagues (Feldman, Griffiths, & Morgan, 2009; Kronrod, 2014; Kronrod, Coppess, & Feldman, 2016), who suggest that adult listeners primarily focus on phonetic detail unless they can and need to access the underlying categories, which can change their listening strategy in speech-sound discrimination tasks. We extend and adapt that line of reasoning for a case where adult listeners are exposed to a series of novel stimuli and subsequently have to discriminate them. In the remainder of this section we propose how two different ways of processing distributional input might affect the speech-sound perception of adult listeners in discrimination tasks. This proposal is a reflection on the trends observed in the current, exploratory study. As the trends might be merely coincidental, the plausibility of any language-specific distributional learning mechanisms needs to be addressed in future work.

Thus, *if* the Czech and the Greek listeners differed in how they were affected by the distributional exposure, the difference might have lain in whether or not the phonetic variation in the training input was informative to them about underlying category structures and consequently in how they processed the training input. The Czechs, for whom durational variation in their native language typically signals underlying category structures, may have attempted to attribute (some of) the phonetic variation in the novel training sounds to underlying categories. This led the unimodally exposed Czechs to recover a single underlying category and thus consider durational differences at posttest as within-category noise, and it also led the bimodally exposed Czechs to recover two underlying categories and consider durational differences at posttest as between-category changes. The Greeks, for whom durational variation in their native language does not come from underlying categorical structures, adopted the primary strategy that adults supposedly use in speech perception (cf. Kronrod et al., 2016) and became entrained to the phonetic detail. The unimodally exposed Greeks thus became sensitized to the dense durational variation in the central part of the stimulus scale that they heard during training and improved in hearing the durational differences at posttest. Following the rather difficult pretest task that required them to discriminate phonologically meaningless vowel durations, these adult listeners might have also approached the passive exposure situation in a somewhat analytical way, attentively focusing on the stimulus properties and tuning in to the slight variations in duration rather than subconsciously tracking the underlying statistics.

The ways in which categorical versus auditory listening during exposure affects posttest performance can be modeled mathematically as follows. A listener who during distributional exposure recovers underlying category structures will at posttest discriminate a stimulus pair *i–j* with the probability calculated using the formula provided by Pollack and Pisoni (1971):

$$p\ (discr.cat) = \frac{1 + \left(PA_i - PA_j\right)^2}{2}\ , \tag{1}$$

where $p(discr.cat)$ is the probability of discriminating using categorical knowledge, PA_i is the probability with which stimulus i is identified as category A, and PA_j is the probability with which stimulus j is identified as category A. In the bimodal-exposure (or, the two-category learning) scenario, category A could be understood as either "short" or "long," with the other category being mutually exclusive.

A listener who does not recover underlying categories during exposure but instead listens auditorily and tunes into the phonetic detail will at posttest discriminate a stimulus pair *i–j* with discrimination probability calculated as:

$$p\ (discr.aud) = \frac{P_i + P_j}{2}\ , \tag{2}$$

where $p(discr.aud)$ is the probability of discriminating auditorily, and P_i is the probability with which stimulus i occurs during training (and likewise for stimulus j).

Calculating and averaging across discrimination probabilities for each step size and reference point in our stimulus set shows that the category learning mechanism yields larger posttest discrimination in a bimodally than in a unimodally trained hypothetical listener, namely, 0.78 versus 0.5. In contrast, the auditory learning mechanism yields larger posttest discrimination in a unimodally than in a bimodally trained listener, namely, 0.61 versus 0.43. We did not model here the change from pre- to posttest and assumed that the pretest starting performance does not differ across groups. The unimodal versus bimodal differences in the modeled posttest performance show a similar pattern to the obtained posttest data (0.614 vs. 0.576 in bimodal vs. unimodal Czechs, and 0.575 vs. 0.485 in unimodal vs. bimodal Greeks), suggesting that Czechs and Greeks might have differed in the extent to which they employed a category-learning or an auditory-listening mechanism.

Modeling performance only for the pair of stimuli that have typically previously been used to test the effects of distributional learning—namely, the stimulus pair represented by durations 151 and 181 milliseconds (stimuli number 4 vs. 8 in Figure 3)—predicts a different pattern of results. Similarly to the current design with the complete stimulus set, the typical design predicts a posttest advantage for bimodal over unimodal listeners under the category-formation mechanism (1 vs. 0.5), but at the same time it predicts a *null* difference between unimodal and bimodal listeners under the auditory-sensitization mechanism (0.61 vs. 0.61). The present design thus allows researchers to test the effects of the competing mechanisms that might potentially be operating during distributional training exposure, and, by incorporating stimulus variability and nonrepetition, it makes the distributional learning experiment more ecologically valid in nature.

