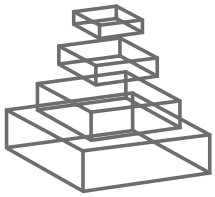
frontiers
RESEARCH TOPICS

THE NAÏVE LANGUAGE EXPERT: HOW INFANTS DISCOVER UNITS AND REGULARITIES IN SPEECH

Topic Editors
Jutta L. Mueller and Claudia Männel

frontiers in PSYCHOLOGY



frontiers

ISSN 1664-8714
ISBN 978-2-88919-329-5
DOI 10.3389/978-2-88919-329-5

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

THE NAÏVE LANGUAGE EXPERT: HOW INFANTS DISCOVER UNITS AND REGULARITIES IN SPEECH

Topic Editors:
Jutta L. Mueller, University of Osnabrueck, Germany
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Infants and brain
Image source and copyright: Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

The advent of behavior-independent measures of cognition and major progress in experimental designs have led to substantial advances in the investigation of infant language learning mechanisms. Research in the last two decades has shown that infants are very efficient users of perceptual and statistical cues in order to extract linguistic units and regular patterns from the speech input. This has lent support for learning-based accounts of language acquisition that challenge traditional nativist views. Still, there are many open questions with respect to when and how specific patterns can be learned and the relevance of different types of input cues. For example, first steps have been made to identify the neural mechanisms supporting on-line extraction of words and statistical regularities from speech. Here, the temporal cortex seems to be a major player. How this region works in concert with other brain areas in order to detect and store new linguistic units is a question of broad interest.

In this Research Topic of Frontiers in Language Sciences, we bring together experimental and review papers across linguistic domains, ranging from phonology to syntax that address on-line language learning in infancy. Specifically, we focused on papers that explore one of the following or related questions: How and when do infants start to segment linguistic units from the speech input and discover the regularities according to which they are related to each other? What is the role of different linguistic cues during these acquisition stages and how do different kinds of information interact? How are these processes reflected in children's behavior, how are they represented in the brain and how do they unfold in time? What are the

characteristics of the acquired representations as they are established, consolidated and stored in long-term memory?

By bringing together behavioral and neurophysiological evidence on language learning mechanisms, we aim to contribute to a more complete picture of the expeditious and highly efficient early stages of language acquisition and their neural implementation.

Table of Contents

frontiers in
PSYCHOLOGY

EDITORIAL
published: 20 August 2013
doi: 10.3389/fpsyg.2013.00526

The naïve language expert: introduction to the research topic

Jutta L. Mueller[1,2]* and Claudia Männel[2]

[1] *Psycho/Neurolinguistics Group, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany*
[2] *Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
**Correspondence: muellerj@cbs.mpg.de*

Edited by:
Manuel Carreiras, Basque Center on Cognition, Brain, and Language, Spain

Since the first seminal reports of young infants' abilities to use both acoustic (e.g., Mandel et al., 1994) and statistical cues (e.g., Saffran et al., 1996) to structure, categorize, and memorize linguistic units from their speech input, the quest for capturing infants' abilities and limitations in the discovery of basic elements and regularities in speech has attracted a lot of attention. While many important findings have been unveiled using sophisticated behavioral methods that allow to measure infants' discrimination of familiar vs. unfamiliar speech sounds, the field has gained a new momentum with the advent of techniques, such as event-related brain potentials (ERPs) or functional near-infrared spectroscopy (fNIRS), which allow to measure discrimination even in the absence of overt behavioral responses. After the first excitement about infants' amazing abilities, new challenges have emerged, for example, the question how different input cues interact, how learner variables, such as bilingual language input, contribute to learning mechanisms, or how low-level learning mechanisms contribute to higher-order language learning, such as word learning or sentence comprehension.

The goal of the current Research Topic is to provide a cutting-edge snapshot of this active research field integrating original research papers using both behavioral and neurophysiological techniques with review articles providing ideas for general frameworks capturing those findings.

Three reviews and one methods article offer global and thought-provoking views on basic principles and computational mechanisms that are operative in early language learning. Seidl and Cristia (2012) provide an overview of research on the discovery of allophony vs. phonemic differences in early infancy and discuss mechanisms supporting this distinction. Bortfeld et al. (2013) make a case for using neurophysiological methods, such as ERPs and fNIRS to investigate two factors they consider basic ingredients for early language learning, namely salience and familiarity. Krogh et al. (2012) provide a timely review of statistical learning across modalities and outline different types of constraints and underlying learning mechanisms. Arciuli and Torkildsen (2012) provide evidence for a close interaction between statistical learning and language processing in normal and impaired language acquisition and call for longitudinal research programs shedding light on this relationship.

Two of the original research papers applied neurophysiological methods. Minagawa-Kawai et al. (2013) report an fNIRS study on the emergence of phonotactic abilities in a cross-linguistic sample of infants. The authors report a null-result and discuss potential methodological pitfalls when using fNIRS. Kooijman et al. (2013) used ERPs measured at the age of 7 months as a predictor of later language skills showing the potential sensitivity and meaningfulness of neurophysiological measures with respect to inter-individual differences across language development.

Another set of research articles focuses on the contribution of prosodic information to the perception of sentential structure. Wellmann et al. (2012) evaluate the role of different prosodic boundary cues in German-learning infants' discrimination of coordinate noun phrases, showing that two out of three cues are sufficient for 8-month-olds to solve this task. For the same age, Bernard and Gervain (2012) show that French-learning infants use prosodic prominence and word frequency as signals to word order in an artificial language.

The largest group of papers deals with specific questions related to speech segmentation. Bosch et al. (2013) investigate word segmentation in 6- and 8-month-olds in previously under-investigated, syllable-timed languages (i.e., Spanish and Catalan) and provide evidence for the early emergence of this ability in monolinguals and bilinguals. For English-learning infants, Thiessen and Erickson (2012) show that this ability emerges even at 5 months if artificial-language units are marked by transitional probabilities and word stress, and that infants' segmentation is guided by transitional probabilities if both information types are placed in conflict. Yurovsky et al. (2012) also study regularities signaling word-like units in child-directed speech, that is, position and onset cues in naming frames. The authors report that in an artificial language either regularity is sufficient to trigger speech segmentation and subsequent word learning in adults. Mintz (2013) is interested in the question when infants are able to segment morphosyntactic endings from verb stems and provides evidence that this happens starting from the first half of the second year of life. Graf Estes (2012) demonstrates that infants at 11 and 17 months recognize words across acoustic variations after successful statistical segmentation and at the older age even apply these generalizations as labels of new objects.

Finally, our Research Topic contains one study which analyzes infant speech production during the second year of life. Yamashita et al. (2013) study English- and Japanese-learning children's phonetic inventory across 15, 20, and 24 months and assume adult-like vocal tract structures to be present by 24 months of age for both languages.

As a compendium of current research efforts in the field of early language learning mechanisms we are confident that this Research Topic offers novel and stimulating ideas for those who are new to the field and would like to get a timely overview as well as for experts who are interested in current developments.

REFERENCES

Arciuli, J., and Torkildsen, J. V. (2012). Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data. *Front. Psychol.* 3:324. doi: 10.3389/fpsyg.2012.00324

Bernard, C., and Gervain, J. (2012). Prosodic cues to word order: what level of representation? *Front. Psychol.* 3:451. doi: 10.3389/fpsyg.2012.00451

Bortfeld, H., Shaw, K., and Depowski, N. (2013). Disentangling the influence of salience and familiarity on infant word learning: methodological advances. *Front. Psychol.* 4:175. doi: 10.3389/fpsyg.2013.00175

Bosch, L., Figueras, M., Teixidó, M., and Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. *Front. Psychol.* 4:106. doi: 10.3389/fpsyg.2013.00106

Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Front. Psychol.* 3:447. doi: 10.3389/fpsyg.2012.00447

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., and Cutler, A. (2013). Predictive brain signals of linguistic development. *Front. Psychol.* 4:25. doi: 10.3389/fpsyg.2013.00025

Krogh, L., Vlach, H. A., and Johnson, S. P. (2012). Statistical learning across development: flexible yet constrained. *Front. Psychol.* 3:598. doi: 10.3389/fpsyg.2012.00598

Mandel, D. R., Jusczyk, P. W., and Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information. *Cognition* 53, 155–180.

Minagawa-Kawai, Y., Cristia, A., Long, B., Vendelin, I., Hakuno, Y., Dutat, M., et al. (2013). Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar. *Front. Psychol.* 4:170. doi: 10.3389/fpsyg.2013.00170

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in english-learning 15-month-olds. *Front. Psychol.* 4:24. doi: 10.3389/fpsyg.2013.00024

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Seidl, A., and Cristia, A. (2012). Infants' learning of phonological status. *Front. Psychol.* 3:448. doi: 10.3389/fpsyg.2012.00448

Thiessen, E. D., and Erickson, L. C. (2012). Discovering words in fluent speech: the contribution of two kinds of statistical information. *Front. Psychol.* 3:590. doi: 10.3389/fpsyg.2012.00590

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., and Höhle, B. (2012). How each prosodic boundary cue matters: evidence from german infants. *Front. Psychol.* 3:580. doi: 10.3389/fpsyg.2012.00580

Yamashita, Y., Nakajima, Y., Ueda, K., Shimada, Y., Hirsh, D., Seno, T., et al. (2013). Acoustic analyses of speech sounds and rhythms in Japanese- and english-learning infants. *Front. Psychol.* 4:57. doi: 10.3389/fpsyg.2013.00057

Yurovsky, D., Yu, C., and Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: scaffolding from child-directed speech. *Front. Psychol.* 3:374. doi: 10.3389/fpsyg.2012.00374

Received: 24 July 2013; accepted: 26 July 2013; published online: 20 August 2013.
Citation: Mueller JL and Männel C (2013) The naïve language expert: introduction to the research topic. Front. Psychol. ***4****:526. doi: 10.3389/fpsyg.2013.00526*
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 02 November 2012
doi: 10.3389/fpsyg.2012.00448

Infants' learning of phonological status

Amanda Seidl*[1]* and *Alejandrina Cristia*[2]

[1] *Purdue University, West Lafayette, IN, USA*
[2] *Neurobiology of Language, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands*

Edited by:
Claudia Männel, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Judit Gervain, CNRS – Université Paris Descartes, France
Nivedita Mani, Georg-August-Universität Göttingen, Germany

***Correspondence:**
Amanda Seidl, Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, IN 47906, USA.
e-mail: aseidl@purdue.edu;
Alejandrina Cristia, Neurobiology of Language, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, Netherlands.
e-mail: alecristia@gmail.com

There is a substantial literature describing how infants become more sensitive to differences between native phonemes (sounds that are both present and meaningful in the input) and less sensitive to differences between non-native phonemes (sounds that are neither present nor meaningful in the input) over the course of development. Here, we review an emergent strand of literature that gives a more nuanced notion of the problem of sound category learning. This research documents infants' discovery of *phonological status*, signaled by a decrease in sensitivity to sounds that map onto the same phonemic category vs. different phonemic categories. The former phones are present in the input, but their difference does not cue meaning distinctions because they are tied to one and the same phoneme. For example, the diphthong *I* in *I'm* should map to the same underlying category as the diphthong in *I'd*, despite the fact that the first vowel is nasal and the second oral. Because such pairs of sounds are processed differently than those than map onto different phonemes by adult speakers, the learner has to come to treat them differently as well. Interestingly, there is some evidence that infants' sensitivity to dimensions that are allophonic in the ambient language declines as early as 11 months. We lay out behavioral research, corpora analyses, and computational work which sheds light on how infants achieve this feat at such a young age. Collectively, this work suggests that the computation of complementary distribution and the calculation of phonetic similarity operate in concert to guide infants toward a functional interpretation of sounds that are present in the input, yet not lexically contrastive. In addition to reviewing this literature, we discuss broader implications for other fundamental theoretical and empirical questions.

Keywords: infants, perception, phonology, phonemes, speech

INTRODUCTION

There is a large literature on how infants become more sensitive to differences between phones that map onto different native phonemes (sounds that are both present and meaningful in the input) and less sensitive to differences between phones that map onto different non-native phonemes (sounds that are neither present nor meaningful in the input) as they mature. This literature shows that infants begin to zero-in on the phonemes present in their native language sometime between 4 and 12 months of age (Werker and Tees, 1984; Polka and Werker, 1994). However, categorizing sounds as either present in, as opposed to absent from, the input is only one step in language acquisition. Certainly this helps the infant to focus on her specific language's properties and ignore other language's properties, and this ability to build language-specific phonetics may even be fundamental in building a lexicon (Kuhl et al., 2008). However, the child must also learn to categorize sounds which are present, but not meaningful in the input language. This task is likely to recruit the same mechanisms as the native/non-native task. Specifically, in every language, there are sounds that are present in the input but do not map onto different native phonemes, since their different forms do not cue meaning distinctions and the child must learn to map these to a unified phonemic representation. For example, the diphthong *I* in *I'm* should map to the same underlying category as the diphthong in *I'd*, despite the fact that one vowel is nasal and the other oral. For ease of expression and reading, we will use the shorthand of "allophones" for phones that map onto the same phonemic category, and "phonemes" for phones that map onto different phonemic categories[1]. In this paper we summarize evidence on the acquisition of allophones to answer two key questions: When and how does the learner determine whether two sounds are allophones or phonemes in the target language?

Before turning to the evidence on the acquisition of allophones and the mechanisms underlying their acquisition, it is important to discuss both how allophones and phonemes are defined within the linguistic, descriptive literature (Section "What are allophones"); and how they are processed by individuals with a fully developed grammar according to the psycholinguistic literature (Section "The end state"). We then review an emergent strand of

[1]"Allophone" is used somewhat variably across papers. For example, some use the word to denote the more marginal pronunciations of a sound (e.g., if vowels are nasalized before nasal vowels and are oral elsewhere, then some would call the nasal alternate an allophone and the oral one a phoneme, Peperkamp et al., 2006). In more traditional phonological terms, all sounds are allophones, surface representations that map onto some phoneme (abstract representation). Following this definition, for example, one should state that in English oral and nasal [i] are allophones of the same phoneme, whereas oral [i] and oral [e] are allophones of different phonemes. We adopt the latter definition, except that for ease of reading we will refer to cases like the previous one (allophones that map onto the *same* phoneme) as "allophones," and to the latter (allophones that map onto *different* phonemes) as "phonemes."

literature that documents when infants begin to apply a differential processing of allophones and phonemes (Section "Infants' processing of allophones") and how they might have learned to make such a distinction between allophones and phonemes (Section "Mechanisms for learning allophones"). The final section ("Implications") discusses how research on infants' learning of phonological status can inform, and be informed by, other areas of investigation. Throughout this article, we identify areas where answers are still lacking. We hope that this review serves as a springboard for such work and helps to point to clear areas out of which this future work can grow.

WHAT ARE ALLOPHONES?

There are two classical cases of allophony, which are commonly discussed in introductory linguistics courses (Trubetzkoy, 1939/1969; Kenstowicz, 1994). The first involves "sounds in complementary distribution." Two sounds are in complementary distribution if the sound which should be used is completely predictable from the context; put differently, the contexts in which each sound can occur are completely non-overlapping. For example, in most varieties of American English, dark /l/ occurs syllable-finally ("ball"), whereas light /l/ occurs in all other positions ("lab"). Notice that no two words in American English differ only on whether they have a light or dark /l/. In other words, sounds in complementary distribution do not cue meaning distinctions. Finally, a third criterion for allophony in this case is that the two sounds must be somehow acoustically related, such that they may be interpreted as the "same" sound, on some abstract level. For instance, although /ŋ/ and /h/ are in complementary distribution in English (the former occurs only in syllable codas, the latter only in syllable onsets), phonologists would not want to posit that they are allophones since they are highly acoustically distinct (Bazell, 1954).

The second classical case of allophony relates to sounds in "free variation." In this case, speakers can produce two or more different sounds in the exact same environment (e.g., ri[ɾ]er versus ri[d]er in American English); however, these differences are not lexically relevant. Much work debates the name "free," since in many such cases the variant which is selected appears to be explained, to a considerable extent, by a number of structural, sociolinguistic, and idiolectal variables (e.g., Fischer, 1958). Nonetheless, it remains the case that two sounds which can be thus exchanged without semantic changes can be viewed as allophones. The traditional way of establishing whether two sounds are in free variation is by carrying out a minimal pair test. Minimal pairs are two word forms that differ in only one sound; if this sound swap results in meaning change or loss, then the two sounds are phonemes, but if it does not, they are allophones in free variation.

In phonology, as in life, things can sometimes get more complicated, and for the definition of allophony this is true in a number of ways. To begin with, there are cases of complementary distribution and free variation that are true in certain phonological and lexical contexts, but not others. For example, one could state that voiceless unaspirated and voiceless aspirated stops are in complementary distribution in American English, with the former occurring e.g., after /s/ (as in "stop") and the latter e.g., in the onset of monosyllables (as in "top"). However, voiceless unaspirated stops are minimally different from the surface realizations of voiced stops in syllable-initial position when following a word ending in /s/, to such an extent that one-year-olds fail to discriminate them (Pegg and Werker, 1997)[2].

Moreover, sometimes two pronunciations are possible without meaning change in some structural positions (e.g., [i]conomy vs. [ə]conomy) but not in others (e.g., wom[ɪ]n vs. wom[ə]n; though perhaps a clearer example is tense and lax vowels, both of which can occur in closed syllables, but only tense vowels occur in open syllables). Additionally, some sounds would fit the definition of phonemes, but may be present in only a handful of loanwords (such as a pronunciation of the composer Bach as Ba[x] versus Ba[k]; Halle, 1964); whereas for others there may be no minimal pairs, even though the linguist's intuition indicates that two sounds are contrastive because they are both active (used in a phonological constraint/rule) and prominent (involved in some type of phonological, morphological, or even long distance effect; Clements, 2001)[3]. Scobbie et al. (1999) and Scobbie and Stewart-Smith (2008), among others, have discussed extensively another ambiguous case from Scottish Standard English, where some vowels have long and short variants that are contextually determined, yet for which some minimal pairs, with specific morpholexical characteristics, can nevertheless be found. This is the case for long and short variants of /ai/, which contrast minimally in "side" and "sighed," with the long version being found in morphologically complex items. In spite of the existence of such minimal pairs, the two sounds are in free variation across speakers in some lexical items, such as "crisis."

In view of such cases both within and across languages, Pierrehumbert (2003) proposes to do away with the distinction between phonemes and allophones and instead attempt an explanation of learners' acquisition of positional allophones, defined as clusters of tokens in acoustic space. A comparable proposal was made in Ladd (2006), who goes further by pointing out that allophones are sometimes very meaningful sociolinguistically, and are thus highly perceptually salient to native speakers. Scobbie and Stewart-Smith (2008) argue, instead, that while the concepts of allophones

[2]One anonymous reviewer points out that this problem only exists in the case that the speech stream is segmented, since only segmentation into syllables would lead to the conclusion that the allophone of /t/ that occurs in "st" clusters does not group with the phonetically similar [d], but rather with the phonetically dissimilar [t]. It is unclear when exactly infants segment into syllables. Some data point to syllables being the basic unit of analysis even for newborns, allowing the discrimination of /atspa-apsta/ but not that of /tsp-pst/ (Bertoncini et al., 1988); while other data suggests a protracted development, as infants do not use the syllable-determined allophones of /r/ in "night rate" versus "nitrate" to segment these words from running speech until about 10 months of age. Thus this is certainly a question that warrants further exploration.

[3]This sort of "active" contrast is eminently common in sign languages, which have very few clear minimal pairs. Thus while a few clear minimal pairs exist e.g., in the domain of handshape (Brentari, 1998), most are cases of near-minimal pairs or cases where a contrast exists in one area of the lexicon, but looks distinct in another area of the lexicon (Brentari and Eccarius, 2012). For example, according to Diane Brentari (p.c.) The ASL sign THOUSAND was originally borrowed from the initialized French Sign Language sign MILLE and had a 3-finger "M" handshape. During the process of nativization the "M" (3 fingers) became "B" (all 4 fingers). That is the more marked 3 fingers became the less marked 4 fingers handshape. However, in this location with this movement there are no minimal pairs with either of these handshapes.

and phonemes may be useful end points, a continuum could exist between allophones and phonemes, and propose that speakers/listeners' grammars could well be fuzzy. More recently, Hall (2009) makes specific proposals as to how to predict perceptibility from gradient versions of an allophony/phonemicness scale. Clearly the limitations of the classical definitions of allophones and phonemes are not new (see e.g., Pike, 1947), but they are just now beginning to gain a unique combination of linguistic *and* psycholinguistic attention as it becomes increasingly clear that such phenomena are not marginal, and that such gradience is relevant to both language learning and processing. Indeed, a look through **Table 1** reveals a window into the scope of this gradience. While it is not the aim of this paper to debate phonological theory, nor to enumerate cases along this continuum, we keep the question of gradience in mind when considering how infant learners may approach the phonological system, and what types of allophones versus phonemes (i.e., at what point of the continuum) have been studied in previous experimental work. With this enriched view of allophony, we now turn to adults' perception of these two (or more) "classes" of sounds.

THE END STATE: ADULTS' PROCESSING OF ALLOPHONES

It should be noted from the outset that the study of whether listeners' sensitivity for contrasts that are allophonic is lower than those that are phonemic faces certain methodological roadblocks, which are worthy of discussion here. One way to interpret this hypothesis is the following: Holding the listener and language constant, one would compare a contrast A that is phonemic against a contrast B that is allophonic, to find that A is processed better (discriminated more speedily and accurately; used for tracking phonological patterns; recruited for coding lexical contrasts). Much of the initial literature we review uses this design (e.g., Whalen et al., 1997; McLennan et al., 2003). When using this design, there is an obvious interpretation alternative to phonological status affecting perception: perhaps there is an intrinsic discriminability difference between A and B. A safeguard against this state of affairs is to test two sets of participants, who have different native languages, and hold the contrast constant, an approach that is also common in the literature (e.g., Johnson and Babel, 2010). In this scenario, intrinsic differences in discriminability of contrasts are irrelevant, since only one contrast is used. However, another problem arises, namely that the stimuli must be recorded in some language. If they are recorded in the language where the contrast is allophonic, they may be pronounced less clearly (provided that speakers tend to neutralize such contrasts), which is not desirable. But if they are recorded in the language where they are phonemic, then the test may facilitate the performance of listeners of that language, who will find the stimuli native. The solution that researchers are increasingly adopting is to use a *third* language where the contrast is phonemic, such that the stimuli are equally foreign to both sets of participants. Results from the latter approach actually fit in perfectly with conclusions derived from the two other (e.g., Boomershine et al., 2008), lending further credence to this body of literature.

In brief summary, previous work suggests that adults do not discriminate allophonic alternates as well as phonemes both behaviorally (Whalen et al., 1997; Peperkamp et al., 2003; Boomershine et al., 2008; Shea and Curtin, 2011) and electrophysiologically (Kazanina et al., 2006; Hacquard et al., 2007). Furthermore, adults rate allophones as more similar to each other than phonemes (Johnson and Babel, 2010). Additionally, words differing in sounds that are allophonic prime each other, but words differing in sounds that are phonemic do not (McLennan et al., 2003). These differences in processing come about as the result of native language exposure and thus second language learners have a hard time gaining sensitivity to sounds that are phonemic in the target language even when those same sounds are present allophonically in the learners' native language (Kondaurova and Francis, 2008).

Thus, overall, perceptual evidence in adults confirms that allophonic and phonemic sounds are not treated similarly. Given that there may be a continuum of allophones to phonemes, as mentioned above, it is worthwhile to evaluate whether this differential behavior arises only for the categorical phonemic/allophonic stages, or also for intermediate cases. This is especially true given recent findings that listeners treat sounds differently based on the reliability of their distributions (Dahan et al., 2008). Specifically, in this study Dahan et al. (2008) examined adults' perception of tensed and laxed variants of /æ/ in the environment of /k/ and /g/. When /æ/ was consistently tensed before /g/, but not /k/ they found a training effect in the experiment such that listeners, upon hearing e.g., the non-tensed /æ/ came to anticipate the following segment as /k/. Thus, when segments vary allophonically in a reliable way this can lead to differential processing very quickly. This is not the case when the variation is not predictable.

In **Table 1**, we reclassify adult perceptual studies in terms of the type of contrast that has been examined. There are several studies which explore one of the endpoints (e.g., Whalen et al., 1997 examines a case of clear complementary distribution) and only a few studies which explore points in between. For example, in English [ð] can never map onto /d/, and therefore they form a phonemic contrast. Whereas [ɾ] is a possible realization of /d/ in word-medial context, there are also quite a few lexical exceptions where they occur in near overlapping distributions (e.g., rider vs. writer). Thus, the comparison of English adults' perception of the [ɾ] and [d], on the one hand, against [ð] and [d], on the other, represents the study of an intermediate case of allophony. This was undertaken in Boomershine et al. (2008), who found poorer discrimination of the former than the latter. Results cannot be attributed to the actual acoustic items used, since the same mapped onto different types for a second group of adults, whose native language was Spanish. In Spanish [ð] and [d] are in complementary distribution (classic allophony) and [ɾ] and [ð] are mostly in overlapping distribution (classic phonemic). Despite the fact that not all items fell on the extremes of the phonemicness/allophony continuum, perceptual results were the opposite across language groups in all tests but a measure of reaction time. Thus, this work seems to suggest that differences between phonemic and allophonic processing are found even when non-extreme points of the continuum are investigated.

Nonetheless, a different pattern emerges from work using electrophysiology. Hacquard et al. (2007) and Kazanina et al.

Table 1 | Perceptible: native speakers report hearing the difference; Unpredictable: the phone cannot be predicted by its phonological context; Lexical: there are many examples of minimal pairs sustaining the contrast.

Type	Perceptible	Unpredictable	Lexical	Example	First author of relevant study
Phonemic	Yes	Yes	Yes	AmE [b-d]	Whalen; Hacquard
	Yes	Usually	Yes	AmE [l-i][a]	
	Yes	Mildly	Marginal	ScE [x][b]	
	Yes	No	Several	Sc [ai-ai:][c]	
	Yes	No	No	German ich-ach	Hacquard
	Yes	Rarely	Yes	AmE [ɾ-d][d]	Boomershine
	No	Rarely	No	AmE [æ-æ̃][e]	
Allophonic	No	No	No	[p-p^h]	Whalen

(Many examples below are from Scobbie and Stewart-Smith, 2008.)

[a] *Typically contrastive, but neutralized in some positions, e.g., seat vs. sit, but see vs. */sl/.*

[b] *Only present in a handful of lexical items, e.g., lock vs. loch.*

[c] *Only present in a handful of items, and predicted by syllable structure e.g., side vs.sighed.*

[d] *[d] is typically realized as [ɾ] in specific contexts, so they are usually predictable from the context, but there are some cases where they contrast, e.g., rider-writer.*

[e] *Some talkers use more heavily while others do not. E.g., some nasalize in non-nasal environments, or fail to nasalize in nasal environments.*

(2006) both explore cases of complementary distribution, free variation, and overlapping distributions. While there are other effects in these studies (e.g., inventory size), overall, using an oddball paradigm, they find different processing results for complementary distribution (which patterns like overlapping distribution) and free variation (which patterns differently). For example, Kazanina et al. (2006), using Russian and Korean manipulated stimuli, find that while Russian listeners (for whom t/d are phonemic) show a significant mismatch response, Korean listeners (for whom t/d are allophonic) show no such response to the exact same stimuli. Hacquard et al. (2007) also examine whether the amplitude to the mismatch response in an oddball paradigm is related to the phonological status of the sounds in question using synthesized stimuli on vowel tenseness in continental French, Argentinean Spanish, and Puerto Rican Spanish listeners. Tense and lax [e]/[ɛ] are phonemic in continental French, allophonic in Argentinean Spanish and in free variation in Puerto Rican Spanish and this is reflected in the size of the mismatch responses. Furthermore, Argentinean listeners seem to discriminate the allophonic differences as well as they discriminate the phonemic ones based on the size of the mismatch response, but Puerto Rican listeners seem to discriminate the allophonic/free variation contrast more poorly than a phonemic contrast ([e]/[a]). Thus, theoretical descriptions and psycholinguistic evidence suggests that allophones and phonemes are different and that typology of the allophones may also be a factor in processing at least at some level. The next section assesses when these differences in processing come about over the course of development.

INFANTS' PROCESSING OF ALLOPHONES

As mentioned above, a considerable body of literature suggests that perception of non-native (absent) sounds declines, whereas perception of native phonemes improves toward the end of the first year of life (Polka et al., 2001; Kuhl et al., 2006; Narayan, 2006). This improvement likely relates to the much richer and more abundant cues for the former: The child will accumulate more passive, phonetic exposure to the former; she may attempt these sounds; she may learn some words that have them, and so forth. The first question we would like to answer is when listeners become less sensitive to allophonic distinctions and more sensitive to phonemic ones. We review evidence from discrimination, phonotactic learning, phonotactic processing, and word learning suggesting that infants are sensitive to phonological status.

English-learning 2-month-olds discriminate allophonic variants (e.g., /t/ in "night rate" versus "nitrate"; Hohne and Jusczyk, 1994) showing an initial sensitivity to sounds that will eventually be treated as allophones later in life. Recent work suggests that, while young infants are sensitive to sounds that are allophones in their ambient language, this sensitivity declines with maturation and language-specialization. Specifically, Seidl et al. (2009) briefly familiarized English- and Quebec French-learning infants with a pattern that depended upon vowel nasality. Note that as mentioned earlier vowel nasality is phonemic in French, but allophonic in English. Infants in this study heard syllables in which nasal vowels were followed by fricatives, but oral vowels were followed by stops. Then they were tested on their ability to generalize this pattern to new syllables. English-learning 4-month-old infants were able to learn this novel phonotactic dependency involving vocalic allophones and behaved like French-learning 11-month-old infants, for whom nasality is phonemic. However, by 11 months of age English-learners were no longer able to encode this abstract phonotactic regularity and showed no evidence of learning. It should be noted that these older infants are not completely impervious to allophones, since they use them to extract words from running speech at 10.5 months (Jusczyk et al., 1999). Rather, these results suggest that the same exact sounds no longer function in the same manner across languages which use them as phonemes versus allophones.

It might be suggested that some of the contrasts that have been studied as allophones could be more perceptually difficult than ones that have been explored as phonemes. Specifically,

allophonic alternates may simply be more difficult to discriminate because they represent subtle changes. For example, Pegg and Werker (1997) found that two phones that map onto different phonemes /t/ and /d/, but are extremely similar, are not discriminable by one-year-olds. In their study, they measured sensitivity to the word-initial realization of /d/ against the post-/s/ realization of /t/, which differ very subtly. However, an important point is that simple acoustic distance between the tokens used in any given test cannot explain developmental changes in allophonic sensitivity, since this sensitivity changes with age and language exposure. Even in the Pegg and Werker (1997) study, 6-month-old were, in fact, able to distinguish the very similar surface realizations of /t/ and /d/. Similarly, Dietrich et al. (2007) and Seidl et al. (2009) show that attention to the same contrast declines in languages for which they are allophonic, but not in languages in which they are phonemic. For example, Dietrich et al. (2007) show that 18-month-old Dutch-, but not English-learning toddlers interpret vowel length as lexically contrastive. Thus, it appears that while sensitivity to allophonic sounds initially exists in infancy, it appears to decline by 11 months of age (Seidl et al., 2009) as infants converge on the native phonemic contrasts present in their input language and come to ignore the non-native ones which are not present in their input language (Werker and Tees, 1984).

It is worthy of note that most of the studies cited above have been conducted on English-learning infants (albeit with two exceptions, Dietrich et al., 2007; Seidl et al., 2009). If we are to draw any clear conclusions concerning the time course of allophonic sensitivity, we will need to expand this work cross-linguistically, since it may be that the time course is different across languages and may also be impacted by the kind of sound distinction explored. Unfortunately, such single language studies only allow for certain allophonic sounds to be tested, and confound potential differences in discriminability with phonological status.

Also worthy of mention is that many allophonic alternates in the studies mentioned above are predictable from the phonological context. For example, the aspiration of /t/ studied in Hohne and Jusczyk (1994) represents a clear case of complementary distribution or classic allophony. Exceptions to this are the cases of vowel nasalization utilized in Seidl et al. (2009) and the case of vowel length in Dietrich et al. (2007). Specifically, although vowels are nasalized before tautosyllabic nasal consonants in English, they are also often nasalized in other locations (e.g., within a word with another nasalized vowel), so complementary distribution does not entirely hold. Thus, although there are cases where nasalization of vowels is completely predictable on phonotactic grounds (before nasal Cs in the same syllable), we also see nasalization in other locations for coarticulatory reasons. To add more complexity to this picture, variation in nasalization has been reported across American English dialects, such that nasalization could become a sociolinguistically relevant feature (e.g., a marker of African American Vernacular English), more than a phonotactically relevant one. Similarly, in Dutch we see a case where vowel length is difficult to classify using our classic definition of allophony. Although there are minimal pairs with vowel length in Dutch, the presence if minimal pairs occurs unevenly across the inventory. For example, /ɔ/ has long and short minimal pairs that differ mostly in length (although there are slight vowel quality differences). All other vowels that have been described as contrastive in length show considerable changes of vowel quality with the addition of length, much as we see in English tense-lax pairs. Certainly, the infant literature is not rich enough to conclude that there are no differences among the different degrees of allophony. Nonetheless, current research suggests that in infants, as in adults, even degrees of contrastiveness may make a difference, with more allophonic pairs being processed less well than more phonemic pairs. Across all studies, however, it appears that younger infants attend to salient distinctions more than older infants when the distinctions are allophonic in the target language.

MECHANISMS FOR LEARNING ALLOPHONES

Young toddlers treat allophones as distinct from phonemes. Further, some of the evidence reviewed suggests that they come to do so within the first year of life. How does such a young toddler come to treat allophones as distinct given that they clearly vary from language to language? Or more specifically, how do they come to attend less to allophonic sound pairs and attend more to phonemic sound pairs? There are several possible answers. Below, we describe computational models and laboratory studies documenting the ways by which allophonic treatment could come to be distinct from phonemic treatment.

PHONETIC MECHANISMS

One possibility for learning the difference between allophones and phonemes is that phonological status may be partially coded in the acoustic signal. Specifically, it may be that allophonic alternates are less distant from each other than phonemic ones; this difference could ensue because speakers produce them less clearly since their listeners pay little attention to them and thus communication is not compromised by their lack of distinctiveness; or simply because speakers themselves do not hear the difference very clearly, and thus never hyperarticulate these sounds. Such a strategy appears to be a cheap and sensible one, since infants are extremely sensitive to the acoustic properties of phonemes in their input (Maye et al., 2002; McMurray and Aslin, 2005; Cristia et al., 2011).

Corpus studies confirm that phonological status is, indeed, coded in the acoustic signal. Yuan and Liberman (2011) measured the Mel-frequency cepstral coefficients (MFCCs) of nasal and oral vowels in three languages (Mandarin, Portuguese, English) and after training used a classifier to sort the vowels into either nasal or oral classes. Results revealed that classification was easier for Portuguese, a language with phonemic nasality, than in either English or Mandarin, languages in which nasality is allophonic. Thus, these data support the idea that there may be acoustic cues to the classification of either phoneme or allophone, such that phonemes are more distinct and hence more easily classified using MFCCs.

Similar findings may obtain in infant-directed speech. In recent work, Cristia et al. (2010) measured two different phonemic and allophonic contrasts in infant- and adult-directed speech in corpora of Quebec French and American English. Specifically, they explored tenseness which is phonemic in English ("bit" vs "beet"), but allophonic in Quebec French: In Quebec French tense vowels

are laxed in closed syllables. They also explored vowel nasality which is phonemic in Quebec French ("mode" vs "monde"), but allophonic in English: In English vowels are nasalized before tautosyllabic nasal consonants. After collecting corpora of both tense and lax, and nasal and oral vowel pairs in each language in phonologically controlled environments and in both infant- and adult-directed registers, they conducted acoustic measures of Euclidean distance between vowel-specific alternates (nasal/oral, tense/lax) using traditional acoustic measures of tenseness and nasality. Results revealed that in terms of acoustic distance the tense/lax pairs of vowels were closer in the allophonic language than in the phonemic one regardless of the specific vowels explored. Nasality, on the other hand, was equally marked in both the phonemic (French) and the allophonic (English) language. While it may be the case that this unevenness was found because nasality is simply more difficult to measure acoustically than tenseness, if we take this data at surface value it appears that the phonemic vs allophonic distinction is better marked in some areas of acoustic space than others.

