Predicting Individual Variation in Language From Infant Speech Perception Measures

Alejandrina Cristia  
Laboratoire de Sciences Cognitives et Psycholinguistique and  
Max Planck Institute for Psycholinguistics

Amanda Seidl  
Purdue University

Caroline Junge  
University of Amsterdam

Melanie Soderstrom  
University of Manitoba

Peter Hagoort  
Max Planck Institute for Psycholinguistics and Radboud University Nijmegen

There are increasing reports that individual variation in behavioral and neurophysiological measures of infant speech processing predicts later language outcomes, and specifically concurrent or subsequent vocabulary size. If such findings are held up under scrutiny, they could both illuminate theoretical models of language development and contribute to the prediction of communicative disorders. A qualitative, systematic review of this emergent literature illustrated the variety of approaches that have been used and highlighted some conceptual problems regarding the measurements. A quantitative analysis of the same data established that the bivariate relation was significant, with correlations of similar strength to those found for well-established non-linguistic predictors of language. Further exploration of infant speech perception predictors, particularly from a methodological perspective, is recommended.

Research over the past five decades demonstrates that infants tune their speech perception abilities to language-specific properties of the speech they hear before their first birthday (for a recent review, see Werker & Gervain, 2013). Although most of this past work has sought to pinpoint milestones by using group averages, increasing attention is being paid to correlations between behavioral and neurophysiological measures gathered within the 1st year of life, on the one hand, and individual differences in language development, on the other (e.g., see the collection in Colombo, McCordle, & Freund, 2008; Colombo et al., 2008). The study of individual variation in infancy is far from new. For example, habituation and dishabituation measures have been used to assess the fit of different models of the architecture of cognitive skills (e.g., Bornstein, 1998) and to predict language, IQ, and educational outcomes longitudinally, even after several other variables have been controlled for (e.g., Bornstein et al., 2006; see also the Discussion section). Here, we discuss the potential of infant speech processing measures to predict variation in language outcomes over this backdrop of the larger field of cognitive development.

Evaluating this emergent strand of research is important for two reasons. First, from a theoretical perspective, focusing on individual differences in speech perception may provide crucial insights into the relative merit of competing theoretical models. For instance, there are two large classes of theories bearing on the emergence of phonology in the 1st year of life (see review in Räsänen, 2012). According to one, infants begin to learn phonology by determining which speech sound categories (which we call phones) their ambient language uses, and they later employ these units to build word forms

We are grateful to our colleagues at the Laboratoire de Sciences Cognitives et Psycholinguistique, and at both the Psychology of Language and the Genetics of Language Departments at the MPI for insightful discussions. We thank particularly the reviewers for their extremely constructive comments. Some of this work was done with the institutional support of ANR-10-LABX-0087 and ANR-10-IDEX-0001-02.

Correspondence concerning this article should be addressed to Alejandrina Cristia, Laboratoire de Sciences Cognitives et Psycholinguistique, 29 rue d’Ulm, 75005, Paris, France. Electronic mail may be sent to alecristia@gmail.com.

© 2013 The Authors  
Child Development © 2013 Society for Research in Child Development, Inc.  
All rights reserved. 0009-3920/2014/8504-0002  
DOI: 10.1111/cdev.12193
(the acoustic part of the word, which is not necessarily associated with word meaning). The other class of theories proposes that the groundwork of early phonological acquisition is accomplished by encoding words or word forms, which in turn aid in the development of the phone inventory. Both models assume that infants have some knowledge at both the level of phones and the level of word forms, but they differ crucially in the hypothesized interaction between the two levels. The best way to adjudicate between the two models would be to assess phone-level knowledge and word-form-level knowledge at multiple points in development for the same infants, and determine which variables best predict the others. Such a solution can only be possible if meaningful individual differences at each of these levels can be measured. The present review speaks to the question of whether variability in infant speech processing measures indeed reflects meaningful individual differences, so that they may eventually be used to disambiguate competing theoretical models.

Second, from a clinical perspective, having an accurate measurement of stable individual differences in typically developing infants may provide important insights allowing for the early detection of speech-language delays, disorders, and impairments. While the importance of early intervention is widely understood, it is difficult to diagnose language impairments until the age of 4 years, or even later (Leonard, 2000). Reliable infant predictors of specific language impairment (SLI), dyslexia, and/or autism spectrum disorders (ASDs) would enable earlier interventions.

In view of the important theoretical and applied questions that could be addressed through measures of language development in individual infants, it is key to evaluate this emergent literature and establish its strengths and weaknesses. First, we report on a systematic review first qualitatively and then using meta-analytic methods. The latter allows us to estimate statistically the strength of the bivariate relation between language and specific speech perception measures gathered in the 1st year of life and a measure of language development (i.e., vocabulary size) sampled concurrently or longitudinally. We further compare these effect sizes with those of well-established predictors. Second, we lay out the conceptual and empirical challenges facing this individual differences approach applied to language. We argue that a theoretical interpretation in terms of simple continuity or causality is premature, and that making the jump from correlation to causality will likely require multidisciplinary approaches and improvement of the measurement instruments. Finally, we examine the complexities that may exist when attempting to apply these measures to the early diagnosis or identification of individuals at risk for speech–language delays, disorders, and impairments.

**Predicting Language from Infant Speech Perception Skills**

As language emerges from the interaction of many factors, the later one starts to track individual variation, the more complex the interdependencies are likely to be. Thus, we concentrated on measures of speech perception gathered from infants 12 months and younger. In addition, we targeted this age group because early detection of communicative impairments is a prerequisite for applying early intervention.

For the same reasons, we prioritized tasks that plausibly reflected language-specific knowledge, by which we mean knowledge that infants have acquired from the ambient speech, rather than tasks that are more likely to capture general cognitive or auditory skills (but see the later discussion regarding the impossibility of designing "pure" measures). These measures hold the promise of being more informative to the theoretical and applied enterprises mentioned earlier.

We focused our attention on tasks that could plausibly tap language-specific knowledge of three fundamental linguistic levels whose fast-paced development in the 1st year has been well described: (a) the level of phones (e.g., involved in distinguishing between /dat/ and /dat/; the distinction between speech sounds, without assuming an abstract and rich phonemic representation), (b) word forms (e.g., involved in the recognition of *mommy* as a speech sequence often encountered in the past; we concentrate on word forms without assuming that these map onto actual words that have a referent), and (c) prosody (e.g., involved in the chunking of utterances into smaller units; we concentrate on suprasegmentals).

Inspecting all these levels was important for two reasons. First, as it is a common theoretical position that infants process speech using precursors of the adult linguistic representations, they can be considered as useful theoretical constructs with which to organize our analysis. Second, it is necessary to capture all three levels because their development might not be independent. For example, a laboratory learning study shows that 9-month-old English learners exposed to /dat/ paired with one object
Qualitative Overview

This overview organizes previous research on how measures at each of the three levels predict vocabulary size in childhood (see below for details on the literature selection). Instead of being exhaustive, it highlights key methodological and conceptual questions.

