Cognition 133 (2014) 408–413

Contents lists available at ScienceDirect

Cognition

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

Distributional learning has immediate and long-lasting effects

Paola Escudero [a,*], Daniel Williams [b]

[a] MARCS Institute, University of Western Sydney, Bankstown Campus, Locked Bag 1797, Penrith, NSW 2751, Australia
[b] Institute of Humanities & Creative Arts, University of Worcester, Henwick Grove, Worcester WR2 6AJ, United Kingdom

ARTICLE INFO

Article history:
Received 25 July 2013
Revised 3 July 2014
Accepted 8 July 2014
Available online 15 August 2014

Keywords:
Statistical learning
Distributional learning
Longitudinal development
Non-native sound discrimination

ABSTRACT

Evidence of distributional learning, a statistical learning mechanism centered on relative frequency of exposure to different tokens, has mainly come from short-term learning and therefore does not ostensibly address the development of important learning processes. The present longitudinal study examines both short- and long-term effects of distributional learning of phonetic categories on non-native sound discrimination over a 12-month period. Two groups of listeners were exposed to a two-minute distribution of auditory stimuli in which the most frequently presented tokens either approximated or exaggerated the natural production of the speech sounds, whereas a control group listened to a piece of classical music for the same length of time. Discrimination by listeners in the two distribution groups improved immediately after the short exposure, replicating previous results. Crucially, this improvement was maintained after six and 12 months, demonstrating that distributional learning has long-lasting effects.

1. Introduction

It is well known that young infants are able to discriminate virtually all speech sounds and that this ability declines as they become more attuned to their native language. By the end of their first year, infants' discrimination of speech sounds resembles more closely that of adults of the same native language. While the change in infants' speech perception may be related to or at least coincide with the onset of word learning, word learning alone cannot account for the change because speech perception becomes specific to the infants' native language before they can reliably distinguish words. Infants' learning of speech sound contrasts is likely to be underpinned by their sensitivity to the statistical regularity of the auditory input in their ambient spoken language (for a review, see Krogh, Vlach, and Johnson, 2013). In that respect, infants have been shown to harness relative frequency distributions in the continuous auditory signal to learn to discriminate speech sounds, a domain-general mechanism known as *distributional learning* (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002; Werker, Yeung, & Yoshida, 2012; Yoshida, Pons, Maye, & Werker, 2010).

In distributional learning experiments, listeners are presented with auditory stimuli that form a continuum and vary in equal steps along a particular acoustic-phonetic dimension. The stimuli are either presented with frequencies that constitute a *bimodal distribution*, in which tokens near the endpoints of the continuum are most frequent, or a *unimodal distribution*, in which tokens around the middle are most frequent. A bimodal distribution mimics how the speech sounds appear in a binary sound contrast in a language because they tend to be produced with properties that place them near the edges of an acoustic-phonetic continuum. For instance, Maye et al. (2002) used the contrast "da" and "ta", where "t" did not have the aspiration of English "t" but mimicked how "t" as pronounced in languages such as Spanish or Dutch. They presented eight tokens along a continuum from "da" to "ta"

* Corresponding author. Tel.: +61 2 9772 6493.
E-mail address: paola.escudero@uws.edu.au (P. Escudero).

http://dx.doi.org/10.1016/j.cognition.2014.07.002
0010-0277/

that varied in eight equal steps on the acoustic-phonetic dimension of voice onset time. Infants were exposed to either a bimodal or a unimodal distribution of these tokens. Their findings show that only infants exposed to the bimodal distribution reliably discriminated stimuli from the endpoints of the continuum during a test phase, providing support that distributional learning is potentially a powerful mechanism in phonetic learning.