Although some information on phonemic status is clearly present in the signal, corpora studies cannot reveal whether the infant learner actually uses this acoustic information about the "closeness" of sounds in her phonological processing. Further work is necessary to answer this question.

While the argument of phonetic similarity is convincing for some cases of allophony, it is unlikely that it could explain perceptual desensitization for all sounds that adults treat as allophones. An intuitive case in point is that of /t/ allophones in English varieties, which can sometimes (albeit rarely) be realized as glottal stops. There is *a priori* no reason to imagine that [t] and [ʔ] are similar; and certainly not more similar than [k] and [ʔ] (that is, if [ʔ] has to be the allophone of some sound, phonetically it is much closer to /k/ than /t/). In view of such arguments, researchers have also explored other mechanisms, to which we turn.

DISTRIBUTIONAL MECHANISMS

An additional possibility is that infants use distributional cues, meaning the context in which a phone occurs, to discern between allophones and phonemes. For example, in English aspirated /t/ and unaspirated /t/ do not occur in the same location, so complementary distribution can effectively be used as a key to the allophonic categorization of sounds in classical phonemic versus allophonic cases. This strategy seems a sensible one since evidence suggests that young babies may be sensitive to distributions of syllables (e.g., Saffran et al., 1996) and sounds (e.g., Chambers et al., 2003; Seidl and Buckley, 2005; Cristià and Seidl, 2008; Seidl et al., 2009).

These distributional mechanisms have received support from a recent artificial grammar learning study. White et al. (2008) explored the effects on infants' perception of exposure to an artificial grammar that could be described as having morphophonologically conditioned allophony. Specifically, they familiarized 8- and 12-month-old infants with a grammar containing "determiners" followed by "content" words in which voicing of the initial C of the content word alternated as a function of the voicing of the final segment of the function word, but only with consonants of certain manners. Note that this represents a slightly different sort of allophony than the sorts discussed above, since the "complementary distribution" did not apply within the "content" words, but it was nonetheless still predictable. While 8-month-olds were able to learn these patterns, only 12-month-olds seemed to have grouped the alternate variants into a single functional category.

In addition, computational modeling also provides some support to the complementary distribution strategy. Peperkamp et al. (2006) investigated the performance of a model that categorized sounds in complementary distribution as allophones, and sounds with overlapping distributions as phonemes. This algorithm was tested on both an artificial language as well as a simplified corpus of phonetically transcribed French. While the algorithm did well in correctly tagging allophones in the artificial grammar, its performance was more error-prone in the French language corpus. Specifically, it over-generated, generating allophonic alternates that were not actually present in French. Errors of this kind were reduced to a certain extent if phonetic proximity was also taken into account.

Peperkamp et al. (2006) also suggest that these errors occur because of the presence of many near-complementary distributions, as mentioned above. Specifically, it is the cases that exist along the continuum between allophones and phonemes, but not at the edges of this continuum, which are difficult for the algorithm to correctly classify. These may be problematic to all learning algorithms of this kind (and, though evidence does not yet support this, to infants as well!). However, since near-complementary distributions are present in natural languages and there is no clear cut-off point along the continuum that has been found, it may be that until we discover how humans process these cases along the continuum we will not be able to create algorithms to do so.

In concert, experimental and modeling results support the contribution of distributional information for learning of certain cases of allophony. They also underline that distributional information alone is not sufficient, but must be packaged together with acoustic similarity. This is not a limitation, as it is likely that multiple mechanisms work in concert for the discovery of phonological status.

LEXICAL MECHANISMS

The most informed, or high-level, source of information for phonological status involves semantic knowledge. Jakobson (1966) proposed that children use semantic cues, essentially using minimal pairs to discern which phonemes are crucial to the input language and which are not. Thus, a child might hear palatalization in English before [j,i,e]. Thus, she will hear at least two different alternate pronunciations of the word *hit*. Specifically, she will hear *hi[c] you* for "hit you," but also hear *hi[ɾ] him* for "hit him." Both of the utterances will be uttered on occasions where hitting takes place. On a lexical account, the child would decide that these two instances of *hit* must map to the same underlying structure, /hɪt/. In addition, the child will be at the same time learning which sounds are phonemes by calculating minimal pairs. Thus, the child will learn that /s/ and /h/ are distinct phonemes of English because *sit* and *hit* map onto different semantic representations. Indeed, Yeung and Werker (2009) experimentally demonstrated that infants regain attention to a non-native contrast after seeing

the members of the contrast paired with different visual referents. These two processes, one of semantic overlap and one of semantic distinction, may occur together and drive children's developing phonological representations. In a certain sense we can rule out the strong version of this hypothesis as the sole method of learning given that infants at 11 months in Seidl et al. (2009) treated allophones as distinct from phonemes. Specifically, because infants at 11 months (and likely even older: Dietrich et al., 2007) do not have many minimal pairs (Caselli et al., 1995) it seems unlikely that they can use lexical cues as the sole driving factor in their phonological category learning.

Thus, the old-style lexical hypothesis seems not to hold much promise. However, a new version of lexical bootstrapping has emerged in recent years. This work is based on the finding that minimal pairs can be insufficient for the learner to maintain a phonological distinction, and that near-minimal pairs are more useful for deciding on phonemic dimensions. Thiessen (2007) documents that 14-month-olds presented with a perfect minimal pair based on stop voicing (such as taw-daw) fail to discriminate two syllables differing along that feature, whereas toddlers exposed to near-minimal pairs (such as tawbow and dawgoo) have an easier time. Swingley (2009) goes further to propose that infants could use commonalities in the pronunciation across otherwise completely different forms (such as the first vowel in *yellow* and *better*, something one could describe as "maximal pairs") to extract sound categories, and argues that this new type of lexical bootstrapping could make a considerable contribution to infants' phonological acquisition. Swingley and collaborators have recently bolstered this case by reporting that 6-month-olds have referential knowledge of several words (Bergelson and Swingley, 2012), such that their lexicon could be slightly larger than previously thought (Tincoff and Jusczyk, 1999). Moreover, corpora analyses showed that infant-directed speech offers few true minimal pairs, but rich maximal pairs structure, which an informed machine learner can profit from to learn about the phonemes of her input language (Swingley, 2009). We expect that a similar training study with infants is underway, which would constitute the final pre-requisite for this view of lexical bootstrapping. These new versions of lexical bootstrapping assume that infants can use semantic information to pull apart phonological categories. It should follow, then, that in the absence of such separating forces, infants could collapse allophonic sounds. More specifically, if maximal pairs are necessary to establish sounds as contrastive then the absence of such pairs may aid in establishing similarities between structures and assigning phonological alternations/allophonic relationships. Thus, this same mechanism might help the toddler establish that the *I* in *I'd* and *I'm* map onto the same representation. To our knowledge, the latter argument has not been made by proponents of lexical bootstrapping of phonology, but we foresee such a theoretical development within that promising line of work.

A second strain of models of phonological acquisition does not assume rich semantic representations to separate the sounds, but proposes that infants hold a pseudo-lexicon, a dictionary of frequently encountered wordforms (Martin et al., in press). In this proposal, wordform minimal pairs are used to detect allophones, such that if the child's lexicon contains two (long) sequences of sounds that are identical except for one sound, then the two sounds that differ across the two stored sequences should be considered allophones of the same phoneme. Using such an algorithm, phones could be classified as allophones and phonemes with a much greater accuracy than with other algorithms using only distributional information, or a combination of distributional information and acoustics (detailed in the Distributional mechanisms section). A pre-requisite for this type of lexical bootstrapping is that the child has a proto-lexicon, a wordform repository. Recent experimental work corroborates this: 11-month-olds showed no preference between sequences of phones that were frequent in their input, but which did not form real words, and actual real words (Ngon et al., in press). In contrast, they do prefer frequent words over infrequent words (Hallé and de Boysson Bardies, 1994), and frequent sequences over infrequent wordforms, even when phonotactics had been controlled for (Ngon et al., in press). The next step in the exploration of this potential explanation for phonological acquisition involves showing that infants use minimal wordform pairs to *collapse* across the distinction, rather than separate it. If this prediction holds, it would demonstrate that minimal wordform pairs and true, lexical minimal pairs do not operate in the same fashion at all.

A variant of the latter hypothesis could be proposed where long-term storage and the assumption of different mechanisms governing wordform and lexical minimal pairs are unnecessary. It is well known that infant-directed speech abounds in repetition, with a much greater narrowness of focus than adult-directed speech (McRoberts et al., 2009). In other words, it appears that infant-directed speech exaggerates "burstiness" (Baayen, 2001), the tendency for lexical items to recur within the same conversational interaction, in a way that could influence phonological acquisition (Skoruppa et al., 2012). A smart learner may be able to use variation across two wordforms experienced in close succession to derive probabilities of non-contrastiveness. For example, if the child hears "dad," "da[d]y," "da[ɾ]y" in the same conversational interaction, she may be able to store that [d] and [ɾ] could be variants of the same phoneme. The latter extension has not yet been espoused by modelers, but we expect it may be just around the corner. The predictions from this hypothesis could also be easily tested using an artificial grammar design.

Whereas the combination of acoustics and distributional cues seemed to gain the learner-model quite a bit, some work suggests that a learner-model combining distributional and lexical mechanisms, or all three together, may only be subtly improved (Boruta, 2011). It is of theoretical and empirical interest to thoroughly investigate the effects and interactions emerging from the integration of all 3 types of mechanisms in the future.

IMPLICATIONS

Collectively, this work suggests that multiple mechanisms, likely including the computation of complementary distribution and the calculation of phonetic similarity, operate in concert to guide infants toward their functional interpretation of sounds that are present in the input, yet not contrastive. This review also bears on the more general question of how infants cope with phonetic variability that is not lexically meaningful such as variation between talkers' voices and accents. Interestingly, infants become resilient

to talker and accent changes also toward the end of the first year of life (Houston and Jusczyk, 2000; Schmale et al., 2010). Future work should investigate whether this similarity is merely superficial, or whether it is indicative of a perceptual reorganization allowing toddlers to recognize wordforms in the presence of lexically irrelevant variation. To answer this question, research should focus on how infants cope with deviations from canonical productions and how predictable those productions are. Moreover, the question of allophones is a categorical one, but many sources of variance are gradient and future work should explore whether these different kinds of variation are more or less learnable since it may be that gradient changes to the acoustic character of a sound are more variable.

A second consideration relates to the nuances in the concepts of phonemes and allophones laid out above, and predictions that can be stated on their learnability. Recent artificial grammar learning work suggests that infants tend to attend more to regular, neither entirely predictable nor entirely unpredictable, patterns (Gerken et al., 2011). In the domain of allophonic learning this might translate to different attention being allotted to patterns that are halfway between allophones and phonemes because of their very irregularity, a matter that could be investigated by assessing infants' acquisition of different types of phonemes/allophones. Additionally, one could imagine that for infants the areas of the grammar in which the irregularity resides may be very important. For example, if the irregularity is lexically or morphologically based the language learning infant may not be immediately aware of it, and so would initially treat the pattern as if it were regular.

Additionally, differential processing of allophones and phonemes could inform translational research. For example, some work suggests that inappropriate learning of sounds in terms of these sound classes (e.g., perceiving equally well different phonemes and different allophonic alternates) correlates with reading ability and differs between normally developing and dyslexic children (Serniclaes et al., 2004). If we can pinpoint these differences in early development it may be possible to intervene while these infants are still at a very plastic stage of development. Thus, longitudinal studies exploring allophonic and phonemic processing may well contribute to early intervention at some point in the future.

Although we have steered clear of production in this review, it is certainly the case that accurately representing sounds as mapping onto distinct phonemes or the same phoneme should relate to production, since the target phonology for production will require the child to use the underlying sound in different ways in different environments. All signs indicate that this is a process that occurs quite early in development (Fikkert and Freitas, 2006). Still it is unclear to what degree the continuum between allophones and phonemes relates to production of those categories. We leave that question for a future review, but mention here that it is crucial to unite these two processes within the infant in order to truly understand the course of infant development.

It remains unclear how infants might make use of "phonetic similarity" in discovering allophones and distinguishing them from phonemes. For example, all vowels are more similar when compared with consonants, yet even young infants do not appear to have difficulty in distinguishing one vowel from another. It may be crucial to discern how acoustic similarity is judged vis-a-vis the infant. It is possible that lexical factors may also play a role in infant learning of phonological categories in a greater way than has been shown in learning models (Swingley, 2009).

Finally, it is clear that allophones may be relevant not just to phonological learning, but also to syntactic learning since allophonic alternates may mark phrasal edges (Selkirk, 1984; Nespor and Vogel, 1986; Seidl, 2000) and this marking may help infants to learn their syntactic structure if they are attentive to these edges (Nespor et al., 1996; Christophe et al., 1998). For example, if there is strengthening of contact at domain edges (Keating et al., 2003) or specific phonological processes at domain edges as mentioned above, e.g., a greater degree of aspiration or longer linguo-palatal contact the higher up you go in the prosodic hierarchy, then if infants are aware of the prosodic cues that they use for syntactic bootstrapping, this knowledge should inform or a least interact with their acquisition of the knowledge of allophones.

REFERENCES

Baayen, R. H. (2001). *Word Frequency Distributions.* Dordrecht: Kluwer Academic Publishers.

Bazell, C. E. (1954). "The choice of criteria in structural linguistics," in *Linguistics Today* eds A. Martinet and U. Weinreich (New York: Linguistic Circle of New York).

Bergelson, E., and Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., and Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *J. Exp. Psychol. Gen.* 117, 21–33.

Boomershine, A., Hall, K. C., Hume, E., and Johnson, K. (2008). "The impact of allophony vs. contrast on speech perception," in *Contrast in Phonology* eds P. Avery, E. Dresher, and K. Rice (de Gruyter: Berlin), 143–172.

Boruta, L. (2011). "Combining indicators of allophony," in *Proceedings of the ACL 2011 Student Session* (Portland, OR: Association for Computational Linguistics), 88–93.

Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology*. Cambridge: MIT Press.

Brentari, D. and Eccarius, P. (2012). "When does a system become phonological: Possible sources for phonological contrast in handshape," in *Formational Units in the Analysis of Signs*, eds H. van der Hulst and R Channon (Nijmegan: Ishara Press), 305–60.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A cross-linguistic study of early lexical development. *Cogn. Dev.* 10, 159–199.

Chambers, K., Onishi, K., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87, B69–B77.

Christophe, A., Guasti, M. T., Nespor, M., and van Ooyen, B. (1998). Prosodic structure and syntactic acquisition: the case of the head-complement parameter. *Dev. Sci.* 6, 213–222.

Clements, G. N. (2001). "Representational economy in constraint-based phonology," in *Distinctive Feature Theory*, ed. T. A. Hall (Berlin: Mouton de Gruyter), 71–146.

Cristia, A., McGuire, G., Seidl, A., and Francis, A. (2011). Effects of the distribution of cues on infants' perception of speech sounds. *J. Phon.* 39, 388–402.

Cristià, A., and Seidl, A. (2008). Is infants' learning of sound patterns constrained by phonological features? *Lang. Learn. Dev.* 4, 203–227.

Cristia, A., Seidl, A., and Onishi, K. H. (2010). Indices acoustiques de phonémicité et d'allophonie dans la parole adressée aux enfants. *Actes des Journées d'Etude sur la Parole* 28, 277–280.

Dahan, D., Drucker, S., and Scarborough, R. (2008). Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition* 108, 710–718.

Dietrich, C., Swingley, D., and Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proc. Natl. Acad. Sci. U.S.A.* 104, 454–464.

Fikkert, P., and Freitas, M. J. (2006). Allophony and allomorphy cue phonological development: evidence from the European Portuguese vowel system. *J. Catal. Ling.* 5, 83–108.

Fischer, J. L. (1958). Social influence in the choice of a linguistic variant. *Word* 14, 47–56.

Gerken, L., Balcomb, F., and Minton, J. (2011). Infants avoid "labouring in vain" by attending more to learnable than unlearnable linguistic patterns. *Dev. Sci.* 14, 972–979.

Hacquard, V., Walter, M. A., and Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: an MEG study. *Brain Lang.* 100, 295–300.

Hall, K. C. (2009). *A Probabilistic Model of Phonological Relationships from Contrast to Allophony.* PhD thesis, The Ohio State University, Columbus.

Halle, M. (1964). "On the bases of phonology," in *The structure of language: readings in the philosophy of language*, eds J. A. Fodor and J. J. Katz (Englewood Cliffs, NJ: Prentice-Hall), 324–333.

Hallé, P., and de Boysson Bardies, B. (1994). Emergence of an early lexicon: infantsõ recognition of words. *Infant Behav. Dev.* 17, 119–129.

Hohne, E. A., and Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Percept. Psychophys.* 56, 613–623.

Houston, D., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1570–1582.

Jakobson, R. (1966). "Beitrag zur allgemeinen," in *Readings in Linguistics II* (Chicago: University of Chicago Press), 51–89.

Johnson, K., and Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *J. Phon.* 38, 127–136.

Jusczyk, P. W., Hohne, E., and Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Percept. Psychophys.* 61, 1465–1476.

Kazanina, N., Phillips, C., and Idsardi, W. (2006). The influence of meaning on the perception of speech sound contrasts. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11381–11386.

Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (2003). "Domain-initial articulatory strengthening in four languages," in *Papers in Laboratory Phonology VI*, eds J. Local, R. Ogden, and R. Temple (Cambridge: Cambridge University Press), 143–161.

Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Cambridge, MA: Blackwell.

Kondaurova, M., and Francis, A. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *J. Acoust. Soc. Am.* 124, 3959–3971.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000.

Kuhl, P. K., Stevens, E., Hayashi, A., Kiritani, S., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9, 13–21.

Ladd, R. (2006). "Distinctive phonesó in surface representation," in *Laboratory Phonology* Vol. 8, eds L. M. Goldstein, D. H. Whalen, and C. T. Best (Mouton de Gruyter), 3–26.

Martin, A., Peperkamp, S., and Dupoux, E. (in press). Learning phonemes with a proto-lexicon. *Cogn. Sci.*

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can effect phonetic discrimination. *Cognition* 82, B101–B111.

McLennan, C. T., Luce, P. A., and Charles-Luce, J. (2003). Representation of lexical form. *J. Exp. Psychol. Learn. Mem. Cogn.* 4, 129–134.

McMurray, R., and Aslin, D. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition* 95, B15–B26.

McRoberts, G. W., McDonough, C., and Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy* 14, 162–194.

Narayan, C. (2006). *Acoustic-Perceptual Salience and Developmental Speech Perception.* PhD thesis, University of Michigan.

Nespor, M., Guasti, M., and Christophe, A. (1996). "Selecting word order: The Rhythmic Activation Principle,"in *Interfaces in Phonology* ed. U. Kleinhenz (Berlin: Academie Verlag), 1–26.

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology.* Dordrecht: Foris.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., and Peperkamp, S. (in press). (Non)words, (non)words, (non)words: evidence for a protolexicon during the first year of life. *Dev. Sci.*

Pegg, J., and Werker, J. F. (1997). Adult and infant perception of two English phones. *J. Acoust. Soc. Am.* 3742–3753.

Peperkamp, S., LeCalvez, R., Nadal, J. P., and Dupoux, E. (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition* 101, B31–B41.

Peperkamp, S., Pettinato, M., and Dupoux, E. (2003). "Allophonic variation and the acquisition of phoneme categories," in *Proceedings of the 27th Annual Boston University Conference on Language Development*, eds B. Beachley, A. Brown, and F. Conlin pages (Boston: Cascadilla Press), 650–661.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech* 3, 115–154.

Pike, K. (1947). *Phonemics.* The University of Michigan Press, Ann Arbor.

Polka, L., Colantonio, C., and Sundara, M. (2001). A cross-language comparison of /d/-/ð/perception: evidence for a new developmental pattern. *J. Acoust. Soc. Am.* 109, 2190–2201.

Polka, L., and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 421–435.

Saffran, J., Aslin, R., and Newport, E. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Schmale, R., Cristià, A., Seidl, A., and Johnson, E. K. (2010). Infants' word segmentation across dialects. *Infancy* 15, 650–662.

Scobbie, J., Turk, A., and Hewlett, N. (1999). "Morphemes, phonetics and lexical items: the case of the Scottish vowel length rule," in *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, 1617–1620.

Scobbie, J. and Stuart-Smith, J. (2008). "Quasi-phonemic contrast and the indeterminacy of the segmental inventory: Examples from Scottish English," in *Contrast in phonology: Perception and Acquisition*, eds P. Avery, B. E. Dresher, and K. Rice (Mouton: Berlin).

Seidl, A. (2000). *Minimal Indirect Reference: A Theory of the Syntax-Phonology Interface.* New York: Routledge.

Seidl, A., and Buckley, E. (2005). On the learning of arbitrary phonological rules. *Lang. Learn. Dev.* 1, 289–316.

Seidl, A., Cristià, A., Onishi, K., and Bernard, A. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Lang. Learn. Dev.* 5, 191–202.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure.* MIT Press, Cambridge, MA.

Serniclaes, W., Heghe, S. V., Mousty, P., Carre, R., and Spreger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *J. Exp. Child. Psychol.* 87, 336–361.

Shea, C., and Curtin, S. (2011). Experience, representations and the production of second language allophones. *Cognition* 27, 229–250.

Skoruppa, K., Mani, N., and Peperkamp, S. (2012). Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Dev. Sci.*

Swingley, D. (2009). Contributions of infant word learning to language development. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3617–3622.

Thiessen, E. (2007). The effect of distributional information on childrenõs use of phonemic contrasts. *J. Mem. Lang.* 56, 16–34.

Tincoff, R., and Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.* 10, 172–175.

Trubetzkoy, N. S. (1939/1969). *Principles of Phonology*, eds A. Christiane and M. Baltaxe, Trans (Berkeley: University of California Press).

Werker, J. F., and Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63.

Whalen, D., Best, C., and Irwin, J. (1997). Lexical effects in the perception and production of American English /p/ allophones. *J. Phon.* 25, 501–528.

White, K. S., Peperkamp, S., Kirk, C., and Morgan, J. (2008). Rapid acquisition of phonological alternations by infants. *Cognition* 107, 238–265.

Yeung, H., and Werker, J. (2009). Learning words' sounds before learning how words sound: 9-month-old infants use distinct objects as cues to categorize speech information. *Cognition* 113, 234–243.

Yuan, J. and Liberman, M. (2011). "Automatic measurement and comparison and vowel nasalization across languages," in *Proceedings of ICPhS* XVII (Hong Kong: ICPHS), 2244–2247.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 July 2012; accepted: 05 October 2012; published online: 02 November 2012.

Citation: Seidl A and Cristia A (2012) Infants' learning of phonological status. Front. Psychology ***3****:448. doi: 10.3389/fpsyg.2012.00448*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

METHODS ARTICLE
published: 17 April 2013
doi: 10.3389/fpsyg.2013.00175

Disentangling the influence of salience and familiarity on infant word learning: methodological advances

***Heather Bortfeld*[1,2]*, *Katie Shaw*[1] *and Nicole Depowski*[1]**

[1] *Department of Psychology, University of Connecticut, Storrs, CT, USA*
[2] *Haskins Laboratories, New Haven, CT, USA*

Edited by:
Jutta L. Mueller, University of Osnabrueck, Germany

Reviewed by:
Debbie L. Mills, Bangor University, UK
Jessica Hay, University of Tennessee, USA

***Correspondence:**
Heather Bortfeld, Department of Psychology, University of Connecticut, 416 Babbidge Rd., Unit 1020, Storrs, CT 06269, USA.
e-mail: heather.bortfeld@uconn.edu

The initial stages of language learning involve a critical interaction between infants' environmental experience and their developing brains. The past several decades of research have produced important behavioral evidence of the many factors influencing this process, both on the part of the child and on the part of the environment that the child is in. The application of neurophysiological techniques to the study of early development has been augmenting these findings at a rapid pace. While the result is an accrual of data bridging the gap between brain and behavior, much work remains to make the link between behavioral evidence of infants' emerging sensitivities and neurophysiological evidence of changes in how their brains process information. Here we review the background behavioral data on how salience and familiarity in the auditory signal shape initial language learning. We follow this with a summary of more recent evidence of changes in infants' brain activity in response to specific aspects of speech. Our goal is to examine language learning through the lens of brain/environment interactions, ultimately focusing on changes in cortical processing of speech across the first year of life. We will ground our examination of recent brain data in the two auditory features initially outlined: salience and familiarity. Our own and others' findings on the influence of these two features reveal that they are key parameters in infants' emerging recognition of structure in the speech signal. Importantly, the evidence we review makes the critical link between behavioral and brain data. We discuss the importance of future work that makes this bridge as a means of moving the study of language development solidly into the domain of brain science.

Keywords: near-infrared spectroscopy (NIRS), MMN (mismatch negativity), MMR (mismatch response), acoustic salience, acoustic familiarity, infant speech perception, repetition suppression effect

SALIENCE AND FAMILIARITY AS GUIDES TO SEGMENTING THE SPEECH SIGNAL

Here we synthesize recent findings on changes in infants' brain activity in response to specific aspects of speech. In particular, we focus on two aspects of the speech signal that have been shown to influence infants' emerging sensitivities: acoustic salience and the familiarity of auditory form. These are not easy things to disentangle. Whether or not this is, indeed, possible, recent findings on the influence of these features on infant auditory processing reveal that they are both fundamental to infants finding structure in the speech signal. Importantly, the evidence we review makes the critical link between behavioral and brain data, and hints at a resolution to the debate about whether familiarity and salience are distinct acoustic features or one and the same. At present, we define acoustic salience (or salience) as a construct that is based on factors external to the language learner; salience can be conceived of as those physical characteristics inherent to the stimulus itself that make it salient. We consider this kind of salience distinct from preferences based on prior exposure or experience. Prior experience is what establishes familiarity of auditory form (or familiarity). Familiarity is what the infant brings to the perceptual process. Based on their experience with language and the environment, infants should develop the initial basis for mental representations (which are not necessarily conscious) of strings of sounds. The emergence of a mental lexicon no doubt influences how acoustic stimuli are subsequently processed (e.g., eventually as words). Here we review key behavioral findings demonstrating how these two features interact to influence infant speech processing. We then link these findings to more recent brain-based measures of infant processing. Finally, we review a methodological advance in use in our own lab, near-infrared spectroscopy (NIRS), including some initial data obtained using NIRS, that point to the utility of this method for teasing apart the relative influences of acoustic salience and familiarity in early language learning.

BEHAVIORAL EVIDENCE OF THE INFLUENCE OF SALIENCE AND FAMILIARITY

One of the most common examples of salience is inherent in the acoustic features that characterize infant directed speech (henceforth, IDS). Speech directed to infants generally consists of patterns of exaggerated pitch and rhythm, causing infants to prefer it to adult-directed speech (ADS) (Fernald, 1985; Fernald and Kuhl, 1987; Cooper and Aslin, 1990). This preference on the part of infants is well established. Early tests of this typically relied on a visual-fixation procedure to establish whether infants

preferred to listen to infant-directed over ADS. For example, in a test of newborns and 12-month-olds (Cooper and Aslin, 1990), results indicated that both the newborns and the 12-month-olds demonstrated increased visual fixation during IDS trials, suggesting a preference for IDS over ADS. More recently, Thiessen et al. (2005) sought to determine whether the preference for IDS over ADS serves an infant's learning needs, particularly in the language domain. These researchers also used a familiarization procedure and found that 6-to 8-month-old infants were better able to segment statistically instantiated items out of a speech stream when they had originally been presented to the infants in IDS. This suggests that, not only do infants prefer IDS, but that the salient characteristics of this form of speech help them recognize individual items within running speech. This is so even before infants are able to exhibit stable language production behaviors themselves.

Despite infants' general preference for IDS, their preference for specific affective content within this form is not constant and has been demonstrated to shift during the first year. For example, 3-month-olds were found to prefer IDS that was pre-rated as soothing or comforting, 6-month-olds preferred IDS pre-rated as approving, and 9-month-olds preferred "directive affective" speech, a type of speech indicating that the infant should behave in a certain manner (Kitamura and Lam, 2009). This developmental shift in affective preference suggests that infants and caretakers influence one another over the course of the infant's first year in determining which type of speech will be most salient for the infant; caregivers are more likely to use a particular affective tone if the child is more likely to attend to it. Presumably, this facilitates language learning, since it promotes the mutual give-and-take that is the basis for communication in general.

If infant-directed speech is caregivers' way of making speech salient to infants, when does familiarity begin to play a role in their processing of the speech stream? As infants hear more and more IDS, it should become a familiar auditory form, both in terms of its general prosodic structure and, perhaps, in terms of the particular words it most frequently contains (e.g., an infant's own name). Indeed, familiarity can be derived from salience, as hypothesized by Snow (1972) based on her observation that the hallmarks of IDS include simple utterances and redundancy. While both of these characteristics increase the overall salience of this form of speech, such salience also provides a scaffold for word learning by highlighting particular forms within the acoustic stream. Eventually, those forms become familiar, thus facilitating additional structural learning. Support for this view comes from a study designed to investigate developmental differences in infants' preference for different aspects of IDS, in which 6-month old infants were found to prefer the repetitive structure, rather than their earlier preference for its prosodic elements (McRoberts et al., 2009). This shift in preference from the prosodic elements to the repetitive elements may be an indication of when infants transition from processing the general characteristics of the speech stream to recognizing components (i.e., words) within it. Such a view is consistent with findings showing that infants discriminate among words relatively early in life (Tincoff and Jusczyk, 1999; Bortfeld et al., 2005; Bergelson and Swingley, 2012).

INTERACTION OF SALIENCE AND FAMILIARITY

Important evidence of the interaction of acoustic salience and familiarity in infants' speech processing comes from a study by Barker and Newman (2004). These researchers found that infants not only showed a preference for words spoken by their mother, but that they were able to attend to her voice in the presence of background noise that consisted of an unfamiliar female speaker. One implication of this finding is that, since infants can attend to their mothers' voices even in the presence of noise, they may be able to learn acoustic structure better from their mother (or from some other highly familiar individual) than when the words are being produced by an unfamiliar speaker. Regardless of who the speaker is, however, IDS contains important cues that appear to facilitate language learning. It also parallels patterns of speech between adults. When adults engage in conversation, the first instance of a word's utterance is typically more enunciated (clear) and longer than in subsequent references, suggesting that the speaker assumes a common ground between him or herself and the listener (Fowler and Housum, 1987). Repeated words are acoustically truncated in IDS as well, suggesting that such truncations may help draw infant attention to previously unsaid words (new information) in sentences that contain predominantly old information (Fisher and Tokura, 1995).

Additional research has revealed that this pattern is not stable when adults speak to infants, highlighting an interesting and important characteristic of language input. Bortfeld and Morgan (2010) investigated the given-new contract in infants by measuring the (several subsequent) repetitions of words produced by mothers in a single instance of speaking. Of note, such repetition is not something that adults would do when speaking with other adults; rather, the focal word is typically referred to with a pronoun after its initial one or two mentions. When speaking to infants, however, adults will repeat a word multiple times, providing an interesting pattern on which to perform acoustic analyses. Therefore, in their study, Bortfeld and Morgan (2010) did just this, finding that the second utterance of a word, when directed to infants was, indeed, truncated, and produced less emphatically, a finding that mirrors Fisher and Tokura's (1995) earlier results. However, when looking beyond the first two mentions of a word, it became clear that mothers revert to emphatically stressing that word all over again, followed again by de-emphasis. Given that adults will repeat a word to an infant sometimes six or eight or ten times, this points to a rhythmic production pattern that, while mirroring adults' speech, exaggerates it through repetition. Although the second (and subsequent) sets of repetitions may be less stressed and enunciated overall in comparison to the first, mothers nonetheless appear to revert to the same pattern of emphasis/de-emphasis. This is the case at least until they change the focus of their speech. Together, these findings provide support for the view that acoustic salience provides the foundation for familiarity. And familiarity, often in concert with salience, facilitates language learning. Of course, a form may become more salient to an infant as its familiarity increases (e.g., by taking on semantic meaning). But if we constrain our characterization of salience to acoustic salience, the directionality of influence implied here makes sense, and appears to hold for individual speech sounds as well. For example, a recent study (Narayan et al.,

2010) challenges the long-standing view that infants can discriminate all functionally discriminable (i.e., categorically distinct) sounds. Instead, Narayan et al. (2010) observed a case in which acoustic salience (in the form of more versus less discriminability) interacts with an infant's environmental exposure. Their work suggests that differential discriminability is not entirely consistent with the all-to-some view of perceptual tuning patterns across the first year of life. Specifically, the researchers focused on Filipino, a language in which there is a subtle difference in nasalization between /na/ and /ŋa/ that does not exist in English; on the other hand, the contrast between /ma/ and /na/ exists in both languages and is much more salient to the listener. English-exposed infants were shown to discriminate /ma/ from /na/ at both 6-to-8 and 10-to-12 months of age, but they were not able to discriminate non-native and less acoustically salient /na/ vs. /ŋa/ contrast at either of these ages. Even very young (e.g., 4-to-5 months of age) English-exposed infants showed discrimination of only the former (/ma/ vs. /na/) and not the latter contrast (/na/ vs. /ŋa/). Notably, Filipino-exposed infants showed discrimination of their native [na]-[ŋa] between 10- and 12-months, but not between 6- and 8-months. This pattern of findings suggests that experience is necessary to establish long-term discrimination of two very similar speech sounds (e.g., /na/ and /ŋa/), while acoustic salience enhances perception of very different sounds (e.g., /ma/ and /na/), providing a more nuanced view of how early perceptual reorganization unfolds.

Of course, different languages are characterized by differences well beyond the phoneme level. For example, it has been hypothesized that infants from different language backgrounds develop preferences for their particular (native-language) stress pattern early in life. To test this hypothesis, Hohle et al. (2009) conducted four experiments with German- and French-exposed infants at both 4- and 6-months. These languages have a notable contrast in stress, with German showing a strong trochaic (strong-weak) pattern that French does not have. At 4-months, German-exposed infants showed no preference for stress pattern; however, at 6-months, they began to show a preference for the trochaic pattern. On the other hand, French-exposed infants did not show a preference for one or the other pattern at 6-months, but were able to discriminate between the two. As with the phoneme discrimination findings, these results suggest that infants' sensitivities are shaped both by their environmental exposure and the absolute salience of the acoustic characteristic in question. Where trochees are quite salient in German, French's syllable timing rendered the trochaic form less salient to French-exposed infants.

Although acoustic salience is a useful tool for infants who are initially learning language, it can present problems as well. For example, if infants pay attention to their world based only on the physical salience of an object (auditory or otherwise), they may be missing other important aspects of the environment. When the item in question is a visually presented object, this can also affect the likelihood that infants will learn about other objects. In a clever study, Pruden et al. (2006) exposed 10-month-olds to a salient object (e.g., a glittery wand) and a less salient object (e.g., a beige bottle opener), while pairing each with a unique and novel auditory label. Despite being asked to identify the non-salient object, infants tended to look more at the salient object. Clearly, infants' tendency to attend to the salient things in the world around them doesn't always facilitate language learning.

Behavioral research has revealed several other sensitivities that infants bring to the learning environment. Consistent with the trochaic bias observed by Hohle et al. (2009) in German- but not French-exposed 6-month-olds, different languages have different units of segmentation. French tends toward the syllable, English and German use stress, and Japanese uses the *mora* (a subsyllabic unit) for segmentation (Cutler and Mehler, 1993). Indeed, earlier research demonstrated that infants exposed to each of these languages approach speech segmentation differently. A French-exposed infant, upon hearing Japanese, will segment the speech stream based on syllables, when the *mora* would actually be more appropriate (Cutler and Mehler, 1993). Although the means of segmentation are different in each language, the methods are similar: infants appear to recognize ambient rhythmic patterns early on, and use these patterns to segment the speech stream, thereby developing more precise awareness of the sounds within those segments. Again, although the familiar structures differ across languages, the general pattern is for those aspects of the environment which are the most salient to infants to become the most familiar (or at least to become familiar faster).

