Predicting individual variation in language from infant speech perception measures

Alejandrina Cristia,[1,2] Amanda Seidl,[3] Caroline Junge,[4] Melanie Soderstrom,[5] and Peter Hagoort[2,6]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, IEC-ENS, EHESS

[2] Max Planck Institute for Psycholinguistics, Neurobiology of Language

[3] Purdue University, SLHS

[4] University of Amsterdam, Faculty of Social and Behavioural Sciences

[5] University of Manitoba, Psychology Department

[5] Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour

Please cite as: Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (in press) Infant predictors of language. Child Development.

Abstract

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 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 relationship 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.

Keywords: Individual differences; Speech perception; Vocabulary development.

Predicting individual variation in language from infant speech perception measures

1. Introduction

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). While 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 first year of life, on the one hand, and individual differences in language development, on the other (for example, see the collection in Colombo, McCardle, & Freund, 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 Section 3). Here, we discuss the potential of infant speech processing measures to predict variation in language outcomes over this successful backdrop in 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 first year of life (a review in Räsänen, 2012). According to one, infants begin to learn phonology by determining which speech sound categories (which we will call phones) their ambient language uses, and later using these units to build wordforms (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 wordforms, which in turn aid in the development of the phone inventory. Both models assume infants have some knowledge at both the level of phones and the level of wordforms, 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 wordform 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. In Section 2, we report on a systematic review first qualitatively (Section 2.1) and then using meta-analytic methods (Section 2.2). The latter allows us to estimate statistically the strength of the bivariate relationship between language-specific speech perception measures gathered in the first 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. In Section 3, 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 (Section 4).

2. Predicting Language from Infant Speech Perception Skills

Since 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 pre-requisite to 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 Section 3 regarding the impossibility of designing 'pure' measures). These measures hold the promise of being more informative to the theoretical and applied enterprises mentioned above.

We focused our attention on tasks that could plausibly tap language-specific knowledge of three fundamental linguistic levels whose fast-paced development in the first year has been well-described: 1) the level of phones (e.g., involved in distinguishing between /dat/ and /ḍat/; the distinction between speech sounds, without assuming an abstract and rich phonemic representation), 2) wordforms (e.g., involved in the recognition of ‘mommy’ as a speech sequence often encountered in the past; we concentrate on wordforms, without assuming that these map onto actual words that have a referent), and 3) 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, since it is a common theoretical position that infants process speech using precursors of the adult linguistic representations, they can be considered 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 and /d̪at/ paired with a very different object show gains in their sensitivity to that foreign contrast (Yeung & Werker, 2009), suggesting that word-level knowledge can impact sound sensitivity. The same can be said of all other interactions (e.g., wordform recognition and prosodic processing; Seidl & Johnson, 2006).

2.1 Qualitative overview

This overview organizes previous research on how measures at each of the three levels predicts vocabulary size in childhood (see Section 2.3.1 for details on the literature selection). Instead of being exhaustive, it highlights key methodological and conceptual questions.

2.1.1 Phones

There is considerable variability in how this level has been studied. For instance, much of this work employs the Conditioned Headturn 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 headturn to 80% of the sound changes, and refrained from producing a headturn in at least 80% of the non-change 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 non-native contrast (Finnish [u-y]). Fewer trials to criterion can be interpreted in at least two ways: as faster learning of the contingency (a non-linguistic 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 headturns 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 and colleagues attribute these correlations to a more advanced speech perception mechanism, negative correlations instead could also have been expected, since it is widely believed that as infants accrue experience with their native language, their discrimination of *native* contrasts *improves*, while discrimination of *non-native* sound contrasts *deteriorates* simultaneously between 6 and 12 months of age (a recent summary in Kuhl, Conboy, Padden, Nelson & Pruit, 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, 2008). In both studies, vocabulary size correlated *positively* with sensitivity for a *native* sound contrast, but *negatively* with *non-native* sensitivity. In principle, this pattern of results cannot be explained by reference to general auditory acuity, since, all else being equal, acuity should impact native and non-native contrasts equally. However, the native and non-native contrasts used here were markedly different in acoustic salience (e.g., in Kuhl et al., 2005, the native syllables were [ta-pa], whereas the non-native syllables were a Mandarin alveolo-palatal fricative and affricate [ɕi-tɕi]), 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 more well-matched set of native and non-native speech contrasts stimuli differing in voice onset time (Conboy, Rivera-Gaxiola, Klarman, 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 non-native discrimination and later language outcome is worthy of further consideration. Continued non-native sensitivity in older infants is generally interpreted as a sign of immaturity. For example, Jansson-Verkasalo et al. (2010) reach a similar conclusion, having found a negative correlation between non-native 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 for that negative correlation. Rivera-Gaxiola, Klarman, Garcia-Sierra, and Kuhl (2005) found that those 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 non-native speech contrasts (an earlier P150-250 thought to relate to acoustic processing). In contrast, those 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 treat irrelevant acoustic details still 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.