Phones. There is considerable variability in how this level has been studied. For instance, much of this work employs the conditioned head turn procedure (CHT). In CHT, infants are trained to turn their heads in response to a native or nonnative sound change by providing a contingent reward, and training this contingent response and/or testing as a whole can be made to stop when infants reach a certain criterion. Therefore, there are several possible measures that can be inspected. For instance, Talay-Ongan (1996) predicted later vocabulary through a binary, and rather strict, classification. Infants “passed” the test if they responded with a head turn to 80% of the sound changes, and refrained from producing a head turn in at least 80% of the nonchange trials; otherwise, they were classified as having “failed.” In another CHT study, Tsao, Liu, and Kuhl (2004) reported a negative correlation between the number of trials a child needed to achieve contingency training criterion and vocabulary size at 13 and 24 months, in a group of American English-learning infants tested on their ability to discriminate a difficult nonnative contrast (Finnish [u–y]). Fewer trials to criterion can be interpreted in at least two ways: as faster learning of the contingency (a nonlinguistic skill) or as easier discrimination of the phones (an auditory and/or linguistic skill). In this study, there was also a positive correlation between percent correct head turns during the test phase and expressive vocabulary size at some of the ages. It is difficult to understand why correlations were significant at some ages and for some measures (trials to criterion or percent correct) but not others. Note, moreover, that although Tsao et al. attribute these correlations to a more advanced speech perception mechanism, negative correlations instead could also have been expected, as it is widely believed that as infants accrue experience with their native language, their discrimination of native contrasts improves, whereas discrimination of nonnative sound contrasts deteriorates simultaneously between 6 and 12 months of age (see the recent summary in Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005).

Other studies by the same lab have circumvented this ambiguity by testing the same group of infants on two contrasts, one that is native and the other not. Kuhl and colleagues document an inverse correlation between discrimination of the two contrasts within the same infants, both behaviorally and electrophysiologically (Kuhl et al., 2005; Kuhl et al., 2008). In both studies, vocabulary size correlated positively with sensitivity for a native sound contrast, but negatively with nonnative sensitivity. In principle, this pattern of results cannot be explained by reference to general auditory acuity, as, all else being equal, acuity should impact native and nonnative contrasts equally. However, the native and nonnative contrasts used here were markedly different in acoustic salience (e.g., in Kuhl et al., 2005, the native syllables were [a-pa], whereas the nonnative syllables were a Mandarin alveolo-palatal fricative and affricate [i-tsi]), which leaves open the possibility that infants used different auditory skills for these two contrasts. Nevertheless, the same general pattern of results was obtained in studies that used a better matched set of native and nonnative speech contrasts stimuli differing in voice onset time (Conboy, Rivera-Gaxiola, Klarmann, Akseylu, & Kuhl, 2005; Conboy, Sommerville, & Kuhl, 2008). These findings provide stronger support for the idea that the correlation found cannot be reduced to general auditory perceptual salience.

The negative correlation between nonnative discrimination and later language outcome is worthy of further consideration. Continued nonnative sensitivity in older infants is generally interpreted as a sign of immaturity. For example, Jansson-Verkasalo, Ruusuvirta, Huotilainen, Alku, and Kushnerenko (2010) reach a similar conclusion, having found a negative correlation between nonnative vowel discrimination in late infancy (as indexed by a mismatch response [MMR]) and vocabulary outcomes a year later in a sample including both full-term and preterm infants. Another event-related potential (ERP) study has suggested a more precise explanation of that negative correlation. Rivera-Gaxiola, Klarmann, Garcia-Sierra, and Kuhl (2005) found that 11-month-olds who went on to have larger vocabularies at 30 months showed initially different ERP patterns for native (a later N250-500 hypothesized to relate to phonetic processing) and
nonnative speech contrasts (an earlier P150-250 
thought to relate to acoustic processing). In con-
trast, 11-month-olds with smaller vocabularies at
30 months exhibited the same late ERP pattern
(N250-550) for both types of contrasts. The authors
interpret this similarity as indicating that infants
with poorer future language skills still treat irrele-
vant acoustic details as phonetically relevant at
11 months.

In short, tasks tapping sound discrimination are
unique in their use of well-matched sound contrasts. 
Undoubtedly, further work is needed to understand
the mechanisms through which such associations
arise, as argued more extensively in Section 3. It
may also be relevant to point out that most work at
this level has emerged from a single group, and thus
independent replications are desirable.

Words. Most studies reported in our analysis
focus at this level on word forms, but there are two
studies that do not: The predictor in Singer (2008)
came from own-name preference in noise, and that
in Swingley (2005) came from a preference for
correctly pronounced highly frequent words (e.g., been,
“leg” in Dutch) over mispronounced versions of the
same words (e.g., beem). The other studies use a
type of segmentation task in which infants are ini-
tially exposed to the word forms spoken in isola-
tion and subsequently hear passages that either
include or do not include the familiarized words.
Typically, longer attention times during passages
containing the familiarized words versus unfamiliar
words suggest successful recognition of the word
forms from running speech. In all cases, preference
was measured in the head turn preference proce-
dure (HPP), where sound presentation is contingent
on infant attention (evidenced by orienting their
head toward the sound source).

Newman, Bernstein Ratner, Jusczyk, Jusczyk,
and Dow (2006) classified children into high- and
low-vocabulary groups on the basis of vocabulary
size at 2 years and retrospectively examined the
correlation of vocabulary with their prior perform-
ance in a range of speech perception tasks. New-
man et al. found that the high- and low-vocabulary
groups in their sample differed most markedly in
terms of their performance in word-segmentation
tasks performed at 7–9 months of age (and less so
in other tasks, such as language discrimination).

Other research has refined the conclusions from
that early report, suggesting how low- and high-
vocabulary children may differ. For example, Junge,
Hagoort, Kooijman, and Cutler (2010) documented
that 7-month-olds who, in advance of their peers,
displayed ERPs typical of 10-month-olds during
word form recognition went on to have larger
vocabularies by 3 years. Interestingly, Junge (2011)
reported that 10-month-olds who had ERPs typical
of 7-month-olds did not have poorer language skills
at 5 years, suggesting that the predictive value of
word-segmentation measures may be limited to
7 months. Alternatively, the infant or the outcome
measure in this study may have been suboptimal
for detecting a difference that may exist at
10 months. In other words, level of difficulty may
need to be adapted to assess individual differences
in the same construct across age groups.

