Infant Behavior & Development 43 (2016) 44–57

Contents lists available at ScienceDirect

Infant Behavior and Development

Full length article

Semantics guide infants' vowel learning: Computational and experimental evidence

S.M.M. ter Schure [a],*, C.M.M. Junge [b], P.P.G. Boersma [a]

[a] Amsterdam Center for Language and Communication, University of Amsterdam, the Netherlands
[b] Department of Social and Behavioral Sciences, Utrecht University, the Netherlands

ARTICLE INFO

Article history:
Received 6 July 2015
Received in revised form
27 November 2015
Accepted 15 January 2016
Available online 28 April 2016

Keywords:
Infant cognition
Associative learning
Auditory discrimination
Phoneme acquisition
Categorization
Computational modeling

ABSTRACT

In their first year, infants' perceptual abilities zoom in on only those speech sound contrasts that are relevant for their language. Infants' lexicons do not yet contain sufficient minimal pairs to explain this phonetic categorization process. Therefore, researchers suggested a bottom-up learning mechanism: infants create categories aligned with the frequency distributions of sounds in their input. Recent evidence shows that this bottom-up mechanism may be complemented by the semantic context in which speech sounds occur, such as simultaneously present objects. To test this hypothesis, we investigated whether discrimination of a non-native vowel contrast improves when sounds from the contrast were paired consistently or randomly with two distinct visually presented objects, while the distribution of speech tokens suggested a single broad category. This was assessed in two ways: computationally, namely in a neural network simulation, and experimentally, namely in a group of 8-month-old infants. The neural network, trained with a large set of sound–meaning pairs, revealed that two categories emerge only if sounds are consistently paired with objects. A group of 49 real 8-month-old infants did not immediately show sensitivity to the pairing condition; a later test at 18 months with some of the same infants, however, showed that this sensitivity at 8 months interacted with their vocabulary size at 18 months. This interaction can be explained by the idea that infants with larger future vocabularies are more positively influenced by consistent training (and/or more negatively influenced by inconsistent training) than infants with smaller future vocabularies. This suggests that consistent pairing with distinct visual objects can help infants to discriminate speech sounds even when the auditory information does not signal a distinction. Together our results give computational as well as experimental support for the idea that semantic context plays a role in disambiguating phonetic auditory input.

1. Introduction

Languages vary in their phoneme inventories. Hence, two sounds that differ in their phonetic characteristics may belong to the same phoneme category in one language but to two different phoneme categories in another. It is therefore vital that infants learn which sounds they should perceive as belonging to the same phoneme in their native language and which

---

* Corresponding author at: University of Amsterdam, Amsterdam Center for Language and Communication, Spuistraat 210, 1012 VT Amsterdam, the Netherlands.

E-mail address: sophieterschure@gmail.com (S.M.M. ter Schure).

http://dx.doi.org/10.1016/j.infbeh.2016.01.002
0163-6383/

they should perceive as distinct phonemes (Cutler, 2012; Kuhl et al., 2008). For example, in English, there is a difference in voice onset time between the two instances of /p/ in 'perceptual', but an English child will learn to ignore this difference, whereas she will learn not to ignore the meaningful difference between the voice onset times in the initial sounds in 'pear' and 'bear'. Despite the apparent difficulty of this learning task, infants have already learned their native phonetic contrasts before their first birthday (vowels by six months: Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; consonants by ten months: Werker & Tees, 1984). It remains unclear, however, *how* infants start building such optimally restricted categories, that is, how they learn to focus on only those contrasts that are relevant for their native language (Werker & Tees, 1984). In the past decades, researchers have focused on two possible mechanisms that could account for this phonetic learning. One account focuses on infants' sensitivity to the frequency distributions of sounds (e.g., Maye, Werker, & Gerken, 2002), while another focuses on the possibility that infants learn phonetic contrasts from contrastive lexical items (e.g., Feldman, Griffiths, Goldwater, & Morgan, 2013).