Infants are also sensitive to statistical regularities in their environment, using them as a guide to structure (e.g., Saffran et al., 1996). Beyond basic sensitivities, infants can then map these regularities to simple visual objects, demonstrating the first step in making label-object associations. For example, in a recent study (e.g., Shukla et al., 2011), infants were presented with a continuous speech stream and were able to recognize relationships between co-occurring segments (e.g., statistical "words") and objects in the environment, but only if there was a high probability for co-occurring syllables (see also Graf Estes et al., 2007; Hay et al., 2011). This ability was extinguished when these statistically co-occurring segments crossed prosodic boundaries. These results are consistent with other work showing that prosody is a salient cue to infants by 6-months of age (see Kitamura and Lam, 2009; McRoberts et al., 2009) and that it interacts with their emerging sensitivity to structure. The fact that infants can map newly recognized structure onto simple visual objects (or at least associate them) demonstrates that the interaction of perceptual salience and familiarity forms the basis for active learning about relationships in the environment.

This happens at a more granular level as well. For example, the statistical likelihood of a sound string like "bref" is relatively high in English; one like "febr" is quite low. Mattys and Jusczyk (2001) observed that American English-exposed 9-month-olds segmented words as a result of the likelihood of the phoneme sequences in their language of exposure (in this case, American English). In other words, their familiarity with their own language's phonotactic structure actively influenced what infants found perceptually salient by the end of the first year. Graf Estes et al. (2011) expanded on this work by using a looking-while-listening paradigm with 18-month-olds. In this, infants were first presented with two object labels that were paired with novel objects. These labels were either legal (contained sound sequences that frequently occur in English) or illegal (contained sound sequences that never occur in English). At test, infants looked at

the correct object when presented with the legal label; they did not look at the correct object when presented with the illegal label. These results demonstrate that phonotactic sensitivities have the power to shape learning.

In earlier work (Bortfeld et al., 2005), my colleagues and I demonstrated that infants can use existing words to scaffold their learning of new words. Specifically, we found that 6-month-olds can learn a new word if they had been familiarized with it while it was consistently preceded by either their own name or some other highly familiar name (e.g., mommy/mamma, depending on which term the mother used to refer to herself). Names for important individuals (e.g., oneself, one's primary caregiver) are highly frequent and thus become very familiar. This study shows that such familiarity can serve as a tool for subsequent segmentation of the speech stream, thereby facilitating progressive language learning. In this case, it is unclear which comes first, salience or familiarity. Presumably the semantic meaning associated with the familiar sound string is what brings the salience to the word, an important caveat to the argument laid out earlier about salience leading familiarity. And familiarity can sometimes undermine learning. In a clever study, Houston and Jusczyk (2000) familiarized infants with words produced by one speaker and then tested whether they could generalize their learning to unfamiliar speakers and to unfamiliar contexts (an ability that would reveal a more abstract form of representation). Results suggested that such abstraction did not happen, at least initially. Of course, speaker-specific representation of words is not a very functional way to learn language; fortunately for everybody, infants' retention of indexical information about individual speakers attenuates by about 10.5-months of age.

We have reviewed just a smattering of the behavioral evidence supporting the role of salience and familiarity in language development. Whether conceptualized as one or two identifiable characteristics of acoustic form, many questions remain. In particular, it is not always clear whether familiarity and/or salience act in a top-down or bottom-up manner. Salience may enter the system, at least initially, in a bottom-up manner (e.g., from the environment; from biologically established biases toward the environment) and thereby shape developing representations. Then again, it may not.

In a final example of the complex interaction between new and learned information in the process of language learning, Mersad and Nazzi (2012) used statistical learning in combination with familiar form. In a tweak of the usual approach to testing statistical learning, these researchers used non-uniform length novel words instead of the standard uniform-length novel "words" from the audio stream. Eight-month-olds were hindered in their ability to segment these non-uniform length novel words when presented with no other cues. However, they could segment the non-uniform length novel words when the words were preceded with a familiar word (*maman*, French for mom). In other words, what had become salient ("*maman*") through initial familiarization provided infants with top-down guidance for parsing a complex (bottom-up) signal. This is just another demonstration of the degree to which top-down and bottom-up processes are interacting in complicated ways—from an early age and all along—to influence language processing. Ultimately, these data highlight the challenge inherent in characterizing which came first in any form of infant perception, salience, or familiarity.

A WAY FORWARD? BRAIN ACTIVITY DISTINGUISHES THE INFLUENCE OF SALIENCE AND FAMILIARITY

Thus far, we have focused exclusively on studies in which behavioral measures were used to investigate how infants process speech. Indeed, infants' overt gaze and sucking behaviors have provided us with important insights into their perceptual experiences, and behavioral measures are foundational in our understanding of how humans begin learning language. However, limitations to the interpretations that can be made based on these measures remain. For example, it is often difficult to tell with certainty what exactly both the looking time and the looks themselves signify (for a cogent review of the issues, see Aslin, 2007). Increasingly, researchers are turning to the growing array of neurophysiological methods that can be used with infants to better understand what those looks mean. Neurophysiological techniques have aided our ability to assess and measure language development through the first year of life and beyond. Although some are still gaining ground in developmental studies (e.g., NIRS), other techniques [e.g., electroencephalography (EEG)] form the basis for our understanding of both the timing and neural correlates underlying language milestones. The continued integration of behavioral methods with one or more of these techniques holds great promise for the advancement of language learning research, in particular, and developmental research, in general.

EEGs AND EVENT-RELATED POTENTIALS

One well-established technique for use with infant populations is EEG, a non-invasive tool with excellent temporal resolution and mild to moderate spatial resolution (for a review, see Fava et al., 2011). The application of this technique to research with preverbal infants has allowed researchers to pinpoint, in tens of milliseconds (ms), when sensory processing is occurring. It also provides information about different processing stages. The non-invasive nature of EEG makes it a relatively safe procedure to use when studying infants, and a multitude of event-related potentials (ERPs) can be assessed, even in neonates (Korotchikova et al., 2009). In addition, EEG can provide data without requiring a behavioral response. This is especially valuable when testing very young infants, who often are unable to produce reliable behavior in response to perceptual stimuli, and when the goal is to determine when an infant notices a stimulus change.

The workhorse of ERP research, the Mismatch Negativity (MMN) component, is one that has been widely used with both infants and adults. The MMN is measured in the 150–250 ms window of time, post-stimulus onset. When presented with a sequential list of identical exemplars, the adult MMN has been found to have higher amplitude for deviant stimuli (e.g., an oddball) (Naatanen, 1995). One of the hallmarks of the MMN is that it is relatively impervious to conscious modulations in attention and thus can be found even when a person is not focusing on the stimuli (Luck, 2005). In adults, the MMN has been observed in response to auditory stimuli even while the individual is engaging in an unassociated cognitive task, such as reading.

This has led to the view that the MMN reflects processing that is pre-attentive and passive (Alho et al., 1992), making it an ideal candidate for use with infants. There has been considerable debate over whether the early time window of the MMN and the factors shown to modulate it are the result of bottom-up perceptual processing alone, particularly in low-level acoustic change detection tasks (Kenemans and Kahkonen, 2011). Several studies have demonstrated a dynamic interaction between salience (bottom-up effects) and familiarity (top-down effects) in MMN amplitudes (for review, see Garrido et al., 2009). The possibility that the measure may get at the interplay between features such as salience and familiarity in early processing underlies its promise for additional infant research on precisely this issue. Thus far, however, much of the infant-specific research has focused on stimulus familiarity as the basis for the change in voltage amplitudes.

In an influential early study, behavioral techniques revealed that infants prefer to listen to their mother's voice relative to that of a stranger (DeCasper and Fifer, 1980). Indeed, and as noted earlier, they can even distinguish their mother's voice in the presence of noise (Barker and Newman, 2004). Beauchemin et al. (2011) sought to better understand the basis for this preference by using the mismatch response (or MMR), a developmental precursor of the mismatch negativity response seen in adults, and source analyses (for cortical localization) during infants' processing of familiar voices. The researchers tested neonates between the ages of 8- and 27-h while they were exposed to a concatenated stream of the French vowel "a" (as in "allo," the French pronunciation of "hello") produced by an unfamiliar female speaker. Two types of auditory oddballs were inserted into the speech stream fifteen percent of the time, either a different unfamiliar female producing "a" or the infant's own mother producing "a." They found that when presented with the mother's voice as an oddball stimulus, MMR amplitudes were significantly greater than MMR amplitudes measured when the second stranger's voice was an oddball stimulus. This finding suggests that familiarity (in this case, with the mother's voice) is in play from birth, thereby influencing auditory processing beyond simple acoustic change detection.

In addition to analyzing the MMR Beauchemin et al. (2011) also conducted source analyses to better gauge not only when but where these modulations were occurring neurophysiologically. They found that the mother's voice activated the left posterior temporal lobe throughout the first 300 ms of exposure, while the stranger's voice activated the right temporal lobe (∼100 ms), followed by a switch to the left temporal areas (200 ms), and then a reversion back to the right temporal lobe (∼300 ms). The authors interpret the lateralized response to the mother's voice as demonstrating earlier recognition of the stimulus as being a language component, as well as evidence that the tuning of voice specific recognition in the brain occurs within the first 24 h after birth. Of course, there remains some skepticism about the accuracy of EEG-based source localization (see Plummer et al., 2008), so these results should be interpreted with caution.

As we have observed based on our review of behavioral data, multiple forms of familiarity may influence infant language learning, well beyond the mother's voice. Familiarity, and thus preference, for a number of aspects of the signal may help the infant begin to segment fluent speech and to learn new words. For example, focusing on sensitivity to stress patterns, Weber et al. (2004) compared 4- and 5-month-old infants German-exposed infants with native German speaking adults. Specifically, they looked at participants' MMR to consonant-vowel-consonant-vowel (CVCV) sequences produced with either trochaic stress (e.g., stress placed on the initial syllable and typical of the German language) or iambic stress (e.g., stress placed on the second syllable and atypical in German). Half of the participants experienced the trochaically stressed words as "standards" and the iambically stressed words as the MMR-dependent "deviants." The reverse was true for the other half of the participants. For the adults, an MMR occurred whether the deviant was either a trochaic or iambic string, suggesting that adults were sensitive to both stress patterns when they were novel relative to the ongoing auditory stream. However, for infants, an MMR was observed in the 5-month-olds for deviant trochaic stimuli only, while neither stress type provoked a significant MMR in the 4-month-olds. This suggests that between 4- and 5-months of age, infants become increasingly tuned to the most common stress patterns of their exposure language, though they have yet to reach adult-like discrimination abilities for unfamiliar stress patterns. This is consistent with the behavioral findings (e.g., Hohle et al., 2009), allowing us to infer that sensitivity to stress patterns are experience-dependent and emerge during the course of preverbal language exposure.

In-line with behavioral studies investigating the influence of familiarity on infant speech segmentation (e.g., Bortfeld et al., 2005), ERP studies have also demonstrated a privileged role for familiar words presented in continuous speech. Kooijman et al. (2005) familiarized 10-month-olds to bisyllabic words, presented in isolation, following the stress pattern of their native language (Dutch) and then presented in sentences at test. During the familiarization phase, enhanced ERP responses were found during word presentation in the frontal, fronto-central, and fronto-temporal regions while at test they were more left lateralized, suggesting different underlying neural processing mechanisms. Importantly, these effects were found prior to word offset, suggesting that infants were recognizing the newly familiarized words based on the first syllable and stress pattern. These findings demonstrate the neural underpinnings involved in speech stream segmentation and provide further evidence of word familiarity influencing said segmentation. In a follow-up study, Junge et al. (2012) further examined the relationship between word familiarization and vocabulary development by longitudinally assessing ERPs at 10-months-old as being predictive of vocabulary development at 12- and 24-months-old. They found that infants who demonstrated better segmentation abilities at 10-months of age also had higher vocabularies at 12- and 24-months-old, suggesting that rapid recognition of words is an integral part of language development and may be useful in understanding individual differences in vocabulary acquisition.

These results, while compelling evidence of the utility of the MMN in infant research, all serve as additional support for the importance of familiarity in infant processing and thus move us no closer to our goal of understanding the interplay between that and acoustic salience. Another common ERP component,

the N400, may highlight a way forward. The N400 has been used extensively in language research in both infants and adults (de Haan et al., 2003). This component is characterized by a negative peak amplitude around 400 ms post-stimulus-onset, although the time window ranges from 250 to 500 ms. Higher N400 amplitudes have been found in adults for sentential semantic violations (e.g., *Bill is lactose intolerant therefore he drinks milk*), although violations within individual words have also resulted in higher amplitudes (Kutas and Federmeier, 2011). The N400 is also influenced by semantic priming in adults (Kutas and Hillyard, 1980), a response elicited by a level of processing typically unexpected in infant research.

However, in a recent study, Parise and Csibra (2012) investigated whether the N400 could be modulated in 9-month-olds by presenting a spoken referent that was inconsistent with a visually presented object. The researchers hypothesized that if an N400 was evident for a mismatch between the auditory and visual modalities, then it would represent infants' association of the heard label with a particular visual stimulus. They further reasoned that if an N400 was not found for a mismatch between object and label, then this would demonstrate that infants may be relying on temporal associations when pairing words with objects and *not* semantic representation. In the study, a mother or a stranger produced a familiar object label. Two seconds later, an occluder was removed, displaying an object. Results showed that when the object did not match the label as spoken by the mother, the N400 response was greater in amplitude, suggesting that infants processed the discrepancy at a semantic level. In contrast, the N400 was attenuated in both match and mismatch trials for the stranger's production of the object label, suggesting that the semantic representation was specific to the mother's voice, and that infants were not yet abstracting their representation across exemplars of the word. These results are consistent with other demonstrations of the important role of a consistent acoustic source (e.g., the mother) in infant language development. But they also hint at a way of getting at the dynamic interplay between familiarity and salience in early word learning: one could argue that the infants' initial semantic representations for the familiar objects were based in the salient acoustic form (e.g., the mother's voice). While it can still be argued that the mother's voice is salient precisely because of its familiarity, it should be clear that the addition of a semantic-level component to the infant ERP toolkit is an important step toward our ability to tease apart the relative influence of familiarity and salience.

In another study investigating the N400 in early language development, Friedrich and Friederici (2005a) compared response activation to phonotactically legal (pseudowords) and illegal words (nonsense words) in 12-month-olds, 19-month-olds, and adults paired with objects. Pseudowords followed the phonotactic rules of the participant's native language (German) while nonsense words violated phonotactic constraints. These researchers found strong evidence of an N400 effect in 19-month-olds for pseudowords over nonsense words when paired with an object, suggesting that prior knowledge of the phonotactic constraints of the native language influence which words can be used as object referents. In contrast, 12-month-olds did not show differences in N400 amplitude based on legality of the words, which the authors assert may reflect a lack of maturity in the N400 ERP. Overall, their study provides additional evidence that familiarity with phonotactic rules of the native language influence word processing and object referencing, particularly in the second year of life, a finding that is consistent with other findings from these researchers (Friedrich and Friederici, 2004, 2005b). Still others have observed enhanced ERPs for newly-learned words in 20-month-old infants, mirroring their response to previously known words in object-pairings, albeit at an earlier time-window (N200–N500; Mills et al., 2005).

The only clear examination of salience as it interacts with familiarity in infant speech processing comes from a study using both early and late time-course ERPs in combination. Specifically, Zangl and Mills (2007) investigated how familiar and unfamiliar words presented in IDS or ADS affected the N200–N400 time-window amplitude and the Nc component in 6- and 13-month-olds. The Nc component is a mid-latency, negative-going waveform characteristic of the fronto-central scalp regions (Richards, 2003). Importantly, it is considered an endogenous attentional component, reflecting top-down influences on attentional orienting and perceptual processing (Richards, 2003), and thus is relevant for understanding how previous experience may facilitate subsequent processing. The researchers found that 13-month-olds, but not 6-month-olds, showed enhanced N200–N400 amplitudes for familiar words presented in IDS over familiar words presented in ADS, but showed such no difference for unfamiliar words. Regardless of age, the Nc component was greater in amplitude for IDS over ADS, suggesting that infants increased attention to the speech stream as a result of the more salient speech register. Together, these findings suggest that exposure format (e.g., more or less salient speech type) and exposure form (e.g., word familiarity) interact in driving infant attention toward speech in the first year of life. More research along this line is sorely needed.

Clearly, EEG (and accompanying ERPs) is an established and important tool for assessing infant perception without requiring explicit behavior. Electrophysiological studies have provided a bridge to better understanding of the neural basis for a variety of behavioral findings. Source localization techniques notwithstanding, the limited spatial resolution of this particular methodology constrains the inferences that can be made about which areas of the brain are developing when, and what their role in early speech processing is. More recently, novel hemodynamic-based techniques (e.g., NIRS) have emerged for application with infant populations, as has the application of established hemodynamic-based techniques (e.g., fMRI) to infant populations. To better understand how neural development facilitates the integration of salience and familiarity in the service of language learning, it is worth examining data from this domain of infant research as well.

HEMODYNAMIC-BASED MEASURES

In an influential early developmental imaging study, Dehaene-Lambertz et al. (2002) tracked changes in cerebral blood flow in 2-to-3-month-old infants using fMRI while the infants were exposed to samples of forward and backward speech in their

native French. Infants were tightly swaddled prior to being placed in the core, so as to restrain their movement. They were presented with recordings of a woman reading passages from a children's book. The passages were either presented normally (e.g., forward speech) or the recordings were time-reversed (e.g., backward speech). The researchers hypothesized that brain regions associated with segmental and suprasegmental speech processing would be more highly active during exposure to typical, forward speech. In contrast, the backward speech condition should violate phonological properties of the infants' native language, and thus, activation in the brain regions sensitive to speech structure should be less active in response to it. Results revealed that, indeed, brain regions were differentially activated as a result of speech condition. During exposure to forward speech, infants' left angular gyrus and left precuneus were significantly activated, suggesting that infants were not only recognizing the familiar acoustic structure during the forward segments (see Démonet et al., 1992 for adult comparison of left angular gyrus), but also engaging in early memory retrieval (see Cavanna and Trimble, 2006 for adult comparison of left precuneus). Of course, because the infants were swaddled, they fell asleep during much of the testing in this study. The researchers coded for sleep state based on their observations of infants' faces during testing. Although many of the results were not influenced by sleep state, it is worth noting that there was some variability in the data based on it that will require additional research to better understand.

Functional studies have likewise provided evidence of infants' sensitivity to a familiar speaker (e.g., Dehaene-Lambertz et al., 2010). In this study, the researchers used fMRI to investigate the neural correlates of speech perception in 2-to-3-month-old infants, specifically comparing speech produced by their own mother to that produced by a stranger, as well as speech versus music. Results revealed that, even by 2-months of age, infants showed left-lateralized processing of speech relative to music, and that this lateralization of activation was modulated by whether the voice was familiar or not. During exposure to their mothers' voice, infants' left posterior temporal region was more highly activate than during exposure to a stranger's voice, suggesting that low-level acoustic familiarity enhances speech-specific processing. These results are consistent with the behavioral findings reviewed earlier from Barker and Newman (2004), as well as recent ERP results from Parise and Csibra (2012), showing an interaction of voice familiarity and semantic representation.

The feasibility of using functional magnetic resonance imaging and other motion sensitive techniques with very young populations is necessarily limited. While fMRI has excellent spatial resolution, it is generally quite noisy and also susceptible to motion artifacts. Researchers have to adjust study designs to account for the challenges of working with infant participants when planning and conducting studies. However, NIRS is a more infant-friendly hemodynamic-based measurement tool; it is non-invasive, less vulnerable to motion artifacts, and safe to use even with newborns (Sakatani et al., 1999; see Aslin, 2012 for a comprehensive review of this technique and its application in infant research).

Near-infrared spectroscopy is providing important insight into the dynamic interaction of a number of factors on how preverbal infants process speech and how this changes in developmental time. For example, using NIRS, Homae et al. (2006), (2007) investigated developmental changes in cortical activation specific to prosody in 3- and 10-month-old infants. They sought to determine when the right lateralization that is typical of prosodic processing in adults (Baum and Pell, 1999) is evident in infants. In their study, infants were presented with both normal and flattened speech, in which the flattened speech was void of pitch contours. They found that 3-month-olds displayed bilateral activation in the temporoparietal, temporal, and frontal regions for both speech types and enhanced activation in the right temporoparietal regions for natural speech (Homae et al., 2006). These findings suggest that even by 3-months of age, infants are sensitive to the prosodic information available in the speech signal. In addition, a follow-up study with 10-month-olds (Homae et al., 2007) using the same methodology, found greater activation in the right temporoparietal and temporal regions for prosodically flattened speech in comparison to natural speech, mirroring adult patterns. The authors assert that the differences between their two findings demonstrate a developmental shift in pitch processing mechanisms as a result of greater experience with the prosody of the child's native language.

COMBINING BRAIN AND BEHAVIOR: REPETITION SUPPRESSION

To assess the cortical changes that underlie advances in language in the first and second years of life, my colleagues and I have been using another hemodynamic-based measurement technique, NIRS (Bortfeld et al., 2007, 2009). Specific to the current focus on how the infant brain is shaped by salience and familiarity, we have been using NIRS with a well-established behavioral protocol. The results, which we will review here, are promising.

As should be apparent from this review, a common tool for studying infants' sensitivity to stimuli (or specific characteristics of stimuli) is to establish response habituation based on looking times. This is something that can likewise be used to study brain responses (e.g., Turk-Browne et al., 2008). In the fMRI literature, habituation to stimulus characteristics is observed in the form of repetition suppression (Grill-Spector et al., 2006), whereby prior exposure to stimuli (or stimulus attributes) decreases the level of activation elicited during subsequent exposure to identical stimuli. Although the underlying neuronal mechanisms remain unclear (for review and discussion, see Henson, 2003; Henson and Rugg, 2003), repetition suppression has been interpreted as the fMRI analog of neuronal response suppression observed using single cell recording (Desimone, 1996). This reduction in brain activation with repeated exposure presents an ideal scenario for establishing whether infants' brains show a decrease in hemodynamic activation concomitant with a decrease in looking (i.e., over the course of habituation), a demonstration of increased familiarity.

When repetition effects are present in a brain region in human adults, they indicate that the particular region (showing a reduction in activation) is supporting the representation of the stimulus, and variants of the paradigm have been used to monitor the abstractness of a particular representation (Grill-Spector and Malach, 2001; Naccache and Dehaene, 2001). For example, the left inferior frontal region appears to be quite

sensitive to sentence repetition, suggesting that it is part of the network supporting early verbal working memory, at least in adults (Dehaene-Lambertz et al., 2006a). In newborns, the repetition of a syllable every 600 ms produced a decrease in ERP amplitudes (Dehaene-Lambertz and Dehaene, 1994; Dehaene-Lambertz and Peña, 2001) and in a more recent study (Dehaene-Lambertz et al., 2010), repetition suppression was observed in 2-month-olds exposed to repetition of the same sentence at 4 s intervals. In infants, this repetition suppression was observed in the left superior temporal gyrus, extending toward the superior temporal sulcus and the middle temporal gyrus. However, a slow event-related paradigm where a single sentence was repeated at much longer (e.g., 14 s) intervals did not produce any repetition suppression (Dehaene-Lambertz et al., 2006b), which may point to the limits of the early verbal working memory window. Of course, the absence of a repetition suppression effect in this case could have been related to any number of factors (e.g., unique characteristics of the BOLD response in infants, complexity of the sentence, or, indeed, the extended time-lag erasing the echoic buffer of the temporal regions).

These findings do, however, highlight a way forward. Importantly, repetition suppression was observed with immediate repetition in these infants, providing a methodological vehicle for clarifying characteristics of auditory representation in infants. More recently, repetition suppression has been observed in infant blood flow data collected using NIRS (e.g., Nakano et al., 2009). In our own work, the utility of repetition suppression has been tested using a mixed stimulus presentation combining aspects of both event-related and block designs. In this approach, we presented infants with individual stimuli repeatedly and with relatively short ISIs (e.g., 3 s). Test blocks were intermixed with control blocks (e.g., sets of comparable but variable stimuli). Initial data from a single (9-month-old) infant (see **Figure 1**) show a repetition suppression effect for the auditory repetitions of an individual word. That is, as a single word was repeated, the activation pattern over the left temporal region decreased with each subsequent repetition (e.g., as seen in the overall hemodynamic response reduction from the first 15 s of word repetition in Trial 1 to the final repetition in Trial 5). Furthermore, novel words that were matched for stress pattern, syllable count, and overall length in control blocks elicited a relatively sustained hemodynamic response in the same cortical location, highlighting the selectivity of the effect.

While these data speak to the brain's changing response with increasing familiarity, one can imagine more complex designs that would work toward differentiating response to both familiarity and salience in the same brain. And really, a robust hemodynamic response to a novel stimulus is an indicator of salience, particularly when compared to the same region's response after multiple repetitions of exposure. One approach using NIRS alone to resolve the salience/familiarity puzzle would be to introduce variations of form (e.g., changes in speaker; changes in pitch) to monitor a "release" from repetition suppression. Such a result would reveal in real time the brain's response to salient changes in the environment and, thus, to salience.

Together with the MMR approach outlined earlier, which pinpoints low-level responses to salient characteristics of the signal, the repetition suppression effect in hemodynamic based measures highlights a way forward. Importantly, NIRS very often reveals such effects on a trial-by-trial basis and in a single subject,

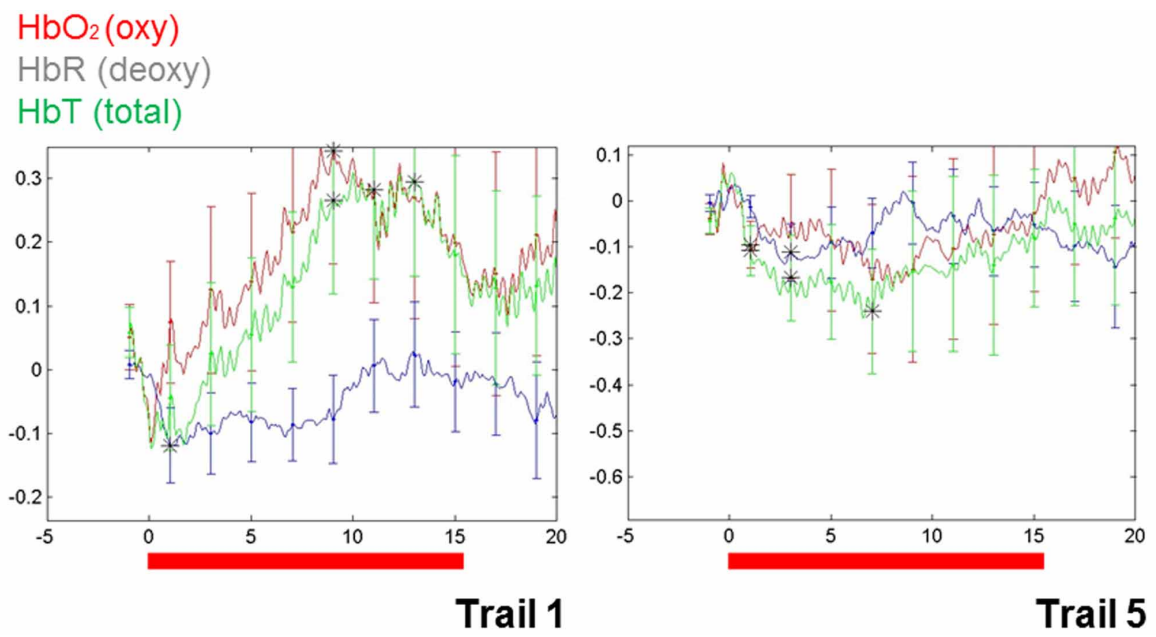


FIGURE 1 | Changes in blood flow in a single 9-month-old infant during the first and the final (fifth) series of repetitions of a single, monosyllabic word. The Y-axis is relative changes in concentration (micromolar) and the X-axis is time. Area of recording is left superior temporal gyrus, with the optode centered over T3 (of the 10–20 system).

something EEG data would be hard-pressed to do. Regardless, the stimulus selectivity of each measure makes them both useful tools for assessing early language processing. In particular, the repetition suppression effect can reveal the point at which a stimulus becomes familiar (or at least begins transitioning toward that state) and (presumably) what changes in that stimulus make it salient again. If familiarity is the basis for the development of representations of words, then a child's failure to show a typical repetition suppression effect may highlight a corresponding failure to encode relevant features of that word (e.g., the temporal order of individual sounds within it; its prosodic form). Such an effect can thus be exploited in a clinical setting as well, potentially providing important diagnostic information into the degree to which a child is (or is not) developing robust lexical representations. It could also be used to establish which feature changes in a stimulus make it salient again. All of these are possibilities that at least hint at a way forward in disentangling influences of salience and familiarity is early learning.

Ultimately, large scale, within-subject data collection will establish the utility of both the MMN and repetition suppression effects in research on infant perceptual processing. For example, blood flow measures collected in a canonical repetition suppression task and electrophysiological measures collected during a canonical mismatched negativity task could be related to subsequent language outcome on a child-by-child basis. For now, we can at least appreciate the complimentary nature of these neurophysiological techniques, both with one another and with the long history of careful behavioral testing that is critical to understanding infant perceptual development. These tools may yet reveal how salience begets familiarity (and vice versa).

Certainly there are limitations in the application of NIRS in infant research, and these should be taken into account when designing and conducting experiments (see Aslin, 2012, for review). Although NIRS is similar to fMRI in that it relies on measuring hemodynamic responses, it is severely more limited in its ability to gauge response from deeper brain structures (e.g., below the level of the cortex). It is optimally suited for examining structures near the cortical surface, ideally with probe design controlling for scalp-surface distance (Beauchamp et al., 2011). Additionally, because NIRS relies on changes in blood oxygenation levels, it has poor temporal resolution. Although the sampling rate for NIRS can far surpass that of fMRI, due to the inherent constraints on blood flow timing it is, for practical purposes, on par with that of fMRI. Finally, best practices for the application of NIRS research include attention to the development of approaches to signal processing and statistical analysis, as well as to probe design, all of which are needed to facilitate replication and cross-study validation of results. Nevertheless, the puzzle of how the developing brain integrates and assigns meaning to auditory information on its way to language is an important one to keep struggling with. The techniques reviewed here will no doubt contribute to our finding the solution.

REFERENCES

Alho, K., Woods, D. L., Algazi, A., and Naatanen, R. (1992). Intermodal selective attention II: effects of attentional load on processing of auditory and visual stimuli in central space. *Electroencephalogr. Clin. Neurophysiol.* 82, 356–368.

Aslin, R. N. (2007). What's in a look? *Dev. Sci.* 10, 48–53.

Aslin, R. N. (2012). Questioning the questions that have been asked about the infant brain using near-infrared spectroscopy. *Cogn. Neuropsychol.* 29, 7–33.

Barker, B. A., and Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition* 94, B45–B53.

Baum, S. R., and Pell, M. D. (1999). The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology* 13, 581–608.

Beauchamp, M. S., Beurlot, M. R., Fava, E., Nath, A. R., Parikh, N. A., Saad, Z. S., et al. (2011). The developmental trajectory of brain-scalp distance from birth through childhood: implications for functional neuroimaging. *PLoS ONE* 6:e24981. doi: 10.1371/journal.pone.0024981

Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., et al. (2011). Mother and stranger: an electrophysiological study of voice processing in newborns. *Cereb. Cortex* 21, 1705–1711.

Bergelson, E., and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common words. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258.

Bortfeld, H., Fava, E., and Boas, D. A. (2009). Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. *Dev. Neuropsychol.* 34, 52–65.

Bortfeld, H., and Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cogn. Psychol.* 60, 241–266.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychol. Sci.* 16, 298–304.

Bortfeld, H., Wruck, E., and Boas, D. A. (2007). Assessing infants' cortical response to speech using near-infrared spectroscopy. *Neuroimage* 34, 407–415.

Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583.

Cooper, R. P., and Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Dev.* 61, 1584–1595.

Cutler, A., and Mehler, J. (1993). The periodicity bias. *J. Phon.* 21, 103–108.

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176.

de Haan, M., Johnson, M. H., and Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: a review. *Int. J. Psychophysiol.* 51, 45–58.

Dehaene-Lambertz, G., and Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature* 370, 292–295.

Dehaene-Lambertz, G., Dehaene, S., Anton, J. L., Campagne, A., Ciuciu, P., Dehaene, P., et al. (2006a). Functional segregation of cortical language areas by sentence repetition. *Hum. Brain Mapp.* 27, 360–371.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., et al. (2006b). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14240–14245.

Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015.

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., et al. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain Lang.* 114, 53–65.

Dehaene-Lambertz, G., and Peña, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport* 12, 3155–3158.

Démonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain* 115, 1753–1768.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13494–13499.

Fava, E., Hull, R., and Bortfeld, H. (2011). Linking behavioral and neurophysiological indicators

of perceptual tuning to language. *Front. Psychol.* 2:174. doi: 10.3389/fpsyg.2011.00174

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behav. Dev.* 8, 181–195.

Fernald, A., and Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* 10, 279–293.

Fisher, C., and Tokura, H. (1995). The given-new contract in speech to infants. *J. Mem. Lang.* 34, 287–310.

Fowler, C. A., and Housum, J. (1987). Talker's signaling of "new" and "old" words in speech and listener's perception and use of the distinction. *J. Mem. Lang.* 26, 489–504.

Friedrich, M., and Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *J. Cogn. Neurosci.* 16, 1465–1477.

Friedrich, M., and Friederici, A. D. (2005a). Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and nonsense words in picture contexts. *J. Cogn. Neurosci.* 17, 1785–1802.

Friedrich, M., and Friederici, A. D. (2005b). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *Neuroreport* 16, 1801–1804.

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463.

Graf Estes, K., Edwards, J., and Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy* 16, 180–197.

Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23.

Grill-Spector, K., and Malach, R. (2001). fMRI-adaptation: a tool for studying the functional-properties of human cortical neurons. *Acta Psychol. (Amst.)* 107, 293–321.

Hay, J., Pelucchi, B., Graf-Estes, K., and Saffran, J. R. (2011). Linking sounds to meanings: infantstatistical learning in a natural language. *Cognition* 63, 93–106.

Henson, R. (2003). Neuroimaging studies of priming. *Prog. Neurobiol.* 70, 53–81.

Henson, R. N., and Rugg, M. D. (2003). Neural response suppression, hemodynamic repetitioneffects and behavioral priming. *Neuropsychologia* 41, 263–270.

Hohle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Dev.* 32, 262–274.

Homae, F., Watanabe, H., Nakajima, K., Miyashita, Y., and Sakai, K. L. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci. Res.* 54, 276–280.

Homae, F., Watanabe, H., Nakano, T., and Taga, G. (2007). Prosodic processing in the developing brain. *Neurosci. Res.* 59, 29–39.

Houston, D. M., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1570–1582.

Junge, C., Kooijman, V., Hagoort, P., and Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* 15, 463–473.

Kenemans, J. L., and Kahkonen, S. (2011). How human electrophysiology informs psychopharmacology: from bottom-up driven processing to top-down control. *Neuropsychopharmacology* 36, 26–51.

Kitamura, C., and Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy* 14, 77–100.

Kooijman, V., Hagoort, P., and Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cogn. Brain Res.* 24, 109–116.

Korotchikova, I., Connolly, S., Ryan, C. A., Murray, D. M., Temko, A., Greene, B. R., et al. (2009). EEG in the healthy term newborn within 12 hours of birth. *Clin. Neurophysiol.* 120, 1046–1053.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Mattys, S. L., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121.

McRoberts, G. W., McDonough, C., and Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy* 14, 162–194.

Mersad, K., and Nazzi, T. (2012). When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Lang. Learn. Dev.* 8, 303–315.

Mills, D. L., Plunkett, K., Prat, C., and Schafer, G. (2005). Watching the infant brain learn words: effects of vocabulary size and experience. *Cogn. Dev.* 20, 19–31.

Naatanen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear.* 16, 6–18.

Naccache, L., and Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition* 80, 215–229.

Nakano, T., Watanabe, H., Homae, F., and Taga, G. (2009). Prefrontal cortical involvement in young infants' analysis of novelty. *Cereb. Cortex* 19, 455–463.