2.1.2 Words

Most studies reported in our analysis focus at this level on wordforms, 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) from a preference for correctly pronounced highly frequent words (e.g., *been*) over mispronounced versions of the same words (e.g., *beem*). The other papers use a type of segmentation task in which infants are initially exposed to the wordforms spoken in isolation, 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 suggests successful recognition of the wordforms from running speech. In all cases, preference was measured in the Headturn Preference Procedure (HPP), where sound presentation is contingent on infant attention (evidenced by orienting their head towards 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 two years and retrospectively examined the correlation of vocabulary with their prior performance in a range of speech perception tasks. Newman et al. found that high and low vocabulary groups in their sample differed most markedly in terms of their performance in word segmentation tasks performed at 7 to 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, Kooijman, Hagoort and Cutler (2010) documented that 7-month-olds who, in advance of their peers, displayed ERPs typical of Dutch 10-month-olds during wordform recognition went on to have larger vocabularies by three 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 specific aspects of wordform recognition are better predictors. Junge, Kooijman, Hagoort, and Cutler (2012) separated *memory* (the matching of a stimulus being experienced with internalized forms) from *segmentation* (the process by which infants pull out a word from running speech). In the memory condition, 10-month-olds were familiarized with a wordform in isolation; in the segmentation condition, the wordform was presented within its original sentential context. For both conditions, recognition 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 wordform recognition indices as predictors is interesting because researchers are beginning to decompose word recognition into sub-skills, and to explore which skills are more or less strongly associated with outcomes.

2.1.3 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 et al. (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 wordform 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, 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.

2.3 Quantitative summary

The qualitative review above 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 to assess 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 non-linguistic 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 in order 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.

2.3.1 Method

The objective of this systematic review was to evaluate the predictive value of measures of language-specific processing gathered in the first 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). In order to be included, an experiment had to meet three criteria: (1) Report an individual variation analysis (via correlations or subgroup comparisons) of the relationship between an infant speech perception measure and a measure of language outcome. (2) 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 in the supplementary materials. (3) 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 on 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 non-unique 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, Herold and Friederici (2009) contains a rich analysis of electroencephalographic data.

INSERT TABLE 1 ABOUT HERE

Specific details pertaining to each *predictor* task were given in the qualitative review above (Section 2.1). After an initial inspection of the included studies, it was decided that only vocabulary size would be used as the language *outcome* measure, since 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 or before 16 months while expressive vocabulary from the CDI was more regularly reported for older toddlers. Other vocabulary-based instruments were also accepted, and their details are available in the supplementary materials. Non-vocabulary 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 re-view.

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 relationship. It will be argued below (Section 2.3.3) that additional research is necessary to establish non-circular 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 paper 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 boxplots did not indicate any outliers that could bias results. Additionally, funnel plots gave no evidence of a publication bias. All methodological details are provided on the website.

2.3.2 Results

To assess whether correlations are statistically higher than zero, we calculated median weighted effect sizes. The answer was positive: The median correlation coefficients was .31, and the 95% confidence interval was above zero, [.22, .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 across 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 with the supplementary materials.

INSERT FIGURE 1 ABOUT HERE

By comparison, Figure 2 shows that correlation coefficients for the non-linguistic predictors have median of .43 and a 95% confidence interval of [.35, .5]. These act as a benchmark for the emergent literature on infant speech perception, although we stress again that the latter have been 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 non-linguistic tasks.

2.3.3 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. Additionally, the strength of this bivariate relationship does not differ significantly from that of more established measures of individual variation in non-linguistic processing. In the remainder of this discussion, we address several questions that are prompted by these results.

2.3.3.1 Why are the speech predictors no better than the non-linguistic 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 non-linguistic 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 to 8 months, 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 while 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) report 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.

2.3.3.2 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 non-linguistic 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 versus non-native sounds). The example from Colombo et al. (1988) discussed above 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.

INSERT TABLE 2 ABOUT HERE

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 is 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 MMR being positive (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 above for the non-native [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 crosslinguistic validation. While some of the measures above have undergone norming (e.g., the [d-t-t^h] 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.

3. Going from Correlation to Causality

The 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 first year, and continue being used to code lexical items in the second year; the wordforms 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. In Section 3.1, we discuss potential confounding variables to argue that a multivariate approach is indispensable. In Section 3.2 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.

3.1 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 relationship 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, 2011). 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* cognitive abilities may predict language development. In this quest, individual variation in selective attention (Colombo, Shaddy, Blaga, Anderson, Kannass, & Richman, 2008) as well as visual recognition memory, imitation, and crossmodal constancy (Rose, Feldman, & Jankowski, 2009) have been found to correlate with vocabulary size at two 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 crossmodal constancy would tag the abilities required to cement the wordform-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 relationship 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 et al., 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 longitudinally 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-ʃ/ discrimination, Cristia, 2011) or more broadly (caregivers' vowel space size predicts trials to criterion in a CHT using the native contrast [ɕi-tɕi], Liu, Tsao & Kuhl, 2003).

