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

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).