Another line of research has explored which spe-
cific aspects of word form recognition are better pre-
dictors. Junge, Kooijman, Hagoort, and Cutler
(2012) separated memory (the matching of a stimulus
being experienced with internalized forms) from seg-
mentation (the process by which infants pull out a
word from running speech). In the memory
condition, 10-month-olds were familiarized with a
word form in isolation; in the segmentation
condition, the word form was presented within its
original sentential context. For both conditions, rec-
ognition was subsequently assessed by comparing
ERPs for familiarized and unfamiliarized single
words. Children were classified as high- versus low-
vocabulary size when they reached 24 months of
age. The two groups showed the same ERP for the
memory task, but differed in the segmentation task,
with high-vocabulary children showing larger ERP
recognition responses in the segmentation task. This
suggests that segmentation may be more predictive
of language acquisition than memory. Another
interpretation is that the segmentation task was
more demanding (required successful memory and
segmentation). As suggested earlier, task complexity
may be a key factor, and at 10 months, a composite
task is better able to assess individual differences.

In sum, the work on word form recognition indi-
ces as predictors is interesting because researchers
are beginning to decompose word recognition into
subskills and to explore which skills are more or
less strongly associated with outcomes.

Prosody. The literature on prosodic predictors is
extremely sparse. Results focusing on prosodic words
appear stable, as two studies with different samples
and tasks—albeit coming from the same research
group—report positive results. Both Weber, Hahne,
Friedrich, and Friederici (2005) and Friedrich, Her-
old, and Friederici (2009) investigated processing of
different word-level stress patterns at 5 months
using ERPs. Children were later classified into high
versus low vocabulary based on their language
outcomes (at 12 and 24 months in Weber et al.,
2005, and at 30 months in Friedrich et al., 2009). The two groups differed in their ERPs to the common stress-initial pattern, but not to the uncommon stress-final one.

At first sight, results focusing on larger prosodic units beyond the word form level appear more mixed. On the one hand, Cristia and Seidl (2011) report that 6-month-olds who preferred well-formed over ill-formed intonational phrases in the HPP developed larger vocabularies at 24 months than infants who showed the opposite preference. On the other hand, in Newman et al. (2006) children who had larger versus smaller vocabularies did not differ significantly in their ability to segment large prosodic units (intonational or phonological phrases) at 6–9 months, although the authors remark that results are clearly in the “right” direction.

Quantitative Summary

The qualitative review earlier appears to suggest that there is some predictive value for infant speech perception measures, although null results have also been reported. It thus becomes imperative to approach the same question quantitatively, which allows assessment of whether correlations are indeed significant overall, and for each level separately. An additional question is how the predictive value of speech perception measures compares to that of other, better studied infant predictors. To answer it, we inspected effect sizes for three nonlinguistic predictors: habituation, dishabituation, and rapid auditory processing (RAP). The former two have already been meta-analyzed (see Kavsek, 2004, for details and references). RAP has been hypothesized to play a role in language development, and estimates of their predictive value could be gathered from Benasich, Thomas, Choudhury, and Leppänen (2002) and Choudhury, Leppänen, Leevers, and Benasich (2007). While this comparison is informative, it is important to keep in mind that the latter have been developed over many years to measure individual variation, whereas no such development has occurred for speech tasks. Therefore, the language predictors are unlikely to be as strong as the others at the present stage.

Method

The objective of this systematic review was to evaluate the predictive value of measures of language-specific processing gathered in the 1st year of life (4–12 months of age). The survey protocol, PRISMA checklist, methodological details, full tables, additional analyses, and analysis script are available on Cristia (2013). To be included, an experiment had to meet three criteria: (a) Report an individual variation analysis (via correlations or subgroup comparisons) of the relation between an infant speech perception measure and a measure of language outcome. (b) Contain a speech perception measure before 12 months, but after the youngest age at which language-specific perception had been documented for that level, to increase the odds that acquired knowledge would be involved. In fact, no study was excluded because of this criterion alone; further information can be found on the project website (Cristia, 2013). (c) Contain a speech perception measure that recruits language-specific knowledge; that is, it must rely on abilities documented to vary between groups of infants exposed to different languages/inputs.

The initial list, containing 15 journal articles and two theses, was put together based on the authors' knowledge of the literature on infant predictors of language. This list was further enhanced through exhaustive searches carried out in English on scholar.google.com, Pubmed, Science Direct, and Proquest from November 16 to 20, 2012. Note that scholar.google.com inspects both journal-published and unpublished research, which is relevant for assessing the possibility of bias in reporting. Titles and, if necessary, abstracts were consulted for the screening, and eligibility was determined by retrieving the full article. Inspection of the 567 nonunique results revealed two additional journal articles and one additional thesis that were missing in the original list. The complete list of items is shown in Table 1; in addition, Friedrich et al. (2009) provide a rich analysis of electroencephalographic data.