Distributional learning can also lead to a generalization of a non-native sound distinction. For instance, Maye and colleagues were not only able to replicate their original finding with the "da"-"ta" contrast (Maye et al., 2008), but also found that infants exposed to a bimodal distribution of "da"-"ta" reliably discriminated endpoint tokens of "ga"-"ka", which differ on the same acoustic-phonetic dimension of voice onset time but exhibit a different place of articulation. This suggests that exposure to the "da"-"ta" bimodal distribution facilitated the discrimination of the "ga"-"ka" contrast as it is distinguished by the same phonetic feature of voice onset time.

Distributional learning may be most effective earlier in life, as 10-month-olds required longer exposure to a bimodal distribution than six- or eight-month-old infants to be able to discriminate a contrast (Yoshida et al., 2010). Additionally, although adults can exploit distributional learning for non-native sound contrasts (Gulian, Escudero, & Boersma, 2007; Hayes-Harb, 2007; Maye & Gerken, 2000; Maye & Gerken, 2001), exposure time to bimodal distributions is at least double that in infant studies (nine or five versus 2 min). However, a more recent study by Escudero, Benders, and Wanrooij (2011) demonstrates that just two minutes of exposure to bimodal distributions, as in infant studies, can lead to improvement in non-native sound discrimination, a finding which was replicated by Wanrooij, Escudero, and Raijmakers (2013). Moreover, in these two studies, the distributions were created using synthetic vowel tokens, while listeners' discrimination was tested using naturally produced vowel tokens, demonstrating that distributional learning from synthetic distributions can be generalized to naturally produced speech sounds.

Thus, infants and adults can display improved discrimination of speech sound contrasts after exposure to bimodal distributions. However, from these previous studies it is not yet clear whether distributional learning has long-lasting effects or whether it is constrained to a short laboratory session, since its working has only been shown *immediately* after exposure (Krogh et al., 2013). The present longitudinal study examines whether improvement via distributional learning is displayed over time, that is, *beyond* a single training session and over a period of 12 months.

2. Method

2.1. Participants

Participants were 79 Spanish listeners who were living in the Netherlands and learning Dutch during the 12 months of the study. They were a subset of the participants reported in Wanrooij et al. (2013). Their ages ranged between 24 and 63 years and at the time of the final test they had been living the Netherlands between one and 21 years. Listeners were assigned to one of three groups (Bimodal: $N = 24$, Enhanced: $N = 30$, or Music: $N = 25$, described below) and remained in that group across the three sessions. As shown in Table 1, the average age (AaT), age of arrival (AoA), length of residence (LoR) in the Netherlands and proficiency in the Dutch language was comparable across the three groups. The lack of differences was confirmed by one-way analyses of variance (ANOVAs) on AaT, AoA and LoR and by an analogous Kruskal–Wallis H test on the ordinal variable of Dutch proficiency scores (1 = lowest proficiency, 6 = highest proficiency), all of which showed no significant differences between the three groups (all $p > 0.1$).

2.2. Stimuli and procedure

Listeners attended three sessions over 12 months, separated by a period of six months. Sessions 1 and 2 were identical and consisted of a categorization test (pre-test), followed by exposure to vowel distributions and then a categorization test once more (post-test). In Session 3, listeners performed a categorization test once and did not receive further exposure to distributions. Exposure and test were the same as those reported in Wanrooij et al. (2013). The three groups of listeners differed only in the stimuli they heard during exposure in Sessions 1 and 2. That is, only the Bimodal and Enhanced groups were presented with two-minute long vowel distributions, whereas the Music group listened to classical music for the same amount of time.

The categorization test was a two-alternative forced-choice categorization task in an XAB format that tested listeners' categorization accuracy of the vowels in the Dutch /aː-ɑ/ contrast. On each trial, listeners heard three vowel sounds and were asked to decide whether the first sound (X) was more like the second (A) or the third (B).

