Cognition 128 (2013) 82–102

Contents lists available at SciVerse ScienceDirect

Cognition

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

Is statistical learning constrained by lower level perceptual organization?

Lauren L. Emberson [a,*], Ran Liu [b], Jason D. Zevin [c,d]

[a] Brain and Cognitive Sciences Department, University of Rochester, United States
[b] Psychology Department, Carnegie Mellon University, United States
[c] Sackler Institute for Developmental Psychobiology, Weill-Cornell Medical School, United States
[d] Haskins Laboratories, United States

ARTICLE INFO

Article history:
Received 28 September 2010
Revised 15 November 2012
Accepted 21 December 2012
Available online 22 April 2013

Keywords:
Language learning
Cognitive development
Statistical learning
Bayesian learning
Speech perception
Speech categorization
Perceptual variability

ABSTRACT

In order for statistical information to aid in complex developmental processes such as language acquisition, learning from higher-order statistics (e.g. across successive syllables in a speech stream to support segmentation) must be possible while perceptual abilities (e.g. speech categorization) are still developing. The current study examines how perceptual organization interacts with statistical learning. Adult participants were presented with multiple exemplars from novel, complex sound categories designed to reflect some of the spectral complexity and variability of speech. These categories were organized into sequential pairs and presented such that higher-order statistics, defined based on sound categories, could support stream segmentation. Perceptual similarity judgments and multi-dimensional scaling revealed that participants only perceived three perceptual clusters of sounds and thus did not distinguish the four experimenter-defined categories, creating a tension between lower level perceptual organization and higher-order statistical information. We examined whether the resulting pattern of learning is more consistent with statistical learning being "bottom-up," constrained by the lower levels of organization, or "top-down," such that higher-order statistical information of the stimulus stream takes priority over perceptual organization and perhaps influences perceptual organization. We consistently find evidence that learning is constrained by perceptual organization. Moreover, participants generalize their learning to novel sounds that occupy a similar perceptual space, suggesting that statistical learning occurs based on regions of or clusters in perceptual space. Overall, these results reveal a constraint on learning of sound sequences such that statistical information is determined based on lower level organization. These findings have important implications for the role of statistical learning in language acquisition.

1. Introduction

Starting in infancy, and continuing into adulthood, humans are highly sensitive to statistical regularities in their environment. From these regularities, it is possible to learn a large amount about the structure of the world without explicit feedback or innate knowledge. In a little over a decade, statistical learning has been implicated in the processing and acquisition of a variety of perceptual and cognitive skills, including knowledge of causal structure and human action (Baldwin, Anderrson, Saffran, & Meyer, 2008; Oakes & Cohen, 1990, 1995; Sobel, Tenenbaum, & Gopnik, 2004), visual processing (Brady & Oliva, 2008; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Yuille & Kersten, 2006), and, the focus of this paper,

* Corresponding author. Address: Brain and Cognitive Sciences Department, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268, United States. Tel.: +1 607 342 1690; fax: +1 585 442 9216.
E-mail address: lemberson@bcs.rochester.edu (L.L. Emberson).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.12.006

language learning (Goldwater, Griffiths, & Johnson, 2009; Saffran, Aslin, & Newport, 1996).

It is common practice for the learning tasks and models that populate this literature to investigate learning from statistical information at a single level of description. Studies either investigate statistical learning available at a relatively low level of the sensory signal (e.g. acquiring the internal structure of categories through distributional statistics) or characterize relationships at higher levels of description (e.g. segmenting streams of categorizable units, such as syllables, into higher-order structures, such as words). As a consequence, most of the existing statistical learning literature implicitly assumes that statistical information at lower levels is resolved *prior* to learning of statistics at higher levels of description. However, developmental trajectories suggest that this assumption is not always borne out and that there may be substantial interactions between statistical information at multiple levels of description. These interactions have important implications for our understanding of the specific mechanisms that support learning and development at each level.