Specific details pertaining to each predictor task were given in the qualitative review earlier. After an initial inspection of the included studies, it was decided that only vocabulary size would be used as the language outcome measure, as it was available in all studies. The most common outcome instrument used was the MacArthur Communicative Developmental Inventory (CDI; Fenson, Dale, Reznick, Bates, & Pethick, 1994), whose reliability, stability, and validity have been extensively studied (e.g., see references in Fernald & Marchman, 2012). The CDI is a vocabulary checklist filled out by a primary caregiver, which includes counts for gestures and words understood and produced (infants 8–16 months), or counts for receptive and expressive vocabulary (16–36 months). Receptive vocabulary was always reported for outcomes gathered at
<table>
<thead>
<tr>
<th>Author and year</th>
<th>Design</th>
<th>Age I</th>
<th>N</th>
<th>Age V</th>
<th>Vocab</th>
<th>E</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phones</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talay-Ongan (1996)</td>
<td>Contrast</td>
<td>8 to 18 CHT: several native consonants</td>
<td>6 &amp; 6</td>
<td>55 to 62 PPVT</td>
<td>+</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>16 U</td>
<td>–</td>
<td>–0.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>24 P</td>
<td>–</td>
<td>–0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHT: Percent correct [u–y]</td>
<td>20</td>
<td>13 U</td>
<td>+</td>
<td>–0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>16 U</td>
<td>+</td>
<td>–0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>24 P</td>
<td>+</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Conboy et al. (2005)</td>
<td>Correlation</td>
<td>11 CHT: d’ nonnative [t–d]</td>
<td>23</td>
<td>11 U</td>
<td>+</td>
<td>–0.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHT: d’ native [t–th] minus d’ nonnative [d–t]</td>
<td>10</td>
<td>11 U</td>
<td>+</td>
<td>–0.37</td>
<td></td>
</tr>
<tr>
<td>Kuhl et al. (2005)</td>
<td>Correlation</td>
<td>7 CHT: d’ native [ta–pa]</td>
<td>17</td>
<td>18 P</td>
<td>+</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHT: d’ nonnative [ci–tsi]</td>
<td>17</td>
<td>18 P</td>
<td>–</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>24 P</td>
<td>+</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rivera-Gaxiola et al. (2005)</td>
<td>Contrast</td>
<td>11 ERP: nonnative [t–d]</td>
<td>13 &amp; 11</td>
<td>18 to 30 P</td>
<td>+</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Kuhl et al. (2008)</td>
<td>Correlation</td>
<td>7.5 ERP: MMN native [ta–pa]</td>
<td>21</td>
<td>18 P</td>
<td>–</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERP: MMN nonnative [ci–tsi] or [ta–da]</td>
<td>21</td>
<td>24 P</td>
<td>+</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>Conboy et al. (2008)</td>
<td>Correlation</td>
<td>11 CHT: Number of conditioning trials</td>
<td>17</td>
<td>11 U</td>
<td>–</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHT: d’ native [ta–th’a]</td>
<td>17</td>
<td>11 U</td>
<td>+</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHT: d’ nonnative [ta–da]</td>
<td>17</td>
<td>11 U</td>
<td>–</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Cardillo (2010)</td>
<td>Contrast</td>
<td>7 to 11 CHT: d’ [u–y]</td>
<td>9 &amp; 8</td>
<td>60 PPVT</td>
<td>+</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>7 CHT: Trials to criterion [u–y]</td>
<td>22</td>
<td>18 P</td>
<td>–</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHT: Percent correct [u–y]</td>
<td>22</td>
<td>18 P</td>
<td>+</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>24 P</td>
<td>+</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Jansson-Verkasalo et al. (2010)</td>
<td>Correlation</td>
<td>12 ERP: MMN nonnative vowels</td>
<td>20</td>
<td>24 P</td>
<td>+</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td><strong>Word forms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swingley (2005)</td>
<td>Correlation</td>
<td>11 HPP: Correctly &gt; mispronounced words</td>
<td>17</td>
<td>16 U</td>
<td>+</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Newman et al. (2006)</td>
<td>Contrast</td>
<td>7–12 HPP: Word segmentation</td>
<td>17 &amp; 10</td>
<td>55 TOLD</td>
<td>+</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Singer (2008)</td>
<td>Contrast</td>
<td>5 or 13 HPP: Own name &gt; foil in noise</td>
<td>20 &amp; 21</td>
<td>53–72 PPVT</td>
<td>+</td>
<td>–0.01</td>
<td></td>
</tr>
<tr>
<td>Junge et al. (2010)</td>
<td>Correlation</td>
<td>7 ERP: Word segmentation</td>
<td>23</td>
<td>36 Reynell</td>
<td>–</td>
<td>–0.45</td>
<td></td>
</tr>
<tr>
<td>Junge et al. (2012)</td>
<td>Correlation</td>
<td>10 ERP: Recognition of word originally heard in a passage</td>
<td>28</td>
<td>12 U</td>
<td>–</td>
<td>–0.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ERP: Recognition of word originally heard in isolation</td>
<td>28</td>
<td>24 P</td>
<td>–</td>
<td>–0.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28</td>
<td>12 U</td>
<td>–</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Junge (2011)a</td>
<td>Correlation</td>
<td>10 ERP: Word segmentation</td>
<td>28</td>
<td>24 P</td>
<td>–</td>
<td>–0.11</td>
<td></td>
</tr>
<tr>
<td>Junge (2011)b</td>
<td>Correlation</td>
<td>10 ERP: Word segmentation, pitch matched</td>
<td>28</td>
<td>24 P</td>
<td>+</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Singh, Reznick, and Xuehua (2012)</td>
<td>Correlation</td>
<td>7.5 ERP: Word segmentation, pitch matched</td>
<td>40</td>
<td>24 P</td>
<td>+</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
or before 16 months, whereas expressive vocabulary from the CDI was more regularly reported for older toddlers. Other vocabulary-based instruments were also accepted, and their details are available on the project website (Cristia, 2013). Nonvocabulary language outcomes (such as syntactic, morphological, semantic, and pragmatic processing) were too variable to warrant exploration. As this literature grows, future meta-analysts may be able to explore these outcomes; for our purposes, a single common outcome measure suffices to illustrate the potential and limitations of the infant predictors under review.

Effect sizes not provided in the original text were calculated from means and standard deviations when available, and from exact \( t \) and \( F \) values when this was not possible (Rosenthal & DiMatteo, 2001). For correlations to be combined using meta-analytic methods, they need to be encoded as positive or negative depending on whether they fit a hypothesized relation. It will be argued later (in the Discussion section) that additional research is necessary to establish noncircular criteria. For the present analysis, we accepted authors’ arguments of when correlations should be positive or negative, which naturally inflates the chances of finding significant effect sizes. Only one effect size per experiment (defined as an independent participant group) was considered in subsequent analyses. When a given study reported multiple correlations for a single participant group, a weighted mean \( r \) was calculated (where the weight was based on the sample size in each correlation). Analyses were carried out using the R package meta (Schwarzer, 2007).

Inspection of box plots did not indicate any outliers that could bias results. In addition, funnel plots gave no evidence of a publication bias. All methodological details are provided on the project website (Cristia, 2013).

### Results

To assess whether correlations are statistically higher than zero, we calculated median weighted effect sizes. The answer was positive: The median correlation coefficients were .31, and the 95% confidence interval was above zero, 95% CI [0.22, 0.4] (see Figure 1). Heterogeneity statistics, used to assess whether one of the three levels differed in their predictive value, revealed no significant difference among the three linguistic subtypes, \( Q(2) = 2.16, p = .34 \). Thus, no one level stood out with a consistently higher or lower predictive value. This is probably due to the great deal of variation in the effect sizes registered within the three levels. Importantly, weighted median effect sizes revealed that correlations were significantly higher than zero when each level is considered separately, as evident in the confidence intervals for each of the three subtypes shown in the figure. Descriptive statistics concerning possible moderating factors are provided on the project website (Cristia, 2013).

Figure 2 shows that correlation coefficients for the nonlinguistic predictors have median of .43 and a 95% CI [0.35, .5]. These act as a benchmark for the emergent literature on infant speech perception, although we stress again that the latter have been

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Design</th>
<th>Age I</th>
<th>( N )</th>
<th>Age V</th>
<th>Vocab</th>
<th>E</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosody Weber et al. (2005)</td>
<td>Correlation</td>
<td>5</td>
<td>ERP: MMN trochaic deviant</td>
<td>18</td>
<td>12</td>
<td>ELFRA</td>
<td>-</td>
</tr>
<tr>
<td>Seidl &amp; Cristia (2011)</td>
<td>Contrast</td>
<td>6</td>
<td>HPP: Preference for well-formed phrases</td>
<td>13 &amp; 11</td>
<td>24</td>
<td>P</td>
<td>+</td>
</tr>
</tbody>
</table>

**Note.** The first author and year allow the retrieval of the article in the references. Design indicates whether the study reported a contrast between two groups of infants, or a correlation (see the project website, Cristia, 2013, for details). Age I = age (in months) at which the infant speech perception measure was taken; \( N \) = sample size (of each of the two groups being contrasted, when relevant); Age V = age (in months) at which the vocabulary was measured; Vocab = instrument used to estimate vocabulary (P = words produced in Communicative Development Inventory [CDI]; U = words understood in CDI; always raw scores); E = expected direction of correlation between the infant and outcome measure (+ = positive; – = negative); \( r \) = untransformed \( r \); CHT = conditioned headturn; PPVT = Peabody Picture Vocabulary Test; ERP = event related potentials; MMN = mismatch negativity; HPP = headturn preference procedure.