Narayan, C. R., Werker, J. F., and Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Dev. Sci.* 13, 407–420.

Parise, E., and Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychol. Sci.* 7, 728–733.

Plummer, C., Harvey, A. S., and Cook, M. (2008). EEG source localization in focal epilepsy: where are we now? *Epilepsy* 49, 201–218.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R., and Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Dev.* 77, 266–280.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: an ERP study. *Dev. Sci.* 6, 312–328.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Sakatani, K., Chen, S., Lichty, W., Zuo, H., and Wang, Y.-P. (1999). Cerebral blood oxygenation changes induced by auditory stimulation in newborns infants measured by near infrared spectroscopy. *Early Hum. Dev.* 55, 229–236.

Shukla, M., White, K. S., and Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6038–6043.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Dev.* 43, 549–565.

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71.

Tincoff, R., and Jusczyk, P. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.* 10, 172–175.

Turk-Browne, N., Scholl, B., and Chun, M. (2008). Babies and brains: habituation in infant cognition and functional neuroimaging. *Front. Hum. Neurosci.* 2:16. doi: 10.3389/neuro.09.016.2008

Weber, C., Hahne, A., Friedrich, M., and Friederici, A. D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cogn. Brain Res.* 18, 149–161.

Zangl, R., and Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6- and 13-month-old infants. *Infancy* 11, 31–62.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 October 2012; accepted: 24 March 2013; published online: 17 April 2013.

*Citation: Bortfeld H, Shaw K and Depowski N (2013) Disentangling the influence of salience and familiarity on infant word learning: methodological advances. Front. Psychol. **4**:175. doi: 10.3389/fpsyg.2013.00175*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 11 January 2013
doi: 10.3389/fpsyg.2012.00598

Statistical learning across development: flexible yet constrained

***Lauren Krogh*[1]*, Haley A. Vlach[2] and Scott P. Johnson[1]**

[1] Department of Psychology, University of California, Los Angeles, CA, USA
[2] Department of Educational Psychology, University of Wisconsin, Madison, WI, USA

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Erik D. Thiessen, Carnegie Mellon University, USA
Christopher Conway, Saint Louis University, USA

***Correspondence:**
Lauren Krogh, Department of Psychology, University of California Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563, USA.
e-mail: lkrogh@ucla.edu

Much research in the past two decades has documented infants' and adults' ability to extract statistical regularities from auditory input. Importantly, recent research has extended these findings to the visual domain, demonstrating learners' sensitivity to statistical patterns within visual arrays and sequences of shapes. In this review we discuss both auditory and visual statistical learning to elucidate both the generality of and constraints on statistical learning. The review first outlines the major findings of the statistical learning literature with infants, followed by discussion of statistical learning across domains, modalities, and development. The second part of this review considers constraints on statistical learning. The discussion focuses on two categories of constraint: constraints on the types of input over which statistical learning operates and constraints based on the state of the learner. The review concludes with a discussion of possible mechanisms underlying statistical learning.

Keywords: infants, auditory statistical learning, visual statistical learning, language acquistion, learning constraints, statistical learning mechanisms

INTRODUCTION

To survive, an organism must be capable of organizing and interpreting the constant stream of sensory input it receives. Research in the last two decades has revealed powerful statistical learning abilities in infants and adults, including the developing capacity to extract statistical regularities from a variety of auditory inputs including artificial and natural language (e.g., Saffran et al., 1996a; Saffran et al., 1996b; Pelucchi et al., 2009) and non-linguistic auditory stimuli (Saffran et al., 1999). An independent line of research has extended these findings to the visual domain, demonstrating infants' and adults' sensitivity to statistical patterns within visual arrays and sequences of shapes (e.g., Fiser and Aslin, 2001, 2002a,b; Kirkham et al., 2002, 2007; Bulf et al., 2011).

The current review discusses auditory and visual statistical learning to elucidate both its generality and its constraints. We first outline the major findings of the statistical learning literature with infants, followed by discussion of statistical learning across domains, modalities, and development. The second part of this review considers constraints on statistical learning. The discussion focuses on two categories of constraint: constraints on the types of input over which statistical learning operates, and constraints based on the state of the learner. The review concludes with a discussion of possible mechanisms underlying statistical learning.

AUDITORY STATISTICAL LEARNING

ARTIFICIAL LANGUAGE

Given the richness and complexity of a natural language, how is it that infants acquire vocabulary and structure so rapidly, and seemingly effortlessly, in their first years after birth? For example, one challenge facing young language learners is the fact that speakers do not mark word boundaries with pauses, and listeners must rely on other information to accomplish this task. Early in the "cognitive revolution," researchers hypothesized that the statistical structure of language might be important for word segmentation (Harris, 1955; Hayes and Clark, 1970). For instance, Hayes and Clark (1970) tested adults' ability to segment "words" from a continuous stream of speech analogs in which the only cue to word boundaries was the distribution of the phonemes. Adult participants successfully segmented words, suggesting sensitivity to statistical information in speech. However, Hayes and Clark did not specify a mechanism to account for this result.

Building upon these findings, Saffran et al. (1996a,b) proposed a mechanism for statistical word segmentation: transitional probability (TP) detection. In their experiments, adults, first-graders, and 8-month-olds were presented with a continuous stream of speech from an artificial language in which word boundaries were indicated by differing TPs between syllables within words (high TPs) and across word boundaries (low TPs). After brief exposure to this language, listeners in all three age groups were able to distinguish between high TP syllable sequences ("words") and low TP sequences ("part-words"). Thus, both infant and adult learners appeared sensitive to the TP information contained in the speech stream, suggesting that statistical learning via sensitivity to TPs is a possible mechanism contributing to language acquisition.

Although such early studies in infant statistical learning conceptualized statistical learning as sensitivity to a particular conditional relation, TP, more recent research highlights a variety of other conditional statistics (e.g., mutual information) that could be used to distinguish words from foil items. This point is discussed in greater detail in a subsequent section, however we mention it briefly here to point out that, although several studies are described in terms

of differing TPs, it remains unclear which conditional relations participants rely upon to segment sequences.

One limitation to the design of the aforementioned studies was that frequency information co-varied with conditional probability statistics. That is, high TP words occurred more frequently than low TP part-words in the learning (familiarization) phase of the experiment, and it remained unclear whether participants distinguished syllable sequences based on differences in conditional relations or simply differential frequencies of occurrence during learning. To address this issue, Aslin et al. (1998) conducted a "frequency-balanced" version of their original study, with words and part-words appearing equally frequently, such that only sensitivity to conditional relations could be used to distinguish the two types of sequences. Aslin et al. found that 8-month-old infants were still able to distinguish high and low TP sequences. This result suggests that infants can track conditional probability information independent of co-occurrence frequency and use this information to determine word boundaries. Taken together, this work demonstrated the potential for statistical learning to support early language acquisition.

The possibility that statistical learning is a primary mechanism underlying early language acquisition raises the question of the age at which statistical learning is functional in young infants. Teinonen et al. (2009) examined statistical learning in sleeping newborns by presenting a continuous stream of three-syllable words in an artificial language similar to that employed by Saffran et al. (1996a), in which the only cues to word boundaries were the conditional relations or frequencies of co-occurrence between syllables. Using electroencephalography, they measured newborns' event-related potential (ERP) negativities to the first, second, and third syllables in the words. Teinonen et al. (2009) found a significant difference between the ERP negativity to the first and third syllables, indicating that the neonatal brain is sensitive to word boundaries marked by conditional relations and reacts differently during word onset compared to word offset. This research demonstrates, therefore, that statistical learning is functional even in newborn infants, and perhaps contributes to language acquisition even prior to birth.

For statistical learning to be a primary mechanism underpinning infants' early language acquisition, however, it must be able to scale up to the demands of more complex natural language (Johnson and Tyler, 2010). The aforementioned studies employed artificial speech composed entirely of bisyllabic words or entirely of trisyllabic words. Natural language, in contrast, consists of much more varied word types. To simulate more natural language learning, Johnson and Tyler (2010) investigated infants' ability to segment an artificial language composed of both bi- and trisyllabic words. Interestingly, neither 5.5- nor 8-month-old infants were able to segment this language, suggesting that certain characteristics of natural language, such as varied word length, may make segmentation more difficult compared to segmentation of artificial languages.

Other research, however, suggests that some characteristics of natural language may help to make statistical word segmentation possible. For instance, Thiessen et al. (2005) found that 7-month-olds were able to segment an artificial language containing words of varying length when the language was produced with infant-but not adult-directed prosody. As an artificial language becomes more complex (here, by consisting of words of mixed, as opposed to uniform, length), therefore, other natural speech cues such as exaggerated prosody may be needed to facilitate statistical word segmentation.

Indeed, conditional probabilities have never been posited as the sole cue to word segmentation in natural language. Instead, researchers have suggested that initial sensitivity to conditional probabilities may facilitate language acquisition by bootstrapping sensitivity to other linguistic cues. For instance, in English, lexical stress serves as a cue to word boundaries as a majority of English words are stressed on their first syllable (Thiessen and Saffran, 2003). Statistical segmentation mechanisms may facilitate sensitivity to stress cues by providing infants with an inventory of words from which they can discover the dominant stress pattern of their native language (Thiessen and Saffran, 2003, 2007; Swingley, 2005).

In the next section, we discuss research that provides even stronger support for the possibility that statistical learning contributes to language acquisition by examining infants' statistical learning in natural language.

NATURAL LANGUAGE

The aforementioned research focused on statistical learning in the context of synthesized artificial languages. More recent research has examined more natural language learning contexts, such as sequences of grammatically correct and semantically meaningful sentences in natural speech. Pelucchi et al. (2009) examined 8-month-olds' ability to extract statistical regularities from an unfamiliar natural language (Italian for English-learning infants). Infants were presented with a constant stream of fluent infant-directed Italian speech for approximately 2 min. After this brief exposure, infants provided evidence of discrimination between high- and low-TP bisyllabic words. Importantly, both types of words had occurred equally frequently in the speech stream, indicating that infants were using conditional probability information, not simply frequency information, in discriminating between words.

The Pelucchi et al. (2009) results imply that infants discriminated likely from unlikely sound sequences in natural language, but they leave open the critical question of how learners represent extracted statistical information. Saffran (2001) took an important step in addressing this question by asking whether English-learning infants treat segmented syllable sequences as candidate English words or simply as highly probable sound sequences. In this experiment, 8-month-old infants were familiarized to a continuous stream of artificial speech composed of nonsense words similar to those used in Saffran et al. (1996a). Following familiarization to the stimuli, infants participated in a post-familiarization test. This test compared infants' listening time to speech in which words and part-words were embedded in either simple English (e.g., "I like my *tubido*") or matched nonsense (e.g., "zy fike ny *tubido*") frames. If infants treated the outputs of statistical learning simply as highly probable sound sequences, both the English and nonsense frame conditions should have elicited similar listening preferences. However, if infants treated the outputs of statistical learning as candidate English words, then

they should have shown differential listening preferences when those units were embedded in English versus nonsense frames. Saffran found that infants exposed to English frames listened significantly longer to words in this English context than to part-words, and that this difference in listening preference for words versus part-words did not extend to the nonsense frame condition. These results suggest that the statistical learning mechanisms underlying word segmentation do generate word-like units and raises the question of whether these units are available to support other aspects of language acquisition, such as mapping words to meaning.

Establishing a link between sound and meaning is an essential aspect of language acquisition, particularly for young language learners. Graf Estes et al. (2007) investigated the connection between statistical word segmentation and object-label learning in 17-month-olds. Infants were presented with 2.5 min of fluent speech composed of bisyllabic nonsense words where the only cues to word boundaries were the conditional relations between syllables. Immediately following this segmentation task, infants were habituated to two object-label combinations, presented one at a time. For each combination, infants heard a bisyllabic sound sequence from the segmentation task while viewing a 3D object on a computer screen. For half the infants, the bisyllabic sound sequences were words from the segmentation task, and for the other half, the sound sequences were non-words (Experiment 1) or part-words (Experiment 2). Following habituation to these two object-label pairings, infants were presented with two types of test trials. "Same" test trials presented the same object-label combinations from the habituation phase. "Switch" test trials switched the labels for the two objects such that the label for object 1 was played while the infant viewed object 2. Longer looking on switch trials would suggest that infants were sensitive to the change in word-object pairings and was therefore taken as evidence of acquisition of the object-label associations. Graf Estes et al. found that only infants exposed to words from the segmentation task as object labels looked longer on switch compared to same test trials. This indicates that by 17 months of age, infants may be able to map newly segmented sound sequences ("words") to novel objects as linguistic labels, but are unable to do so with non-words or part-words. These results support the claim that statistically segmented sound sequences are word-like and suggest that the output of auditory statistical learning is represented linguistically.

Recent work has also found associations between statistical learning abilities and natural language processing (Conway et al., 2010; Misyak and Christiansen, 2012). For instance, Misyak and Christiansen (2012) found that even after controlling for measures of short-term and working memory, vocabulary, reading experience, cognitive motivation, and fluid intelligence, performance on statistical learning tasks was the key predictor of comprehension of natural language sentences. Such findings suggest that statistical learning may be relevant to language learning not only because extracted statistical information may be represented linguistically, but also because statistical and language learning might overlap in their underlying mechanisms (Christiansen et al., 2007; Misyak and Christiansen, 2012; see also work on cross-situational statistical learning, e.g., Smith and Yu, 2008).

NON-LINGUISTIC STIMULI

Demonstrations that conditional probability information extracted from auditory input is represented linguistically (Saffran, 2001; Graf Estes et al., 2007) and that learners form associations between auditory statistical learning and language learning (Conway et al., 2010; Misyak and Christiansen, 2012) raise the question whether statistical learning is language-specific, or whether it also operates over non-linguistic stimuli. In the auditory domain, Saffran et al. (1999) found that both infants and adults appeared to detect statistical regularities in non-linguistic sequences of "tone words." The procedure and stimuli used were modeled directly after those used in Saffran et al.'s (1996a,b) studies employing speech, allowing for a direct comparison of participants' performance with tones and syllables. Both adults and infants performed with similar accuracy in discriminating words from part-words, regardless of whether these units were instantiated in syllables or tones. These findings suggest that statistical structure can be extracted from auditory input regardless of the domain in which it is presented (syllables or tones), and raise the possibility that statistical learning might also function over input from other modalities.

VISUAL STATISTICAL LEARNING

Investigating infants' and adults' extraction of statistical structure in visual input addresses the question of domain-generality by asking whether or not statistical learning is limited to auditory input.

INFANTS

Kirkham et al. (2002) examined infants' detection of statistical regularities from sequentially presented visual information. Two-, 5- and 8-month-old infants were habituated to a continuous stream of six looming colored shapes presented one at a time with no breaks or pauses between shapes. The six shapes were organized into three pairs that were presented in random order such that the boundaries between pairs were defined by TPs (TP = 1.0 within pairs, TP = 0.33 between pairs). Following habituation, infants viewed six test displays alternating between the familiar habituation sequence and a novel sequence composed of the same six shapes from habituation presented in random order. Infants at all three ages exhibited a significant novelty preference, suggesting that the infants were sensitive to statistical regularities that defined the visual shape sequences. This was the first published experiment to demonstrate not only infants' sensitivity to statistically defined structure in visual sequences, but also to suggest that statistical learning is a domain-general learning process, capable of identifying statistical structure across modalities.

The Kirkham et al. (2002) study was also the first to investigate the developmental time course of visual statistical learning during the first year after birth. Kirkham et al. found no significant differences in novelty preferences between age groups. This lack of observed development, combined with the finding that statistical structures could be detected after only a few minutes of exposure, suggests visual statistical learning may be functional at or soon after the onset of visual experience. Bulf et al. (2011) explored this possibility by investigating whether infants are capable of extracting statistical regularities from visual sequences at birth. Bulf et al. employed a habituation design similar to that used by Kirkham

et al. (2002), presenting newborn infants (mean age 38 h) with continuous sequences of either four or six looming shapes following a statistically defined structure. Newborns provided evidence of detecting the structure of the shape sequences, though only in sequences composed of four, not six, shapes. Thus, statistical learning appears to be functional at birth, operating over both auditory (Teinonen et al., 2009), and visual input (Bulf et al., 2011), but is constrained, an issue we discuss in greater detail in a subsequent section.

The method of testing employed by Kirkham et al. (2002) and Bulf et al. (2011) demonstrated that infants can discriminate between structured and random sequences. However, it did not indicate what statistical or structural features allowed infants to make this discrimination. Rather than computing conditional statistics, as has been found in studies of auditory statistical learning, infants could have been responding to a variety of other features, such as frequency of shape co-occurrence, which co-varied with conditional probability information. Determining which features infants are sensitive to is important for understanding the extent and utility of statistical learning as detection of different statistical features allow varying degrees of associative learning and inference. For instance, *co-occurrence statistics* inform the observer of the likelihood of two events occurring together, but leave the observer uncertain of the likelihood of an event occurring given that the other has taken place. In contrast, *conditional probability statistics* serve to reduce uncertainty by measuring the predictive power of one event with respect to another. Reducing uncertainty contributes to efficient coding of sensory information and is thought to be essential for associative learning (see Fiser and Aslin, 2002b). Thus, a learning mechanism that allows detection of conditional probability statistics would support more effective learning, including the prediction of the likelihood of future events, relative to co-occurrence frequency.

Fiser and Aslin (2002a) examined whether infants were sensitive to conditional probability statistics in visual input in addition to co-occurrence frequency. They habituated 9-month-olds to looming multi-element scenes, then tested infants' preference for various element pairs that had occurred in the scenes. The researchers found that infants preferred not only element pairs that co-occurred more frequently as embedded elements in scenes, but also pairs that had higher conditional probability (viz., predictability) between elements in the pair. Thus, infants were sensitive to the statistical coherence of the elements within visual scenes in addition to co-occurrence frequency. In sum, this research demonstrates infants' sensitivity to conditional relations in both auditory and visual input, suggesting that statistical learning is a domain-general process. In the next section, we outline research with adults that provides even stronger support for this idea by examining statistical learning of more complex visual stimuli and the generalizability of statistical learning across contexts.

ADULTS

Although research with infants has begun to demonstrate the robustness of statistical learning for detecting statistical structure in visual scenes and sequences, the complexity of the visual structures examined in infant studies are rather simplistic compared to those examined in studies with adults. For example, research with adults has examined learners' sensitivity to first- as well as higher-order statistics, and has employed more complex multi-element scenes and sequences than those used with infants to examine the flexibility of the representations learners extract from such input.

Fiser and Aslin (2001) explored the range of first- and higher-order statistics that adults compute during passive viewing of visual scenes. Participants viewed a total of 12 shapes, which were divided into six base pairs. Three of these pairs appeared at a time in various positions within either a 3×3 or 5×5 grid "scene." The relations between any two shapes in a scene could be described in terms of co-occurrence and conditional probabilities. Each base pair appeared in half of the scenes, such that the probability of co-occurrence of the two shapes in each of the six base pairs was 0.5. Because the two objects composing each base pair always occurred together within a scene, shapes within base pairs had a conditional probability of 1.0. Fiser and Aslin found that adults detected first-order statistics (single-shape frequency) as well as several higher-order statistics from the scenes. Specifically, participants detected absolute shape-position relations within the grid and shape-pair arrangements independent of grid position. Most importantly, even when the probabilities of co-occurrence of some base pairs and non-base pairs were equated, adults were still able to distinguish the familiar base pairs based solely on their (higher) conditional probabilities.

The finding that adults are capable of implicitly extracting higher-order statistics from static spatially presented visual stimuli led Fiser and Aslin (2002b) to probe this ability further with temporally presented stimuli. In this experiment, adult participants viewed 12 shapes organized into four temporal triplets, such that after the first element of the triplet appeared on the screen, the second and then the third elements of the triplet always followed. There were no pauses or breaks between successive shapes such that the triplet structure could only be learned via temporal-order statistics among pairs or triplets of shapes. Just as with spatially presented visual stimuli, participants became sensitive to first-order as well as higher-order statistics in the temporal shape sequences. Participants retained the frequency of individual shapes and distinguished sequences of shapes presented during familiarization from both novel sequences of familiar shapes and sequences of shapes seen during familiarization but presented less frequently. Interestingly, when frequency information and co-occurrence probabilities were equated, adults were still able to distinguish shape sequences based on differing conditional probabilities.

These demonstrations of visual statistical learning with both temporally and spatially presented input raises the question of how such information is represented and whether such representations might generalize to new contexts. Turk-Browne and Scholl (2009) demonstrated that learning of statistical regularities in temporal shape sequences (finding shape "triplets" in a continuous stream of shapes) was expressed in static spatial configurations of these same shape triplets. Similarly, learning of statistically defined spatial configurations (base pairs, as in Fiser and Aslin, 2001) facilitated detection performance in temporal streams (Turk-Browne and Scholl, 2009). Thus, visual statistical learning

in adults appears to produce flexible representations that can be generalized to new situations. Such transferability is likely important for visual statistical learning to be practical in ever-changing real-world visual environments.

CONSTRAINTS ON STATISTICAL LEARNING

The generalizability of statistical learning across tasks and domains raises the important question of whether and what constraints may exist on statistical learning. If one considers the infinite number of possible statistical relations that could be computed at each level of representation, it becomes clear that for statistical learning to be feasible, it must be constrained. What are these constraints?

TYPES OF INPUT

It is unlikely that all statistical regularities are learned equally well, given the infinite number of possible statistics that could be extracted from the environment. Rather, research suggests that statistical learning mechanisms preferentially track statistical regularities in the types of input that occur most frequently in the natural environment (Newport and Aslin, 2004; Conway and Christiansen, 2009; Emberson et al., 2011).

Spatial versus sequential input

Intuitively, there seem to be structured differences in the organization of auditory and visual information in the natural environment. For instance, auditory information is conveyed temporally whereas visual information is arrayed spatially. Moreover, each sensory modality seems to process particular aspects of environmental input. For instance, a brief snapshot is typically enough time to recognize a complex visual scene whereas at least several seconds are needed to recognize a voice or melody (Conway and Christiansen, 2009). These intuitions are supported by studies of perception and memory suggesting that spatial information weighs most prominently in visual cognition, whereas temporal information weighs most prominently in audition (see Conway and Christiansen, 2009 for a discussion). Such modality differences raise the question of whether statistical learning processes might be constrained to preferentially track statistics in input that accords with the auditory-temporal, visual-spatial structure of the environment.

Conway and colleagues (Conway and Christiansen, 2005, 2009; Emberson et al., 2011) examined how modality differences may constrain implicit statistical learning. For example, Conway and Christiansen (2009) investigated whether vision and audition exhibited different constraints on statistical learning of spatially and temporally structured information. Conway and Christiansen compared learning of one statistically defined structure presented in three different formats: auditory information presented temporally (pure tones of various frequencies presented one at a time through headphones), visual information presented temporally (different colored squares presented one at a time in the center of screen), and visual information presented spatially (the same colored squares presented simultaneously left to right in a horizontal row across the center of the screen). The task was an artificial grammar learning (AGL) task in which adult learners were presented with a set of training sequences that adhered to a specific rule-governed finite state grammar. After the learning task, learners were presented with a test on classifying novel sequences as being either legal (generated by the same rules as the training sequences) or illegal. The results demonstrated that participants in the visual-spatial condition classified test sequences with a similar degree of accuracy as participants in the auditory condition. However, participants in the visual-temporal condition were significantly less accurate in their classifications compared to those in the auditory condition. This ability to acquire the structure of spatially arrayed visual input as well as temporally structured auditory, but not visual, input suggests that adults' statistical learning may be constrained to preferentially track statistics in inputs that accord with the auditory-temporal, visual-spatial structure of the environment.

Presentation rate

Of course, human learners, including young infants, provide evidence of detecting statistical patterns in sequential visual input under some circumstances (e.g., Fiser and Aslin, 2002b; Kirkham et al., 2002; Bulf et al., 2011). A recent study by Emberson et al. (2011) helped to reconcile these seemingly contradictory findings by investigating the mediating role of presentation timing in statistical learning of auditory and visual information. Their results suggest that there is an interaction of presentation format (spatial versus sequential) and presentation timing in constraining statistical learning across modalities.

Emberson et al. (2011) compared visual and auditory statistical learning in an interleaved familiarization design. Adult learners were presented with a visual stream of abstract shapes organized into triplets that was interleaved pseudo-randomly with an auditory stream of monosyllabic nonsense words also organized into triplets. Participants were randomly assigned to either attend to the visual stream or the auditory stream, and given a cover task (detecting repeat elements in only that stream) to ensure that attention was allocated to the appropriate stream. Following familiarization, participants were tested on learning in each modality. During test trials, participants judged which of two sequences seemed more familiar: a triplet from familiarization or a foil sequence that did not adhere to the triplet structure. Importantly, this study compared effects of variation in presentation rate. In the "fast" condition, elements were presented for 225 ms with an ISI of 150 ms, resulting in an SOA of 375 ms. In the "slow" condition, elements were presented with an SOA of 750 ms.

Emberson et al. (2011) found that performance in the unattended modality did not differ from chance in any condition. At the fast presentation rate, the statistical relations between adjacent elements were only learned in the attended *auditory* stream. At the slow presentation rate, the opposite effect occurred: only the relations between adjacent elements in the attended *visual* stream were learned. Emberson et al. posited that visual statistical learning improved with the slower rate of presentation because it was less temporally demanding on the visual system. In contrast, auditory statistical learning was impaired at the slower presentation rate because of weaker perceptual grouping cues. That is, when sequential elements were separated by longer intervals, they were less likely to form a single perceptual unit or stream, hindering the detection of statistical information in the stream. Taken together,

these results document complex constraints on statistical learning that accord with the structure of the natural environment, with relatively rapid presentation of temporal information critical for auditory statistical learning, and either static spatial information or relatively slowly presented temporal information critical for visual statistical learning.

Natural language: types of non-adjacent regularities

This interaction of presentation format and timing in statistical learning illustrates one way in which constraints on the types of information over which statistical learning operates may reflect environmental structure. Some researchers have additionally argued that constraints on learning not only reflect, but also help to explain, structural aspects of the environment, such as those found in natural languages (e.g., Christiansen and Chater, 2008). For example, a wide range of adjacent regularities appear throughout natural languages, but the types of non-adjacent regularities languages exhibit are quite constrained.

Newport and Aslin (2004) investigated the intriguing possibility that constraints on the types of non-adjacent statistical computations that learners perform may match and even drive observed constraints on non-adjacent regularities in natural languages. For example, it is common for natural languages to contain non-adjacent regularities relating elements of one kind while skipping over intervening elements of a different kind. In Hebrew and Arabic, word stems are formed out of phonemic segments of one kind (consonants), while intervening segments are of another kind (vowels). In contrast, it is uncommon for natural languages to contain non-adjacent regularities in which intervening items are of the same kind as that in which the non-adjacent regularities occur. Newport and Aslin examined adults' detection of conditional relations among non-adjacent elements that did and did not adhere to this natural language structure: non-adjacent consonants (with one unrelated intervening vocalic segment), non-adjacent vowels (with one unrelated intervening consonantal segment), and non-adjacent syllables (with one intervening syllable that was unrelated). In accord with the structure of natural languages, adults seemed to be unable to track the relations between non-adjacent syllables, where the intervening element was of the same kind (a syllable). Even when the patterns were quite simple and participants were given extensive exposure to the patterns (in one case over 10 days of repeated exposures), participants remained unable to track relations between non-adjacent syllables. In contrast, adults readily learned the relations between non-adjacent consonants and vowels, where the intervening element was a different kind from that in which the non-adjacent regularities occurred. These findings suggest that constraints on statistical learning may help to explain the universal aspects of these patterns in natural languages. Similar to Conway and colleagues' results (Conway and Christiansen, 2009; Emberson et al., 2011), these findings also demonstrate that human learners preferentially track statistical information only in particular types of environmental input. Such findings highlight the importance of considering statistical learning in its broader environmental context, including the nature of the input to which the learner is exposed, as well as the cognitive, developmental, and attentional state of the learner.

THE STATE OF THE LEARNER

Human learners are characterized by perceptual biases and cognitive constraints. Appreciating the influences of learners' biases and developmental state on statistical learning is necessary for a complete understanding of the extent and limits of this domain-general learning process across development.

Spatiotemporal biases and perceptual similarity

Consideration of learners' perceptual biases is especially important for understanding constraints on visual statistical learning, as such biases have been shown to influence the types of statistics learners extract from visual scenes (Fiser et al., 2007). One general perceptual bias exhibited by infants and adults is the bias to perceive objects as moving along specific trajectories given certain visual and/or auditory cues (e.g., Sekuler and Sekuler, 1999; Shimojo et al., 2001). When observing two identical objects moving toward each other, coinciding, then moving away from each other, two interpretations are possible: (1) the two objects streamed past one another (*streaming*), or (2), the two objects bounced off of one another (*bouncing*). Various perceptual features such as the acceleration of the objects (Sekuler and Sekuler, 1999; Fiser et al., 2007) or the presence of a sound at the time of coincidence (Sekuler et al., 1997; Watanabe and Shimojo, 2001) bias observers toward one of these two interpretations.

Fiser et al. (2007) investigated whether this perceptual bias to perceive objects as moving along specific trajectories affected the types of statistics adult learners computed from visual events. Participants observed a single object move behind an occluder and then saw two objects emerge from behind the occluder simultaneously. One object emerged from the occluder following the same trajectory as the first object. The second object emerged from the occluder at a 90° angle to the original trajectory. Thus, presentations could be interpreted two different ways: (1) as an object streaming behind the occluder on a straight trajectory, or (2) as an object bouncing off of a surface behind the occluder and reemerging on the same side that it originated.

To examine whether perceived motion trajectories would bias statistical learning, Fiser et al. (2007) manipulated the acceleration of the objects to bias observers toward one of these two percepts. Objects moving at constant speed produced a streaming percept whereas decelerating-accelerating objects produced a bouncing percept. If visual statistical learning mechanisms compute all available temporal co-occurrences of shape pairs, then learners should acquire transitions from the first shape to each of the two later shapes equally well, regardless of whether observers were biased toward streaming or bouncing percepts. However, this is not what Fiser et al. found. Rather, adults preferentially learned the associations consistent with the perceptual bias of streaming or bouncing they had during familiarization. Thus, this perceptual bias constrained statistical learning to shape pairs consistent with that bias.

The influence of perceptual biases on statistical computations is not limited to statistics in visual scenes. Similar to spatiotemporal biases, Gestalt principles of perception have been shown to constrain the detection of statistical relations in both auditory and visual input (Baker et al., 2004; Creel et al., 2004; Newport and Aslin, 2004; Emberson et al., 2011). For example, Creel et al.

(2004) demonstrated that Gestalt principles of element similarity interact with temporal adjacency in determining what kinds of auditory statistical regularities are learned. In this experiment, adult participants were presented with two interleaved streams of tone triplets such that participants heard the first tone of the first triplet stream, followed by the first tone of the second triplet stream, then the second tone of the first stream, then the second tone of the second stream, and so on (Creel et al., 2004). The result of this interleaving was that triplets could only be detected via sensitivity to non-adjacent conditional relations.

Interestingly, adults showed no learning of the tone triplets, only sensitivity to the less reliable relations between adjacent elements in the stream. However, when Creel et al. (2004) included perceptual grouping cues, by presenting the two interleaved streams in differing pitch ranges or timbres, adults became sensitive to the conditional relations between the similar, yet temporally non-adjacent, elements. This finding suggests that Gestalt principles of similarity interact with temporal adjacency in constraining statistical learning.

Availability of cognitive resources

Thus far, our discussion has highlighted similarities in infants' and adults' sensitivities to statistical information. Researchers hold differing views, however, on how implicit statistical learning abilities may change across development (e.g., Thomas et al., 2004; Janacsek et al., 2012) or remain constant across development (e.g., Reber, 1993; Vinter and Perruchet, 2000).

In some studies reporting developmental differences, older individuals show better learning than younger individuals (e.g., Maybery et al., 1995). Consistent with this possibility, infants provide evidence for tracking increasingly complex statistical regularities in visual sequences with age: 2- 5- and 8-month-old infants distinguished structured from random sequences composed of six looming shapes (Kirkham et al., 2002), but newborn infants only distinguished structured from random sequences when the sequences contained four, not six, items (Bulf et al., 2011).

In other cases, however, younger individuals outperform older individuals (e.g., Jost et al., 2011; Janacsek et al., 2012). Jost et al. (2011) compared the time course of children's and adults' implicit learning by examining participants' ERPs during a visual statistical learning task. Participants observed a series of stimuli presented one at a time on a screen and pressed a button whenever the target stimulus appeared, which was predicted at different levels of probability by the stimuli immediately preceding the target. Jost et al. found that children exhibited learning-related ERP components earlier in the study than adults, suggesting that children required less exposure to the patterns to detect the statistical structure.

To explain differences in statistical learning ability across development, researchers have appealed to domain-general, maturational constraints on perception and memory. Bulf et al. (2011) suggested that newborns' limited attentional and working memory capacities may inhibit statistical learning efficiency. Interestingly, researchers have posited a similar explanation to account for findings of children outperforming adults. In that case, however, researchers have offered the paradoxical idea that maturational constraints on perception and memory confer a computational advantage for some types of learning (e.g., Newport, 1988, 1990; Elman, 1993). In particular, Newport's (1990) "Less is More" hypothesis assumes that children's abilities to perceive and store complex stimuli is reduced compared to those of adults, and suggests that such limitations give children an advantage for tasks requiring componential analysis because children are better able to identify and process component parts. Adults, in contrast, attempt to perceive and store stimulus relations of greater complexity.

Suggestions that maturational constraints on perception and memory can both hurt and help performance in tasks requiring componential analysis appear contradictory. However, most empirical support for Newport's "Less is More" hypothesis (1990; e.g., Kersten and Earles, 2001) comes from child and adult populations, leaving open the possibility that very early increases in infants' relatively limited perception and memory abilities may be positively related to statistical learning ability. To our knowledge, however, Bulf et al.'s (2011) hypothesis that limited cognitive resources limit newborns' statistical learning performance has not yet been confirmed independently. Although visual working memory performance increases roughly linearly across the first postnatal year (Diamond, 1985; see Bell and Morasch, 2007 for a review), a number of other early developments could, in principle, be responsible for changes in statistical learning (e.g., different spatiotemporal biases due to changes in perceptual acuity). An important avenue for future research will be to investigate these possibilities, beginning by examining the relation between the development of infant working memory ability and statistical learning ability.

In addition to maturational constraints on perception and memory, the allocation of attentional resources may also play a role in constraining statistical learning. Although some researchers have argued that statistical learning is an "automatic" (i.e., implicit, rapid) process (e.g. Saffran et al., 1997), other researchers have found reason to suggest that statistical learning both is and is not automatic (e.g., Turk-Browne et al., 2005). It is automatic in that statistical computations seem to be carried out without conscious intent and often without awareness that any structure was learned (e.g., Saffran et al., 1997; Meulemans et al., 1998; Turk-Browne et al., 2005). However, statistical learning is not automatic in that it operates better over attended versus unattended input (e.g., Toro et al., 2005; Turk-Browne et al., 2005; Emberson et al., 2011). For instance, when two interleaved streams of shapes are presented to observers in two different colors, and participants are instructed to attend to only one color, only the statistical relations in the attended color are learned (Turk-Browne et al., 2005). This attentional constraint on statistical learning appears to be one of its most general limitations, likely constraining detection of statistical regularities regardless of input domain or modality (e.g., Emberson et al., 2011).

Prior experience

In addition to maturational changes in cognitive resources, such as working memory capacity and attention, another important aspect of development is learning from experience interacting with the environment. Expectations about the structure of the environment undergo rapid changes in the first years after birth due to experiences interacting with the world (e.g., Campos et al., 1992; Adolph et al., 1993). Such changes in learners' expectations

about the structure of their environment may have the potential to influence statistical learning processes (Thiessen, 2010). For example, years of experience with language may provide adults with strong expectations that words and objects relate to one another (e.g., Namy and Waxman, 1998).