Finally, the relationship between all of 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.

3.2 Towards 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 papers 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 non-linguistic 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 non-native 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 non-linguistic skills recruited in these tasks is to measure both and document that the effect of the non-linguistic predictors (on infant and/or childhood measures) does 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 supplementary materials for analyses of some relevant data).

A multivariate approach can also be informative when applied longitudinally, as they 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 towards 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). Additionally, 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., 2004) with computational and animal models. To begin with, computational models of language development can be used to understand some of these relationships *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). Further, direct manipulations can be done using animal models, who should possess at least some of the prerequisites for certain skills involved in the acquired knowledge discussed above. 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 relationships 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.

4. Infant Predictors for Communicative Disorders

With the exception of Jansson et al. (2010), the research reviewed in Section 3 focused on fullterm, 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 non-linearities. For instance, the onset of stuttering is often associated with somewhat more advanced language development at about two years that is suddenly curtailed by fluency difficulties as the child attempts multi-word 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 in Section 2.3.3.2, 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 in order 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 whose 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 fullterm infants vary in the *quality* of speech to which they have been exposed. Fullterm newborns experience three 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). Since lower frequencies convey information on prosody and vowels, this 'filtered' exposure may focus fullterm infants' learning on these two levels. Thus, there are several reasons why preterm and fullterm infants could differ in terms of speech perception.

And yet the evidence for preterm-fullterm differences is not straightforward. To begin with, Peña and colleagues argue that preterms and fullterms achieve the same landmarks when matched for gestational (i.e., maturational) age. In support of this conclusion, they report no difference between preterms and fullterms when age has been corrected in terms of electrophysiological correlates of language discrimination (between different rhythmic classes at three months, and within the same class at 6 months; Peña, Pittaluga & Mehler, 2010) and sound discrimination (native [ba-da] does not differ from non-native [da-ḍa] at nine 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, Hoehle,Walch, Weber and Obladen (2008) reported that, unlike their fullterm peers, neither four- nor six-month-old preterms exhibit a behavioral preference for prosodic words with their language's predominant stress pattern typical of fullterm four-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 fullterms. Others argue that preterms and fullterms should 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 fullterms when matched in some parameter (gestational age or experience) other work suggests divergent patterns. Figueras-Montiu and Bosch-Galceran (2010) find that preterms and fullterms differed in their discrimination of an acoustically large, native vocalic contrast at four 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, Janson-Verkasalo et al. (2010) reported that preterm 12-month-olds showed *larger* electrophysiological responses for a subtle non-native vocalic contrast than fullterm infants, suggesting greater sensitivity to it; notably, the groups did not differ for the native contrast nor for the non-native contrast when tested at six months.

Which divergences between preterm and fullterms 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.

5. Conclusions

We have described recent findings documenting moderate bivariate correlations between infant speech perception and vocabulary acquisition. While 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

Been, P. H., van Leeuwen, T. H., van Herten, M., Maassen, B., van der Leij, A., & Zwarts, F. (2008). Auditory P2 is reduced in 5 month old infants from dyslexic families. *Brain Research in Language, 1*, 93-110. http://dx.doi.org/10.1007/978-0-387-74980-8_3

Benasich, A., Thomas, J. J., Choudhury, N., & Leppänen, P. H. T. (2002). The importance of auditory processing abilities to early language development. *Developmental Psychology, 40*, 278-292. http://dx.doi.org/10.1002/dev.10032

Bornstein, M. H. (1998). Stability in mental development from early life. In F. Simion & G. Butterworth (Eds.), *The development of sensory, motor and cognitive capacities in early infancy* (pp. 301-332). Hove, England: Psychology Press. (no DOI)

Bornstein, M. H., Hahn, C., Bell, C., Haynes, O. M., Slater, A., Golding, J., Wolke, D. & the ALSPAC Study Team. (2006). Stability in cognition across early childhood. *Psychological Science, 17,* 151-158. http://dx.doi.org/10.1111/j.1467-9280.2006.01678.x

Bosch, L. (2011). Precursors to language in preterm infants. In Oliver Braddick, Janette Atkinson and Giorgio M. Innocenti (editors) *Progress in Brain Research, Vol. 189* (pp. 239-257). Burlington: Academic Press. http://dx.doi.org/10.1016/B978-0-444-53884-0.00028-2

Cardillo, G.C. (2010). *Predicting the predictors.* Unpublished PhD thesis, University of Washington. (no DOI)

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months. *Clinical Neurophysiology, 122*, 320-338. http://dx.doi.org/10.1016/j.clinph.2010.05.035

Choudhury, N., Leppänen, P. H. T., Leevers, H. J., & Benasich, A. A. (2007). Infant information processing and family history of specific language impairment.