*aChapter 4. bChapter 6.*
### Correlation (Based on Fisher's z Transformation)

<table>
<thead>
<tr>
<th>Study</th>
<th>Total</th>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phones</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrillo_2010</td>
<td>22</td>
<td>-0.05</td>
<td>[-0.37; 0.29]</td>
<td>2.4%</td>
</tr>
<tr>
<td>Conway_2005</td>
<td>23</td>
<td>0.41</td>
<td>[0.26; 0.53]</td>
<td>2.5%</td>
</tr>
<tr>
<td>Conway_2008</td>
<td>17</td>
<td>0.09</td>
<td>[-0.37; 0.51]</td>
<td>1.7%</td>
</tr>
<tr>
<td>Jansson_2010</td>
<td>20</td>
<td>0.45</td>
<td>[0.04; 0.72]</td>
<td>2.1%</td>
</tr>
<tr>
<td>Kuhl_2005</td>
<td>17</td>
<td>0.44</td>
<td>[0.08; 0.70]</td>
<td>1.7%</td>
</tr>
<tr>
<td>Kuhl_2008</td>
<td>23</td>
<td>0.48</td>
<td>[0.09; 0.75]</td>
<td>2.5%</td>
</tr>
<tr>
<td>Rivera_2005</td>
<td>24</td>
<td>0.43</td>
<td>[0.07; 0.69]</td>
<td>2.6%</td>
</tr>
<tr>
<td>Talay_1996</td>
<td>12</td>
<td>0.61</td>
<td>[0.18; 0.64]</td>
<td>1.1%</td>
</tr>
<tr>
<td>Tsao_2004</td>
<td>20</td>
<td>0.25</td>
<td>[-0.18; 0.64]</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

**Random Effects Model**

<table>
<thead>
<tr>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35</td>
<td>[0.22; 0.47]</td>
<td>18.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Words</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Junge_2010_exp1</td>
<td>23</td>
<td>0.45</td>
<td>[0.10; 0.70]</td>
<td>2.5%</td>
</tr>
<tr>
<td>Junge_2011_ch4</td>
<td>27</td>
<td>0.04</td>
<td>[-0.35; 0.41]</td>
<td>3.0%</td>
</tr>
<tr>
<td>Junge_2011_ch6</td>
<td>23</td>
<td>0.15</td>
<td>[-0.27; 0.52]</td>
<td>2.5%</td>
</tr>
<tr>
<td>Junge_2012</td>
<td>28</td>
<td>0.24</td>
<td>[-0.07; 0.51]</td>
<td>3.1%</td>
</tr>
<tr>
<td>Newman_2006</td>
<td>27</td>
<td>0.40</td>
<td>[0.11; 0.64]</td>
<td>3.0%</td>
</tr>
<tr>
<td>Singer_2008</td>
<td>41</td>
<td>-0.01</td>
<td>[-0.48; 0.46]</td>
<td>4.7%</td>
</tr>
<tr>
<td>Singh_2012</td>
<td>40</td>
<td>0.34</td>
<td>[-0.28; 0.76]</td>
<td>4.6%</td>
</tr>
<tr>
<td>Swingley_2005</td>
<td>17</td>
<td>0.43</td>
<td>[-0.01; 0.73]</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

**Random Effects Model**

<table>
<thead>
<tr>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>[0.14; 0.40]</td>
<td>25.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Prosody</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cristia_2011</td>
<td>24</td>
<td>0.45</td>
<td>[-0.02; 0.76]</td>
<td>2.6%</td>
</tr>
<tr>
<td>Weber_2005</td>
<td>18</td>
<td>0.41</td>
<td>[0.13; 0.63]</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

**Random Effects Model**

<table>
<thead>
<tr>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.42</td>
<td>[0.18; 0.61]</td>
<td>4.5%</td>
</tr>
</tbody>
</table>

---

**Figure 1.** Summary results for each study within each of the three speech perception levels. Each black row contains all relevant information for a single independent effect size. Total indicates the number of observations in the relevant study (if multiple effect sizes could be calculated for a single study, the largest N is shown here). COR shows the coefficient of correlation effect size, 95% CI the 95% confidence interval, and W(Random) the weight attributed to that study in the random effects model fit. Each gray row contains information on the median effect sizes by speech perception level. The scale for the forest plot is given at the bottom.

### Correlation (Based on Fisher's z Transformation)

<table>
<thead>
<tr>
<th>Study</th>
<th>Total</th>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dishabituation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fagan_1981_exp1</td>
<td>19</td>
<td>0.33</td>
<td>[0.03; 0.57]</td>
<td>2.0%</td>
</tr>
<tr>
<td>Fagan_1981_exp2</td>
<td>20</td>
<td>0.66</td>
<td>[0.28; 0.86]</td>
<td>2.1%</td>
</tr>
<tr>
<td>Fagan_1981_exp3</td>
<td>19</td>
<td>0.46</td>
<td>[0.05; 0.74]</td>
<td>2.0%</td>
</tr>
<tr>
<td>Fagan_1993_exp1</td>
<td>25</td>
<td>0.36</td>
<td>[0.02; 0.69]</td>
<td>2.7%</td>
</tr>
<tr>
<td>Fagan_1993_exp2</td>
<td>12</td>
<td>0.50</td>
<td>[0.11; 0.76]</td>
<td>1.1%</td>
</tr>
<tr>
<td>Fagan_1994</td>
<td>36</td>
<td>0.42</td>
<td>[-0.08; 0.76]</td>
<td>4.1%</td>
</tr>
<tr>
<td>Fagan_1992</td>
<td>149</td>
<td>0.36</td>
<td>[-0.05; 0.67]</td>
<td>18.2%</td>
</tr>
</tbody>
</table>

**Random Effects Model**

<table>
<thead>
<tr>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.42</td>
<td>[0.29; 0.54]</td>
<td>32.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Habitation</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bornstein_1985</td>
<td>18</td>
<td>0.55</td>
<td>[0.13; 0.80]</td>
<td>1.5%</td>
</tr>
<tr>
<td>Miller_1979</td>
<td>29</td>
<td>0.39</td>
<td>[-0.06; 0.71]</td>
<td>3.2%</td>
</tr>
<tr>
<td>Willatts_1998</td>
<td>44</td>
<td>0.41</td>
<td>[-0.05; 0.73]</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