Thiessen (2010) investigated how such expectations influence adults' statistical learning of word-object associations. Adults were presented with paired audio-visual information in which word boundaries as well as word-object associations were statistically defined. Participants tracked both of these statistical relations simultaneously, and word segmentation benefited from the addition of word-object associations. When adults were presented with tonal rather than linguistic stimuli, however, they did not benefit from the regular relations between tone words and objects. Thiessen suggested that experience with language may predispose adults to expect words and objects to relate to one other, such that they are sensitive to these associations in linguistic input, but not in tonal input. This hypothesis leads to the prediction that young infants may not benefit from word-object relations even with linguistic input, because they may not yet have built up the expectation that words relate to objects (e.g., Werker et al., 1998). This is precisely what Thiessen found; similar to adults in the tonal condition, 8-month-old infants' ability to segment words did not benefit from the presence of word-object relations, regardless of whether linguistic or non-linguistic input was used.

Thiessen's (2010) findings demonstrate the role of prior experience and learners' expectations in facilitating computation of previously ignored statistics. Other research, however, indicates that prior experience can impede statistical computations. For example, Gebhart et al. (2009) presented adult learners with auditory sequences of trisyllabic nonsense words defined by the TPs between syllables. When the researchers altered the organization of the nonsense words mid-way through the familiarization stream, participants only learned the first of the two structures. Participants detected words in both structures only when exposure to the second structure was tripled in duration, or when the transition between structures was explicitly marked. Thus, successful extraction of the statistical regularities in one auditory structure inhibited learning of a subsequent auditory structure.

MECHANISMS UNDERLYING STATISTICAL LEARNING

How is it that statistical learning can be so constrained while still adapting flexibility to input across domains and modalities? The reason for both flexibility and constraints on statistical learning is likely because the environment contains both variance and invariance; organisms need a way to flexibly adapt and generalize to different contexts while simultaneously honing in on the types of structures that are most consistent and informative in the environment. What is less clear are the mechanisms by which statistical learning occurs and how these mechanisms are configured to allow for both flexibility and constraints.

We began this review by introducing statistical learning as sensitivity to transitional probabilities (TPs), and this view was predominant in the early days of infant statistical learning research that focused predominantly on word segmentation. However, there is now a wealth of data on infants' and adults' statistical learning across domains, and this calls for a broader view of statistical learning (e.g., Saffran, 2001; Maye et al., 2002; Thiessen and Saffran, 2003; Graf Estes et al., 2007; Smith and Yu, 2008; Frank et al., 2010). For example, consider Saffran's (2001) and Graf Estes et al.'s (2007) findings that the output of statistical learning is entire word-like units, not simply highly probable sound sequences. A mechanism that only tracks probabilistic relations between elements cannot fully account for such a finding (see Thiessen et al., 2012). Moreover, even in segmentation tasks, models designed to track transitional probabilities do not always accord well with human performance (see Frank et al., 2010).

A variety of alternate models of statistical learning have been proposed that do not rely on explicitly computed statistics. It is not yet clear which type of model produces the most valid account of human learning processes across tasks (Frank et al., 2010). A complete review of all such models is beyond the scope of this review; instead, we briefly describe one well-known model, PARSER (Perruchet and Vinter, 1998), to illustrate that there are multiple possible mechanisms to account for statistical learning data.

PARSER (Perruchet and Vinter, 1998) is a type of "chunking" model that produces the same segmentation results as Saffran et al. (1996a,b) by implementing basic laws of attention, memory, and associative learning, rather than by computing statistics such as transitional probabilities. PARSER is modeled on the principle that perception guides internal representation. Briefly, units that are perceived within one attentional focus are "chunked" into a new representational unit. The fate of these new representations depends on fundamental principles of memory: internal representations of chunks that are repeated are progressively strengthened, and representations of chunks that are not repeated are forgotten (Perruchet and Vinter, 1998). Applied to Saffran et al.'s (1996a,b) segmentation task, PARSER would first randomly segment the speech stream into small chunks. Because chunks have a greater chance of being repeated if they are part of the same word than if they span a word boundary, internal representations of words or parts of words will be stronger in memory than representations of non-words and chunks spanning word boundaries. Thus, PARSER can account for Saffran et al.'s (1996a,b) findings of participants' greater sense of familiarity for words than non-words or part-words.

As noted, several models of statistical learning employing quite different mechanisms have been proposed to account for the various findings of the statistical learning literature, but no model has yet been proposed that can account well for human performance across statistical learning tasks (Thiessen et al., 2012). In particular, what is lacking are models that achieve sensitivity to other statistical relations in addition to conditional relations, such as the central tendency of a set of elements (distributional statistical learning; e.g., Maye et al., 2002), as well as models that account for human's learning and generalization based upon similarity across items extracted from the input (e.g., Thiessen and Saffran, 2003). Thiessen et al. (2012) argued that mechanisms designed only to account for the extraction of units, such as segmenting words from a speech stream, cannot account for these other forms of statistical learning.

Thiessen et al. (2012) proposed a framework that attempts to account for these various forms of statistical learning by combining processes of extraction with processes of comparison across extracted segments in an iterative model whereby the discovery of new structures via comparison serves to educate the extraction processes. To illustrate this idea, consider the finding that when syllable stress and statistical cues indicated different word boundaries in a speech stream, 7-month-olds segmented based on statistical cues, whereas 9-month-olds segmented based on stress cues (Thiessen and Saffran, 2003). Models that are only designed to account for segmentation cannot explain these findings without positing additional changing constraints on the learner or on the statistical learning mechanism itself. In contrast, Thiessen et al.'s (2012) framework accounts for such findings without necessitating new or changing constraints; according to this framework, such findings demonstrate initial segmentation based on conditional statistics followed by comparison across segmented words, allowing the discovery of patterns of stress cues in English words, which in turn inform the process of segmentation in the future.

Although Thiessen et al.'s (2012) framework has not yet been implemented into a working computational model, such a framework pushes the field forward by offering a mechanism that accounts for developmental differences in statistical learning. Moreover, this framework is also helpful for thinking about the origins of the constraints on and flexibility of statistical learning. That is, the framework is based on general processes of attention, memory, and comparison that likely govern extraction and generalization across domains. Furthermore, this framework describes a way in which learners may use a constrained, limited-capacity mechanism to flexibly adapt to different characteristics of the input over time.

CONCLUSION

Statistical learning is a means of uncovering structure in complex environmental input. It operates in both auditory and visual domains, and encodes multiple types of statistics simultaneously. Constraints on statistical learning serve to reduce the number of possible associations available, making statistical learning tractable.

A comprehensive model of statistical learning across domains has not yet been reported in the literature, but much progress has been made in uncovering the origins of both the flexibility of and constraints on statistical learning. Specifically, flexibility may be the result of mechanisms built upon domain-general processes, such as attention, memory, and perception, rather than domain- or modality-specific processes. Flexibility may be built into the system as a product of learners' ability to discover new structures via comparison, and use those new structures to influence further extraction (Thiessen et al., 2012). Constraints on statistical learning are driven by a variety of factors: limited attention, perception, and memory capacity, as well as maturational increases in these domain-general processes; learned biases and expectations about the structure of the environment; and ways in which statistical tendencies in language have been shaped to fit the human brain, rather than vice versa.

Thus, while research has revealed numerous influences on the various constraints on statistical learning, the principal contribution to flexibility in statistical learning appears to be its domain-general nature. Nevertheless, the domain-generality of statistical learning mechanisms has been hotly debated. Some researchers interpret demonstrations of statistical learning across domains and modalities as evidence of a single, domain-general statistical learning mechanism (e.g., Kirkham et al., 2002), but others contend that statistical learning cannot be domain-general due to observed modality-specific constraints (Conway and Christiansen, 2005, 2009; Emberson et al., 2011). Specifically, they cite findings such as the auditory-temporal, visual-spatial distinction as evidence for separate statistical learning mechanisms for each modality (Conway and Christiansen, 2009). One limitation of this line of reasoning, however, is that constraints differentially affecting statistical learning of different types of input *within* modalities (e.g., Endress, 2010; Thiessen, 2010) would necessitate multiple statistical learning mechanisms *within* modalities as well as across modalities. Thus, the domain-general view seems to be the most parsimonious account of the data. However, evidence supporting a domain-general account of statistical learning does not exclude the possibility of multiple domain- or modality-specific statistical learning subsystems. Further research is needed to determine which of these views provides the most complete account of statistical learning. Research examining statistical learning performance using comparable tasks across domains and modalities, as well as research comparing the ability of modality-specific and domain-general computational models to fit such human data, may be particularly informative.

Moreover, future research should continue to investigate the type of flexibility in statistical learning documented by Turk-Browne and Scholl (2009), who demonstrated flexibility in the transferability of the representations that emerged from adults' visual statistical learning. Further research should pursue similar lines of research employing other tasks and input types to investigate the generalizability of such findings across modalities. A final important avenue for future research will be to continue working toward developing a comprehensive model that can accommodate the various forms of statistical learning (sensitivity to conditional relations, distributional statistics) across domains as well as developmental changes in such learning. Longitudinal research and research that makes within-subjects comparisons across tasks may be particularly useful in this endeavor.

REFERENCES

Adolph, K. E., Eppler, M. A., and Gibson, E. J. (1993). Crawling versus walking infants' perception of affordances for locomotion over sloping surfaces. *Child Dev.* 64, 1158–1174.

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324.

Baker, C. I., Olson, C. R., and Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychol. Sci.* 15, 460–466.

Bell, M. A., and Morasch, K. C. (2007). "The development of working memory in the first 2 years of life," in *Short and Long-term Memory in Infancy and Early Childhood*, eds L. M. Oakes and P. J. Bauer (New York: Oxford University Press), 27–50.

Bulf, H., Johnson, S. P., and Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition* 121, 127–132.

Campos, J. J., Bertenthal, B. I., and Kermoian, R. (1992). Early experience and emotional development: the emergence of wariness of heights. *Psychol. Sci.* 3, 61–64.

Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–509.

Christiansen, M. H., Conway, C. M., and Onnis, L. (2007). "Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, eds D. S. McNamara and J. G. Trafton (Austin, TX: Cognitive Science Society), 173–178.

Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition* 114, 356–371.

Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39.

Conway, C. M., and Christiansen, M. H. (2009). Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *Eur. J. Cogn. Psychol.* 21, 561–580.

Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol.* 30, 1119–1130.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Dev.* 56, 868–883.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99.

Emberson, L. L., Conway, C. M., and Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Q. J. Exp. Psychol.* 64, 1021–1040.

Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychol. (Amst.)* 135, 182–190.

Fiser, J., and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504.

Fiser, J., and Aslin, R. N. (2002a). Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15822–15826.

Fiser, J., and Aslin, R. N. (2002b). Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467.

Fiser, J., Scholl, B. J., and Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychon. Bull. Rev.* 14, 173–178.

Frank, M. C., Goldwater, S., Griffiths, T., and Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition* 117, 107–125.

Gebhart, A. L., Aslin, R. N., and Newport, E. L. (2009). Changing structures in mid-stream: learning along the statistical garden path. *Cogn. Sci.* 33, 1087–1116.

Graf Estes, K. M., Evans, J., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Harris, Z. (1955). From phoneme to morpheme. *Language* 31, 190–222.

Hayes, J. R., and Clark, H. H. (1970). "Experiments in the segmentation of an artificial speech analog," in *Cognition and the Development of Language*, ed. J. R. Hayes (New York: Wiley), 221–234.

Janacsek, K., Fiser, J., and Nemeth, D. (2012). The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. *Dev. Sci.* 15, 496–505.

Johnson, E. K., and Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Dev. Sci.* 13, 339–345.

Jost, E., Conway, C. M., Purdy, J. D., and Hendricks, M. A. (2011). Neurophysiological correlates of visual statistical learning in adults and children. *Paper Presented at the 33rd Annual meeting of the Cognitive Science Society, July 2011*, Boston.

Kersten, A. W., and Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *J. Mem. Lang.* 44, 250–273.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition* 83, B35–B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., and Johnson, S. P. (2007). Location, location, location: development of spatiotemporal sequence learning in infancy. *Child Dev.* 78, 1559–1571.

Maybery, M., Taylor, M., and O'Brien-Malone, A. (1995). Implicit learning: sensitive to age but not IQ. *Aust. J. Psychol.* 47, 8–17.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

Meulemans, T., Van der Linden, M., and Perruchet, P. (1998). Implicit sequence learning in children. *J. Exp. Child. Psychol.* 69, 199–221.

Misyak, J. B., and Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331.

Namy, L. L., and Waxman, S. R. (1998). Words and gestures: infants' interpretation of different forms of symbolic reference. *Child Dev.* 69, 295–308.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: studies of the acquisition of American Sign Language. *Lang. Sci.* 10, 147–172.

Newport, E. L. (1990). Maturational constraints on language learning. *Cogn. Sci.* 14, 11–28.

Newport, E. L., and Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of nonadjacent dependencies. *Cogn. Psychol.* 48, 127–162.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev.* 80, 674–685.

Perruchet, P., and Vinter, A. (1998). PARSER: a model for word segmentation. *J. Mem. Lang.* 39, 246–263.

Reber, A. R. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious.* New York: Oxford University Press.

Saffran, J. R. (2001). Words in a sea of sounds: the output of statistical learning. *Cognition* 81, 149–169.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* 8, 101–105.

Sekuler, A. B., and Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception* 28, 415–432.

Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385, 308.

Shimojo, S., Watanabe, K., and Scheier, C. (2001). "The resolution of ambiguous motion: attentional modulation and development," in *Visual Attention and Cortical Circuits*, eds J. Braun, C. Koch, and J. Davis (Cambridge, MA: MIT Press), 243–264.

Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cogn. Psychol.* 50, 86–132.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10:21. doi:10.1186/1471-2202-10-21

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cogn. Sci.* 34, 1093–1106.

Thiessen, E. D., Hill, E. E., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71.

Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (2012). The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.*

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.

Thiessen, E. D., and Saffran, J. R. (2007). Learning to learn: infants' acquisition of stress-based strategies for word segmentation. *Lang. Learn. Dev.* 3, 73–100.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., et al. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *J. Cogn. Neurosci.* 16, 1339–1351.

Toro, J. M., Sinnett, S., and Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition* 97, B25–B34.

Turk-Browne, N. B., Junge, J. A., and Scholl, B. J. (2005). The automaticity

of visual statistical learning. *J. Exp. Psychol. Gen.* 134, 552–564.

Turk-Browne, N. B., and Scholl, B. J. (2009). Flexible visual statistical learning: transfer across space and time. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 195–202.

Vinter, A., and Perruchet, P. (2000). Implicit learning in children is not related to age: evidence from drawing behavior. *Child Dev.* 71, 1223–1240.

Watanabe, K., and Shimojo, S. (2001). When sound affects vision: effects of auditory grouping on visual motion perception. *Psychol. Sci.* 12, 109–116.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., and Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Dev. Psychol.* 34, 1289–1309.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 August 2012; accepted: 18 December 2012; published online: 11 January 2013.

*Citation: Krogh L, Vlach HA and Johnson SP (2013) Statistical learning across development: flexible yet constrained. Front. Psychology **3**:598. doi: 10.3389/fpsyg.2012.00598*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 31 August 2012
doi: 10.3389/fpsyg.2012.00324

Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data

Joanne Arciuli*[1] *and* ***Janne von Koss Torkildsen*[2]**

[1] *Faculty of Health Sciences, University of Sydney, Sydney, NSW, Australia*
[2] *Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway*

Edited by:
Jutta L. Mueller, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Christopher Conway, Saint Louis University, USA
Jenny Saffran, University of Wisconsin-Madison, USA

***Correspondence:**
Joanne Arciuli, Faculty of Health Sciences, University of Sydney, PO Box 170, Lidcombe, 1825 NSW, Australia.
e-mail: joanne.arciuli@sydney.edu.au

Mastery of language can be a struggle for some children. Amongst those that succeed in achieving this feat there is variability in proficiency. Cognitive scientists remain intrigued by this variation. A now substantial body of research suggests that language acquisition is underpinned by a child's capacity for statistical learning (SL). Moreover, a growing body of research has demonstrated that variability in SL is associated with variability in language proficiency. Yet, there is a striking lack of longitudinal data. To date, there has been no comprehensive investigation of whether a capacity for SL in young children is, in fact, associated with language proficiency in subsequent years. Here we review key studies that have led to the need for this longitudinal research. Advancing the language acquisition debate via longitudinal research has the potential to transform our understanding of typical development as well as disorders such as autism, specific language impairment, and dyslexia.

Keywords: statistical learning, language acquisition, longitudinal studies, language impairment, language proficiency

Statistical learning (SL) likely plays a role in a large number of perceptual and cognitive activities. For example, "Every time we listen to a blues song or a piano concerto, our brains pick up on the underlying statistics regarding which notes tend to occur together or follow one another in these different styles. We use this accumulated knowledge to appraise unfamiliar pieces of music or different performances of well-known songs. In short, our expectations are an outcome of statistical learning" (Janata, 2006, p. 29). The role of SL during language acquisition has been hotly debated over several decades. Certainly, it is clear that language contains many statistical regularities. It has been suggested that SL operates on these regularities and facilitates processes as varied as word segmentation, vocabulary learning, and syntax (Rowland and Pine, 2000; Finn and Hudson Kam, 2008; Yu, 2008).

Consider word segmentation. Child-directed speech includes utterances such as *prettydolly* and *prettykitty*. Each utterance is composed of two words, usually spoken as a continuous stream without pausing between words. How do children identify the separate words (*pretty*, *dolly*, and *kitty*)? Perhaps they detect implicitly the strength of associations between adjacent syllables; *pre* is often followed by *tty* (high joint probability), however, *tty* is rarely followed by *do* (low joint probability). In natural language, joint probabilities between syllables are highest within words; those spanning word boundaries are lower. Thus, sensitivity to these co-occurrence statistics might assist children to segment the speech stream into words, possibly in conjunction with other cues such as prosody (Hay and Saffran, in press). Newman et al. (2006) discovered a relationship between infants' ability to segment the speech stream into words and language proficiency at 24 months and, later, between 4 and 6 years. It was shown that IQ did not mediate this relationship.

As elegant as this theory of language acquisition is, there remain striking gaps in our understanding of the link between SL and language. To date, there has been no direct investigation of whether a capacity for SL in young children is, in fact, associated with language proficiency in subsequent years. This paper provides a brief introduction to the language acquisition debate, a review of key studies indicating a link between SL ability and language proficiency, and a discussion of the kind of longitudinal research that is needed in order to advance the debate about the role of learning during language acquisition. We argue that this kind of longitudinal research will enhance our understanding of language development in typically developing children and, potentially, transform our understanding of disorders such as autism, specific language impairment (SLI), and dyslexia.

DEBATE ABOUT THE NATURE OF LANGUAGE ACQUISITION

There has been a long-standing debate about the link between learning and language acquisition. Chomsky and others have speculated that language is too complex and the learning environment too impoverished to be assisted by a general learning mechanism (e.g., Chomsky, 1975; Pinker, 1989; Crain, 1991; special issue edited by Ritter, 2002). This led to the suggestion that children come into the world already equipped with a great deal of linguistic knowledge. The innateness hypothesis incorporates multiple and intertwined notions including both linguistic universals and modularity. While these cannot be covered adequately here, Evans and Levinson (2009) and Hulme and Snowling (2009)

provide contemporary discussion of linguistic universals and of modularity in relation to children's development, respectively.

The language acquisition debate has been reinvigorated by the emergence of large language databases in combination with powerful computing resources which have revealed surprisingly rich statistical structure in natural language. Moreover, a now substantial body of research indicates that, from a very young age, the brain can detect these statistical regularities. This appears to occur even under challenging learning conditions (e.g., when only positive evidence is available; when stimuli are presented briefly; when there are irregularities in the input). Key studies from these bodies of research are reviewed in subsequent sections of this paper.

For some, the debate does not center on a distinction between innateness versus learning, but rather on the relative contributions of these (Gould and Marler, 1987; Yang, 2004; Gervain and Mehler, 2010). Yang (2006) suggested that "A somewhat curious response to the mystery of grammar learning is to say that there is basically *no* learning. . .for the unfortunate few who do experience language learning problems, getting a detailed understanding of how language learning takes place is probably well worthwhile" (pp. 150–152). An unresolved question is whether some aspects of language, such as grammatical structure, are less learnable and more heavily underpinned by innate knowledge than others (e.g., Nowak et al., 2002; Peña et al., 2002; Seidenberg et al., 2002). Bayesian models have representational flexibility allowing a move away from some conventional dichotomies that have shaped language acquisition research. For example, Perfors et al. (2011) explored the learning of phrase structure in the context of typical child-directed speech and innate domain-general capacities. Others have suggested that there is a shift from general learning mechanisms to language specific processes across development (Namy, 2012).

STATISTICAL LEARNING

Statistical learning has been described as "automatic," "incidental," and "spontaneous." Perruchet and Pacton (2006) argued that SL is a form of implicit learning in that participants in SL experiments are presented with structured material and are not given any instructed regarding learning; they learn from exposure to positive instances.

Statistical learning of regularities can be assessed in a number of ways. One method is the long-established sequential learning paradigm which utilizes embedded triplets to determine sensitivity to adjacent dependencies. The paradigm can be used with either auditory or visual stimuli. For instance, a child may be asked to watch a continuous sequence of evenly paced individually presented items (represented here by letters). Typically, each item appears for around 400 ms. The sequence contains embedded triplets such as A–P–K; for example, . . . L–A–P–K–G–H–D–A–P–K–X . . .. After several minutes of watching this *familiarization stream*, the experimenter surprises the child with *test phase*: during forced-choice trials two triplets are presented in succession (the three component items of the triplet are displayed individually, one triplet then the other). One of these triplets, the embedded triplet, had been repeated during the continuous sequence while the other, a foil, had never appeared. The child judges which of the triplets is familiar (e.g., APK or AXG?). Most identify the embedded triplets as familiar, even though there was no advance warning of patterns and no reinforcement. Generally, participants have no conscious sense of familiarity. Data are analyzed to determine whether performance is significantly different from chance (using a one-sample *t*-test comparing the group average for percentage of correctly identified embedded triplets against chance, which is 50%).

Studies focusing on infants have utilized this paradigm or similar ones, but require a different kind of responding during the test phase (e.g., headturn preference). A seminal study in *Science* revealed that 8-month-olds can learn the strength of sequential associations between syllables in pseudospeech after only 2 min of exposure (Saffran et al., 1996). Recently, another study demonstrated that 8-month-olds are able to track such transitional probabilities also in *natural language* (Pelucchi et al., 2009). A study using sequences of visually presented shapes showed SL in 2-month-olds (Kirkham et al., 2002).

Many studies of SL have examined the ability to detect associations among adjacent items that are presented sequentially; however, natural language also contains *non-adjacent patterns* (e.g., syntactic structure can involve dependencies among elements that are distant from one another). SL can also operate on non-adjacent patterns (Newport and Aslin, 2004). Recent studies using the event-related potential (ERP) technique have found that the ability to extract statistical dependencies between adjacent elements in the speech stream appears to be present from birth, and that infants can learn non-adjacent dependencies in a natural, non-native language by 4 months of age (Teinonen et al., 2009; Friederici et al., 2011). Some aspects of language processing may require *spatial* rather than sequential learning (e.g., certain aspects of orthography; aspects of sign language). SL has been shown to operate on spatial regularities (e.g., Fiser and Aslin, 2005).

Statistical learning does not decay rapidly. Kim et al. (2009) exposed participants to statistical regularities present in a familiarization stream 24 h before the test phase and showed significant learning despite the delay. Arciuli and Simpson (2012a) replicated this finding. In addition, they demonstrated that SL is remarkably consistent regardless of whether familiarization and test phase are separated by 30 min, 1, 2, 4, or 24 h. Participants still showed significant learning. Neuroscientific evidence has confirmed that SL operates without instruction to learn, and in those who had no conscious sense of familiarity during the test phase (Turk-Browne et al., 2009).

Many SL paradigms, such as the triplet paradigm described above, measure participants' recognition of the exact stimuli that was used during familiarization. SL studies have also included measures of generalization; that is, whether participants can learn regularities from one set or stimuli and subsequently apply their implicit knowledge to stimuli they have never encountered before (Gómez and Gerken, 2000). Generalization indicates that learners have moved beyond recognition of specific items to an understanding of the underlying patterns they represent. Studies of infants have shown that the ability to generalize regularities in language is already present in the first year of life (Marcus et al., 1999; Gerken and Bollt, 2008) and is so robust that infants can

generalize a predominant grammatical pattern even when they are faced with inconsistent input (Gómez and Lakusta, 2004).

Gómez and Lakusta found that 12-month-olds were able to abstract form-based categories in an artificial grammar where only 83% of the training strings represented the correct grammar and the remaining 17% represented a different structure which was inconsistent with the grammar. Furthermore, results from studies using generalization paradigms indicate that variability (e.g., in terms of the number of different exemplars representing a structure) is a key factor which facilitates generalization (Wonnacott et al., 2012). It is still unclear, however, whether this finding carries over to complex natural language settings (van Heugten and Johnson, 2010).

The role of sleep is particularly interesting with regard to the difference between SL paradigms that test recognition and SL paradigms that test generalization. Arciuli and Simpson (2012a) found that adults' recognition of the exact stimuli presented during familiarization was not affected by sleep. This result is consistent with a study by Nemeth et al. (2010) showing no effect of sleep on subjects' ability to learn specific motor sequences in an implicit learning task. However, a different picture emerges from the studies that have investigated subjects' ability to generalize their learning to novel cases. Gómez et al. (2006) compared non-adjacent dependency learning in infants who napped between familiarization and testing to infants who did not sleep. Results showed that the no-nap group preferred listening to familiar over unfamiliar trials, consistent with veridical memory of specific non-adjacent phrases. Infants in the nap group, however, listened longer to sentences conforming to the grammar, but did not distinguish between familiar and unfamiliar items, suggesting that they had abstracted away from particular stimulus items. A follow-up study by Hupbach et al. (2009) found that infants had forgotten specific stimulus sentences 24 h after exposure to the grammar. The abstract information, on the other hand, was retained, but only if a nap had followed shortly after language exposure. Converging evidence from adults comes from a study by Durrant et al. (2011) which showed improved abstraction of statistical patterns underlying tone sequences after a night's sleep or a brief daytime nap when compared to equivalent periods of wakefulness. The above findings suggest that sleep may contribute to abstraction of statistical regularities, perhaps by promoting a qualitative change in memory which enables greater flexibility in learning (Gómez et al., 2006). Such cognitive flexibility is critical in the process of language acquisition as generalization plays a major role in linguistic productivity.

Statistical learning tasks used to study language learning typically employ artificial miniature languages. The advantage of these languages is that they enable the experimenter to constrain the input in such a way that learning can be attributed solely to the use of those cues directly under experimental control. The main problem with these materials is their low ecological validity. The input participants are exposed to typically lacks the complexity of natural language on a number of dimensions (e.g., acoustic variability, number of words, and frequency of repetition). Thus, it is unclear to what degree findings from the SL literature can be applied to language learning "in the wild." A goal for future SL studies of language acquisition will be to simulate the complexity of a natural language task while controlling for pre-existing linguistic knowledge as well as the properties of the input. Ideally, such studies should use paradigms where the listener is exposed to auditory and visual stimuli simultaneously to mimic naturalistic learning conditions. Studies incorporating these qualities are beginning to emerge (Gullberg et al., 2010; Hay et al., 2011; Lew-Williams et al., 2011), but at present we need to base our hypotheses about the relationship between language acquisition and SL on studies that have used artificial miniature languages.

While there has been ever increasing interest in SL for more than a decade, it is only in the last few years that researchers have begun focusing on individual differences in this ability and on demonstrating a direct relationship between SL and performance on other cognitive tasks. There is now mounting evidence suggesting that SL is a distinct ability with meaningful individual differences. Kaufman et al. (2010) found variability in SL (using a serial reaction time task) in 153 adolescents aged 16–18 years that was independent of IQ, working memory, and explicit associative learning. The only elementary cognitive task related to SL was processing speed. Arciuli and Simpson (2011) revealed variability in SL in 183 children aged 5–12 years; a finding that is crucial for the argument that SL relates to variability in language *acquisition.*

STATISTICAL LEARNING AND LANGUAGE: A COMMON NEURAL BASIS

There is a growing body of evidence showing that SL recruits the same brain areas as those used in language processing (de Vries et al., 2011; Folia et al., 2011; Petersson et al., 2012). A number of studies using functional magnetic resonance imaging (fMRI) have found that Broca's area, which is one of the classic language areas, is involved in artificial grammar learning paradigms as well as in the implicit learning of structured motor sequences (Lieberman et al., 2004; Forkstam et al., 2006; Clerget et al., 2012). Corroborating evidence comes from a study using diffusion tensor magnetic resonance imaging (DTI) which found that white matter integrity around Broca's area predicted performance in an artificial grammar learning task (Floeel et al., 2009). Furthermore, a recent ERP study demonstrated similar neural correlates for a sequential learning task and a language task using a within-subject design (Christiansen et al., 2012).

Studies using repetitive transcranial stimulation (rTMS) and transcranial direct current stimulation (tDCS) have taken these findings a step further by demonstrating a causal relationship between activation in Broca's area and learning of artificial grammars (Uddén et al., 2008; de Vries et al., 2010). The study by de Vries et al. (2010) is of special interest because it focused on the grammar *acquisition* process rather than the subsequent syntactic judgment. In this experiment three groups of subjects participated in an artificial grammar learning task: one group who received anodal tDCS over Broca's area, one group who received stimulation over an area which has not been implicated in artificial grammar learning, and one group who received sham stimulation. The group who received stimulation in Broca's area during the *acquisition* of the grammar performed better than the two other groups in the subsequent grammatical classification task. Interestingly, tDCS over Broca's area did not significantly enhance working memory, ruling out increased working memory capacity during acquisition

as the explanation for the group difference. However, the study employed a between-subjects design, and although an effort was made to match the subjects on a number of criteria, pre-existing group differences may have contributed to the observed effect.

Additional evidence supporting a common neural basis for SL and language comes from investigations of patients with agrammatic aphasia. Christiansen et al. (2010) tested seven patients diagnosed with agrammatic aphasia on a visual SL task. In the training phase of the experiment, patients and control participants were exposed to strings of non-linguistic symbols conforming to an artificial grammar. Both patients and controls performed well in the cover task which involved judging whether one grammatical string matched the next. However, in the test phase where subjects were asked to classify novel strings as either grammatical or ungrammatical, only control participants performed better than chance. Differences between patients and controls could not be attributed to poor visual-perceptual skills or low visuo-spatial working memory in the agrammatic patients. Thus, the results suggest that the language impairment in agrammatic aphasia is associated with impairment in non-linguistic sequence learning, indicating that domain-general neural mechanisms underlie both language and SL. Converging evidence comes from a study by Patel et al. (2008) showing that Broca's aphasics display impaired processing of structural relations in musical sequences.

Based on this type of evidence, Uddén and Bahlmann (2012) introduced the structured sequence processing perspective which proposes that there are domain general mechanisms in the brain which are common to the processing of structured sequences in language, music, and action. They reviewed a large number of studies which have consistently shown that the left inferior frontal gyrus is engaged in processing of structured sequences independently of whether these are linguistic, musical, or action-related.

THE ASSOCIATION BETWEEN SL AND PROFICIENCY WITH SPOKEN LANGUAGE

There is growing behavioral evidence of an association between SL and language proficiency. Conway et al. (2010) examined the relationship between SL and word predictability in sentence processing in adults. Experiment 1 revealed a positive relationship between visual SL (sequences of colored squares) and auditory sentence processing. Experiment 2 showed a positive relationship between auditory SL (sequences of syllables embedded in pseudospeech) and audiovisual sentence processing. Experiment 3 demonstrated that this relationship was not mediated by immediate verbal recall (digit span) or non-verbal intelligence (Raven's Progressive Matrices). See Misyak and Christiansen (2012) for an investigation of the link between SL and comprehension of natural language sentences in adults that reported a similar outcome: a relationship between SL and language proficiency that exists independently of cognitive motivation, short-term memory, and fluid intelligence. The findings from these two studies suggest that SL is tapping a distinct capacity.

Consistent with these findings, several studies of language impaired adults have shown poor SL, and that generalization of SL to novel cases appears to represent a particular problem for this population (Plante et al., 2002; Grunow et al., 2006; Richardson et al., 2006; Torkildsen et al., in press). In the study by Grunow et al. (2006) adult subjects with and without language-based learning disabilities listened to strings of three non-words where the first and third word had a dependent relationship. Adults without language impairment were able to learn the non-adjacent contingencies and generalize the underlying structure when variability of the middle element was high (24 unique words), but not when it was low (12 unique words). Adults with language impairment did not show any discrimination between grammatical and ungrammatical strings in either variability condition. Torkildsen et al. (in press) examined the effect of exemplar variability on SL in a simpler learning task, involving adjacent dependencies. Half the learners were exposed to three exemplars of each of the open class elements presented 16 times each (low variability condition), while the other half were exposed 24 exemplars twice (high variability condition). Learners with normal language were able to recognize trained items and generalize the grammar to novel non-word strings in both high and low variability conditions, but relative effect sizes suggested that high variability facilitated learning. In the language impaired group, only those exposed to the high variability condition were able to demonstrate generalization of the grammar. Such evidence has led to the proposal that language impairment may result from a general problem in SL (Hsu and Bishop, 2010; but see Dąbrowska, 2010). However, many studies of adults with language impairment have only examined SL in the verbal domain, making it difficult to disentangle the effects of language impairment and a possible impairment in non-verbal SL.

Examination of the link between individual differences in SL and natural language proficiency is clearly a promising endeavor; however, none of the above studies examined children. To date, only a few studies of children and adolescents have examined the relationship between language proficiency and SL. Tomblin et al. (2007) found that grammar impairments in adolescents were directly associated with low performance on a visual sequential pattern learning task. A recent study by Conway et al. (2011) found that visual sequence learning was significantly correlated with language outcomes in deaf children with cochlear implants. The observed correlations between sequence learning and language were especially robust for a language test measuring the ability to formulate semantically and grammatically correct spoken sentences of increasing length and complexity. The correlation between language and sequence learning was not mediated by either working memory or vocabulary knowledge.

A study by Evans et al. (2009) revealed a link between auditory sequential SL and language proficiency in children aged 6–14 years. They used two tests of SL: (i) syllables in pseudospeech and (ii) sequences of musical tones. Children with SLI performed more poorly than controls on both SL tasks. Children with language impairment did show SL, but required longer exposure to stimuli to learn embedded regularities. After controlling for age, SL during the short exposure condition correlated positively with receptive and expressive vocabulary in typically developing children. After controlling for age, SL during the long exposure condition was positively correlated with receptive vocabulary in children with language impairment. SL was not correlated with IQ in either

group of children. In line with Conway et al., this finding suggests that SL is tapping a type of learning that is not assessed by tests of IQ.

As far as we are aware, the only study examining the relationship between an independent test of SL and syntactic acquisition in typically developing children is that reported by Kidd (2012). In this study, 4–6-year-olds were given tests of explicit word pair learning and implicit visual sequence learning in addition to a syntactic priming task. The syntactic priming task included a test phase where children described pictures after they had been primed with a particular syntactic construction (the passive form) and a post-test phase where children described pictures without having been primed. The post-test phase investigated whether priming effects persevered after priming had ceased. Results showed that performance on the implicit SL task predicted maintenance of the syntactic priming effect into the post-test phase of testing. Scores on the explicit learning task, on the other hand, did not predict priming effects. These findings indicate that children's SL abilities are recruited when learning grammatical usage patterns in input.

The findings reported by Kidd (2012) are consistent with comparable studies of adults such as Conway et al. (2010) and Misyak and Christiansen (2012). However, while the passive form is not typically used by 4–6-years-olds, it is likely that participants in Kidd's experiment came to the experiment with at least some experience with this construction. Thus, an investigation of an entirely novel syntactic construction would be needed to make claims about the role SL plays in children's ability to break into the syntactic system that governs their language.