The 40 X stimuli were naturally produced vowel tokens taken from a corpus of 10 male and 10 female speakers of Northern Standard Dutch (Adank, Van Hout, & Smits, 2004).[1] The average duration values were 210 ms for the /aː/ tokens (standard deviation (sd): 29) and 94 ms for the /ɑ/ tokens (sd: 19). Average first (F1) and second formant (F2) frequencies for the /aː/ tokens were 923 Hz (sd: 75) and 1552 Hz (sd: 107), respectively, for females, and 652 Hz (sd: 144) and 1424 Hz (sd: 98), respectively, for males. For the /ɑ/ tokens, average F1 and F2 values were 719 Hz (sd: 100) and 1239 Hz (sd: 168), respectively, for females, and 584 (sd: 99) and 1156 Hz (sd: 127), respectively, for males.

The two A and B auditory response options were synthetic tokens created in the *Praat* program (Boersma & Weenink, 2010) and were based on the acoustic values of natural productions of the Dutch vowels in the words "maan" (moon) and "man" (man), /aː/ and /ɑ/ respectively as reported in Pols, Tromp, and Plomp (1973). The F1 and F2 valued of the A and B stimuli are displayed in Fig. 1.

[1] The goodness of the X stimuli and the challenging nature of the task are demonstrated by 20 native Dutch listeners achieving far from ceiling performance with an average accuracy of 88% (Escudero & Wanrooij, 2010).

Table 1
Mean age at Session 1 (AaT), age of arrival (AoA) and length of residence (LoR) in the Netherlands (years) and Dutch proficiency scores per listener group. Standard deviations indicated in parentheses.

Group	*N*	AaT	AoA	LoR	Dutch proficiency
Enhanced	30	37.0 (8.0)	31.5 (6.0)	5.5 (5.6)	3.8 (2.2)
Bimodal	24	35.2 (7.4)	30.6 (7.3)	4.5 (3.5)	4.1 (2.4)
Music	25	38.4 (6.3)	33.6 (7.4)	5.0 (5.5)	4.7 (1.9)

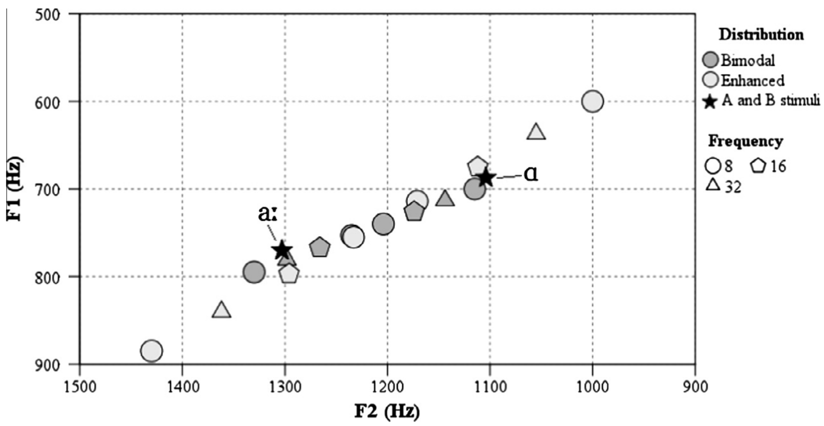

Fig. 1. First (F1) and second formants (F2) of the enhanced (light gray shapes) and bimodal (dark gray shapes) distributions. Frequency of presentation of the enhanced and bimodal tokens is represented by different shapes. Also shown are F1 and F2 of the A and B response stimuli corresponding to Dutch /aː/-/ɑ/ (black stars).

Unlike naturally produced Dutch vowels, the A and B stimuli did not differ in duration as this was at 140 ms for both stimuli. This is motivated by previous studies that show Spanish listeners rely on duration more heavily than vowel formants to discriminate these two Dutch vowels, leading to chance levels of discrimination accuracy, whereas Dutch listeners show the reverse (Escudero, Benders, & Lipski, 2009). The A and B stimuli thus allowed listeners to make their choice based on the acoustic-phonetic dimensions of vowel formants (i.e., vowel quality) rather than simply just duration.