**Random Effects Model**

<table>
<thead>
<tr>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.45</td>
<td>[0.20; 0.65]</td>
<td>10.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Rapid Auditory Processing</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Benasich_2002</td>
<td>43</td>
<td>0.61</td>
<td>[0.28; 0.87]</td>
<td>5.0%</td>
</tr>
<tr>
<td>Choudhury_2007</td>
<td>34</td>
<td>0.35</td>
<td>[-0.27; 0.77]</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

**Random Effects Model**

<table>
<thead>
<tr>
<th>COR</th>
<th>95%–CI</th>
<th>W(Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.54</td>
<td>[0.25; 0.74]</td>
<td>8.9%</td>
</tr>
</tbody>
</table>

---

**Figure 2.** Summary results for the nonlinguistic predictors. Each black row contains all relevant information for a single independent effect size. Total indicates the number of observations in the relevant study (if multiple effect sizes could be calculated for a single study, the largest N is shown here). COR shows the coefficient of correlation effect size, 95% CI the 95% confidence interval, and W(Random) the weight attributed to that study in the random effects model fit. Each gray row contains information on the median effect sizes by speech perception level. The scale for the forest plot is given at the bottom.
developed specifically to measure individual variation, unlike the speech perception measures they are being compared with. Heterogeneity failed to reach significance in this case as well, $Q(5) = 6.1$, $p = .3$. This suggested that the predictive value of linguistic predictors was neither markedly better nor worse than that of established nonlinguistic tasks.

**Discussion**

The first conclusion to be drawn from the meta-analysis is that a variety of measures of infant speech perception significantly predict variance in vocabulary size. In addition, the strength of this bivariate relation does not differ significantly from that of more established measures of individual variation in nonlinguistic processing. In the remainder of this Discussion, we address several questions that are prompted by these results.

*Why Are the Speech Predictors No Better Than the Nonlinguistic Predictors?*

The lack of a difference between the two groups of measures is not to be taken lightly. While the speech perception measures have been, in the immense majority of cases, directly borrowed from paradigms designed to highlight group effects, the nonlinguistic predictors we used as benchmarks have actually been developed for years with the specific purpose of making them sensitive to individual variation. Over the course of the 1980s and 1990s, researchers have tested literally thousands of infants on slightly different versions of tasks aimed at measuring habituation and dishabituation (changing the visual stimuli, the trial duration, the number of diverse tasks included, focusing on one or another infant age, and even testing infants at multiple ages). These methodological variants matter. For example, Rose, Feldman, and Wallace (1988) tested about 100 infants on between 6 and 11 dishabituation tasks at 6–8 months of age, often in multiple visits. In this sample, they estimated correlations with IQ at 3 years of age to vary between .37 and .63 depending on the number of tasks included. Similarly, Colombo, Mitchell, and Horowitz (1988) inspected various aspects of visual attention in over 60 infants, tested several times at either about 4 or 7 months. They concluded that some aspects of visual attention (specifically, shift rate) were reliable individual features at one age tested (7 months), but not at the other one (4 months). For auditory processing, Benasich and Tallal (1996) reported that each stimulus set used in their longitudinal study had been previously pretested on a total of between 88 and 120 infants. Thus, not only did the speech perception tasks measure meaningful individual variation in their under-developed state, but also their predictive power was no worse than that found for three types of constructs, each of which was represented by tasks that have been carefully honed to measure individual variation.

*How May the Infant Speech Perception Tasks Be Improved?*

For speech tasks to achieve their full predictive potential, the very first step should be the development of tasks with moderate to high test–retest reliability. To date, only one study has been published on this topic, communicating the development of a “hybrid procedure” that yielded test–retest $r = .66$ in a sample of only 10 infants (Houston, Horn, Qi, Ting, & Gao, 2007). Although the $r$ is high, the sample size is an order of magnitude lower than the reliability and validity studies on the nonlinguistic predictors above, thus inviting independent replication in larger samples. In habituation–dishabituation research, higher correlations with outcome have been found by compounding multiple problems of a similar nature. This approach has yet to be used in infant speech perception, where participants perform either a single task (e.g., the prosody studies) or multiple tasks thought to be fundamentally different (e.g., processing native vs. nonnative sounds). The example from Colombo et al. (1988) discussed earlier is worth bringing up in this context, as it illustrates how development may interfere with long-term stability. This is particularly relevant for language, as infants continue to master their native language as a function of exposure (Table 2).

Even if high test–retest reliability is achieved, there is a second important problem concerning the psychometric properties of the measures used. To begin with, while both size of $d’$ and amplitude of a mismatch response can capture gradients in sensitivity to a sound contrast, it remains unclear that the same can be said of the size of a familiarity preference in a listening task. In fact, expecting such a preference is unusual in the dishabituation literature, where it is thought that only younger, less experienced, and/or less skilled infants may show a familiarity preference (e.g., Hunter & Ames, 1988). Further work remains to interpret correlations that arise partially due to positive MMR (e.g., Jansson-Verkasalo et al., 2010) or negative $d’$ (e.g.,
Conboy et al., 2005). Furthermore, the direction of association (whether correlations are positive or negative) is sometimes arguable (as noted earlier for the nonnative [y–u] contrast).

Ideally, the properties of the response should be established in independently collected participant samples, mapping out the developmental patterns associated with a given skill, and ensuring that this development relates to language acquisition per se through cross-linguistic validation. Although some of the measures above have undergone norming (e.g., the [d-t-th] contrast used in Rivera-Gaxiola et al., 2005), in others it was the same researchers who documented the “normal” developmental pattern and later reported on individual differences, possibly on partially overlapping infant samples (e.g., Weber et al., 2005). This is a dimension that could be improved upon through increased collaborative efforts.

**Going from Correlation to Causality**

A previous section documented the strength of the bivariate association between infant measures and vocabulary size. In this section we discuss the possible ways in which such correlations can, and cannot, be interpreted. A typical first interpretation involves continuity: Phonemic categories are established in the 1st year and continue being used to code lexical items in the 2nd year; the word forms that infants segment are the seeds for later lexical categories; and so forth. This kind of conclusion makes intuitive sense, but there are several reasons why it is premature. First, we discuss potential confounding variables to argue that a multivariate approach is indispensable. Second, we lay out the challenges facing a multivariate perspective, with special attention to the possibility of separating continuity in linguistic development from the contribution of domain-general processing skills.

**Multifactoriality and Codependence**

It should be noted at the outset that the following problems are not criticisms aimed specifically at speech predictors, but are general to any infant predictor. For example, while the bivariate relation between infant habituation measures and childhood IQ is moderate, the predictive value of the infant measure is small and entirely mediated by intermediate cognitive development that is, itself, predicted to a greater extent by environmental factors than by habituation measures (Bornstein et al., 2006). As it happens with every other measure, early infant speech perception skill and later vocabulary size will share variance because they are both affected by other variables. Next, we go over these potentially hidden variables, from the most general to the most specific to language.