A natural next step to follow up Kidd's finding is to investigate *how* children make use of the output of SL in the language acquisition process. A recent line of research has set out to examine exactly this question (Graf Estes et al., 2007; Lany and Saffran, 2010, 2011). For example, Graf Estes et al. (2007) asked whether SL during word segmentation yields output that can act as word candidates which can be used in subsequent lexical-semantic acquisition. In the first part of an experiment, 17-month-olds were familiarized with an artificial language where transitional probabilities allowed the segmentation of four words. Next, the infants were taught two novel label-object associations where the labels were either words in the artificial language, sequences that crossed word boundaries in the artificial language (part-words), or words that did not appear in the familiarized language at all (non-words). Graf Estes and colleagues found that infants who had been taught labels that were words in the familiarized speech stream were able to learn the label-object pairings, but infants who were taught part-words or non-words did not demonstrate any learning of the pairings. This result suggests that the output of the SL process can function as input to subsequent word learning.

Mirman et al. (2008) extended this finding by showing that the relationship between statistical segmentation and word learning is also present in adults. However, the authors found a difference between infants and adults in the dynamics between statistical segmentation and word learning. In contrast to infants, who could not learn label-object mappings for part-words or non-words they had not been familiarized with, adults learned words in all three conditions, but were faster in acquiring non-words and familiarized words than part-words. This latter finding suggests that for adults SL has an inhibitory role in hindering the learning of novel meanings for labels that violate learned transitional probabilities (part-words), while for infants SL has a facilitative role in assisting the mapping of labels to novel meanings when labels are consistent with learned transitional probabilities.

Evidence pointing in this direction is not restricted to the area of word learning. A recent study of the acquisition of morphosyntax shows that the non-adjacent dependencies which have the most advantageous distributional patterns are the ones that infants first show evidence of knowing when tested with headturn preference procedures (van Heugten and Johnson, 2010). Thus, there is reason to believe that the output from SL mechanisms is used at various levels of linguistic analysis both by infants and children.

Second-language acquisition (L2) learning is different from first-language (L1) learning in a number of critical ways. Still, it is possible that the detection of statistical regularities plays a role in L2 acquisition. Ellis (2002) argued that both L1 and L2 learning is related to input frequency and its detection and argued that while frequency has been all but ignored in applied linguistics for the last 40 years it may be appropriate to revisit it as a causal factor. Interestingly, a recent study of 153 adolescents demonstrated a significant positive relationship between implicit SL and second language learning of French and German (Kaufman et al., 2010). We do not know of any research that has examined SL in infants living in bilingual environments or any studies that have examined a link between a capacity for SL and proficiency of L2 acquisition; although, it would seem worthwhile to pursue these avenues in future research.

SL IN THE CONTEXT OF WRITTEN LANGUAGE

Both reading and spelling involve learning the correspondences between arbitrary visual symbols and the linguistically meaningful sounds of a language. In English the mapping between letters and sounds can be thought of as probabilistic (e.g., Harm and Seidenberg, 2004; Treiman and Kessler, 2006; Deacon et al., 2008; Kessler, 2009; Seva et al., 2009). For example, the letter "c" often maps onto the phoneme/k/. Of course, "c" can be linked with other phonemes (as in "circle" or "cello"). In the absence of explicit instruction, over time, children are likely to detect contextual cues such as many words beginning with the letter "c" followed directly by the letter "i" have/s/as their initial phoneme. The statistical regularities in written language include non-adjacent pairings (such as "a" later followed by "e": "cape" versus "cap"). Children are taught explicitly about some of these mappings (and rightly so). Clearly, they are not taught about every single correspondence and contextual cue in English. Surely, that would be impossible.

Arciuli has examined probabilistic cues to lexical stress contained within orthography. For example, corpus analyses have revealed that around 70% of disyllabic English words ending with the letters "-ure" have first syllable stress, whereas around 80% of words ending with "-uct" have second syllable stress. Adults are sensitive to these probabilities. They tend to assign first syllable stress when reading a non-word such as "lenture," but second syllable stress when reading "feduct" (see Arciuli and Cupples, 2006, regarding cues in word endings and Arciuli and Cupples, 2007, regarding cues in beginnings). A triangulation of (1) corpus analyses of children's age-appropriate reading materials, (2)

behavioral testing across a range of ages, and (3) computational modeling demonstrated that sensitivity to probabilistic cues to lexical stress during reading aloud follows a developmental trajectory in children across the age range of 5–12 years (Arciuli et al., 2010). As children's exposure to written language increases, sensitivity to these probabilities increases. This sensitivity occurs without having to draw children's attention to the probabilities explicitly.

The computational modeling component of the study by Arciuli et al. (2010) drew on a single-route connectionist approach to reading in order to explore how children learn to assign lexical stress. Connectionist models operate on the statistical regularities present in the input to which they are exposed. In these models learning occurs via adjustment of the weights on connections between units in order to approximate a target response. Gradually, these connection weights are altered in order to increase the accuracy of the model's response. Importantly, connectionist models can be trained iteratively enabling us to explore developmental trajectories based on age-appropriate input. Thus, connectionist models embody the principle of SL. For many years cognitive scientists have contrasted connectionist approaches where a system learns regularities with an alternative approach where predetermined rules are utilized. For example, Rastle and Coltheart (2000) reported on a rule-based algorithm for stress assignment as part of the dual-route cascaded model of reading that was designed to simulate the reading aloud of disyllabic non-words. The algorithm involved searching through the letter string of a non-word for morphemes (to identify a specified set of affixes: 54 prefixes and 101 suffixes), and then consulted a database for information concerning whether each morpheme carried stress or not (e.g., the suffix"-ing" does not carry lexical stress). The algorithm successfully simulated some aspects of stress assignment in adults' reading; however, it was difficult to see how children might come to acquire such a system. How might children learn what constitutes a prefix and what constitutes a suffix? How might children end up with a store of knowledge pertaining to whether affixes carry lexical stress or not? More recent instantiations of the dual-route model of the reading aloud of polysyllables have incorporated connectionist principles (e.g., Perry et al., 2010).

The debate about rules versus statistics and whether some kind of hybrid system might best for explaining language acquisition continues (Newport, 2010). Connectionist modeling has a central role in this debate. For example, connectionist modeling has been used by researchers interested in the so-called "more than one mechanism" (MOM) hypothesis of language acquisition. According to the MOM hypothesis language is acquired via both rule-based and statistical mechanisms. Some researchers have used under performance of a connectionist model in simulating human data as evidence in favor of MOM (Endress and Bonatti, 2007) while others have used connectionist modeling to directly rebuke such claims (Laakso and Calvo, 2011).

Arciuli and Paul (2012) examined sensitivity to probabilistic orthographic cues to lexical stress in adolescents with autism compared with matched typically developing peers (all participants were 13–17 years; groups were matched on age, verbal IQ, spoken language, and reading ability). Using the stimuli and silent reading task from Arciuli and Cupples (2006) they demonstrated that adolescents with autism lack sensitivity to these cues. There was no requirement to produce individual words, so it seems unlikely that motor explanations can account for this finding. They discuss the possibility that some individuals with autism lack the ability to "tune in" to the details of ambient language (Shriberg et al., 2011). Arciuli and Paul suggested that this lack of attunement may be related more generally to impaired SL. An fMRI study by Scott-van Zeeland et al. (2010) revealed a lack of SL during exposure to artificial language containing statistical regularities in individuals with ASD (9–16 years). In contrast, behavioral research has indicated that implicit learning is intact in individuals (8–14 years) with autism (Brown et al., 2010). More research is needed to clarify whether SL is impaired in autism. In keeping with what we know about variability of SL in typically developing individuals (e.g., Arciuli and Simpson, 2011), it seems likely that there is also variability in SL ability in the autism population. This may explain why some group studies find impaired SL in autism while others do not. It is worth noting the suggestion that social cues may enhance children's implicit learning by highlighting *what* it is that is to be learned and *when* it ought to be learned (Meltzoff et al., 2009). It may be that some children with autism are not sensitive to the kinds of social cues that support SL (see also Tomasello, 2010).

Arciuli and Simpson (2012b) examined the relationship between SL and reading aloud in typically developing children and healthy adults. SL was assessed using sequences of visually presented items, a variation of the triplet-learning paradigm. Reading accuracy was assessed using a standardized test of single word reading. This constituted a highly conservative test of the hypothesis that an individual's capacity for SL might be related to their reading proficiency: the SL task used non-linguistic stimuli bearing no particular resemblance to the reading process, while the reading task had not been designed with an emphasis on the probabilistic relationship between letters and sounds. The data revealed a significant positive relationship between SL and reading proficiency in children and also in adults, even after age and attention were taken into consideration. Neither phonological working memory nor non-verbal IQ mediated the relationship between SL and reading ability.

Presumably, a capacity for SL could facilitate the acquisition of written language directly (there are many statistical regularities in written language) as well as indirectly via links with oral language proficiency (it is well known that reading and spelling ability is closely related to oral language ability). We are not aware of any research that has examined whether infants' capacity for SL is related to their proficiency with written language in later years.

THE POTENTIAL OF LONGITUDINAL RESEARCH

Solid progress has been made in supplying the kind of empirical evidence required to demonstrate that SL plays a role in language acquisition. Especially helpful in this regard are recent studies that have shown a link between performance on a test of SL and performance on a test of language proficiency, as well as studies demonstrating how infants and adults use the output of the SL process in subsequent lexical acquisition. We have now reached a point where longitudinal research is needed to assist in furthering the language acquisition debate. Longitudinal studies cannot prove causality, but they are a vital step in exploring the nature of

a relationship once an association between variables has been discovered, and a necessary step before intervention studies targeted at those with impairments can be considered.

While we know of no previous studies which have investigated a direct link between SL and later language outcomes, there are longitudinal studies showing that speech segmentation, phonological discrimination, and non-linguistic auditory processing abilities during the first year of life, abilities which may be associated with SL, predict later language outcomes (Newman et al., 2006; Kuhl et al., 2008; Choudhury and Benasich, 2011). There are also longitudinal studies of toddlers in their second or third years demonstrating that lexical processing skills in meaningful contexts predict later language outcomes. Marchman and Fernald (2008) found that speed of spoken word recognition and vocabulary size at 25 months predicted language skills at 8 years of age. In a more recent study, Fernald and Marchman (2012) extended these findings by showing that word recognition at an even younger age, 18 months, predicted vocabulary growth into the second half of the third year in typically developing and late talking toddlers.

These findings demonstrate that longitudinal research beginning in the first or second year has great potential for investigating the influence of various cognitive abilities on language development. However, such longitudinal studies present a number of challenges. One of these is that there is great variability in infants' ability to successfully complete behavioral and electrophysiological assessments at this age. Since longitudinal studies are costly and time-consuming, many researchers are forced to keep the data collection period as short as possible, over only a year or two.

Moreover, some very early language assessments (at age 18–24 months) have shown poor sensitivity and specificity in predicting language outcomes only a year later. Some 2-year-olds turn out to be "late bloomers," and a fair proportion of others who had age-appropriate language at age 2 meet the criteria for language delay at age 3 (Dale et al., 2003; Henrichs et al., 2011). One option is to begin testing a little later, around 3 years of age, and follow up with subsequent testing of oral and written language proficiency thereafter.

Certainly, longitudinal research will need to investigate SL in relation to acquisition in different linguistic domains (e.g., vocabulary and morphosyntax; oral versus written language) and strive to employ more naturalistic stimulus materials than those which have traditionally been used. Ideally, behavioral studies tapping SL and linguistic knowledge at developmentally significant ages need to be combined with corpus analyses to obtain a realistic picture of the input that children receive. This kind of research can be used in conjunction with computational models and neuroimaging to explore possible mechanisms that give rise to developmental changes in behavior.

Longitudinal investigation of whether early SL ability is related to later language proficiency is an important step toward the design of intervention studies which in turn can be used to examine causality. For example, in the area of SLI it has been explicitly stated that "The extent to which deficits in statistical learning could supplement extant theories, such as deficits in working memory, in the literature of SLI requires further empirical examination. . . this line of research can potentially provide useful information for future development of intervention programs" (Hsu and Bishop, 2010, p. 275). In terms of treatment possibilities, increasing participants' exposure to particular linguistic constructions (such as those in some relative clauses) can make them easier to learn (Wells et al., 2009). Another line of research with clinical relevance are studies that have demonstrated the benefit of high variability for learning morpho-syntactic relations (Gómez, 2002; Gómez and Maye, 2005; Torkildsen et al., in press). These studies indicate that the structure of the learning context can determine whether a particular grammar is learned and generalized. This is an especially relevant finding, given that failure to generalize learning has been identified as a significant problem for those with impaired language. Thus, language impairments associated with inefficient SL might potentially be remediated by focusing on the salience, volume, and/or variability of the input provided to learners. Assessment of SL may also assist early identification of risk/impairment so that other evidence-based interventions can be introduced.

In sum, it has been well established that many infants, children, adolescents, and adults are equipped with highly efficient abilities to detect statistical regularities in input. Recent research has brought the knowledge that humans use the output from these statistical mechanisms in language acquisition and that individual differences in SL are related to language proficiency. Longitudinal studies are needed to determine the extent to which SL contributes to the transition from non-linguistic infant to fully fledged language user in typically developing individuals and the extent to which impaired SL presents challenges for those with disorders such as autism, SLI, and dyslexia.

REFERENCES

Arciuli, J., and Cupples, L. (2006). The processing of lexical stress during visual word recognition: typicality effects and orthographic correlates. *Q. J. Exp. Psychol.* 59, 920–948.

Arciuli, J., and Cupples, L. (2007). "Would you rather 'embert a cudsert' or 'cudsert an embert'? How spelling patterns at the beginning of English bisyllables can cue grammatical category," in *Mental States: Language and Cognitive Structure*, eds A. Schalley and D. Khlentzos (Amsterdam: John Benjamins Publishing), 213–237.

Arciuli, J., Monaghan, P., and Seva, N. (2010). Learning to assign lexical stress during reading aloud: corpus, behavioural and computational investigations. *J. Mem. Lang.* 63, 180–196.

Arciuli, J., and Paul, R. (2012). Sensitivity to probabilistic orthographic cues to lexical stress in adolescent speakers with ASD and typical peers. *Q. J. Exp. Psychol.* 65, 1288–1295.

Arciuli, J., and Simpson, I. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Dev. Sci.* 14, 464–473.

Arciuli, J., and Simpson, I. (2012a). Statistical learning is lasting and consistent over time. *Neurosci. Lett.* 517, 133–135.

Arciuli, J., and Simpson, I. (2012b). Statistical learning is related to reading ability in children and adults. *Cogn. Sci.* 36, 286–304.

Brown, J., Aczel, B., Jiminez, L., Kaufman, S., and Grant, K. (2010). Intact implicit learning in autism spectrum conditions. *Q. J. Exp. Psychol.* 63, 1789–1812.

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. London: Plenum Press.

Choudhury, N., and Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: prediction to 3 and 4 year language and cognitive abilities. *Clin. Neuropsychol.* 122, 320–338.

Christiansen, M. H., Conway, C., and Onnis, L. (2012). Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Lang. Cogn. Process.* 27, 231–256.

Christiansen, M. H., Kelly, M. L., Shillcock, R. C., and Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition* 116, 382–393.

Clerget, E., Poncin, W., Fadiga, L., and Olivier, E. (2012). Role of Broca's area in implicit motor skill learning: evidence from continuous theta-burst magnetic stimulation. *J. Cogn. Neurosci.* 24, 80–92.

Conway, C., Bauernschmidt, A., Huang, S., and Pisoni, D. (2010). Implicit learning in language processing: word predictability is the key. *Cognition* 114, 356–371.

Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., and Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Dev. Sci.* 14, 69–82.

Crain, S. (1991). Language acquisition in the absence of experience. *Behav. Brain Sci.* 14, 597–650.

Dąbrowska, E. (2010). Productivity, proceduralisation and SLI: comment on Hsu and Bishop. *Hum. Dev.* 53, 276–284.

Dale, P. S., Price, T. S., Bishop, D. V., and Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *J. Speech Lang. Hear. Res.* 46, 544–560.

de Vries, M. H., Barth, A. C., Maiworm, S., Knecht, S., Zwitserlood, P., and Flöel, A. (2010). Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *J. Cogn. Neurosci.* 22, 2427–2436.

de Vries, M. H., Christiansen, M. H., and Petersson, K. (2011). Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* 5, 10–35.

Deacon, S. H., Conrad, N., and Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Can. Psychol.* 49, 118–124.

Durrant, S. J., Taylor, C., Cairney, S., and Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia* 49, 1322–1331.

Ellis, N. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acq.* 24, 143–188.

Endress, A., and Bonatti, L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105, 247–299.

Evans, J. L., Saffran, J. R., and Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* 52, 321–335.

Evans, N., and Levinson, S. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–492.

Fernald, A., and Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Dev.* 83, 203–222.

Finn, A. S., and Hudson Kam, C. L. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition* 108, 477–499.

Fiser, J., and Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134, 521–537.

Floeel, A., De Vries, M. H., Scholz, J., Breitenstein, C., and Johansen-Berg, H. (2009). White matter integrity around Broca's area predicts grammar learning success. *Neuroimage* 4, 1974–1981.

Folia, V., Uddén, J., de Vries, M., Forkstam, C., and Petersson, K. M. (2011). Artificial language learning in adults and children. *Lang. Learn.* 60, 188–220.

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., and Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *Neuroimage* 32, 956–967.

Friederici, A. D., Mueller, J., and Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS ONE* 6, e17920. doi:10.1371/journal.pone.0017920

Gerken, L., and Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: implications for generalization mechanisms and constraints. *Lang. Learn. Dev.* 4, 228–248.

Gervain, J., and Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annu. Rev. Psychol.* 61, 191–218.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 5, 431–436.

Gómez, R. L., Bootzin, R. R., and Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychol. Sci.* 8, 670–674.

Gómez, R. L., and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci. (Regul. Ed.)* 4, 178–186.

Gómez, R. L., and Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Dev. Sci.* 7, 567–580.

Gómez, R. L., and Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy* 2, 183–206.

Gould, J. L., and Marler, P. (1987). Learning by instinct. *Sci. Am.* 255, 74–85.

Graf Estes, K. M., Evans, J., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Grunow, H., Spaulding, T. J., Gómez, R. L., and Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *J. Commun. Disord.* 39, 158–170.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., and Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Lang. Learn.* 60, 5–24.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Hay, J. F., Pelucchi, B., Graf Estes, K., and Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cogn. Psychol.* 63, 93–106.

Hay, J. F., and Saffran, J. R. (in press). Rhythmic grouping biases constrain infant statistical learning. *Infancy*. doi: 10.1111/j.1532-7078.2011.00110.x

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W., Hofman, A., Raat, H., Verhulst, F. C., and Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: the generation R study. *J. Speech Lang. Hear. Res.* 54, 854–869.

Hsu, H., and Bishop, D. (2010). Grammatical difficulties in children with specific language impairment: is learning deficient? *Hum. Dev.* 53, 264–277.

Hulme, C., and Snowling, M. (2009). *Developmental Disorders of Language Learning and Cognition.* London: Wiley.

Hupbach, A., Gómez, R. L., Bootzin, R. R., and Nadel, L. (2009). Nap-dependent learning in infants. *Dev. Sci.* 12, 1007–1012.

Janata, P. (2006). Hitting the right note. *Nature* 443, 29–30.

Kaufman, S., DeYoung, C., Gray, J., Jiminez, L., Brown, J., and MacKintosh, N. (2010). Implicit learning as an ability. *Cognition* 116, 321–340.

Kessler, B. (2009). Statistical learning of conditional orthographic correspondences. *Writing Syst. Res.* 1, 19–34.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Dev. Psychol.* 48, 171–184.

Kim, R., Seitz, A., Feenstra, H., and Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neurosci. Lett.* 461, 145–149.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83, B35–B42.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000.

Laakso, A., and Calvo, P. (2011). How many mechanisms are needed to analyse speech? A connectionist simulation of structural rule learning in artificial language acquisition. *Cognition* 35, 1243–1281.

Lany, J., and Saffran, J. R. (2010). From statistics to meaning. *Psychol. Sci.* 21, 284–291.

Lany, J., and Saffran, J. R. (2011). Interactions between statistical and semantic information in infant language development. *Dev. Sci.* 14, 1207–1219.

Lew-Williams, C., Pelucchi, B., and Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Dev. Sci.* 14, 1323–1329.

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S., and Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cogn. Neurosci.* 16, 427–438.

Marchman, V. A., and Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Dev. Sci.* 11, F9–F16.

Marcus, G. F., Vijayan, S., Bandi Rao, S., and Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science* 283, 77–80.

Meltzoff, A. N., Kuhl, P., Movellan, J., and Sejnowski, T. (2009). Foundations for a new science of learning. *Science* 325, 284–288.

Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition* 108, 271–280.

Misyak, J. B., and Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331.

Namy, L. L. (2012). Getting specific: early general mechanisms give rise to domain-specific expertise in word learning. *Lang. Learn. Dev.* 8, 47–60.

Nemeth, D., Janacksek, K., Londe, Z., Ullman, M. T., Howard, D. V., and Howard, J. H. (2010). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Exp. Brain Res.* 201, 351–358.

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., and Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Dev. Psychol.* 42, 643–655.

Newport, E. (2010). Plus or minus 30 years in the language sciences. *Top. Cogn. Sci.* 2, 367–373.

Newport, E. L., and Aslin, R. N. (2004). Learning at a distance. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* 48, 127–162.

Nowak, M. A., Komarova, N. L., and Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature* 417, 611–617.

Patel, A. D., Iversen, J. R., Wassenaar, M., and Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* 22, 776–789.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev.* 80, 674–685.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science* 298, 604–607.

Perfors, A., Tenenbaum, J., and Reiger, T. (2011). The learnability of abstract syntactic principles. *Cognition* 118, 306–338.

Perruchet, P., and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci. (Regul. Ed.)* 10, 233–238.

Perry, C., Ziegler, J. C., and Zorzi, M. (2010). Beyond single syllables: large-scale modelling of reading aloud with the connectionist dual process (CDP++) model. *Cogn. Psychol.* 61, 106–151.

Petersson, K. M., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* 120, 83–95.

Pinker, S. (1989). *Learnability and Cognition: the Acquisition of Argument Structure.* Cambridge: MIT Press.

Plante, E., Gómez, R., and Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *J. Commun. Disord.* 35, 453–462.

Rastle, K., and Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *J. Mem. Lang.* 42, 342–364.

Richardson, J., Harris, L., Plante, E., and Gerken, L. (2006). Subcategory learning in normal and language learning-disabled adults: how much information do they need? *J. Speech Lang. Hear. Res.* 49, 1257–1266.

Ritter, N. A. (2002). Introduction. *Linguist. Rev.* 19, 1–7.

Rowland, C. F., and Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: 'what children do know?' *J. Child Lang.* 27, 157–181.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science* 274, 1926–1928.

Scott-van Zeeland, A., McNealy, K., Wang, A., Sigman, M., Bookheimer, S., and Dapretto, M. (2010). No neural evidence of statistical learning during exposure to artificial languages in children with autism spectrum disorders. *Biol. Psychiatry* 68, 345–351.

Seidenberg, M. S., MacDonald, M. C., and Saffran, J. R. (2002). Does grammar start where statistics stop? *Science* 298, 553–554.

Seva, N., Monaghan, P., and Arciuli, J. (2009). Stressing what is important: orthographic cues and lexical stress assignment. *J. Neurolinguist.* 22, 237–249.

Shriberg, L. D., Paul, R., Black, L. M., and van Santen, J. P. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *J. Autism Dev. Disord.* 41, 405–426.

Teinonen, T., Fellmann, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10, 21. doi:10.1186/1471-2202-10-21

Tomasello, M. (2010). *Origins of Human Communication.* Cambridge, MA: MIT Press.

Tomblin, J. B., Mainela-Arnold, E., and Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Lang. Learn. Dev.* 3, 269–293.

Torkildsen, J. V. K., Dailey, N. S., Aguilar, J. M., Gómez, R., and Plante, E. (in press). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *J. Speech Lang. Hear. Res.*

Treiman, R., and Kessler, B. (2006). Spelling as statistical learning: using consonantal context to spell vowels. *J. Educ. Psychol.* 98, 642–652.

Turk-Browne, N., Scholl, B., Chun, M., and Johnson, M. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* 21, 1934–1945.

Uddén, J., and Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2023–2032.

Uddén, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P., and Petersson, K. M. (2008). The inferior frontal cortex in artificial syntax processing: an rTMS study. *Brain Res.* 1224, 69–78.

van Heugten, M., and Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *J. Mem. Lang.* 63, 197–209.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., and MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cogn. Psychol.* 58, 250–271.

Wonnacott, E., Boyd, J. K., Thomson, J., and Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *J. Mem. Lang.* 66, 458–478.

Yang, C. (2004). Universal grammar, statistics or both? *Trends Cogn. Sci. (Regul. Ed.)* 8, 451–456.

Yang, C. (2006). *The Infinite Gift: How Children Learn and Unlearn the Languages of the World.* New York: Scribner.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Lang. Learn. Dev.* 4, 32–62.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 July 2012; paper pending published: 18 July 2012; accepted: 13 August 2012; published online: 31 August 2012.

*Citation: Arciuli J and Torkildsen JvK (2012) Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data. Front. Psychology **3**:324. doi: 10.3389/fpsyg.2012.00324*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 16 April 2013
doi: 10.3389/fpsyg.2013.00170

Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar

Yasuyo Minagawa-Kawai[1,2]*, Alejandrina Cristia[3], Bria Long[4], Inga Vendelin[5], Yoko Hakuno[1], Michel Dutat[5], Luca Filippin[5], Dominique Cabrol[6] and Emmanuel Dupoux[5]

[1] Graduate School of Human Relations, Keio University, Tokyo, Japan
[2] Institut d'Etudes de la Cognition, Ecole Normale Supérieur, Paris, France
[3] Neurobiology of Language, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
[4] Department of Psychology, Harvard University, Cambridge, MA, USA
[5] Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, ENS-IES, CNRS, Paris, France
[6] AP-HP Cochin Port Royal, Paris, France

Edited by:
Jutta L. Mueller, University of Osnabrueck, Germany

Reviewed by:
Judit Gervain, CNRS – Universite Paris Descartes, France
Silke Telkemeyer, Free University Berlin, Germany

***Correspondence:**
Yasuyo Minagawa-Kawai, Graduate School of Human Relations, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.
e-mail: myasuyo@bea.hi-ho.ne.jp

Each language has a unique set of phonemic categories and phonotactic rules which determine permissible sound sequences in that language. Behavioral research demonstrates that one's native language shapes the perception of both sound categories and sound sequences in adults, and neuroimaging results further indicate that the processing of native phonemes and phonotactics involves a left-dominant perisylvian brain network. Recent work using a novel technique, functional Near InfraRed Spectroscopy (NIRS), has suggested that a left-dominant network becomes evident toward the end of the first year of life as infants process phonemic contrasts. The present research project attempted to assess whether the same pattern would be seen for native phonotactics. We measured brain responses in Japanese- and French-learning infants to two contrasts: Abuna vs. Abna (a phonotactic contrast that is native in French, but not in Japanese) and Abuna vs. Abuuna (a vowel length contrast that is native in Japanese, but not in French). Results did not show a significant response to either contrast in either group, unlike both previous behavioral research on phonotactic processing and NIRS work on phonemic processing. To understand these null results, we performed similar NIRS experiments with Japanese adult participants. These data suggest that the infant null results arise from an interaction of multiple factors, involving the suitability of the experimental paradigm for NIRS measurements and stimulus perceptibility. We discuss the challenges facing this novel technique, particularly focusing on the optimal stimulus presentation which could yield strong enough hemodynamic responses when using the change detection paradigm.

Keywords: near infrared spectroscopy, phonotactics, phoneme perception, infant, speech perception

INTRODUCTION

When listening to speech, the human brain must process various aspects of auditory signals instantly over a series of levels: beginning with the acoustic and phonemic levels, through lexical access, syntactic integration, and up to the level of semantic interpretation. Infants begin to set the foundations of their language-specific knowledge that allows these computations well before they begin to talk (Kuhl, 2011). One of the early landmarks of language acquisition concerns learning the rules that govern sound sequences, or phonotactics, which differ in important ways across languages. For example, whereas English allows for two or more consonants at the beginning, middle, or end of the word, Japanese does not tolerate such consonant clusters. By the age of 9 months, infants show a preference for words that follow the phonotactics of their ambient language (Jusczyk et al., 1993), indicating that by this age they have begun to acquire their native phonological grammar. This phonotactic knowledge affects the formation of abstract sound categories (since infants decide whether two sounds map onto a single or two phonemic categories; White et al., 2008), the extraction of word forms from running speech (as illegal clusters are treated as word boundaries; Friederici and Wessels, 1993; Mattys and Jusczyk, 2001), and the acquisition of word form-meaning associations (because toddlers learn to associate meaning more easily to items with high-frequency phonotactics, compared to low-frequency ones; Graf Estes et al., 2011).

Another example of the impact of phonotactics on perception comes from perceptual repair, the process in which listeners report hearing a legal sequence of sounds even when they had been presented with an illegal sequence. For example, adult Japanese speakers tend to report hearing/abuna/when presented with/abna/, because consonant sequences (clusters) are illegal in Japanese (Dupoux et al., 1999). A recent behavioral study with infants has documented the developmental emergence of this effect (Mazuka et al., 2011): at 8 months of age, Japanese infants were able to discriminate between Abna-type and Abuna-type words, whereas by 14 months they had lost this ability. In contrast, French infants succeeded at both ages. These findings indicate that language-specific phonotactic constraints can affect perception

even before infants have learned to speak, a timeline that coincides with the emergence of native perception for phonemic categories (e.g., Werker and Tees, 1984).

Studies with a variety of neuroimaging methods have only begun to reveal the neurophysiological underpinnings of the development of language networks in the infant brain (Minagawa-Kawai et al., 2008; Gervain et al., 2010; Kuhl, 2011). A particularly fruitful avenue of research combines a change detection paradigm and a hemodynamically based, child-friendly method called Near InfraRed Spectroscopy (NIRS). In *baseline* blocks, infants are presented with a repeated background stimulus (e.g., itta itta itta . . .). In *target* blocks, infants are presented with alternating items (e.g., itta itte itta itte . . .). The contrast between baseline and target blocks is thought to reveal the areas engaged in the discrimination of the two types presented during target blocks. Research using this method reveals that brain activation to phonemic categories becomes left-lateralized toward the end of the first year (e.g., Sato et al., 2003, 2010; Minagawa-Kawai et al., 2007; a recent summary in Minagawa-Kawai et al., 2011a). However, there is no data on the neural network subserving infants' processing of native phonotactics.

In fact, adult fMRI research suggests that there is a considerable overlap between the network recruited for native phonotactics and that involved in native phonemic processing. Jacquemot et al. (2003) presented Japanese and French adults with pseudo-word triplets, which were drawn from three possible types:/abna/(containing a cluster),/abuna/(containing a short vowel), and/abuuna/(containing a long vowel). Some trials contained identical triplets, others contained a contrast between cluster and short vowel, and yet others contained a contrast between short and long vowel. Participants' task was to decide whether or not the last item in the triplet was physically identical to the preceding two. Notice that the duration contrast contained a phonological change for Japanese listeners but not for French adults, whereas the converse was true for the cluster contrast. Results showed that the phonological contrast in one's native language activated the left superior temporal gyrus (STG) and left supra marginal gyrus (SMG) to a greater extent than the non-phonological contrast for both French participants (i.e., /abna, abuna/ > /abuna, abuuna/) and Japanese participants (i.e., /abuna, abuuna/ > /abna, abuna/). Activation in left STG (including the planum temporale) was interpreted to reflect phonological processing, while SMG activation appeared to be related to the task's loading on phonological short-term memory. Notice that, despite tapping phonotactic and phonemic knowledge respectively, the two contrasts activated a similar cerebral network (see also Friedrich and Friederici, 2005; Rossi et al., 2011, for lexical tasks tapping phonotactic knowledge).

In summary, a wealth of behavioral research has shown that both phonotactic and phonemic knowledge emerge toward the end of the first year. Moreover, contrasts between native sound categories come to involve a left-dominant brain network around this age as well. Finally, adult neuroimaging work suggests that there is an overlap between the network processing phonemic and phonotactic dimensions, although the crucial data on this is missing in infancy. The present study thus set out to complete this picture. We used a change detection paradigm similar to those previously used in NIRS research to study the brain network involved in processing two types of contrasts: one relying on phonotactic knowledge and the other on phonemic sound categories. The present investigation is to our knowledge the first cross-linguistic NIRS study, as we tested both Japanese and French infants on their perception of clusters and vowel duration contrasts. In spite of being theoretically well-motivated, however, we forewarn readers that we found very weak evidence of change detection for either contrast or population. Although in the following section we show the cerebral response data to phonological grammar in two different language groups, this paper will mainly discuss the factors likely led to these null results. We are certain that these null results are not due to simple low-level factors (such as a malfunctioning NIRS machine or low stimuli quality), as we assured that our NIRS machines successfully captured Hb responses in the infant brain in our previous work with the same probe pads and basic NIRS paradigm (Japan: Minagawa-Kawai et al., 2007, 2011c, France: Minagawa-Kawai et al., 2011c; Cristia et al., 2013). Other variables considered were: the acoustic salience of the contrasts when embedded in a word, participants, the design of the NIRS probe pads, and the particular experimental paradigm. After careful reflection, we reasoned that one methodological parameter of the stimulus presentation was the most likely cause, and therefore conducted an additional adult NIRS experiment to directly examine this factor. Together with these additional data, we discuss the optimal method to evoke strong enough Hemoglobin (Hb) responses while taking into account the relative limitations of this novel technique. As previous studies have rarely reported null results, the present report will contribute to more efficient experiment planning for infant NIRS studies.

MATERIALS AND METHODS

PARTICIPANTS

There is considerable variability in the exact timeline of the emergence of language-specific effects, sometimes reported as early as 6–8 months or as late as 27 months (a review in Tsuji and Cristia, submitted). Moreover, this emergence is not always stable. For example, discrimination responses to a duration vowel contrast showed W-shaped changes in infants across 3- to 14-months-old (Minagawa-Kawai et al., 2007). Therefore, we tested Japanese infants at a wide range of ages, from 3 to 14 months, in order to explore the stability of neural bases of attunement to the phonological grammar. We also made age groups similar to those of Minagawa-Kawai et al. (2007). Specifically, the following numbers of infants were included in the analyses: 15 within the group of 3–5 months [3–5 m] (9 males; $M = 4{:}5$; range 3:4–5:12); 11 6–7 m (8 males; $M = 7{:}3$; range 6:5–7:28); 15 8–9 m (8 males; $M = 9{:}1$; range 8:4–9:30); 15 10–11 m (12 males; $M = 10{:}29$; range 10:1–11:22); 10 12–14 m (6 males; $M = 13{:}16$; range 12:8–14:27). Thus, to examine the developmental change of neural response to native and non-native phonotactic contrasts, we focused on Japanese infants by measuring them at various ages from 3- to 14-months-old. Furthermore, a previous behavioral study using similar phonotactic stimuli for Japanese and French infants reported the language-specific difference at the age of 14-months-old. Therefore we also tested 14-months-old French infants to contrast to Japanese infants at the same age. Twenty 14-month-olds were included (12 males; age $M = 14{:}3$

range 13:19–14:14). All participants had been born full-term being raised in a largely monolingual home, with no exposure to the other language under analysis here (i.e., no Japanese exposure in French infants; no French exposure in Japanese infants). Japanese infants were tested in Tokyo, and French infants in Paris. In addition, 11 French and 35 Japanese infants participated but their data were excluded from analysis for the following reasons: not enough data (i.e., less than four good trials in each condition, $N = 29$), cried or were fussy ($N = 10$), exposure to the other language being tested here ($N = 2$), and technical error ($N = 5$). Consent forms were obtained from parents before the infants' participation. This study was approved in Japan by the ethic committee of Keio University, Faculty of Letters (No. 09049); and in France through the Ile de France III Ethics Committee (No. ID RCB (AFSSAPS) 2007-A01142-51).