There were two repetitions of the 20 X stimuli counterbalanced across the two A and B response options, yielding a total of 80 trials. The inter-stimulus interval (ISI) between the three sounds in each XAB-trial was 1.2 s. Listeners were told that the next trial would only appear after their response and were encouraged to respond as quickly as possible.

During the exposure phase, the Enhanced and Bimodal groups heard an enhanced or a bimodal vowel continuum, respectively, which consisted of eight synthetic vowel tokens generated using the same procedure as for the A and B stimuli[2] and based on the Dutch /aː/-/ɑ/ contrast. As shown in Fig. 1, the near-endpoints of the enhanced vowel continuum exaggerate the quality (F1 and F2 frequencies of the vowel) difference between the two Dutch vowels in comparison to the bimodal continuum. The acoustic values of the Bimodal distribution were chosen to reflect average values of the Dutch /aː/-/ɑ/ vowels (Fig. 1), while the values of the Enhanced distribution enlarged the spectral difference between the two vowels, similar to those of infant-directed speech and foreigner-directed speech (Escudero et al., 2011). We reasoned that a larger spectral difference between the vowels in the Enhanced distributions will be perceptually more salient for non-native listeners and may lead to a larger effect of training than the smaller difference in the Bimodal distribution.

Along with the number of steps in each continuum, the frequency distributions of the stimuli matched those used in previous distributional learning studies with adults and infants. That is, tokens 2 and 7 (the near-endpoints) were presented four times as often as the tokens 4 and 5 (the middle). Fig. 1 (adapted from Fig. 1 in Escudero et al., 2011) displays the F1 and F2 values of the eight vowel tokens in the enhanced and bimodal distributions and the frequency with which each token was presented.

As in previous infant studies, the exposure phase contained 128 stimuli separated by an ISI of 750 ms and lasted only two minutes. Listeners in the Bimodal and Enhanced groups were asked to listen to the vowels carefully because they would perform another test afterward, whereas participants in the Music group were asked to relax while listening to classical music and were also told that they would perform a second test afterward.

3. Results

Fig. 2 shows group accuracy scores (percentage of correct responses) for each of the five tests listeners

[2] The specific acoustic values of the stimuli in the distributions and how they were calculated can be found in Escudero et al. (2011) and Wanrooij et al. (2013).

Table 2
Improvement measures by Group (standard errors in parenthesis).

	Short 1	Short 2	Long 1	Long 2	Highest	Average
Enhanced	5.8 (1.5)	1.4 (1.4)	4.3 (1.5)	1.5 (1.9)	11.6 (1.8)	4.9 (1.2)
Bimodal	2.4 (1.6)	8.2 (1.5)	1.2 (1.6)	4.2 (2.1)	13.2 (2.0)	5.8 (1.3)
Music	0.6 (1.6)	3.3 (1.5)	−1.9 (1.6)	−1.5 (2.1)	5.8 (1.9)	1.3 (1.3)

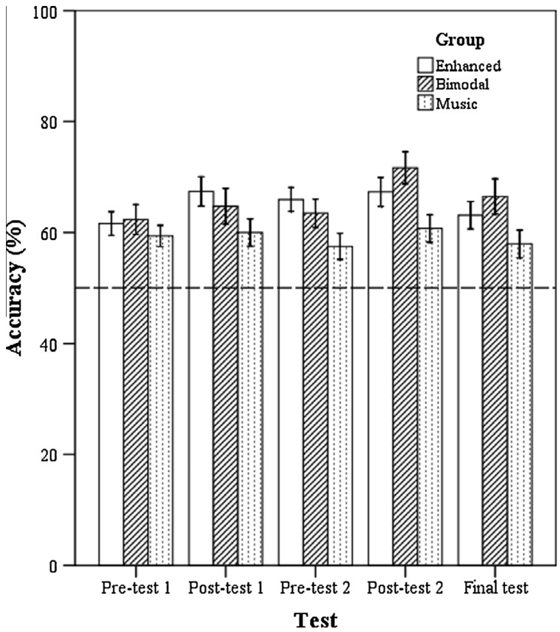

Fig. 2. Accuracy scores (%) across the five tests. Error bars represent the standard error, chance (50%) is marked by a dashed line.