Developmental Science, 10, 213-236. http://dx.doi.org/10.1111/j.1467-7687.2007.00546.x

Colombo, J., McCardle, P., & Freund, L. (2008*). Infant pathways to language*. Hove, UK: Psychology Press. (no DOI)

Colombo, J., Mitchell, D. W., & Horowitz, F. D. (1988). Infant visual attention in the paired-comparison paradigm: Test-retest and attention-performance relations. *Child Development, 59,* 1198-1210.http://dx.doi.org/10.2307/1130483

Colombo, J., Shaddy J. D., Blaga, O. M., Anderson, C. J., Kannass, K. N. & Richman, W. A. (2008). Early attentional predictors of vocabulary in childhood. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language* (pp. 143-168). Hove, UK: Psychology Press. (no DOI)

Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology, 44*, 1505-1512. http://dx.doi.org/10.1037/a0012975

Conboy, B., Rivera-Gaxiola, M., Klarman, L., Akseylu, E. & Kuhl, P. K. (2005). Associations between native and nonnative speech sound discrimination and language development at the end of the first year. In A., Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Supplement to the Proceedings of the 29th Boston University Conference on Language Development.* Retrieved from http://ilabs.washington.edu/kuhl/pdf/ConboyBUCLD2005.pdf. (no DOI)

Cristia, A. (2009). *Individual variation in infant speech perception.* Unpublished PhD Thesis, Purdue University.

Cristia, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *Journal of the Acoustical Society of America, 129*, 3271-3280. http://dx.doi.org/10.1121/1.3562562

Cristia, A. (2013, August 15). Individual Variability in Infancy project (website). Retrieved from https://sites.google.com/site/invarinf/

Cristia, A. & Seidl, A. (2011). Sensitivity to prosody at 6 months predicts vocabulary at 24 months. In N. Denis, K. Mesh, & H. Sung (Eds.) *Proceedings of the 35th BUCLD* (pp. 145-156). Somerville: Cascadilla Press. (no DOI)

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E. T., D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*, 1-173. http://dx.doi.org/10.2307/1166093

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83,* 203-222. http://dx.doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding. *Developmental Psychology, 42*, 98-116. http://dx.doi.org/10.1037/0012-1649.42.1.98

Figueras Montiu, M., & Bosch Galceran, L. (2010). Capacidades de discriminación fonética de un contraste vocálico nativo en el prematuro. *Psicothema*. *22*, 669-676.

Friedrich, M, Herold, B., & Friederici, A. D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. *Cortex, 45*, 662-676. http://dx.doi.org/10.1016/j.cortex.2008.06.014

Gauthier, B. & Shi, R. (2011). A connectionist study on the role of pitch in infant-directed speech. *Journal of the Acoustical Society of America, 130*, EL380-EL386. http://dx.doi.org/10.1121/1.3653546

Gonzalez-Gomez, N. & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science, 15,* 885-894. http://dx.doi.org/10.1111/j.1467-7687.2012.01186.x

Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science, 14*, 336-352. http://dx.doi.org/10.1111/j.1467-7687.2010.00978.x

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes Publishing Co.

Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine & Child Neurology, 50*, 678-683. http://dx.doi.org/10.1111/j.1469-8749.2008.03055.x

Houston, D. M., Horn, D. L., Qi, R., Ting, J. Y., & Gao, S. (2007). Assessing speech discrimination in individual infants. *Infancy, 12*, 119-145. http://dx.doi.org/10.1111/j.1532-7078.2007.tb00237.x

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5,* 69-95.

Jansson-Verkasalo, E., Ruusuvirta, T., Huotilainen, M., Alku, P., Kushnerenko, E., et al. (2010). Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age. *BMC Neuroscience, 11*, 88. http://dx.doi.org/10.1186/1471-2202-11-88

Junge, C. (2011). *The relevance of early word recognition: Insights from the infant brain.* PhD dissertation, Radboud University Nijmegen, MPI series in Psycholinguistics 67, Nijmegen.

Junge, C., Hagoort, P., Kooijman, V., & Cutler, A. (2010). Brain potentials for word segmentation at seven months predict later language development. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *Proceedings of the 34th Annual Boston University Conference on Language Development* (pp. 209-220). Somerville, MA: Cascadilla Press.

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science, 15*, 463-473. http://dx.doi.org/10.1111/j.1467-7687.2012.1144.x

Kavsek, M. (2004). Predicting later IQ from infant visual habituation and dishabituation. *Applied Developmental Psychology, 25*, 381-392. http://dx.doi.org/10.1016/j.appdev.2004.04.006

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language. *Philosophical Transactions of the Royal Society B, 363*, 979-1000. http://dx.doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T. & Pruitt, J. (2005). Early speech perception and later language development. *Language Learning and Development, 1*, 237-264.