STIMULI

Three types of non-words/abna/,/abuna/and/abuuna/were used as stimulus words. Three tokens of each word (i.e., a total of nine tokens) were chosen from recordings made by a female bilingual speaker of Japanese and French so that the vowels and consonants in the stimuli were good tokens of the category in both languages. These tokens were clearly pronounced in an infant-directed speech fashion and their acoustic details are shown on **Table 1**. As shown in the Table, we selected tokens for/abna/that had no vowel-like waveform between the two consonants. The present procedure used three exemplars for each word to make the phonetic variability higher and thus closer to that of a natural context (Mazuka et al., 2011). The word/abuna/was used as a baseline stimulus, and the other words served as two target conditions as contrast to the baseline: cluster (/abna/) and vowel duration (/abuuna/) conditions. Following the general change detection paradigm widely used in NIRS studies (e.g., Furuya and Mori, 2003; Sato et al., 2003, 2010), we did not use silence period as a baseline. Instead we presented/abuna/between the target blocks. Thus in the stimulus presentation, a baseline block (9–18 s) and a target block (9 s) are alternated so that we could measure Hb change during the target block in contrast to the baseline block. To exclude the systemic vascular effects from the Hb signal, the duration of the baseline block was jittered, while the length of the target block was fixed as in a typical block design paradigm. Participants first heard a baseline block consisting of three variations of/abuna/for 9–18 s with a stimulus onset asynchrony (SOA) of 1.5 s; thus, one baseline block contains 6–12/abuna/tokens. In a cluster target block,/abna/and/abuna/were pseudo-randomly presented for 9 s with the same SOA. Similarly, in a vowel lengthening target block,/abuuna/and/abuna/were pseudo-randomly presented. Baseline blocks (9–18 s) and the target blocks were alternated for at least 16 times (8 times per condition) and a maximum of 30 times. This resulted in a total duration with a minimum of about 6.5 min to a maximum of 11.5 min. The order of the two different target conditions was also presented pseudo-randomly.

PROCEDURES

Infants were tested either in Paris or Tokyo. The actual NIRS machines differed across the two labs (Paris: UCL-NTS, Department of Medical Physics and Bioengineering, UCL, London, UK; Tokyo: ETG-7000, Hitachi Medical Co., Japan) (Everdell et al., 2005). Both systems provide estimates of Hb concentration changes of the optical paths in the brain between the nearest pairs of incident and detection optodes. Both systems emit two wavelengths (approximately 780 and 830 nm for ETG-7000, 670 and 850 nm for UCL-NTS) of continuous near infrared lasers, modulated at different frequencies depending on the channels and the wavelengths, and detected with the sharp frequency filters of lock-in amplifiers (Watanabe et al., 1996). The same probe geometry was used in both labs, which is represented in **Figure 1**. There was one pad over each temporal area, which was placed using anatomical landmarks to align the bottom of the pad with the T3-T5 line in the international 10–20 system (Jasper, 1958). Each pad contained four emission and four detection optodes, arranged in a 2×4 rectangular lattice. These optodes were placed in their respective temporal areas with a source-detector separation length of 25 mm (Watanabe et al., 1996; Yamashita et al., 1996). This separation enables us to measure hemodynamic changes up to 2.5–3 cm deep from the head surface, which traverses the gray matter on the outer surface of the brain (Fukui et al., 2003). Given this geometry, measurements from each hemisphere can be derived from 10 channels

Table 1 | Acoustic information of the stimuli.

Duration (ms)	Abuna					Abna				Abuuna				
	/a/	/b/	/u/	/n/	/a/	/a/	/b/	/n/	/a/	/a/	/b/	/u:/	/n/	/a/
Phoneme	115.4	88.8	128.2	99.9	151.8	112.1	127.6	78.1	140.8	126.3	88.8	369.1	88.3	150.0
SD	2.5	23.4	27.9	4.4	5.3	4.3	3.0	2.6	2.6	5.7	18.5	25.3	10.3	6.2
Word			584.0 (70.2)					458.6 (7.9)				822.4 (18.0)		
Pitch (Hz)	Minimum		187.2 (10.7)			Minimum		190.3 (5.5)		Minimum		191.3 (1.5)		
Range	max		239.3 (20.1)			max		243.0 (2.6)		max		241.3 (1.5)		
Average			213.6 (10.5)					225.3 (14.0)				221.0 (2.0)		

Averaged duration of phonemes and words, pitch range and averaged pitch values for each word type are shown. Values inside parenthesis are standard deviation. No accentuation is assigned to these stimuli.

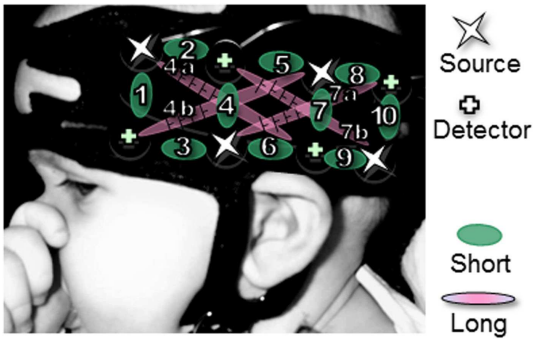


FIGURE 1 | Pad placement. Location of the pads relative to surface landmarks on a typical 5-month-old infant's head. Channels with a long distance (4a,b,7a,b) are only applicable to the French results.

between adjacent sources and detectors, and – only in the UCL system used in Paris – 4 channels between non-adjacent sources and detectors, at a distance of 56 mm. For ease of expression, we will call the former channels "short-distance" (since they are defined by two optodes at a shorter separation) and the latter "long distance."

Once the cap was fit on an infant participant, the stimuli were presented from a loudspeaker. The stimulus sounds played by a PC were presented at an amplitude of about 70 dB SPL, measured at the approximate location of the infant's head sitting on a caregiver's lap in the center of a sound-attenuated booth. To reduce motion artifacts, an experimenter entertained the infant with silent toys during the recording.

DATA ANALYSIS

Artifact detection, baselines, and detrending

Intensity signals were converted into oxygenated hemoglobin (oxy-Hb) and deoxygenated (deoxy-Hb) hemoglobin concentration using the modified Beer-Lambert Law. The state-of-the-art methods in infant NIRS analyses profit from insights that have been gained in more established hemodynamic methods, including the use of General Linear Models (GLM; for example, GLM was applied in the infant studies reported in Telkemeyer et al., 2009; Kotilahti et al., 2010; Minagawa-Kawai et al., 2011a,b). In such state-of-the-art analyses, artifacts are assessed at the level of probes (rather than channels; see e.g., Kotilahti et al., 2010, for a discussion of advantages); and by using the criterion of changes larger than 1.5 mM/mm (millimolars per millimeter) within 100 ms in band-passed (0.02–0.7) filtered total hemoglobin (Pena et al., 2003; Gervain et al., 2008; see Minagawa-Kawai et al., 2011a for discussion). Artifacted stretches of the signal are excluded from the analyses by giving them a weight of zero in the GLM. Additional regressors accounted for baseline changes following major artifacts (through boxcars), and slow non-linear trends (sine and cosine regressors with periods of 2, 3, . . . *n* min, up to the duration of the session).

Activation levels were estimated by assessing the correlation between the signal observed and the signal predicted by convolving the canonical hemodynamic response with boxcars for two experimental regressors, one for each of the two conditions (Cluster, Duration). Individual channels were judged as responding if their degree of correlation was higher than expected by chance, using Monte Carlo bootstrap resampling to correct for multiple comparisons (Westfall and Young, 1993; $N = 10{,}000$). Planned analyses involved entering the beta values from responding channels (and their hemispheric counterparts) into an Analyses of Variance (ANOVA) to assess effects of Condition (Cluster, Duration), Hemisphere (Left, Right), and their interactions, for different age groups. Such an analysis has been used to document the emergence of left-dominant responses to sound changes, and right-dominance to prosodic changes, in Sato et al. (2010). Applying the same analysis to the current study, we predicted that Japanese infants at 12–14 months of age would exhibit greater responses in left channels for the Duration versus the Cluster blocks, whereas both Duration and Cluster blocks would lead to largely bilateral responses in younger Japanese children. In contrast, the French 14-month-olds should exhibit greater responses in left than right cortices for Cluster blocks, but not Duration blocks.

Another way of measuring lateralization involves the calculation of laterality indices, estimated as the difference in activation in left compared to right channels, divided by the total activation. Laterality indices have most frequently been estimated in previous infant NIRS work using the maximum absolute total-Hb observed within a range of channels defined as a Region of Interest (ROI) due to their likelihood of tapping auditory cortices (Furuya and Mori, 2003; Sato et al., 2003; Minagawa-Kawai et al., 2007). To calculate laterality indices, we reconstructed the hemodynamic response function for each infant, channel, and condition by fitting a linear model with 20 one-second boxcar regressors time-shifted by 0, 1, . . ., 20 s respectively from the onset of the target block. We then extracted the value of the maximum absolute total-Hb among the resulting betas for 0–9s within channels 4, 6, and 7 (in the left and right hemisphere), and inputted these values into the formula [L-R]/[L + R]. As with the ANOVA analyses, we expected laterality indices to be significantly above zero only for the Duration blocks in Japanese 12- to 14-month-olds, and for Cluster blocks in French 14-month-olds.

RESULTS

As evident in **Figure 2**, none of the 10 channels measured from Japanese infants (top panel) or the 14 channels measured from French infants (bottom panel) responded significantly to the stimuli. Inspection of these waveforms clearly shows that the low ßwere not due to infants' responses deviating from adults' responses along documented dimensions of variation (such as the width of the response or the extent of the subsequent undershoot; e.g., Handwerker et al., 2004), but rather because of an overall lack of response. That is, when observing individual channels and infants, it was not the case that oxy-Hb levels increased after the onset of the target block while deoxy-Hb decreased during that time. Instead, levels increased or decreased in both oxy-Hb and deoxy-Hb in tandem, or (most frequently) increased and decreased more or less randomly. For reference, **Figure 3** compares the average HRF recovered from the current Japanese and French data with the average HRF recovered from another study (Minagawa-Kawai et al., 2009) using essentially identical equipment and procedure in our respective labs. Clearly, the HRF responses measured in the

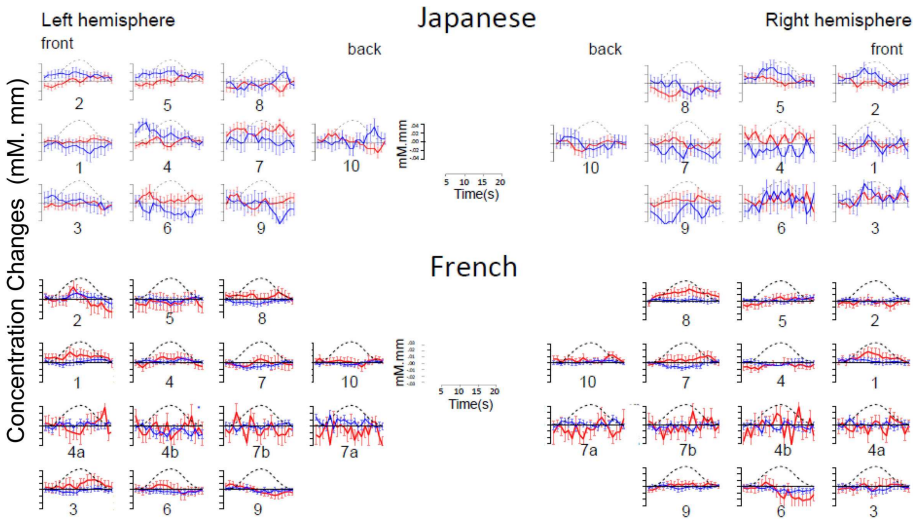


FIGURE 2 | Time course of hemoglobin responses. The top panel shows the time courses of oxy-Hb (red) and deoxy-Hb (blue) in the 10 left and 10 right channels recorded in the Japanese infants (collapsing across all ages, $N = 66$). The bottom panel shows the same in the 14 left and 14 right channels recorded in the French infants ($N = 20$). For both panels, we have collapsed across conditions for ease of inspection (see **Figures A1** and **A2** in Appendix for time courses separating by condition).

other studies were much more pronounced than those recorded in the current study.

Given the low overall responding level, no channels could be included in ANOVAs. Although the laterality index calculations remain theoretically possible, any departure from zero would be rather surprising given the lack of clear hemodynamic response. **Figure 4** shows laterality indices by age group using maximum absolute total-Hb in the calculation. In this Figure, there is a trend for left-dominant activations for 6- to 7-month-old Japanese infants in response to Cluster blocks. A trend for left-dominance appears at 12- to 14-month-olds for Duration. The data from French infants does little to clarify the picture. There are no asymmetrical responses to any of Cluster and Duration blocks. Thus, laterality indices using maximum total-Hb lead to unreliable results in the present study, likely due to the lack of a clear hemodynamic response to the stimuli being tested.

DISCUSSION

The present study was the first to breach the question of the emergence of language-specific responses to phonotactic regularities. In this quest, we adopted a standard paradigm with some minor modifications, and used the same equipment and setup as in previous studies focusing on sound contrasts. Unlike previous studies, no response was apparent to contrasts in vowel duration, or contrasts between a bisyllable with a cluster versus a trisyllable. Furthermore, laterality indices were unreliable and variable, likely due to weak and variable Hb values, with no clear evidence for stable bilaterality early on, eventually replaced by left-dominance in conditions that were language-specific for the infant listeners. Although older Japanese infants showed left-dominant activations to the native Japanese vowel duration contrast in accordance with previous results (Minagawa-Kawai et al., 2007), this result may not be reliable as we did not have a clear Hb response to this contrast (**Figure A1** in Appendix). In the remainder of this Discussion, we raise potential explanations for this null result, and argue that the most likely one takes into account both the low perceptibility of the phonotactic contrasts and the paradigm which is likely suboptimally suited to measure small hemodynamic responses.

Two possible explanations can be ruled out as unconvincing, namely insufficient sample size and inaccurate probe placement. Both in comparison with the general body of previous NIRS work and with published studies using the same method adopted here, sufficiently large numbers of infants were included. A recent systematic review of published infant NIRS research shows that the median number of infants per group in infant NIRS research is 15 (Cristia et al., 2013). More relevant to the current study, Sato et al. (2003) included between six and seven infants in each age group, and the smallest sample sizes per age group included in Minagawa-Kawai et al. (2007) was eight (at 25–28 months; sample sizes were larger for younger age groups: 3–4 months $N = 15$, 6–7 months $N = 14$, 10–11 months $N = 11$; at 13–15 months $2N = 9$). In the present study, the smallest sample size was 10, with as many as 20 infants being included (French 14-month-olds). Pad location is also unlikely to have been a contributing factor, since we have previously been able to register responses using these precise pad locations to a variety of auditory stimuli in previous studies (Minagawa-Kawai et al., 2007, 2009, 2011a,b; Arimitsu et al., 2011). Indeed, while the Paris setup allowed for an even denser sampling through the use of multiple interoptode distances, no clear response to either change was evident in the French data.

One salient difference between the present study and previous NIRS work concerns the position of the contrast under study within words. All previous NIRS work on infants' sound discrimination has made use of bisyllables, with the relevant contrast

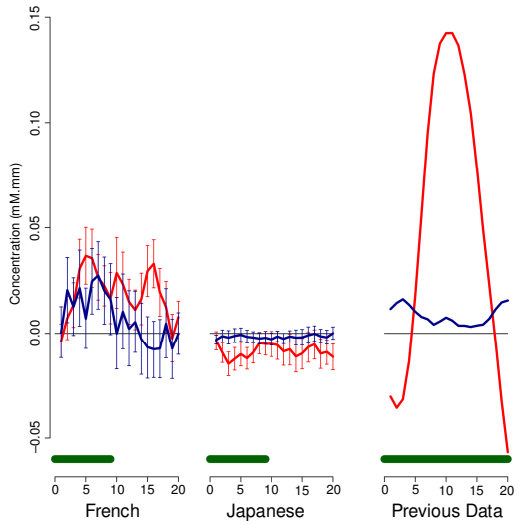


FIGURE 3 | Comparison of average time courses of hemoglobin responses with a previous study. Each panel shows average and standard error for oxy-Hb (red) and deoxy-Hb (blue) within an auditory Region of Interest; the green bar at the bottom indicates the stimulation period. The left panel shows responses to duration block in Japanese infants in the present study (collapsing across ages); the middle panel responses to cluster blocks in French infants in the present study; and the right panel responses to a vowel quality change reported in previous research (Minagawa-Kawai et al., 2009).

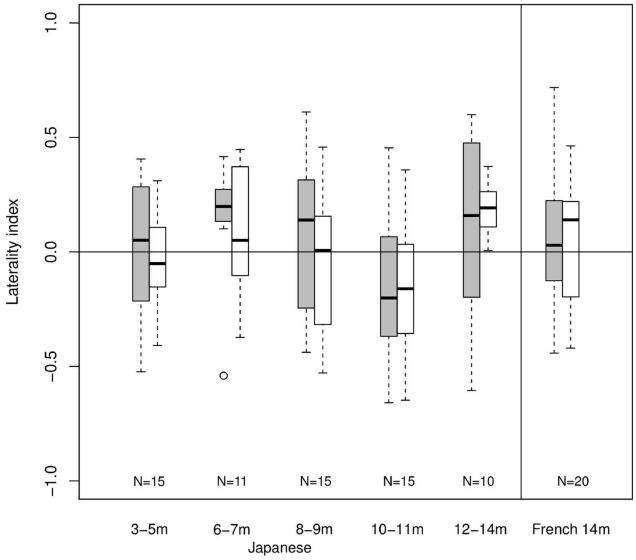


FIGURE 4 | Laterality indices. Each bar indicates mean (black line), first and third quartile (edges of the box), and range (edges of the whiskers) in laterality indices within each age group and condition (gray for cluster, white for duration; *N* shown at the bottom).

occurring in the final syllable (Sato et al., 2003; Minagawa-Kawai et al., 2007; Arimitsu et al., 2011). Our interest was in phonotactics; since word-medial position is where Japanese and French differ the most in terms of the sequences that are tolerated, we embedded both relevant contrasts in a middle syllable. However, some behavioral research in both toddlers (Nazzi and Bertoncini, 2009) and adults (Endress and Mehler, 2010) suggests that the perception of word edges is more accurate than the perception of word middles. By embedding the relevant contrasts in non-salient positions, we might have rendered the task more difficult for infants.

A possible argument against this explanation is that similar stimuli successfully elicited cross-linguistic differences in phonotactic perception patterns using behavioral methods (Mazuka et al., 2011). However, the infants in Mazuka et al. (2011) were actively attending to the sounds in order to control their presentation, whereas in the present study infants were being distracted with silent toys. It is well-known that even if a change is automatically detected, attention can greatly modulate the size of the response (Imaizumi et al., 1998). Although both ERP and fMRI have been effective in detecting cross-linguistic differences in adults of the precise type used here (Dehaene-Lambertz et al., 2000; Jacquemot et al., 2003), participants in those studies were also actively listening to and performing a task with the stimuli. However, distraction alone cannot account for the null result, given that the same procedure has been used in all previous infant NIRS studies that focused on the processing of native sound categories.

Two additional explanations likely played a key role in preventing the reliable measurement of Hb responses, namely insufficient block duration and token variability. Longer stimulation periods (the median is 15 s) are preferred in NIRS experiments as they are thought to increase the likelihood that increases in Hb concentration will accumulate to a point measurable beyond noise. Although event-related paradigms can in principle be used with NIRS (and fMRI), they are extremely rare, making up less than 10% of published infant NIRS studies (Cristia et al., 2013). This problem could be aggravated when using an oddball paradigm, like the one used in the present study, since the baseline period is not defined as the absence of stimulation but only the absence of change. Although one previous infant study has used block durations similar to the ones employed in the present study (Sato et al., 2010, 10 s), their SOA was 1 s, resulting in the presentation of 10 words versus 6 words (SOA = 1.5 s in 9 s block) in our study. Thus, the presentation of stimuli utilized in the present study may have contributed to the weaker response. The second explanation concerns the use of multiple tokens per stimulus word. Traditionally, both behavioral and neuroimaging studies that have used an oddball paradigm to study speech perception have used a single token per category. That is, a single sound recording is presented as the background stimulus, which is repeated over and over. This facilitates the construction of the auditory memory trace as participants can easily process and encode the precise acoustic representation of a single token. As a result, the contrast with between the background stimuli and the single token that represents the oddball becomes more salient. However, since the previous behavioral study showed that Japanese 14-month-olds could detect a vowel deletion change (cluster) with the use of a single token, we chose to use multiple tokens. Unfortunately, this may have ultimately weakened

the automatic change detection response measured in our study. In order to examine the validity of these two explanations, we performed an additional NIRS experiment with Japanese adult participants.

Eight Japanese adult participants (five female; averaged age 35.6) were tested with the Hitachi system. In this study, we varied the procedure in order to investigate two methodological factors that could lead to weak signal-to-noise ratios: (A) the duration of stimulus block, and (B) token variability. Specifically, we compared two target block durations, 9 s (as in the current infant studies) and 15 s (which could allow further response build up). We also compared multiple tokens per stimulus word (as in the current infant studies) against a single token of each word (which should facilitate the discrimination task). Thus, we used the same stimulus and presentation of block design as employed in the infant study, but we manipulated target duration and token variability (in a 2 × 2 factorial design) to gather four sessions in each participant, counterbalanced in order. In these four sessions, the stimuli in the target block were (1) single tokens for 9 s (short-single), (2) single tokens for 15 s (long-single), (3) multiple tokens for 9 s (short-multi) and (4) multiple tokens for 15 s (long-multi). For the baseline block, we applied the same criteria of token variability used in the relevant target block (i.e., single tokens for 1 and 2, multiple tokens for 3 and 4); and the durations were jittered within 9–18 s for the short conditions (1 and 3) and 15–22 s for the long conditions (2 and 4). Each target block was presented 4–5 times, for a total session duration of 4–5 min. The experimental procedure differed from that of infants in the following ways: (1) adult participants paid attention to the stimuli without any distractions (e.g., toys); and (2) the distance between the optical probes was increased to 30 mm to take into account the difference in scalp and skull thickness of adults. Because our aim was to compare the response amplitude between the four conditions, we focused on the maximum Hb change from four channels corresponding to the vicinity of auditory regions, which typically show auditory evoked responses (Minagawa-Kawai et al., 2008). Furthermore, we only analyzed the data for the duration target blocks, containing the change "abuna-abuuna," since the cluster target blocks should not elicit any strong responses due to the lack of consonant clusters in Japanese phonology.

The grand average of Hb time course is indicated in **Figure 5**. Clearly, the "long-single" condition elicited the largest and clear response among the four conditions. In contrast, "long-multi" and "short-single" evoked a weaker Hb response, with the weakest activation for "short-multi," precisely the combination we implemented with infants. To confirm this tendency in **Figure 5**, an ANOVA with duration and token variability as two within subject factors was performed using Z-scores obtained from the GLM analysis. Results support the tendencies observed, showing main effects of these two factors [duration $F(1,31) = 5.84$, $p = 0.046$; token variability $F(1,31) = 18.67$, $p = 0.004$][1]. These

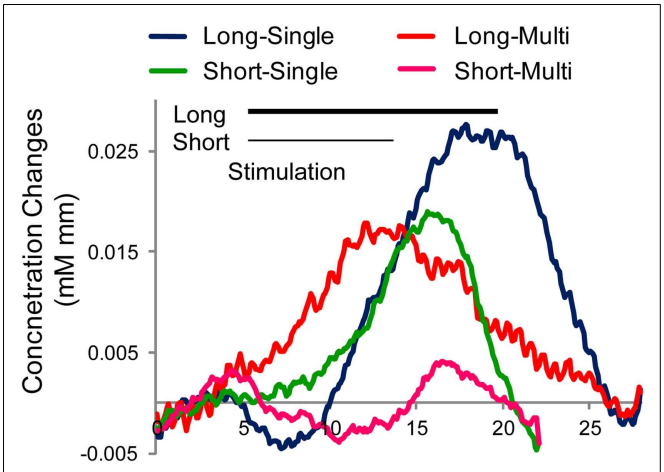


FIGURE 5 | Averaged time course of hemoglobin (oxy-Hb) responses around left auditory area by condition in Japanese adults ($N = 8$). Four stimulus conditions differ in target duration (Long or Short) and token variation (Single or Multiple tokens). Stimulation period is indicated by black line for the long stimulation (15 s) and gray line for the short stimulation (9 s).

results confirmed our predictions that shorter target block duration and greater token variability negatively affect the amplitude of the evoked Hb response, which could constitute key factors in the infant study reported above. Moreover, the adult participants were paying attention to the stimuli, whereas infants were being distracted with silent toys, a factor that, as noted, may have further decreased the evoked responses.

Therefore, we can conclude that the present null results probably result from an interaction of multiple factors with overly short block duration and overly large token variability playing a main role, in addition to the other factors noted above (long SOA leading to few tokens in the target block, attention being drawn away from the stimuli, low salience of the change when embedded in trisyllabic words). Overall, these infant and adult results delineate some limitations of NIRS as a technique to measure infant perception. It has previously been pointed out that less repetition is required for NIRS measurements, and that this should be one of the advantages of NIRS over electro-encephalography (EEG) (Imaizumi et al., 1998; Furuya and Mori, 2003; Minagawa-Kawai et al., 2008). This intuition emerges from the fact that NIRS relies on a vascular response, which should be relatively stronger and more stable than a fast, event-related electrophysiological response. However, it would behoove NIRS researchers not be overly optimistic, and to bear in mind that the physical saliency of the stimuli, the choice of the experimental paradigm, and the stimulus presentation parameters are in fact crucially important and must be carefully selected to allow the vascular response to occur and the event-related Hb response to build up. This is particularly important for

[1]A reviewer suggested that analyses methods could also play a role, in that simple averaging could yield stronger signal-to-noise ratios than the ones we reported. We investigated this possibility in the adult data by also calculating averaged Hb within time windows of 7–17 s after the stimulus onset for the long stimulation and 7–14 s for the short one. Although we had similar results to those from GLM [duration $F(1,31) = 4.93$, $p = 0.061$; token variability $F(1,31) = 18.90$, $p = 0.003$], the F-values were not much larger using averaging.

infant studies, where stable attention to the stimuli is less likely realistic.

This series of studies with both infants and adults have revealed a vulnerability of the change detection paradigm, frequently employed in the NIRS literature (Furuya and Mori, 2003; Sato et al., 2003, 2010; Minagawa-Kawai et al., 2007, 2011a). This paradigm has been widely used to assess the discrimination of phonemic and prosodic categories in both adults and infants (for a review, see Minagawa-Kawai et al., 2011a). Both previous adult fMRI work (Jacquemot et al., 2003) and infant behavioral research (Mazuka et al., 2011) that focused on the processing of the same kinds of contrasts studied here employed multiple tokens as stimuli. However, as we confirmed in a separate study, using multiple tokens in the NIRS-based change detection paradigm reduced the Hb signals. Thus, when all the evidence is taken into account, it appears that the change detection paradigm implemented in NIRS is a less sensitive index of discrimination abilities in infants than behavioral measures. While token variability reduced the observed NIRS responses, it did not prevent young Japanese infants or French toddlers from discriminating the exact same kinds of contrasts (Mazuka et al., 2011). Our follow-up adult study suggested that this noxious effect did not completely eliminate Hb responses, as indicated by weak but reliable response to the multiple stimuli condition in adults. This further suggests a significant role for attention during the change detection procedure. As a final remark, it should be pointed out that the change detection procedure which presents alternating and non-alternating stimuli is still a robust paradigm for various types of stimuli in NIRS studies (Minagawa-Kawai et al., 2011a). What we suggest here is that there is an optimal method to elicit strong Hb responses. We hope that this knowledge may strengthen future infant studies using NIRS.

CONCLUSION

In summary, the present study sought to shed light on how the infant brain comes to code native phonotactics, and compare the resulting network with that found for native sound categories. Unfortunately, we were unable to observe a discrimination response for either phonotactics or a duration contrast that had been used in previous NIRS research. We have argued that the most likely explanations for the null result relate to an unfortunate combination of short target blocks, low stimulus perceptibility due to the use of multiple tokens and target position in the stimulus words, and low signal-to-noise ratio due to the lack of a task involving the stimuli. Future work may be wise to avoid such an outcome by carefully choosing the experimental parameters to obtain a strong enough hemodynamic response by varying stimulus saliency and/or directing infants' attention to the stimuli.

ACKNOWLEDGMENTS

This work was supported in part by Grant-in-Aid for Scientific Research (KAKENHI) (B) (Project No.24118508) and Grant-in-Aid for Scientific Research on Innovative Areas (Project No.24300105) to Yasuyo Minagawa-Kawai, a grant from the Agence Nationale pour la Recherche (ANR-09-BLAN-0327 SOCODEV), and a grant from the Foundation de France to Emmanuel Dupoux, a fellowship from the Ecole de Neurosciences de Paris to Emmanuel Dupoux and Alejandrina Cristia, and a grant from the Fyssen Foundation to Alejandrina Cristia. We thank Yutaka Sato for his help with the previous NIRS data, and Sayaka Ishii, and Noriko Morisawa for helping with NIRS experiment. We gratefully acknowledge Sylvie Margules for carrying out the infant recruitment and Anne Bachmann for designing and constructing the probe pad.

REFERENCES

Arimitsu, T., Uchida-Ota, M., Yagihashi, T., Kojima, S., Watanabe, S., Hokuto, I., et al. (2011). Functional hemispheric specialization in processing phonemic and prosodic auditory changes in neonates. *Front. Psychol.* 2:202. doi:10.3389/fpsyg.2011.00202

Cristia, A., Dupoux, E., Hakuno, Y., Lloyd-Fox, S., Schuetze, M., Kivits, J., et al. (2013). An online database of infant functional Near InfraRed Spectroscopy studies: a community-augmented systematic review. *PLoS ONE* 8:e58906. doi:10.1371/journal.pone.0058906

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1568–1578.

Endress, A. D., and Mehler, J. (2010). Perceptual constraints in phonotactic learning. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 235–250.

Everdell, N. L., Gibson, P., Tullis, I. D., Vaithianathan, T., Hebden, J. C., and Delpy, D. T. (2005). A frequency multiplexed near-infrared topography system for imaging functional activation in the brain. *Rev. Sci. Instrm.* 73, 093705–093735.

Friederici, A. D., and Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Percept. Psychophys.* 54, 287–295.

Friedrich, M., and Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and nonsense words in picture contexts. *J. Cogn. Neurosci.* 17, 1785–1802.

Fukui, Y., Ajichi, Y., and Okada, E. (2003). Monte Carlo prediction of near-infrared light propagation in realistic adult and neonatal head models. *Appl. Opt.* 42, 2881–2887.

Furuya, I., and Mori, K. (2003). Cerebral lateralization in spoken language processing measured by multi-channel near-infrared spectroscopy (NIRS). *No To Shinkei* 55, 226–231.

Gervain, J., Macagno, F., Cogoi, S., Pena, M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14222–14227.

Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., et al. (2010). Near-infrared spectroscopy: a report from the McDonnell infant methodology consortium. *Dev. Cogn. Neurosci* 1, 22–46.

Graf Estes, K., Edwards, J., and Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy* 16, 180–197.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651.

Imaizumi, S., Mori, K., Kiritani, S., Hosoi, H., and Tonoike, M. (1998). Task-dependent laterality for cue decoding during spoken language processing. *Neuroreport* 9, 899–903.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 23, 9541–9546.

Jasper, H. H. (1958). The ten–twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 367–380.

Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687.

Kotilahti, K., Nissila, I., Nasi, T., Lipiainen, L., Noponen, T., Merilainen, P., et al. (2010). Hemodynamic responses to speech and music in newborn infants. *Hum. Brain Mapp.* 31, 595–603.

Kuhl, P. K. (2011). Early language learning and literacy: neuroscience implications for education. *Mind Brain Educ.* 5, 128–142.

Mattys, S. L., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121.

Mazuka, R., Cao, Y., Dupoux, E., and Christophe, A. (2011). The development of a phonological illusion: a cross-linguistic study with Japanese and French infants. *Dev. Sci.* 14, 693–699.

Minagawa-Kawai, Y., Cristia, A., and Dupoux, E. (2011a). Cerebral lateralization and early speech acquisition: a developmental scenario. *Dev. Cogn. Neurosci.* 1, 217–232.

Minagawa-Kawai, Y., Cristia, A., Vendelin, I., Cabrol, D., and Dupoux, E. (2011b). Assessing signal-driven mechanisms in neonates: brain responses to temporally and spectrally different sounds. *Front. Psychol.* 2:135. doi:10.3389/fpsyg.2011.00135

Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., and Dupoux, E. (2011c). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cereb. Cortex* 21, 254–261.

Minagawa-Kawai, Y., Mori, K., Hebden, J. C., and Dupoux, E. (2008). Optical imaging of infants' neurocognitive development: recent advances and perspectives. *Dev. Neurobiol.* 68, 712–728.

Minagawa-Kawai, Y., Mori, K., Naoi, N., and Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *J. Neurosci.* 27, 315–321.

Minagawa-Kawai, Y., Naoi, N., and Kojima, S. (2009). *New Approach to Functional Neuroimaging: Near Infrared Spectroscopy*. Tokyo: Keio University Press.

Nazzi, T., and Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. *Lang. Speech* 52, 463–480.

Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11702–11705.

Rossi, S., Jurgenson, I. B., Hanulikova, A., Telkemeyer, S., Wartenburger, I., and Obrig, H. (2011). Implicit processing of phonotactic cues: evidence from electrophysiological and vascular responses. *J. Cogn. Neurosci.* 23, 1752–1764.

Sato, Y., Mori, K., Furuya, I., Hayashi, R., Minagawa-Kawai, Y., and Koizumi, T. (2003). Developmental changes in cerebral lateralization to spoken language in infants: measured by near-infrared spectroscopy. *Jpn. J. Logoped Phoniatr.* 44, 165–171.

Sato, Y., Sogabe, Y., and Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *J. Cogn. Neurosci.* 22, 2503–2513.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* 29, 14726–14733.

Watanabe, E., Yamashita, Y., Maki, A., Ito, Y., and Koizumi, H. (1996). Non-invasive functional mapping with multi-channel near infrared spectroscopic topography in humans. *Neurosci. Lett.* 205, 41–44.

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 25, 121–133.

White, K. S., Peperkamp, S., Kirk, C., and Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition* 107, 238–265.

Yamashita, Y., Maki, A., and Koizumi, H. (1996). Near-infrared topographic measurement system: imaging of absorbers localized in a scattering medium. *Rev. Sci. Instrum.* 67, 730–732.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 October 2012; accepted: 20 March 2013; published online: 16 April 2013.

*Citation: Minagawa-Kawai Y, Cristia A, Long B, Vendelin I, Hakuno Y, Dutat M, Filippin L, Cabrol D and Dupoux E (2013) Insights on NIRS sensitivity from a cross-linguistic study on the emergence of phonological grammar. Front. Psychol. **4**:170. doi: 10.3389/fpsyg.2013.00170*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

APPENDIX

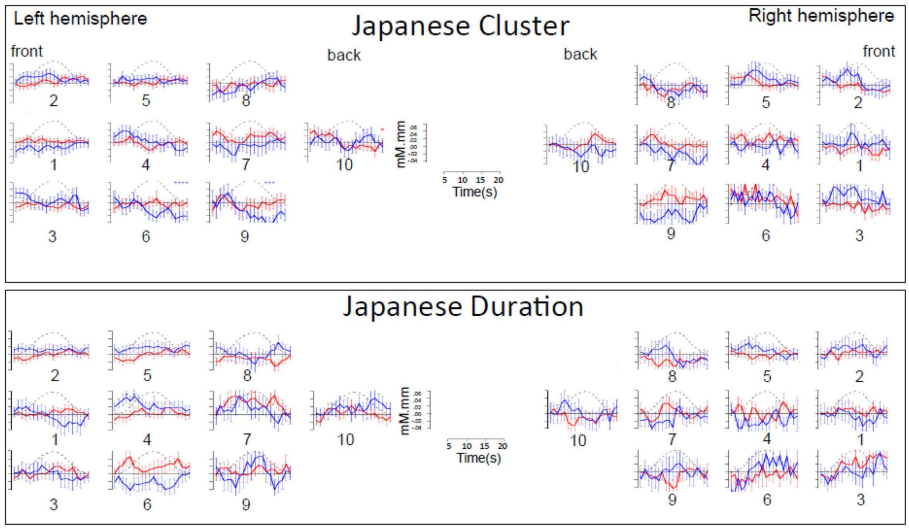


FIGURE A1 | Time course of hemoglobin responses by condition in Japanese infants (*N* = 66). The top panel shows the time courses of oxy-Hb (red) and deoxy-Hb (blue) in the 10 left and 10 right channels recorded in the Japanese infants (collapsing across all ages) during Cluster blocks. The bottom panel shows the same for Duration blocks.