performed during the 12-month period. Following Wanrooij et al. (2013), a mixed design analysis was run on accuracy scores with Test as a within-subjects factor (five levels) and Group as a between-subjects factor (Bimodal, Enhanced and Music). We found no main effect of Group ($F(2,76) = 2.52$, $p = 0.087$), supporting the overall homogeneity of the groups, and a main effect of Test ($F(4\varepsilon, 304\varepsilon, \varepsilon = 0.858) = 7.19$, $p < 0.001$),[3] indicating that the five tests had different accuracy levels. The analysis also yielded a significant Test * Group interaction ($F(8\varepsilon, 304\varepsilon, \varepsilon = 0.858) = 2.35$, $p = 0.025$), indicating that some groups improved more than others between the five tests. This interaction prompted further analysis on the amount of improvement between tests and groups to address more directly the short- and long-term effects of distributional learning (Table 2).

Short-term improvement, i.e., improvement immediately after exposure to distributions, was computed as the difference between post- and pre-tests in Sessions 1 and 2 (*Short 1* = post-test 1 minus pre-test 1, *Short 2* = post-test 2 minus pre-test 2). In order to test for differences between the three groups on these two measures, separate one-way ANOVAs were run. For Short 1 improvement, the differences between the three groups approached significance ($F(2,76) = 3.01$, $p = 0.055$) and planned comparisons show that the Enhanced group's improvement scores were significantly higher than those of the Music group ($p = 0.019$). A one-sample *t*-test revealed the Enhanced group's improvement was significantly different from zero which represents no improvement ($t(29) = 3.99$, $p < 0.001$). For Short 2 improvement, the differences were significant ($F(2,76) = 5.78$, $p = 0.005$) and planned comparisons reveal that the Bimodal group improved more than both the Enhanced ($p = 0.001$) and Music groups ($p = 0.023$). A one-sample *t*-test showed that the Bimodal group's Short 2 improvement was significantly different from zero ($t(23) = 4.47$, $p < 0.001$).

Long-term improvement, was computed as the difference between improvement after six and 12 months, respectively (*Long 1* (after six months) = pre-test 2 minus pre-test 1, *Long 2* (after 12 months) = final test minus pre-test 1). One-way ANOVAs on the long-term improvement scores revealed significant differences between the groups for Long 1 improvement ($F(2,76) = 4.15$, $p = 0.019$) but not for Long 2 improvement ($F(2,76) = 1.80$, $p = 0.173$). Planned comparisons show that the Enhanced group improved more than the Music group for the Long 1 improvement ($p = 0.005$) and, though non-significant, the trend of Long 2 improvement was most reliable for the Bimodal group ($p = 0.062$). Additional one-sample *t*-tests revealed significant differences from zero (i.e., no improvement) for the Enhanced group's Long 1 improvement scores ($t(29) = 2.55$, $p = 0.016$) and the Bimodal group's Long 2 improvement scores ($t(23) = 2.11$, $p = 0.046$).

Lastly, as a general measure of improvement between tests – irrespective or short- or long-term effects – *highest improvement* was examined, which was the highest score a listener achieved at any point in 12 months to their pre-test 1 score. A one-way ANOVA on these scores showed significant differences between the three groups ($F(2,76) = 4.03$, $p = 0.022$) and planned comparisons revealed that both the Bimodal and Enhanced groups' scores were significantly greater than those of the Music group ($p = 0.009$ and $p = 0.032$, respectively).

4. Discussion

Across the five tests in the 12-month period, only the groups exposed to sound distributions showed reliable short- and long-term improvement. Specifically, these groups improved immediately after the first (Enhanced) and second (Bimodal) exposure sessions, while the Music group did not improve after either session, which is in line

[3] Huynh-Feldt corrections were applied to reduce the number of degrees of freedom by ε, as variance in the accuracy scores for the five tests was not equal (Mauchly's test of sphericity was violated).

with previous results showing that listening to an enhanced or a bimodal distribution for two minutes leads to immediate improvement in non-native sound discrimination (Escudero et al., 2011; Wanrooij et al., 2013).