Some such variables have general effects on brain development or temperament: Children vary with respect to biological rhythms, such as sleeping and feeding, which have been strongly linked to outcomes across domains (e.g., Pivik, Andres, & Badger, 2012). It is conceivable that children with certain biological rhythms simply mature faster, showing advanced speech perception and garnering larger vocabularies as they grow.

Moreover, all measures of language-specific speech perception used to date involve, to a greater or lesser extent, a host of skills that are not necessarily specific to the native language. Vocabulary size had long been an indirect index of verbal IQ, and thus an outcome measure of choice for habituation and dishabituation studies (see Kavsek, 2004). Recent work attempts to understand which specific

### Table 2

**Summary of Correlations in Individual Variation Across Two Speech Perception Measures**

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Age</th>
<th>N</th>
<th>r</th>
<th>Measures and tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phones</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conboy et al. (2005)</td>
<td>7</td>
<td>16</td>
<td>0.48</td>
<td>d’ native and nonnative contrasts</td>
</tr>
<tr>
<td>Kuhl et al. (2005)</td>
<td>7</td>
<td>16</td>
<td>0.44</td>
<td>MMN and d’ for native contrast</td>
</tr>
<tr>
<td>Kuhl et al. (2008)</td>
<td>7.5</td>
<td>15</td>
<td>0.58</td>
<td>d’ native and nonnative contrasts</td>
</tr>
<tr>
<td>Cardillo (2010)</td>
<td>7–11</td>
<td>20</td>
<td>0.19</td>
<td>d’ at 7 and 11 months</td>
</tr>
<tr>
<td>Junge et al. (2012)</td>
<td>10</td>
<td>28</td>
<td>0.36</td>
<td>Word recognition in isolation versus sentence</td>
</tr>
<tr>
<td>Word forms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singh et al. (2012)</td>
<td>7.5</td>
<td>40</td>
<td>0.27</td>
<td>Word recognition with and without pitch change</td>
</tr>
<tr>
<td>Houston et al. (2007)</td>
<td>9</td>
<td>10</td>
<td>0.65</td>
<td>Novelty preference for a novel audiovisual word in 2 separate days</td>
</tr>
</tbody>
</table>

*Note. MMN = mismatch negativity.*
cognitive abilities may predict language development. In this quest, individual variation in selective attention (Colombo, Shaddy, et al., 2008) as well as visual recognition memory, imitation, and cross-modal constancy (Rose, Feldman, & Jankowski, 2009) have been found to correlate with vocabulary size at 2 years of age. These predictors fit well within a causal explanation. For example, selective attention is necessary to isolate referents of words, visual memory is involved in the recognition of these referents, and cross-modal constancy would be required to cement the word form–referent association. Nonetheless, causality tying such constructs with language cannot be demonstrated at present any more than they can for infant speech perception and later language, as neither of these studies incorporated measures for all other factors that could mediate the relation found (see also the next section).

Another likely candidate for a hidden variable that would inflate bivariate correlations is auditory processing, which is necessarily involved in speech perception because speech is based on sounds. A sizable literature ties language outcomes to a variety of measures of auditory processing, including sound detection (Been et al., 2008), sound encoding (Molfese & Molfese, 1985), change detection (Leppänen, Hämäläinen, Guttorm, Eklund, & Salmi- nen, 2012), and informational masking (Choudhury & Benasich, 2011). Much of this literature has arisen in the quest for infant predictors for language impairments. We do not claim that auditory skills matter more or less than any other skills at this point, but we add this to the catalog of confounding factors.

Moreover, the infant’s linguistic environment plays a clear role in both speech perception and later language. Children’s language development has been longitudinal tied to several characteristics of the input provided by primary caregivers (e.g., Hart & Risley, 1995). Similarly, speech perception skills relate to caregivers’ articulation of phones, either in a highly specific manner (caregivers’ /s/ clarity predicts infants’ native /s/-/ discrimination; Cristia, 2011) or more broadly (caregivers’ vowel space size predicts trials to criterion in a CHT using the native contrast [i:]-[i]; Liu, Kuhl, & Tsoa, 2003).

Finally, the relation between all these factors could be complex and multidirectional. For instance, caregivers’ clarity of speech in turn appears to be affected by infant attention (Lam & Kitamura, 2012) and at least some improvements in infant speech perception could be partially due to caregiving attitudes that correlate with the acoustic properties of maternal speech (see, e.g., Lam & Kitamura, 2010). In such a scenario, a correlation between infant speech perception skills and outcome arises from multiple sources, which are difficult to separate.

Toward Multivariate Models

One way to address these problems is by measuring multiple variables in the same infants (provided that all tasks are equally sensitive), and factoring out constructs that are not of primary interest. Some of the studies included in our systematic review have tested infants on multiple tasks. On this basis, we address one pressing question that is likely in the readers’ mind: Can the predictive value of infant speech perception tasks be reduced to the nonlinguistic components involved in those tasks?

Here, we have tried to avoid this problem by concentrating on measures that rely on acquired linguistic knowledge (e.g., infants’ differential processing of native and nonnative sound contrasts must necessarily stem from their acquisition of the native sound system) over, for example, sound discrimination measured at birth (Molfese & Molfese, 1985). Still, the only way to rule out that the predictive value of infant language processing measures cannot possibly be explained away by nonlinguistic skills recruited in these tasks is to both measure and document that the effect of the nonlinguistic predictors (on infant and/or childhood measures) not disappear when other factors (e.g., environmental and emotional variables) are included in the model. Unfortunately, this desirable multivariate evidence is simply too scarce at present (see the project website, Cristia, 2013, for analyses of some relevant data).

A multivariate approach can also be informative when applied longitudinally, as it can document a more specific cascade of effects through intermediate stages of development. In another strand of research on individual differences in toddlers, Fernald and colleagues are exploring the predictive value of a lab-based measure of toddlers’ lexical retrieval (a recent review of this research, and their work on variation in language, can be found in Fernald & Marchman, 2012). In the “looking-while-listening” paradigm, toddlers hear a word and see two objects on a screen, one of which matches the word heard. The speed with which 25-month-olds orient toward the correct object correlates concurrently and predictively with vocabulary size (Fernald, Perfors, & Marchman, 2006; see also Werker, Fennell, Corcoran, & Stager, 2002, for a related research line). In addition, the amount of time looking at the correct object, compared to the incorrect
object, explains substantial variance in childhood vocabulary and can predict language development beyond such childhood vocabulary measures (Fernald & Marchman, 2012).