An anonymous reviewer considers it unnecessary to attribute the results to two different learning mechanisms, suggesting instead that it is "possible that the mechanisms of distributional learning are exactly the same for the two groups, but just operating on different underlying representations." We agree that, indeed, for both groups, experience gained in the course of the distributional training actually produced an increase in listeners' overall sensitivity to the variable durational information. We could further assume that categorical learning from this durational information was then enhanced when the training distributions matched the listeners' native-language expectations (which explains the bimodal advantage in Czechs), but was curbed when the training distributions could not be matched with the listeners' expectations. This alternative proposal, unlike most of the distributional-training literature to date, including the present study, would thus predict that distributional exposure elicits primarily an increase in overall sensitivity, which then is either modulated by existing underlying categories or not modulated where such categories are lacking. It would be interesting for future research to address whether distributional learning is at all a default universal mechanism by which adults process novel (categorically clustered) stimuli, or whether it can be constrained by language-specific structures.

Relevance to Theories of Adult (Language) Learning

The observed trending interaction of language, training, and test has relevance for theories of second language acquisition. The interaction suggests that the outcomes of statistical speech-sound learning depend on one's native language. The very outcomes of learning seem different, irrespective of what kind of processes the different learner groups engaged in (i.e., whether one group experienced distributional learning and the other group auditory sensitization, or whether both groups engaged primarily in auditory sensitization with category learning as a bonus for only one of them, or whether yet another different mechanism was at play). The pairwise comparisons of pretest versus posttest performance indicate that the Greeks did not learn from bimodal exposure as much as the Czechs did (or perhaps did not learn categories at all), which runs contrary to an earlier proposed theory that distributional category learning in adults might be available primarily for previously uncategorized, blank-slate dimensions (Bohn, 1995; Escudero, 2005; Escudero & Boersma, 2004). The result more likely accords with recent findings on cue-reweighting, statistical sequential learning, and distributional learning of nonnative tone contrast, which point toward the facilitating role of prior linguistic experience (Ong et al., 2017; Schertz et al., 2016; Siegelman et al., 2018).

Unlike the findings of previous studies, the present differences between pretest and posttest performance suggest a potentially opposite effect in the Greek listeners, namely, larger improvement after unimodal than after bimodal exposure. It is likely that such an effect, if it exists, might be obscured in studies using the traditional pretest and posttest design in which participants are tested on just one (or at most two) stimulus difference(s), namely, a stimulus pair spanning the maximal, or another large, distance, whose discrimination likely improves simply due to task or stimulus familiarity. Future research should compare these two test designs, and perhaps others, and test whether and to what extent these can uncover the changes in listeners' perceptual sensitivities.

Our study bears relevance to research on general learning abilities in adults. If future work confirms our proposal that distributional exposure is beneficial (i.e., has the expected category-recovery outcomes) only (or, at least, to a greater extent) for listeners who already have prior categorical experience with the tested dimension, it would imply that distributional learning might not operate as an unsupervised, bottom-up mechanism throughout the lifetime. It has been reported that immediate explicit feedback can facilitate learning from stimulus probability distributions (Ashby, Queller, & Berretty, 1999; Goudbeek et al., 2008). In an unattended exposure paradigm, such as the one employed here, some form of, at least *implicit*, supervision from higher-level representations might be necessary for speech-sound distributional learning to take place in adults. Participants trained with a bimodal distribution might be able to improve their discrimination of a novel boundary region only (or to a greater extent) if they can associate the novel stimuli with some existing knowledge, which thus modulates, or supervises, the learning. In contrast, without the availability of any underlying categorical structures, distributional training might have rather unexpected effects: Exposure to probability distributions could result in entrainment to the most frequently presented physical properties of stimuli. Such effects of prior abstract knowledge on adults' ability to learn implicitly from input have been reported for linguistic levels other than phonology, such as word order or mapping between a grammatical category and meaning (Leung & Williams, 2014; Onnis & Thiessen, 2013; Williams, 2020).