In the current paper, we examine how statistical learning proceeds when information at lower and higher levels of description must be resolved *simultaneously*, a situation we argue is more akin to the natural task faced by language learners. The lower level learning challenge we investigate is the acquisition of auditory categories. Spoken language learners face the problem of having to categorize acoustically variable vocal utterances into functionally equivalent items (e.g. phonetic categories). The higher level learning challenge we chose to investigate is the acquisition of transitional probabilities defined *across* auditory categories that contain multiple acoustically-varying exemplars. To this end, we used a set of auditory categories (Wade & Holt, 2005), each containing acoustically-varying non-speech exemplars that are novel to adult listeners. Participants heard familiarization streams characterized by transitional probabilities defined across these *experimenter-defined categories*. We then probed the nature of statistical learning resulting from this exposure. We contrast two broad theoretical possibilities:

(1) statistical learning is primarily "top-down" – that is, statistics at higher levels can be optimally learned (see Section 3), regardless of how lower level information is processed;
(2) statistical learning is primarily "bottom-up" – that is, the processing of lower level information determines the statistics of higher levels regardless of whether those statistics are optimal at higher levels.

1.1. Statistical learning across levels of description in language

Statistical information has the potential to aid language learning at many different levels of description, including speech sound categorization, word segmentation and lexical development, and syntactic processing (for reviews, see Romberg & Saffran, 2010; Saffran & Thiessen, 2007). Most of the research identified by the term "statistical learning" has focused on the use of transitional probabilities to accomplish word segmentation and lexical development. To illustrate, take the phrase "pretty baby" (Saffran et al., 1996) which would typically be produced as a continuous utterance /prI'tibeI'bi/. In this phrase, as well as the ambient language, transitional probabilities—the probability that one will encounter Y given X—and other statistical regularities (e.g. co-occurrence frequency) are higher for syllables that cohere to form a word (e.g. "pre" and "ty") than syllables that cross word boundaries (e.g. "ty" and "ba"). Not only can infants use transitional probabilities to segment words from a speech stream (Aslin, Saffran, & Newport, 1998; Saffran et al., 1996), but they are also more likely to use syllables linked by high transitional probability as lexical labels (Graf-Estes, Evans, Alibali, & Saffran, 2007). Thus, learning from statistical regularities likely contributes to lexical development, characterized in part by the vocabulary explosion beginning around 14-months (Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006; Romberg & Saffran, 2010; Saffran & Thiessen, 2007).

However, the experiments that populate this literature and the corresponding models typically ignore a key challenge infants face in dealing with natural speech input: transitional probabilities are necessarily accumulated over many different utterances of the same phrase (e.g. "pretty baby" or /prI'tibeI'bi/), and it is well known that, even within the productions of a single speaker, there exists a large amount of acoustic variability across the multiple utterances of any given linguistic unit (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Peterson & Barney, 1952).[1] As such, the utterance "pretty baby" can be described as a sequence of unique exemplars drawn from the functional categories that specify the words at a more abstract, linguistic level of description.

Most statistical learning studies eliminate the perceptual variability across successive experiences by employing a corpus of acoustically identical repetitions of sounds (e.g. Saffran et al., 1996). Thus, this work includes the assumption that acoustic variation is already resolved, likely through the process of speech categorization—a phenomenon in which variable utterances within the same functional speech categories are treated equivalently—before transitional probabilities across the speech categories are learned.

Although literate adults can easily transcribe words and phrases as strings of symbols reflecting these functional categories, it is not at all clear that these are the perceptual abilities that infants bring to language learning (e.g. that might enable them to treat the /i/ at the end of "pretty" as an instance of the same category as the /i/ at the end of "baby," or even as the end of "pretty" spoken by a different speaker). This issue is complicated by broader questions in the field of whether phonemes are the appropriate "unit" of speech processing. In the case of most speech sounds, there is a lack of sufficient acoustic cues available to reliably categorize and discriminate phonemes (Lotto, 2000; Peterson & Barney, 1952; Shankweiler,

[1] While this paper focuses on acoustic variability, there are many other sources of information that vary across utterances including contextual and visual information and interaction with a caregiver. All of these factors have been shown to modify cognitive processing in infancy and thus variation of these factors will alter the informational content of each utterance.