The short-term results also suggest that the patterns of improvement may not be uniform, given that the immediate improvement was reliable for the Enhanced group in Session 1 but for the Bimodal group in Session 2. This can be explained by the type of stimuli in the two distributions. That is, exposure to the exaggerated properties of the non-native contrast in the enhanced vowel distribution may have led to more rapid or efficient distributional learning for the Enhanced group, whereas more than one session with the bimodal distribution was required to bring about a reliable short-term improvement for the Bimodal group.

Regarding long-term improvement, the groups exposed to sound distributions achieved higher average accuracy and improvement scores than the Music group, clearly indicating a general effect of distributional learning on non-native sound discrimination across the five tests. The distribution groups maintained similar accuracy levels at the beginning of Session 2 (pre-test 2) to those at the end of Session 1 (post-test 1), including the significant improvement made by the Enhanced group after exposure in Session 1, which suggests that accuracy remained broadly similar six months after exposure and that there was a *long-lasting* effect of distributional learning. Additionally, from the first to final test over the 12 months of the study, the Bimodal group showed a reliable improvement, demonstrating a *consistent and long-lasting* effect of distributional learning.

Interestingly, the Enhanced group's accuracy improvement from Session 1 did not increase after further exposure to the distribution in Session 2 because their initial improvement was simply maintained. This may suggest that improvement may plateau up to a certain accuracy level. This is perhaps not too unexpected as this particular vowel contrast is notoriously difficult for Spanish learners of Dutch (Escudero et al., 2009).

While it cannot be ruled out that natural exposure to Dutch outside the testing sessions may have contributed to the *lasting* effect of learning in the Enhanced and Bimodal groups in some way, it was clearly not responsible for the improvement in non-native sound discrimination. This is because the Music group, who was not exposed to sound distributions during the testing sessions but was exposed to Dutch outside of the laboratory, showed no signs of learning throughout the 12 months.

The present results have further implications on distributional learning as a statistical learning mechanism. In their review of distributional learning, Werker et al. (2012) highlight two main limitations. Firstly, distributional learning seems to be less effective for the learning of difficult non-native sound contrasts that vary on more than one acoustic-phonetic dimension (Cristià, McGuire, Seidl, & Francis, 2011; Goudbeek, Cutler, & Smits, 2008). Secondly, distributional learning appears to be a much more important learning mechanism earlier in life because improvements in young infants' discrimination accuracy are quantitatively larger and require shorter exposure to sound distributions than for older infants and adults (Hayes-Harb, 2007; Maye & Gerken, 2000; Yoshida et al., 2010).

With respect to the first limitation, the Dutch /aː/-/ɑ/ contrast is notoriously difficult for Spanish learners to discriminate, even after many years of learning Dutch (Escudero & Wanrooij, 2010). This is because Spanish listeners rely more heavily on an acoustic cue that is less relevant for this contrast than on the primary cues of F1 or F2 (Escudero et al., 2009). While the distributions in the training phase of the present study did indeed vary on more than one acoustic-phonetic dimension, i.e., F1 as well as F2, they did not vary on the dimension of duration. By keeping duration constant in the continua, we were able to draw listeners' attention to the dimensions of F1 and F2, which are successfully used by native Dutch listeners to discriminate the contrast (Escudero et al., 2009). Thus, in addition to the number of acoustic-phonetic dimensions, the effectiveness of distributional learning may depend on the relative salience of particular dimensions for distinguishing a contrast. After all, successful discrimination of speech sounds depends on listeners' relative sensitivity to the available acoustic cues found in variable speech input (Escudero & Boersma, 2004; Iverson et al., 2003). While the present study shows that distributional learning is possible for contrasts differing on multiple acoustic-phonetic dimensions, listeners were directed to specific dimensions that are perceptually most relevant (i.e., F1 and F2) by diminishing the competition from an acoustically more salient but perceptually less relevant or reliable acoustic cue (e.g., vowel duration).