Limitations and Future Research

The triple interaction of language, training, and test, which directly addresses the question of language-specific training outcomes, only trended toward statistical significance given an alpha .05 (the *p*-value was .065). The subsequent pairwise comparisons of means and 95% confidence intervals suggested that all groups had somewhat better discrimination scores at posttest than at pretest (which is in line with the initial main significant effects of test), and these improvements were significant for bimodally trained Czech listeners and unimodally trained Greek listeners. Given the marginal statistical significance of the triple interaction and the overlapping confidence intervals, as well as the somewhat unexpected significant pretest-to-posttest improvement in the unimodally trained Greeks, a replication of the current experiment is needed before one can conclude that prior experience affects the mechanism that listeners employ in distributional training tasks.

A serious limitation of the current study is its low power. We had not properly considered the type and size of potential outcomes, and we underestimated a priori power calculations. Using the *simr* package (Green & MacLeod, 2016), and assuming an effect size of −0.6 (which is slightly smaller than, but roughly comparable to, the effect observed here) for the triple interaction of language, training, and test, we additionally calculated that a properly powered study (i.e., with power of 80%) would need to include more than twice as many participants as the present one (that is, between 106 and 159 participants in total, which would yield power of 66% and 84%, respectively). The present results, given the experiment's low power, should thus be taken as exploratory, requiring a better-powered replication.

A potential critique of our stimulus design might be that our test material was more familiar to the unimodally trained listeners than to the bimodally trained listeners. However, the degree of familiarity is questionable because training and test stimuli were *not* identical: They were produced by different voices (a synthetic male voice in training, an edited natural female voice in test) and also did not overlap in durational properties (with continuous sampling yielding 210 varied training stimuli, versus a different set of 11 stimuli at test). Unimodally trained listeners were, indeed, exposed to more stimuli in the critical test region than the bimodally trained listeners. Crucially, however, our results suggest that this larger degree of familiarity with the critical region in unimodally trained listeners influenced posttest perception in a language-specific way, strengthening sensitivity to vowel duration in Greek listeners but not in Czech listeners. Thus, it is plausible that exposure to the statistical distributions of speech sounds could have affected perceptual learning in language-specific ways.

We argued that adult listeners with no prior experience of categorizing a novel, trained dimension might, instead of forming new categories, become sensitized to acoustic detail. This proposal, if it is confirmed by future replications, will most likely hold only for the initial stages of learning because adult second language learners are often able to eventually acquire new speech sounds (Casillas, 2020). We would speculate that bottom-up (distributional) learning may not be as readily available in adults as it is in infants, such that the learning of novel units needs to be supervised by previously acquired knowledge. Future work should test at which stage of learning adults can begin forming novel categories for nonnative speech sounds and what information is necessary for this process to begin.

Another way of thinking about such outcomes is in terms of attention. Thanks to their prior experience with categorical structures on the durational dimension, the unimodally exposed Czechs could have inhibited assigning categorical structure to the unimodal variation because it conflicted with the bimodal length structure familiar from their native language. However, the (unimodallytrained) Greek listeners, who did not recover any categorical structures during the exposure phase, assigned relevance to fine phonetic differences that they heard most often during training (i.e., the center of the unimodal distribution), and focused on discriminating those at posttest. The role of attention in adult distributional learning of speech sounds is certainly one area for future research.

Conclusion

In adults, exposure to statistical distributions of speech sounds may not necessarily evoke a homogenous distributional learning mechanism by which all listeners infer the underlying categorical make-up of the stimulus space in the same way. The present study indicates that listeners' prior language experience might affect what they learn from distributional exposure. Our current, admittedly limited, data suggest that adults with prior expectations about the categorical structure of the stimulus space seemed to benefit more from the distributional information and uncover the underlying categories than listeners without such prior categorical bias. Listeners with no existing categorical bias might not primarily attribute the encountered stimulus variation to underlying categories and might instead become sensitized to the frequently heard auditory information. Future, properly powered studies could employ the novel distributional training design presented here to test whether or not the trends observed here can be confirmed.

Final revised version accepted 28 May 2020

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at http://www.iris-database.org and https://osf.io/xdq79/. All proprietary materials have been precisely identified in the manuscript.