While these forays examining multiple factors are important, the leap to causality will necessarily be elusive. Causality cannot be established fully with correlational studies, nor is it possible to directly manipulate factors in language learning to observe their long-term effects. We suggest that the field stands to gain a great deal by complementing large-scale, longitudinal samples (as in Bornstein et al., 2006) with computational and animal models. To begin with, computational models of language development can be used to understand some of these relations in vitro. For example, the potential impact of caregivers’ clarity of articulation on category learning has been studied by applying learning algorithms to data drawn from infant-directed speech (e.g., Gauthier & Shi, 2011, and citations therein). Furthermore, direct manipulations can be done using animal models, which should possess at least some of the prerequisites for certain skills involved in the acquired knowledge discussed earlier in the article. An example of their application to general auditory skills can be found in the work of Threlkeld, McClure, Rosen, and Fitch (2006), who found that local lesions in specific gyri soon after birth gave rise to rapid auditory processing deficits in mice. Naturally, neither computational nor animal models could ever prove that the same causal relations hold for human infants. In general, the latter two approaches are more rare in the study of infant predictors of later language and cognition, and none has been applied to the link between infant speech perception and childhood language. Nonetheless, the combination of large, multivariate, longitudinal studies with controlled experiments and modeling have potential to provide complementary insights on whether a causal interpretation could be entertained.

Infant Predictors for Communicative Disorders

With the exception of Jansson-Verkasalo et al. (2010), the research reviewed previously focused on full-term, healthy, typically developing infants, with no familial history of language impairments. In this section, we point out some problems facing the clinical translation of such approaches. Two roadblocks stand in the way of developing similar measures that could be used for early screening and/or diagnosis of communication delays and disorders.

First, it might be particularly challenging to develop reliable infant predictors (whether based on infant speech perception or in general) in pathologies known to involve developmental discontinuities or reversals, or whose cognitive locus is unclear. For example, although lack of response to name may be a predictor of ASDs, infants at risk of autism differ from controls on this measure at 12 months of age, but not at 6 months (Nadig et al., 2007). Another example can be found in the area of fluency, as stuttering also exhibits developmental nonlinearities. For instance, the onset of stuttering is often associated with somewhat more advanced language development at about 2 years that is suddenly curtailed by fluency difficulties as the child attempts multiword utterances; notably, for half of these children, fluency issues resolve before school age (Reilly et al., 2009). Along a similar line, it is remarkably difficult to establish a specific locus of impairment. To take just one example, Ramus and Ahissar (2012) review over a dozen theories, each pinpointing a different process or cognitive construct as the source of dyslexia, and spanning from low-level visual or auditory acuity to very high-level, metalinguistic manipulations. If it is already difficult to isolate a single cause in children or adults who exhibit the disorder, it will be even more challenging to develop reliable predictors measured in infants at risk.

The second, and perhaps most important, cluster of difficulties in translating these findings to clinical practice relates to the methodologies involved. Infant speech perception tasks were developed to maximize finding small differences across groups and are therefore not well designed for detecting or assessing individual capacities. As discussed above, there is a great deal of research necessary to determine whether an infant’s performance in a given speech perception task using HPP, CHT, and ERPs is reliable across repeated testing. There is even more work to do to assure that tasks are equated for (irrelevant) difficulty across populations.

Indeed, the field must first establish clear developmental patterns and document whether certain infant populations typically deviate from them. We illustrate this with preterm infants, who often display language delays for which language-specific versus cognitive sources are still debated (for a discussion, see Bosch, 2011). Premature birth is comorbid with a range of conditions that could lead to slower neurological development. In addition to this potential maturational difference, preterm and full-term infants vary in the quality of speech to which they have been exposed. Full-term newborns experience 3 full months of speech input in utero, where information carried by higher spectral frequencies is...
attenuated to a greater extent than information at lower frequencies (Granier-Deferre, Ribeiro, Jacquet, & Bassereau, 2011). As lower frequencies convey information on prosody and vowels, this “filtered” exposure may focus full-term infants’ learning on these two levels. Thus, there are several reasons why preterm and full-term infants could differ in terms of speech perception.

And yet the evidence for preterm/full-term differences is not straightforward. To begin with, Peña and colleagues argue that preterms and full terms achieve the same landmarks when matched for gestational (i.e., maturational) age. In support of this conclusion, they report no difference between preterms and full terms when age has been corrected in terms of electrophysiologically correlates of language discrimination (between different rhythmic classes at 3 months, and within the same class at 6 months; Peña, Pittaluga, & Mehler, 2010) and sound discrimination (native [ba-da] does not differ from nonnative [da-da] at 9 months, but it does at 12 months; Peña, Werker, & Dehaene-Lambertz, 2012). The former result had also been obtained using behavioral methods in an independent lab (Bosch, 2011). As for prosodic words, Herold, Höhle, Walch, Weber, and Obladen (2008) reported that, unlike their full-term peers, neither 4- nor 6-month-old preterms exhibit a behavioral preference for prosodic words with their language’s predominant stress pattern typical of full-term 4-month-olds. However, a direct contrast between the groups was not significant in Bosch (2011), partially due to high within-group variability in preterms and full terms. Others argue that preterms and full terms should be best matched in terms of experience, based on preferences for common consonant sequences (Gonzalez-Gomez & Nazzi, 2012; contra the conclusions in Peña et al., 2012).

While such results suggest broadly similar patterns of development in preterms and full terms when matched in some parameter (gestational age or experience), other work suggests divergent patterns. Figueras Montiu and Bosch Galceran (2010) find that preterms and full terms differed in their discrimination of an acoustically large, native vocalic contrast at 4 months because the preterms exhibited no sign of discriminating it. By 8 months, both groups were equally successful in distinguishing the contrast, although in other work Bosch (2011) finds them to differ in word segmentation. This pattern of results could indicate reduced phonetic sensitivity in preterms (among many other explanations). Contrastingly, Jansson-Verkasalo et al. (2010) reported that preterm 12-month-olds showed larger electrophysiological responses for a subtle nonnative vocalic contrast than full-term infants, suggesting greater sensitivity to it; notably, the groups did not differ for the native contrast or for the nonnative contrast when tested at 6 months.

Which divergences between preterm and full terms are reliable? When they exist, are they indicative of qualitatively diverse developmental patterns, and/or are they attributable to motor, cognitive, auditory, and/or linguistic sources? Can individual variation among preterms be used to guide speech-language interventions in those preterms that need it the most? Such are the questions that need to be addressed when trying to use infant speech perception measures as predictors. We expect that more precise and robust measures will greatly aid in this enterprise.

Conclusions

We have described recent findings documenting moderate bivariate correlations between infant speech perception and vocabulary acquisition. Although we are optimistic regarding the promise of such measures, throughout this article we have pointed to specific areas ripe for improvement or development. We highlight the need to go beyond bivariate correlations, and attempt to clarify the hypothesized causal links between infant measures and toddler language within specific theoretical models. We hope that later work will explore methodologies that are both reliable and simple to administer so that these measures in infancy can be used as starting points for early interventions.

References


Talay-Ongan, A. (1996). Infants’ phonemic awareness and receptive language profiles. *Child Language Teach-
ing and Therapy, 12, 99–112. doi:10.1177/02656909601200201