Lam, C., & Kitamura, C. (2010). Maternal interactions with a hearing and hearing-impaired twin: Similarities and differences in speech input, interaction quality, and word production. *Journal of Speech, Language and Hearing Research, 53,* 543. http://dx.doi.org/10.1044/1092-4388(2010/09-0126)

Lam, C. & Kitamura, C. (2012). Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation. Developmental Science, 15, 212-221. http://dx.doi.org/10.1111/j.1467-7687.2011.01118.x

Leonard, L. (2000). *Children with Specific Language Impairment.* Cambridge, MA: MIT Press.

Leppänen, P. H., Hämäläinen, J. A., Guttorm, T. K., Eklund, K. M., Salminen, H., et al. (2012). Infant brain responses associated with reading-related skills before school and at school age. *Neurophysiologie Clinique, 42,* 35-41. http://dx.doi.org/10.1016/j.neucli.2011.08.005

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*, F1-F10. http://dx.doi.org/10.1111/1467-7687.00275

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants. *Infant Behavior and Development, 8*, 197-211. http://dx.doi.org/10.1016/S0163-6383(85)80006-0

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatric and Adolescent Medicine, 161,* 378-383. http://dx.doi.org/10.1001/archpedi.161.4.378

Newman, R. S., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P.W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development. *Developmental Psychology, 42,* 337-367. http://dx.doi.org/10.1037/0012-1649.42.4.643

Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences USA*, *107*, 3823-3828. http://dx.doi.org/10.1073/pnas.0914326107

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *Journal of Neuroscience, 32,* 11159-11163. http://dx.doi.org/10.1523/JNEUROSCI.6516-11.2012

Pivik, R. T., Andres, A., & Badger, T. (2012). Effects of diet on early stage cortical perception and discrimination of syllables differing in voice-onset time. *Brain and Language, 120*, 27-41. http://dx.doi.org/10.1016/j.bandl.2011.08.004

Räsänen, O. (2012). Computational modeling of phonetic and lexical learning in early language acquisition. *Speech Communication, 54*, 975-997. http://dx.doi.org/10.1016/j.specom.2012.05.001

Reilly, S., Onslow, M., Packman, A., Wake, M., Bavin, E. L., Prior, M., Eadie, P., Cini, E., Bolzonello C., & Ukoumunne, O. (2009). Predicting stuttering onset by the age of 3 years. *Pediatrics*, *123*, 270. http://dx.doi.org/10.1542/peds.2007-3219

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *NeuroReport (Developmental Neuroscience), 16,* 495-498. http://dx.doi.org/10.1097/00001756-200504040-00015

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. *Child Development, 80*, 134-150. http://dx.doi.org/10.1111/j.1467-8624.2008.01250.x

Rose, S. A., Feldman, J. F., & Wallace, I. F. (1988). Individual differences in infants' information processing. *Child Development*, 1177-1197. http://dx.doi.org/10.2307/1130482

Rosenthal, R. & DiMatteo, M. R. (2001). Meta-analysis. *Annual Review of Psychology, 52*, 59-82. http://dx.doi.org/10.1146/annurev.psych.52.1.59

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R news, 7,* 40-45.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*, 565-573. 10.1111/j.1467-7687.2006.00534.x

Singer, E. R. (2008). *Infant speech perception in noise and vocabulary outcomes*. Unpublished master's thesis, University of Maryland.

Singh, L., Reznick, J. R., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development. *Developmental Science, 15*, 482-495. *http://dx.doi.org/10.1111/j.1467-7687.2012.01141.x*

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*, 432-443. http://dx.doi.org/10.1111/j.1467-

7687.2005.00432.x

Talay-Ongan, A. (1996). Infants' phonemic awareness and receptive language profiles. *Child Language Teaching and Therapy, 12,* 99-112. http://dx.doi.org/10.1177/026565909601200201

Threlkeld, S. W., McClure, M. M., Rosen, G. D., & Fitch, R. H. (2006). Developmental timeframes for induction of microgyria and rapid auditory processing deficits in the rat. *Brain Research*, *1109*, 22-31. http://dx.doi.org/10.1016/j.brainres.2006.06.022

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life. *Child Development, 75,* 1067-1084. http://dx.doi.org/10.1111/j.1467-8624.2004.00726.x

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment. *Cognitive Brain Research, 25*, 180-187. http://dx.doi.org/10.1016/j.cogbrainres.2005.05.007

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words. *Infancy 3,* 1-30.

Werker, J.F. & Gervain, J. (2013). Language Acquisition: Perceptual Foundations in Infancy. In P. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology* (pp. 909-925). Oxford UK: Oxford University Press.