References

Arvaniti, A. (1999). Standard modern Greek. *Journal of the International Phonetic Association*, *29*, 167–172. https://doi.org/10.1017/S0025100300006538

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*, 1178–1199. https://doi.org/10.3758/BF03207622

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, the Netherlands: John Benjamins.

Boersma, P., & Weenink, D. (1992–2020). Praat: Doing phonetics by computer [Computer software]. Retrieved from http://www.praat.org.

Bohn, O.-S. (1995). Cross language speech production in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Timonium, MD: York Press.

Carlson, R., & Granström, B. (1975). Perception of segmental duration. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 90–106). Berlin, Heidelberg: Springer.

Casillas, J. V. (2020). The longitudinal development of fine–phonetic detail: Stop production in a domestic immersion program. *Language Learning*, *70*, 768–806, , Retrieved from https://doi.org/10.1111/lang.12392

Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Amsterdam, the Netherlands: LOT, Netherlands Graduate School of Linguistics.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America*, *130*, EL206–EL212. https://doi.org/10.1121/1.3629144

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585. https://doi.org/10.1017/S0272263104040021

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*, 752–782. https://doi.org/10.1037/a0017196

Fourakis, M., Botinis, A., & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica*, *56*, 28–43. https://doi.org/10.1159/000028439

Garrido, M. I., Teng, C. L. J., Taylor, J. A., Rowe, E. G., & Mattingley, J. B. (2016). Surprise responses in the human brain demonstrate statistical learning under high concurrent cognitive demand. *npj Science of Learning*, *1*, 16006. https://doi.org/10.1038/npjscilearn.2016.6

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

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498. https://doi.org/10.1111/2041-210X.12504

Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 1893–1896). Saarbrücken: University of Saarbrücken. Retrieved from https://hdl.handle.net/11245/1.275271

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*, 65–94. https://doi.org/10.1177/0267658307071601

Holt, L. L., & Lotto, A. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*, 1218–1227. https://doi.org/10.3758/APP.72.5.1218

Kleinschmidt, D. F., & Jaeger, T. F. (2016). What do you expect from an unfamiliar talker? In J. Trueswell, A. Papafragou, D. Grodner, & D. Mirman (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2351–2356). Austin, TX: Cognitive Science Society.

Kronrod, Y. (2014). *Bayesian model of categorical effects in L1 and L2 speech processing* (PhD dissertation). Retrieved from http://ling.umd.edu/publications/285/

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review*, *23*, 1681–1712. https://doi.org/10.3758/s13423-016-1049-y

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). R package *emmeans* (Version 1.2.4) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/emmeans/index.html

Leung, J. H., & Williams, J. N. (2014). Crosslinguistic differences in implicit language learning. *Studies in Second Language Acquisition*, *36*, 733–755. https://doi.org/10.1017/S0272263114000333

Love, B. C. (2003). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin & Review*, *10*, 190–197. https://doi.org/10.3758/bf03196484

Maye, J., & Gerken, L. A. (2001). Learning phonemes: How far can the input take us? In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *BUCLD 25 Proceedings* (pp. 480–490). Somerville, MA: Cascadilla Press.

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

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, *92*, 26–32. https://doi.org/10.1016/j.bandl.2004.05.005

Ong, J. H., Burnham, D., & Escudero, P. (2015). Distributional learning of lexical tones: A comparison of attended vs. unattended listening. *PLoS One*, *10*(7), e0133446. https://doi.org/10.1371/journal.pone.0133446

Ong, J. H., Burnham, D., Escudero, P., & Stevens, C. J. (2017). Effect of linguistic and musical experience on distributional learning of nonnative lexical tones. *Journal of Speech, Language, and Hearing Research*, *60*, 2769–2780. https://doi.org/10.1044/2016_JSLHR-S-16-0080

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*, 268–284. https://doi.org/10.1016/j.cognition.2012.10.008

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning additional languages as hierarchical probabilistic inference: Insights from first language processing. *Language Learning*, *66*, 900–944. https://doi.org/10.1111/lang.12168

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*, 328–333. https://doi.org/10.1121/1.1914506

Pollack, I., & Pisoni, D. B. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, *24*, 299–300. https://doi.org/10.3758/BF03329012

R Core Team (2016). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.r-project.org

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language*, *10*, 243–335. https://doi.org/10.1016/B978-0-12-608610-2.50012-1