1.1. Distribution-driven learning of perception

Although it was initially hypothesized that infants learn sounds from contrastive meanings, i.e., *minimal pairs* (Werker & Tees, 1984), this idea was challenged by the finding that infants are sensitive to language-specific phonetic detail at an age at which they hardly know any words, let alone enough minimal pairs to allow for all contrasts (e.g., Caselli et al., 1995; Dietrich, Swingley, & Werker, 2007). Instead, current theories of first language acquisition argue that perceptual reorganization occurs mainly through bottom-up learning from speech input (e.g., Kuhl et al., 2008; Pierrehumbert, 2003; Werker & Curtin, 2005). One such learning mechanism is that infants keep track of the frequency distributions of sounds in their input, and create categories for these speech sounds accordingly. For example, on an F1 (first formant) continuum from 400 to 800 Hz, Spanish distinguishes just two front vowel phonemes (/e/, /a/), with prototypical instances of /e/ and /a/ occurring more frequently than sounds in between. Observing this two-peaked frequency distribution, a Spanish infant could create two phonemes in her personal inventory. Portuguese, on the other hand, has three categories (/e/, /ε/, /a/) on the same continuum, hence a three-peaked distribution, so that a Portuguese infant can create three phoneme categories in the same area where a Spanish infant creates only two.

Most theories argue that infants' phonetic categories emerge from observing these frequency peaks in their input, while the adult perceptual system may also incorporate feedback from other levels of representation (e.g., Pierrehumbert, 2003: 138; Werker & Curtin, 2005). In this view, infants develop phonetic categories before they start to store word forms and add meaning. This entails that infants' initial phonetic perception is not affected by the auditory or visual contexts of the speech sounds. There is computational as well as experimental support for the view that native phonetic categorization begins with infants' sensitivity to such phonetic distributions, without requiring higher-level linguistic knowledge.

Computational modeling shows that language-specific perceptual behavior can arise in a neural network containing nothing more than a general learning mechanism that connects particular sensory inputs to patterns of activation at a higher level (Guenther & Gjaja, 1996). The distribution of sounds in the output of adult speakers (which is the chief input for infants) is determined by the number of phoneme categories in the language that they speak. If one exposes a neural network to these sounds, certain patterns of activation emerge that correspond to the peaks in the distributions. Recent models have tested whether infant-directed speech indeed contains sufficiently clear peaks for such a distributional learning mechanism to succeed. Indeed, this appears to be the case for both consonants (at least for VOT contrasts, McMurray, Aslin, & Toscano, 2009) and vowels (Vallabha, McClelland, Pons, Werker, & Amano, 2007; Benders, 2013). In short, computational models of first language acquisition provide evidence that infants' input contains sufficient information to learn phonetic contrasts without requiring lexical knowledge.

Experimental evidence shows that real infants can indeed learn a novel phonetic contrast from only auditory input, even within several minutes (Cristia, McGuire, Seidl, & Francis, 2011; Maye et al., 2002; Maye, Weiss, & Aslin, 2008; Yoshida, Pons, Maye, & Werker, 2010; Wanrooij, Boersma, & van Zuijen, 2014). For example, Maye et al. (2002, 2008) presented infants with a continuum of a phonetic contrast. In a 2.5-min training phase, one group of infants heard a large number of stimuli from the center of this continuum and fewer stimuli from the two edges (a one-peaked frequency distribution). Another group of infants heard mostly stimuli from near the edges of the continuum and fewer from the center (a two-peaked distribution). Subsequently, all infants were tested on their discrimination of the phonetic contrast. Infants who had heard the two-peaked distribution during training discriminated the contrast better than infants who had heard the one-peaked distribution.[1] Apparently, the shape of the phonetic distribution that infants hear rapidly affects their sound categorization.

---

[1] Although true experimental support for the effect of training distribution can only follow from a direct comparison between two-peaked and one-peaked groups, many distributional learning studies only report a significant discrimination within the two-peaked group and an absence of significance in the one-peaked group. As the number of such results has increased, the existence of the effect has become more plausible. Also, some studies do report significant group differences (Maye et al., 2008; Wanrooij et al., 2014). Together, we take this as sufficient evidence for an effect of distributional learning.