Yeung, H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound. *Cognition, 113*, 234-243. http://dx.doi.org/10.1016/j.cognition.2009.08.010

Authors' Note

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, alecristia@gmail.com

Tables

Table 1. *Summary of infant predictors of language.* 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 online materials 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 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*. Notes: [1]Chapter 4; [2]Chapter 6.

Author	Year	Design	Age I		N	Age V	Vocab	E	r
				Phones					
Talay	1996	Contrast	8 to 18	CHT: several native consonants	6 & 6	55 to 62	PPVT	+	0.66
Tsao	2004	Correlation	6	CHT: Trials to criterion [u-y]	20	13	U	-	-0.7
					16	16	U	-	-0.47
					13	24	P	-	-0.48
				CHT: Percent correct [u-y]	20	13	U	+	-0.05
					16	16	U	+	-0.17
					13	24	P	+	0.05
Conboy	2005	Correlation	11	CHT: d' non-native [t-d]	23	11	U	+	-0.37
				CHT: d' native [t-t^{h}] minus d' non-native [d-t]	10	11	U	+	-0.37
Kuhl	2005	Correlation	7	CHT: d' native [ta-pa]	17	18	P	+	0.49
					16	24	P	+	0.49
				CHT: d' non-native [ɕi-tɕi]	17	18	P	-	0.5
					16	24	P	-	0.22
Rivera	2005	Contrast	11	ERP: non-native [t-d]	13 & 11	18 to 30	P	+	-0.53
Kuhl	2008	Correlation	7.5	ERP: MMN native [ta-pa]	21	18	P	-	0.43
					23	24	P		-0.43

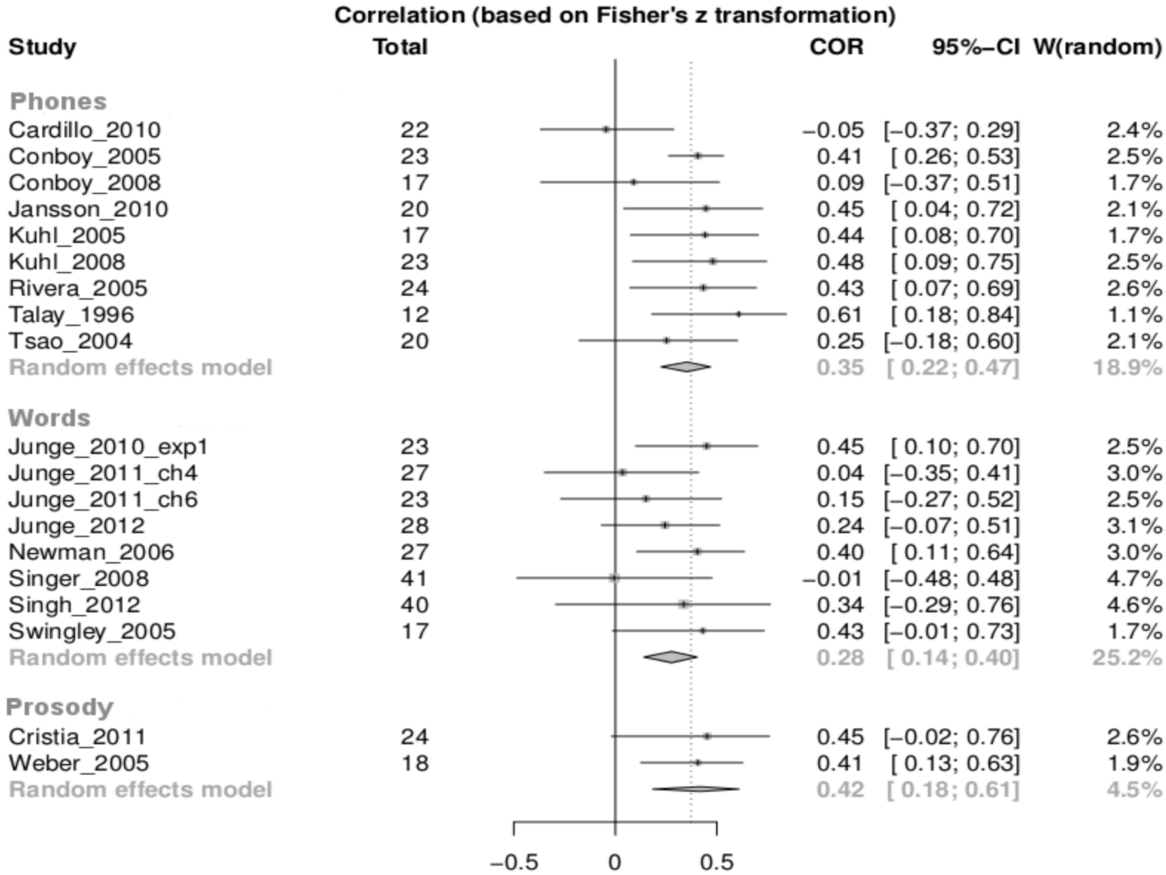
				ERP: MMN non-native [ɕi-tɕi] or [ta-da]	21	24	P	+	-0.61
Conboy	2008	Correlation	11	CHT: Number of conditioning trials	17	11	U	-	0.39
				CHT: d' native [ta-t^{h}a]	17	11	U	+	0.05
				CHT: d' non-native [ta-da]	17	11	U	-	0.43
Cardillo	2010	Contrast	7 to 11	CHT: d' [u-y]	9 & 8	60	PPVT	+	0.23
		Correlation	7	CHT: Trials to criterion [u-y]	22	18	P	-	0.26
					20	24	P		0.37
				CHT: Percent correct [u-y]	22	18	P	+	0.05
					20	24	P		0.18
Jansson	2010	Correlation	12	ERP: MMN non-native vowels	20	24	P	+	0.45
				Wordforms					
Swingley	2005	Correlation	11	HPP: Correctly > mispronounced words	17	16	U	+	0.43
Newman	2006	Contrast	7 to 12	HPP: Word segmentation	17 & 10	55	TOLD	+	0.4
Singer	2008	Contrast	5 or 13	HPP: Own name > foil in noise	20 & 21	53 to 72	PPVT	+	-0.01
Junge	2010	Correlation	7	ERP: Word segmentation	23	36	Reynell	-	-0.45
Junge	2012	Correlation	10	ERP: Recognition of word originally heard in a passage	28	12	U	-	-0.56
					28	24	P	-	-0.38
				ERP: Recognition of word originally heard in isolation	28	12	U	-	0.08
					28	24	P	-	-0.11
Junge	2011[1]	Correlation	10	ERP: Word segmentation	27	11	U	-	-0.09
					27	16	U	-	0.02
Junge	2011[2]	Contrast	10	ERP: Word segmentation	9 & 14	36	Reynell	-	-0.15
Singh	2012	Correlation	7.5	ERP: Word segmentation, pitch matched	40	24	P	+	0.32
				ERP: Word segmentation, pitch matched 2	40	24	P	+	0.4
				ERP: Word segmentation, pitch mismatched	40	24	P	+	0.29
				Prosody					
Weber	2005	Correlation	5	ERP: MMN trochaic deviant	18	12	ELFRA	-	-0.45
						24	ELFRA		-0.37
Cristia	2011	Contrast	6	HPP: Preference for well-formed phrases	13 & 11	24	P	+	0.45