With respect to the second limitation mentioned above, domain-general statistical learning mechanisms, such as distributional learning, appear prominent in early phases of language acquisition and come to co-exist with more sophisticated or specialized language learning mechanisms, such as word learning (Emberson, Liu, & Zevin, 2013). The task in the test phase of the present study required listeners to make phonological/categorical decisions because none of the XAB stimuli on each individual trial were physically identical and the use of a long ISI (1.2 s) enforced a higher level of abstraction. This type of design differs markedly from those in most other studies on distributional learning, which repeat the same auditory stimuli in training and test phases (e.g., Maye et al., 2002; Maye et al., 2008; Werker et al., 2012; Yoshida et al., 2010) and therefore potentially reduce the burden on speech processing. It remains to be tested whether larger improvements will be observed for infants than for adults in performing the task in the present study. This vowel contrast is known to be particularly troublesome for Spanish adults to discriminate (Escudero et al., 2009), which is reflected in the relatively modest discrimination scores in Fig. 2 as well as the accuracy improvement scores in Table 2. As adults may have other processes available to them, especially with regard to abstract categories, it is possible that they may harness distributional learning in different ways from infants.

Regarding the maintenance of the distributional learning demonstrated by the present longitudinal study, it could be expected that neither the Enhanced nor Bimodal groups would maintain their improvements over time because exposure to the distributions is required for listeners to shift their attention to F1 and F2 dimensions. However, the Enhanced or Bimodal groups maintained their improvements after six and 12 months without additional stimulation in the intervening periods. Thus the frequency with which specific stimuli were presented at the laboratory (i.e., our experimental distributions) seems to have triggered listeners' attention to the relevant acoustic dimensions of F1 and F2 which transferred across sessions that are months apart. The maintenance of improvement suggests that the shifts in listeners' attention to certain relevant acoustic-phonetic dimensions were not temporary but examples of consolidation of learning. It appears therefore that distributional learning is a mechanism that can impact long-term learning (Krogh et al., 2013).

In sum, the present study once again finds significant accuracy improvement in non-native sound discrimination immediately after two-minute exposure to a bimodal or enhanced distribution. Most importantly, accuracy improvements were maintained after six and 12 months, demonstrating an effect beyond a single laboratory session. The long-lasting nature of distributional learning revealed in this study thus shows that this mechanism can be harnessed in important learning processes such as the development of speech sound discrimination.

Acknowledgements

This research was supported by grant 275.75.005 from the Netherlands Organization for Scientific Research (NWO) awarded to the first author. We thank Kateřina Chládková and Polina Vasiliev and two anonymous reviewers for their insightful comments.

References

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America, 116*, 1729–1738.

Boersma, P., Weenink, D. (2010) Praat: Doing phonetics by computer. <http://www.praat.org> Last viewed: July 2010.

Cristià, A., McGuire, G., Seidl, A., & Francis, A. (2011). Effects of the distribution of cues on infants' perception of speech sounds. *Journal of Phonetics, 39*, 388–402.

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition, 128*, 82–102.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German and Spanish listeners. *Journal of Phonetics, 37*, 452–465.

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.

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

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native and L2 vowel perception. *Language and Speech, 53*, 343–365.

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

Gulian, M., Escudero, P., Boersma, P. (2007) Supervision hampers distributional learning of vowel contrasts. In *Proceedings of the international congress of phonetic sciences* (pp. 1893–1896).

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

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47–B57.

Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology, 3*, 598.

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

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *BUCLD24 Proceedings* (pp. 522–533). Somerville, MA: Cascadilla.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–111.

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America, 53*, 1093–1101.

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics, 41*, 307–319.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science, 21*, 221–226.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy, 15*, 420–433.