Rogers, J. C., & Davis, M. H. (2009). Categorical perception of speech without stimulus repetition. In *Proceedings of Interspeech* (pp. 376–379). Conference held at Brighton, UK, 6–10 September 2009.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, *78*, 355–367. https://doi.org/10.3758/s13414-015-0987-1

Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 546–566). Malden, MA: Blackwell.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Šimáčková, Š., Podlipský, V. J., & Chládková, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association*, *42*, 225–232. https://doi.org/10.1017/S0025100312000102

Skarnitzl, R., & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny [Reference values of vowel formants in young adult speakers of standard Czech]. *Akustické listy*, *18*, 7–11. Retrieved from https://www.czakustika.cz/misc/aklisty/18-1.pdf

Smits, R., Sereno, J., & Jongman, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 733–754. https://doi.org/10.1037/0096-1523.32.3.733

Wanrooij, K., & Boersma, P. (2013). Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America*, *133*, EL398–EL404. https://doi.org/10.1121/1.4798618

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014a). Distributional vowel training is less effective for adults than for infants: A study using the mismatch response. *PLoS One*, *9*(10), 1–13. https://doi.org/10.1371/journal.pone.0109806

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014b). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology: Language Sciences*, *5*(Article 77), 1–12. https://doi.org/10.3389/fpsyg.2014.00077

Wanrooij, K., De Vos, J. F., & Boersma, P. (2015). Distributional vowel training may not be effective for Dutch adults. In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 670. 1-5 retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0670.pdf.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*, 35–44. https://doi.org/10.3758/BF03207136

Williams, J. N. (2020). The neuroscience of implicit learning. *Language Learning*, *70*, 255–307. https://doi.org/10.1111/lang.12405

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience*, *226*, 1319–1332. https://doi.org/10.1162/jocn.2009.21272

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Foreign Speech Sounds in Adulthood: Learning From Exposure Seems Easier When Guided by Similarities to the Native Language

What This Research Was About and Why It Is Important

Babies learn the sounds of their native language by mere listening. Long before they can say their first word, they can perceive critical properties of vowels and consonants in their mother tongue. Can adults also learn sounds of an ambient language from unguided (passive) exposure by unconsciously tracking the acoustic properties and variation in the surrounding speech stream? Or are adult learners entrenched in what they already know, and so learn foreign sounds better when the foreign sounds' properties are familiar from their native language experience. We tested these questions in a laboratory experiment involving two different groups of adults: native speakers of Czech and native speakers of Greek. The two languages differ in whether or not they contrast vowel sounds by the duration (length) of the vowel: Czechs use duration in their language to distinguish meanings of words, Greeks do not. In this study, after listening to various short and long foreign-language vowels, Czech adults seemed to form a "short" and a "'long" category better than Greek adults. Our tentative conclusion is that for adults learning completely new sounds from passive listening might be less likely when there is no prior knowledge of that sound characteristic, especially in the initial stages of learning.

What the Researchers Did

- The researchers created a 5-minute training program during which 25 Czech and 28 Greek adults listened to 210 various vowel sounds differing in duration.

- For half of the Czechs and half of the Greeks the variation in duration represented *two* vowel categories—a long and a short one. For the other half of each language group, the variation suggested only a *single* category.
- Listeners were tested before and after training on their ability to tell apart pairs of short and long vowels.

What the Researchers Found

All listeners became better at distinguishing the short and long vowels over time.

- How well they learned this length contrast seemed to depend on their native language.
- Czechs, whose native language uses vowel duration to contrast words, benefited from exposure to two durational categories—it helped them more accurately detect durational differences after training.
- Greeks, whose native language does not use duration to contrast words, rather unexpectedly, improved in detecting durational differences after being exposed to a *single* durational category.

Things to Consider

Intensive exposure to foreign language sounds may not be enough for all adult learners to extract new speech categories, as speech sound perception might depend on what they know from their native language.

- The present study was done with a small number of participants and the results of the statistical analysis did not unequivocally demonstrate native-language effect. This means that the findings are only indicative and should be tested in future work.

Materials and data: Materials and data are publicly available at https://osf.io/xdq79/ and http://www.iris-database.org.

How to cite this summary: Chládková, K., & Šimáčková, Š. (2020). Foreign speech sounds in adulthood: Learning from exposure seems easier when guided by similarities to the native language. *OASIS Summary* of Chládková & Šimáčková (2021) in *Language Learning*. https://oasis-database.org