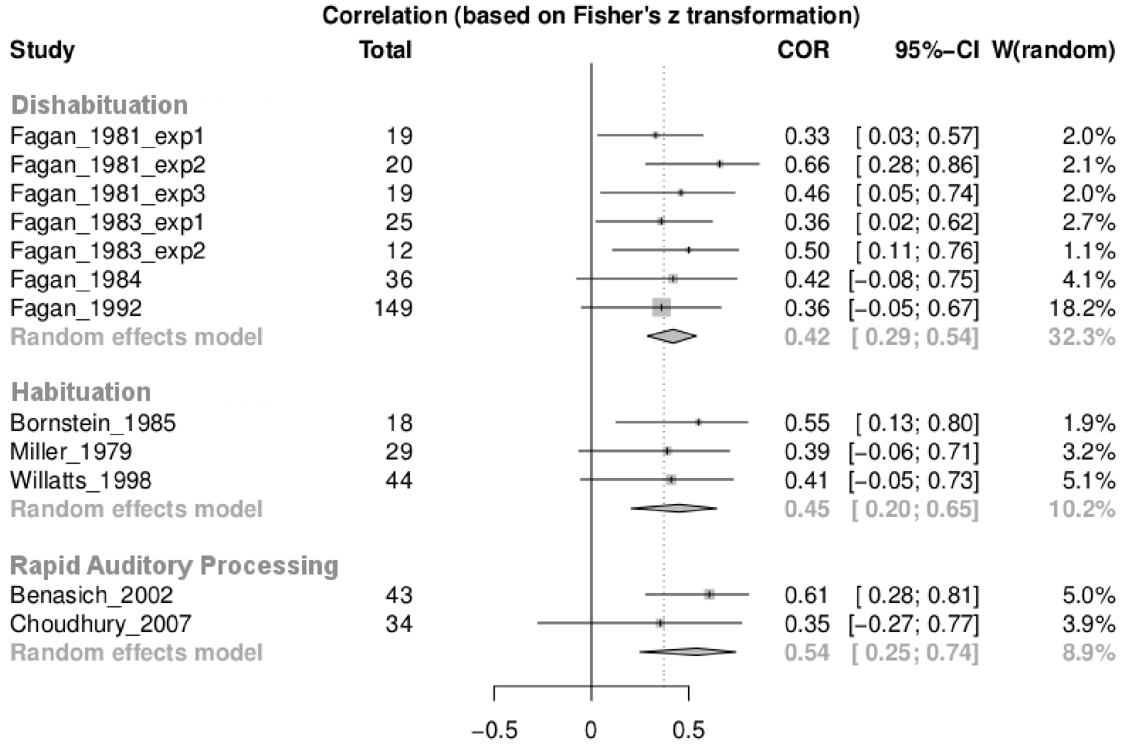
Table 2. ***Summary of correlations in individual variation across two speech perception measures.***

	Author	Year	Age	*N*	*r*	Measures and Tasks
Phones	Conboy	2005	7	16	0.48	d' native and non-native contrasts
	Kuhl	2005	7	16	0.44	MMN and d' for native contrast
			7	13	0.31	MMN and d' for non-native contrast
	Kuhl	2008	7.5	15	0.58	d' native and non-native contrasts
			7.5	6	0.74	d' two native contrasts
	Cardillo	2010	7-11	20	0.19	d' at 7 and 11 months
			7-11	20	0.34	Trials to criterion at 7 and 11 months
	Junge	2012	10	28	0.36	Word recognition in isolation vs. sentence
Wordforms	Singh	2012	7.5	40	0.27	Word recognition with and without pitch change
	Houston	2011	9	10	0.65	Novelty preference for a novel audiovisual word in 2 separate days

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 percent 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.

Figure 2. Summary results for the non-linguistic predictors. See caption in Figure 1 for details.


Correlation (based on Fisher's z transformation)
Study
Total
COR
95%−CI
W(random)
Phones
Cardillo_2010 22 −0.05 [−0.37; 0.29] 2.4%
Conboy_2005 23 0.41 [0.26; 0.53] 2.5%
Conboy_2008 17 0.09 [−0.37; 0.51] 1.7%
Jansson_2010 20 0.45 [0.04; 0.72] 2.1%
Kuhl_2005 17 0.44 [0.08; 0.70] 1.7%
Kuhl_2008 23 0.48 [0.09; 0.75] 2.5%
Rivera_2005 24 0.43 [0.07; 0.69] 2.6%
Talay_1996 12 0.61 [0.18; 0.84] 1.1%
Tsao_2004 20 0.25 [−0.18; 0.60] 2.1%
Random effects model 0.35 [0.22; 0.47] 18.9%
Words
Junge_2010_exp1 23 0.45 [0.10; 0.70] 2.5%
Junge_2011_ch4 27 0.04 [−0.35; 0.41] 3.0%
Junge_2011_ch6 23 0.15 [−0.27; 0.52] 2.5%
Junge_2012 28 0.24 [−0.07; 0.51] 3.1%
Newman_2006 27 0.40 [0.11; 0.64] 3.0%
Singer_2008 41 −0.01 [−0.48; 0.48] 4.7%
Singh_2012 40 0.34 [−0.29; 0.76] 4.6%
Swingley_2005 17 0.43 [−0.01; 0.73] 1.7%
Random effects model 0.28 [0.14; 0.40] 25.2%
Prosody
Cristia_2011 24 0.45 [−0.02; 0.76] 2.6%
Weber_2005 18 0.41 [0.13; 0.63] 1.9%
Random effects model 0.42 [0.18; 0.61] 4.5%
−0.5
0
0.5


Correlation (based on Fisher's z transformation)
Study
Total
COR
95%−CI
W(random)
Dishabituation
Fagan_1981_exp1 19 0.33 [0.03; 0.57] 2.0%
Fagan_1981_exp2 20 0.66 [0.28; 0.86] 2.1%
Fagan_1981_exp3 19 0.46 [0.05; 0.74] 2.0%
Fagan_1983_exp1 25 0.36 [0.02; 0.62] 2.7%
Fagan_1983_exp2 12 0.50 [0.11; 0.76] 1.1%
Fagan_1984 36 0.42 [−0.08; 0.75] 4.1%
Fagan_1992 149 0.36 [−0.05; 0.67] 18.2%
Random effects model 0.42 [0.29; 0.54] 32.3%
Habituation
Bornstein_1985 18 0.55 [0.13; 0.80] 1.9%
Miller_1979 29 0.39 [−0.06; 0.71] 3.2%
Willatts_1998 44 0.41 [−0.05; 0.73] 5.1%
Random effects model 0.45 [0.20; 0.65] 10.2%
Rapid Auditory Processing
Benasich_2002 43 0.61 [0.28; 0.81] 5.0%
Choudhury_2007 34 0.35 [−0.27; 0.77] 3.9%
Random effects model 0.54 [0.25; 0.74] 8.9%
−0.5
0
0.5