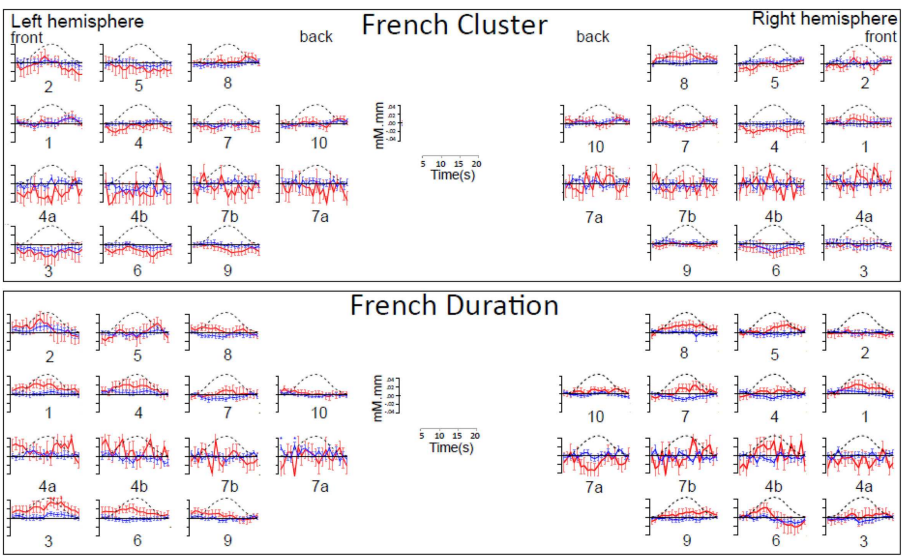


FIGURE A2 | Time course of hemoglobin responses by condition in French infants (*N* = 20). The top panel shows the time courses of oxy-Hb (red) and deoxy-Hb (blue) in the 14 left and 14 right channels recorded in the French infants during Cluster blocks. The bottom panel shows the same for Duration blocks.

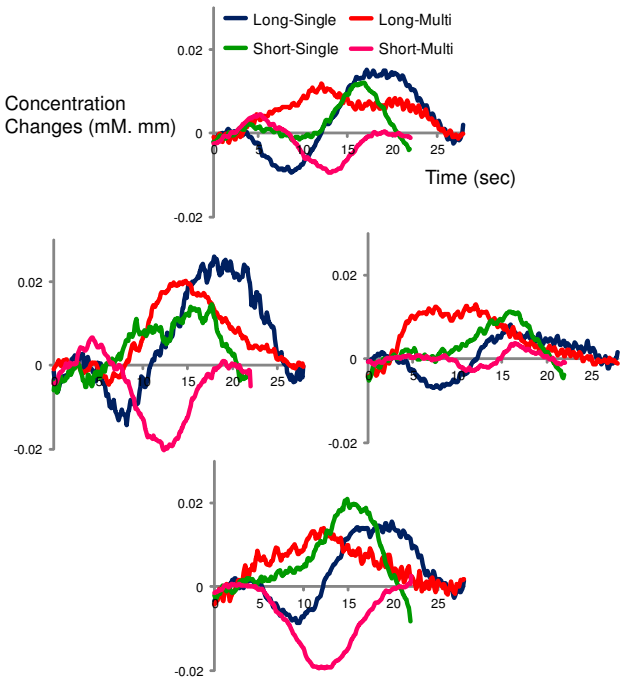


FIGURE A3 | Time course of oxy-Hb responses by condition in Japanese adults ($N = 8$) in four channels (CH4, 5, 6 and 7 in Figure 1) which correspond to vicinity of auditory area on the left side.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 08 February 2013
doi: 10.3389/fpsyg.2013.00025

Predictive brain signals of linguistic development

Valesca Kooijman[1], Caroline Junge[2], Elizabeth K. Johnson[3], Peter Hagoort[4,5] and Anne Cutler[4,5,6]*

[1] *Food and Biobased Research, Wageningen University and Research Centre, Wageningen, Netherlands*
[2] *Department of Psychology, University of Amsterdam, Amsterdam, Netherlands*
[3] *Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada*
[4] *Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands*
[5] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands*
[6] *MARCS Institute, University of Western Sydney, Penrith, NSW, Australia*

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
April A. Benasich, Rutgers University, USA
Leher Singh, National University of Singapore, Singapore

****Correspondence:***
Anne Cutler, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, Netherlands.
e-mail: anne.cutler@mpi.nl

The ability to extract word forms from continuous speech is a prerequisite for constructing a vocabulary and emerges in the first year of life. Electrophysiological (ERP) studies of speech segmentation by 9- to 12-month-old listeners in several languages have found a left-localized negativity linked to word onset as a marker of word detection. We report an ERP study showing significant evidence of speech segmentation in Dutch-learning 7-month-olds. In contrast to the left-localized negative effect reported with older infants, the observed overall mean effect had a positive polarity. Inspection of individual results revealed two participant sub-groups: a majority showing a positive-going response, and a minority showing the left negativity observed in older age groups. We retested participants at age three, on vocabulary comprehension and word and sentence production. On every test, children who at 7 months had shown the negativity associated with segmentation of words from speech outperformed those who had produced positive-going brain responses to the same input. The earlier that infants show the left-localized brain responses typically indicating detection of words in speech, the better their early childhood language skills.

Keywords: infant speech perception, speech segmentation, language skill development, vocabulary size, brain development, brain polarity, ERPs

INTRODUCTION

Spoken language is one of the dimensions of the infant's environment for which perceptual information is available, processed, and stored even before birth (DeCasper et al., 1994). Accordingly, the first year of an infant's life sees steady continuous growth in the skills required to turn a speech signal into a comprehended message (Saffran et al., 2006). Although the first spoken words may be produced only at the end of that year, the perceptual skills that make such production possible develop steadily from birth onward.

This development is not simply a passive result of maturation. The infant's task is to acquire the environmental language(s), and thus to attend to meaningful perceptual variation where it is required to differentiate relevant contrasts (and accordingly to ignore variation that is perceptually detectable, but irrelevant to this particular language). Differences between languages and acoustically salient differences within a language induce differences in the speed and the order with which this phonological task is achieved (Narayan et al., 2010).

One of the most important skills an infant must acquire is the ability to segment speech, i.e., to recognize a word form even though it is embedded in a speech context that may be completely novel. Since speech input to infants consists mainly of multi-word utterances (Van de Weijer, 1999), segmentation is a vital prerequisite of initial vocabulary construction, and infants indeed display segmentation skills before they command a workable vocabulary. This was demonstrated by Jusczyk and Aslin (1995), using a two-phase Headturn Preference Procedure (HPP). Infants were first familiarized with words in isolation, then tested with short texts which did or did not contain the familiarized words. Infants listened longer to the texts containing the target words than to the control texts, showing that they could indeed distinguish between the two – in other words, that they had detected the target words although they were embedded in continuous speech.

Phonological differences between languages also affect the relative appearance of segmentation abilities in the two-phase HPP. English and Dutch are very closely related languages, but evidence of segmentation skills has been seen earlier in English than in Dutch HPP studies, and this difference is ascribed to the relative salience of the cues involved. Across languages, cues to word segmentation can be derived from characteristic rhythmic patterns, and both adults and infants exploit this correspondence in parsing speech (Cutler, 1994). In English, rhythm is stress-based and essentially reduces to the distinction between strong syllables with full vowels and weak syllables with reduced vowels (Fear et al., 1995). This makes for an easy and salient distinction, and English-acquiring infants show segmentation skills in HPP for monosyllabic or bisyllabic initially stressed nouns at least by 7.5 months (Jusczyk and Aslin, 1995; Jusczyk et al., 1999). In Dutch, rhythm is also stress-based, but the strong-weak distinction is less salient than in English, since vowels in unstressed syllables are less frequently reduced (Van der Hulst, 1984). Dutch babies segment speech successfully at 9 months, but fail to show segmentation at

7.5 months (Kuijpers et al., 1998; note that this study was a direct Dutch replication of Jusczyk et al., 1999).

There is also variation across individuals. This variation is related to later language development, as Newman et al. (2006) discovered. They collected vocabulary scores at 2 years for children who had taken part in various HPP segmentation experiments in their first year, and selected from the extensive group the 15% with the largest vocabularies (on average 646 words) and the 15% with the smallest vocabularies (on average 73 words). Members of the former group were significantly more likely as infants to have shown a segmentation effect, in line with the group pattern, than their age mates who now had lower vocabulary scores. Results from experiments that did not involve segmentation, for instance discriminating between languages, were unrelated to later vocabulary size. Newman et al.'s finding is important, as it was the first to underscore the close relationship between being able to segment words from speech and being able to store words in a vocabulary.

Although Newman et al. (2006) drew their conclusion from a comparison of the two outer ends of a large vocabulary size distribution, Singh et al. (2012) showed that the same relationship held across a group of 40 individuals. Singh et al. tested infants at 7.5 months with a simple segmentation task (as used by Jusczyk and Aslin, 1995) and a more complex segmentation task in which the familiarization stimuli could differ from the test stimuli in pitch, then tracked their vocabulary growth to age two. At an individual level, recognition scores (the difference in listening time across trials with familiarized versus unfamiliarized input) on each task correlated significantly with vocabulary size at 24 months, with the more complex task showing stronger correlation. The better the 7.5-month-olds' segmentation skills, the more words they knew at age two.

Recent findings further suggest that a link between early speech segmentation ability and later vocabulary size also holds for preterms: although as a group they do not demonstrate similar evidence of segmentation skill compared to full term 8-month-olds (matched for gestation), those who show similar behavioral responses have higher productive vocabulary at 12 and at 18 months (Bosch, 2011).

In the HPP, the duration of an infant's behavioral response (a headturn to keep listening to an audio input) provides evidence that familiarized words have been detected and thus that segmentation has happened. This is a reliable indicator of segmentation, but it is not a direct view of the segmentation in process. It became possible to track segmentation as it happens, however, once a version of the two-phase segmentation experiment was developed that was suitable for use with measurement of event-related potentials (ERPs). In an ERP study, brain responses time-locked to onset of a familiarized word can be compared with responses to a control word that was not heard before. Kooijman et al. (2005) devised such a method; they tested Dutch 10-month-olds, using low-frequency Dutch bisyllabic words of the kind that Kuijpers et al. (1998) had used in their Dutch HPP study. Familiarization with 10 occurrences of the same word (e.g., *monnik* "monk"; see **Table 1**) in isolation produced a response that became steadily more negative. After familiarization with a word, the infants heard eight sentences, four containing the familiarized word and four a matched control word. Infant brain responses keyed to the onset of familiarized target words were significantly negative in amplitude relative to the responses to the unfamiliarized control words; that is, this difference in the infant brain responses as the spoken sentences were being heard was here the measure showing that a familiar word had been detected.

Table 1 | An example of an experimental trial in the ERP study, with English glosses.

Familiarization	Ten tokens of either ***monnik*** or ***sultan***
Test	*De monnik wiedt zijn tuintje dagelijks*
	"The monk weeds his garden every day"
	De strenge sultan regeert met straffe hand
	"The strict sultan rules with an iron hand"
	De sultan bestuurt het kleine landje
	"The sultan administers the little country"
	Pieter ziet de vriendelijke monnik in het hofje
	"Peter sees the friendly monk in the almshouse"
	Volgend jaar komt de jonge sultan naar Nederland
	"Next year the young sultan is coming to The Netherlands"
	Omar geeft de vriendelijke sultan nog een sigaar
	"Omar gives the friendly sultan another cigar"
	Elke week plukt de jonge monnik verse appels
	"Every week the young monk picks fresh apples"
	De strenge monnik draagt een zware habijt
	"The strict monk wears a heavy habit"

The experimental words are underlined in the sentences; the word that was heard in familiarization was deemed the familiar word, its pair was then the unfamiliar control.

Subsequent studies confirmed that the stress-based segmentation underlying the HPP results also drove the negative-going ERP segmentation response (Kooijman et al., 2009), and showed significant evidence of segmentation by some 10-month-olds even without prior familiarization: presented first with a sentence such as *De strenge monnik draagt een zware habijt* "The strict monk wears a heavy habit," these infants then produced the negative-going recognition response to *monnik* presented later in isolation (in comparison to a control word that had not been part of the preceding sentence; Junge et al., 2012).

Further ERP research on speech segmentation also showed more negative-going brain responses for familiarized words relative to unfamiliar control words in (older) infants acquiring other languages. A negative familiarity effect was observed in 12-month-olds acquiring European French (Goyet et al., 2010; this study used familiarization with isolated words and a test phase of target words in passages, as in Kooijman et al., 2005). The same effect was observed in German 12-month-olds, in a study using familiarization with words within passages and test with isolated words (Männel and Friederici, 2010).

Just as the HPP segmentation response is related to later vocabulary development, so is the ERP segmentation response. Of the 28 infants tested by Junge et al. (2012), 18 showed the ability to achieve segmentation without prior familiarization, while 10 did not. In line with Newman et al.'s (2006) and Singh et al.'s (2012) evidence from HPP studies, a *post hoc* analysis of Junge et al.'s (2012) ERP data showed a relationship between vocabulary size at 12 months and the presence of this segmentation ability at 10 months. A median split was applied to vocabulary measures collected at 12 months via parental questionnaires, yielding a group with larger receptive vocabularies at that age (mean 146 items; range 71–264) and a group with smaller vocabularies (mean 40, range 0–68). In the sentence familiarization task, the former group showed a significant negative recognition response; the latter group did not. In a condition where one isolated word was presented both in familiarization and test, so that word segmentation abilities were not required, each group showed evidence of word recognition. Thus the online ERP measure offers insight into individual differences in success at early word recognition tasks requiring speech segmentation, and how these differences relate to language learning in general.

The ERP studies described so far have shown segmentation at 10–12 months, but HPP studies have shown segmentation to occur earlier, at 7.5 or 8 months (Jusczyk and Aslin, 1995; Polka and Sundara, 2012). The online ERP measure, requiring no behavioral response from infants, may hence allow a more direct reflection of Dutch infants' segmentation capacities, at an earlier age than so far demonstrated with the HPP. However, the literature on infant ERPs shows that responses are quite likely to vary as a function of age. For example, early responses can manifest with different polarity from responses later in life. Kudo et al. (2011) report a positive-going response indicating segmentation of a sequence of tones by neonates, where the same sequences had produced detection negativities in adults (Abla et al., 2008). Männel and Friederici (2010) found that 6-month-old German-learners showed a positivity in a familiarization condition that required segmentation ability, while in 12-month-olds the same condition elicited a clear negative response. Likewise, in an ERP study of phonetic discrimination responses Garcia-Sierra et al. (2011) found that infants acquiring both English and Spanish tended at 6–9 months to show a positive-going response to phonetically deviant stimuli, whereas at 10–12 months the same stimuli elicited negative-going responses. Indeed, in the original Kooijman et al. (2005) ERP study, not all participants showed the negative-going recognition response that constituted the average result. A minority showed, instead, a positive-going response to the target words at test (Junge, 2011).

Polarity differences across age groups in infancy can simply reflect differing relations of a constantly placed reference electrode to a test electrode on a very small versus a larger skull. They can also arise from maturation effects; ERP maturation from birth to the first birthday shows an overall pattern in which the generators responsible for positive amplitudes mature earlier (in the first 6 months) than those responsible for negative amplitudes (from 6 months on; Kushnerenko et al., 2002). In both cases, it is unlikely that observed polarity differences in ERPs to speech signals relate systematically to underlying cognitive processes. In contrast, a third possibility could be that polarity differences reflect differences in relative task demands or in auditory processing (Rivera-Gaxiola et al., 2005b). We will return to this issue in the discussion.

In this paper we report an ERP study of word segmentation from continuous speech by Dutch infants at 7 months. This is a particularly interesting age given that American English learners can segment speech in HPP studies at 7.5 months (Jusczyk and Aslin, 1995; Jusczyk et al., 1999) and their abilities at that age are related to their later vocabulary size (Singh et al., 2012), while Dutch learners at that same age do not demonstrate segmentation ability in HPP (Kuijpers et al., 1998). The ERP paradigm, though, provides a more sensitive view of learners' early responses to language input. We report detailed analysis of ERP patterns associated with segmentation in our study with 7-month-olds, and assessment of the subsequent language abilities of the same participants at 3 years. From this we conclude that early ERP patterns indexing speech segmentation ability directly predict later patterns of language skills.

ERPs AT 7 MONTHS

PARTICIPANTS

Twenty-eight 7-month-old infants from Dutch monolingual families participated (mean age = 7.05 months; age range = 6.11 −7.19 months; 13 female). Twenty-two additional infants were tested, but excluded from data analyses because of fussiness or sleepiness. All infants were reported to have normal development and hearing, and no major problems during pregnancy or birth. All infants were full term, bar one who had been 3.6 weeks premature. There were no neurological or language problems in the immediate families. The parents signed a consent form and received 20 euro for participation.

STIMULI AND DESIGN

We used the same stimuli and design as in Kooijman et al. (2005). Forty low-frequency bisyllabic initially stressed nouns were selected from the CELEX Dutch lexical database (Baayen et al., 1993); examples are *monnik* "monk," *sultan* "sultan." A set of four sentences was constructed for each noun. The nouns were arranged in pairs, with noun position in the sentences, and words preceding the noun, matched across pairs; **Table 1** shows an example noun pair with corresponding sentences. The stimuli (all the sentences, and 10 isolated tokens of each noun) were recorded in a sound-attenuating booth by a female speaker of Dutch in a lively child-directed manner, and sampled to disk at 16 kHz mono. The mean duration of the nouns was 710 ms for the isolated words (range: 373–1269 ms) and 721 ms for the target words in the sentences (range 224–1046 ms). The sentences had a mean duration of 4082 ms (range: 2697–5839 ms).

The experiment contained 20 experimental familiarization + test trials (for an example see **Table 1**), each with 10 tokens of a target noun (familiarization), followed by eight randomized sentences (test). Four of the test sentences contained the word just familiarized (familiarized target words); four contained the unfamiliar noun paired with it (unfamiliar control words). There were four presentation lists, counterbalancing familiarization set (half of the target words were used for familiarization in Lists A and B,

the other half in Lists C and D) and Order of presentation (Lists B and D were as A and C, but with the trials ordered inversely). Each list was heard by seven infants.

PROCEDURE

Infants were seated in a child seat in a sound-attenuating test booth and listened to the stimuli via three loudspeakers situated to the front. Also in front of the infants, a computer screen showed a moving screensaver, not synchronized with the stimuli, and the infants could additionally play with a small silent toy. A parent sat next to each child and listened to a masking CD through closed-ear headphones. Breaks were taken when necessary. Familiarization and test blocks were presented until an infant became too distracted to continue. The experiment lasted on average 32 min; mean block length was 1.6 min, with 2.5 s silence between isolated words and 4.2 s silence between sentences. Subjects heard at least eight blocks (mean: 13, range: 8–20).

EEG RECORDINGS

Electroencephalogram (EEG) measurement was via infant-size Brain-Caps with 27 Ag/AgCl sintered ring electrodes. Twenty-one electrodes were placed according to the American Electroencephalographic Society 10% standard system (midline: Fz, FCz, Cz, Pz, Oz; frontal: F7, F8, F3, F4; fronto-temporal: FT7, FT8; fronto-central: FC3, FC4; central: C3, C4: centro-parietal: CP3, CP4; parietal: P3, P4; and occipital: PO7, PO8). Six electrodes were placed bilaterally on non-standard positions: a temporal pair (LT and RT) at 33% of the interaural distance lateral to Cz, a temporo-parietal pair (LTP and RTP) at 30% of the interaural distance lateral to Cz and 13% of the inion-nasion distance posterior to Cz, and a parietal pair (LP and RP) midway between LTP/RTP and PO7/PO8.

The left mastoid served as online reference for all electrodes. EEG electrodes were referenced to the left mastoid online and re-referenced offline to linked mastoids. Vertical eye movements and blinks were monitored via a supra- to sub-orbital bipolar montage, and horizontal eye movements via a right-to-left canthal bipolar montage. Two occipital electrodes (PO7, PO8) and the midline electrodes Fz, FCz, Cz, Pz, Oz were excluded from analysis either due to excessive artifact (mainly the parietal and occipital electrodes, because the infant's back of the head rested against the child seat) or due to poor cap fit (for some of our subjects we could not get good recordings from FCz and Cz, because all electrodes were bundled together above Cz, creating too much space between the fronto-central electrodes and the skull). Impedances at the remaining electrodes were around 10 kΩ. A BrainAmp DC EEG amplifier recorded EEG and EOG data using a band pass of 0.1–30 Hz and a sample rate of 200 Hz. Excess slow wave activity can often obscure ERP effects in young infants (Weber et al., 2004); to remove it, we filtered the EEG signal offline to 1–30 Hz before further analysis.

Offline, individual trials were aligned 200 ms before acoustic onset of the target words, and screened for artifact from −200 to 800 ms. We rejected trials when amplitude on any electrode channel exceeded ±150 μV or when clear correlations with the eye channels were observed. This resulted in rejection rates of 55.6 and 62.5% of the trials time-locked to the isolated words or to the target words in the sentences, respectively; these are similar rejection rates as in Kooijman et al. (2005). Infants contributed on average 11.4 (SD 3.0) artifact-free trials for the familiarization phase and 19.6 (SD 7.0) for the test phase.

STATISTICAL ANALYSES

We examined the role of word familiarity for the familiarization phase (comparing ERPs for the first two isolated tokens ("unfamiliar") versus the last two isolated tokens of the target noun ("familiarized") and for the test phase (comparing ERPs to the four familiarized target versus the four unfamiliar control words within sentence context). For each condition for each subject, average waveforms were calculated in the −200 to 800 ms window. For illustration purposes, we averaged for each condition the subject average waveforms into grand average waveforms. The number of trials used in each grand average waveform was respectively 332 and 309 for the unfamiliar and familiarized isolated words, and 549 and 548 for the unfamiliar control and familiarized target words in the sentences. Time windows for statistical analyses were chosen based on visual inspection of the data.

Repeated measures analyses of variance were performed for the chosen time windows with Familiarity (two: Familiar; Unfamiliar), Quadrant (four: Left Frontal; Right Frontal; Left Posterior; Right Posterior), and Electrode (five per quadrant; Left Frontal: F7, F3, FT7, FC3, C3; Right Frontal: F8, F4, FT8, FC4, C4; Left Posterior: LT, LTP, CP3, LP, P3; Right Posterior: RT, RTP, CP4, RP, P4) as within-subject variables. The Huynh–Feldt epsilon correction was used for all tests. The original degrees of freedom as well as the adjusted p-values are reported. The onsets of the effects were tested by performing t-tests on the difference waveforms in bins of 50 ms with a 40 ms overlap (i.e., 0–50, 10–60 etc), with significance from zero ($p < 0.05$) on five consecutive bins taken as evidence for onset.

RESULTS: ISOLATED WORDS

The isolated words allow assessment of sensitivity to repetition. We averaged the EEG to token 1 and 2 of the familiarization phase, representing the ERP response to the most unfamiliar isolated words, and the EEG to token 9 and 10, representing the ERP response to the most familiar of the isolated words because by then eight tokens of the same word had already been heard. A difference between these two averages signals an infant's recognition of the repetition. The ERPs to these unfamiliar versus familiarized isolated tokens indeed seem to differ in two time windows, as **Figure 1** shows. First, there is one early peak from 40 to 20 ms that is more negative to the familiarized than to the unfamiliar tokens over a subset of electrodes (FC3, FC4, LT, CP3). Second, familiarized isolated words elicited again a more negative ERP than unfamiliar isolated words in the 200–500 ms time window, mainly over frontal electrodes. This is in the same time window, and with similar distribution and polarity, as the familiarity effect for isolated words reported for the older age group (Kooijman et al., 2005). We analyzed the mean amplitudes in these time windows.

The first time window, the N1, did not show significant differences ($F_{1,27} = 2.43$, $p = 0.13$; no significant interactions with Familiarity). We then examined the same time window (200–500 ms) as in Kooijman et al. (2005) for the familiarization phase. There was an effect of Familiarity that narrowly missed significance

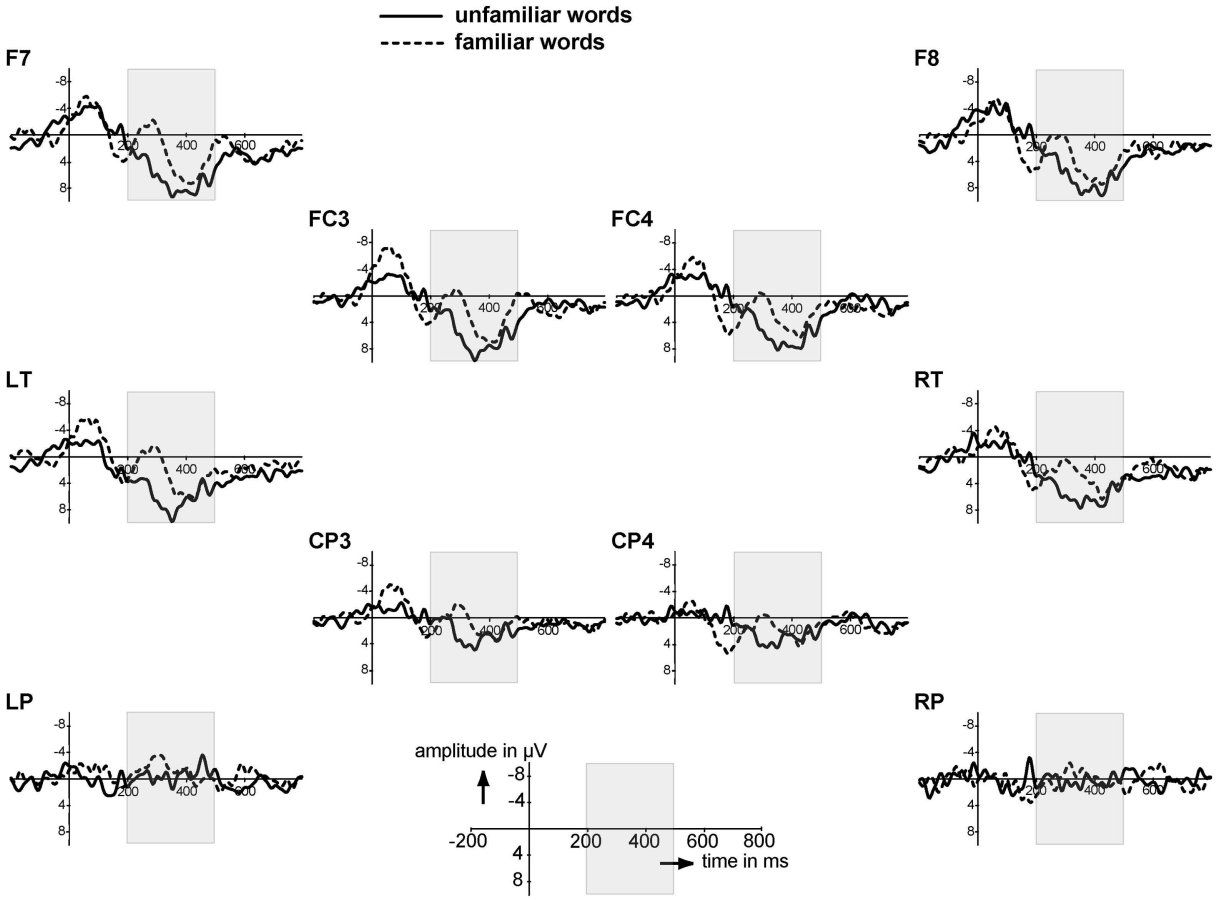


FIGURE 1 | Event-Related Brain Potentials to the unfamiliar (word position 1 & 2) and familiar (word position 9 &10) isolated words on a subset of electrodes; negativity is plotted upwards. Electrodes are laid out as they are on the scalp. The gray area indicates the time window of 200–500 ms.

($F_{1,27} = 3.39$, $p = 0.077$), and a significant interaction of Familiarity with Quadrant ($F_{3,81} = 2.74$, $p = 0.05$). Analyses per quadrant revealed a main effect of Familiarity over the left frontal quadrant only ($F_{1,27} = 5.94$, $p = 0.02$); the right frontal and the posterior quadrants showed no significant effects ($p > 0.10$). Thus, the broad negative ERP effect to the familiar isolated words is strongest over the left frontal area. Onset analyses (see Statistical Analyses) revealed an onset starting at 220 ms for the left frontal electrodes F7 and FT7.

These ERP results thus show a brain response to the repetition of tokens of the same word starting at 220 ms. This familiarity response is similar in polarity and in distribution to that found by Kooijman et al. (2005), but starts 60 ms later; 10-month-olds in that study showed a Familiarity response starting at 160 ms. Like the 10-month-olds, however, the present 7-month-old listeners can recognize repetition of the same form in isolation, a prerequisite for being able to detect repetition of the same form in a speech context.

RESULTS: SENTENCES

Figure 2 shows that the ERPs to the familiarized target and unfamiliar control words in the sentences deviate from each other in two ways. First, familiarized target words elicit a more positive ERP than unfamiliar control words over the frontal areas from 350 to 450 ms, and second, they elicit a more negative ERP than unfamiliar control words over the left posterior area starting at about 430–530 ms. We performed statistical analyses over the mean amplitudes in these time windows.

A significant interaction of Familiarity × Quadrant ($F_{3,81} = 4.05$, $p = 0.018$) was observed for the 350–450 ms window, but there was no main effect of Familiarity ($F_{1,27} < 1$). Analyses per quadrant showed a narrowly missed significant effect of Familiarity over the right frontal quadrant ($F_{1,27} = 3.70$, $p = 0.065$), suggesting a more restricted location of the effect within this quadrant. Further analyses over a subset of four electrodes (F4, F8, FC4, and FT8) in that quadrant indeed revealed a significant main effect of Condition ($F_{1,27} = 4.28$, $p = 0.048$). There were no significant effects in equivalent analyses for the remaining three quadrants. Seventeen participants showed a positive effect on right frontal electrodes. Thus, the early effect of Familiarity is strongest over the right frontal brain area and has a positive polarity. Onset tests revealed a significant effect ($p < 0.05$) at 300 ms for electrode FT8.

In the later time window (430–530) statistical analyses show no significant main effect of Familiarity ($F_{1,27} < 1$) and no interaction

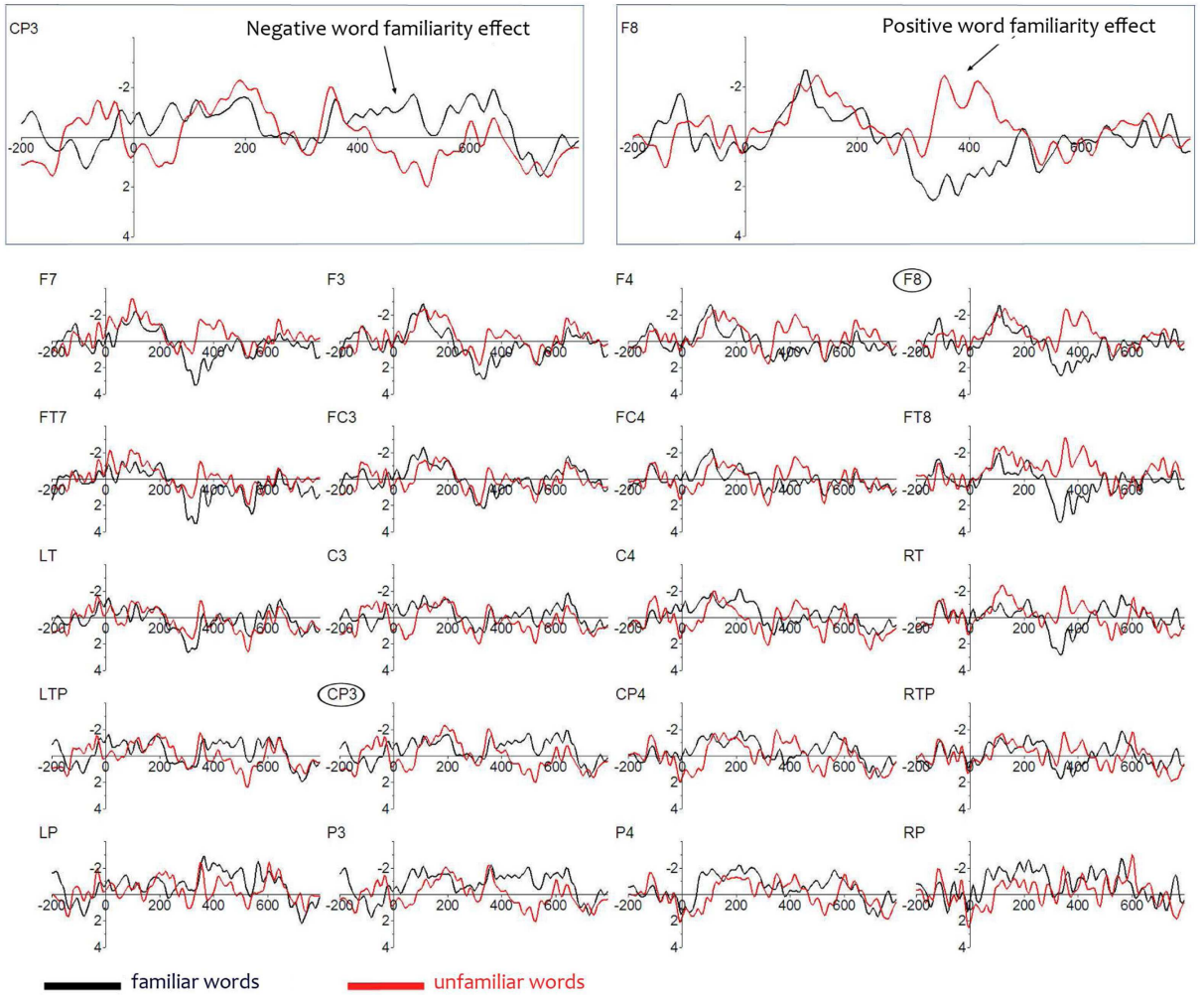


FIGURE 2 | Event-Related Brain Potentials on lateral electrodes to the familiarized target and unfamiliar control words in the sentences; negativity is plotted upwards. Electrodes are laid out as they are on the scalp. Enlarged are a left centro-parietal electrode (CP3) illustrating the later negative familiarity effect, and a right frontal electrode (F8), illustrating the earlier positive familiarity effect.

between Quadrant and Familiarity ($F_{3,81} = 2.31$, $p = 0.10$). Visual inspection of the grand average waveforms reveals that in this window the effect is restricted to electrodes over the left hemisphere at the posterior sites LTP, CP3, and P3. An analysis over only these three left posterior electrodes revealed a significant effect of Familiarity ($F_{1,27} = 4.24$, $p = 0.049$; 14 participants showed this effect). In sum, we observe in the test phase two rather localized effects: a positive right frontal effect and a negative left posterior effect.

These two effects could be equally present in all children, such that the same children who show a positive right frontal effect are also the ones who show a negative left posterior effect. Another possibility could be that there are two subpopulations: some infants show a positive frontal effect yet others a negative left-going effect.

To examine whether the two familiarity effects in the test phase come from distinct or from the same populations, we calculated the correlation between these two effects (i.e., the average difference in amplitude from the four right frontal electrodes in the early time window with the average difference in amplitude from the three left posterior electrodes in the later time window). A (significant) positive correlation would be evidence of two subpopulations, whereas a negative correlation would indicate that the positive and the negative familiarity effects would be nearly simultaneously present within the same population. Indeed, there was a significant positive relationship [$r(28) = +0.41$, $p = 0.03$], suggesting that the two effects are not driven by the same participants: those with an early positive familiarity effect continue to have a positive familiarity effect, and those with a later negative familiarity effect did not have an earlier positive effect. This could also explain why we do not find a significant effect on left fronto-temporal electrodes, which was the site at which the familiarity effects for 10-month-olds were observed (Kooijman et al., 2005; Junge et al., 2012): the different polarities of the familiarity effect on left frontal electrodes for each sub-group would cancel each other out in a grand average.

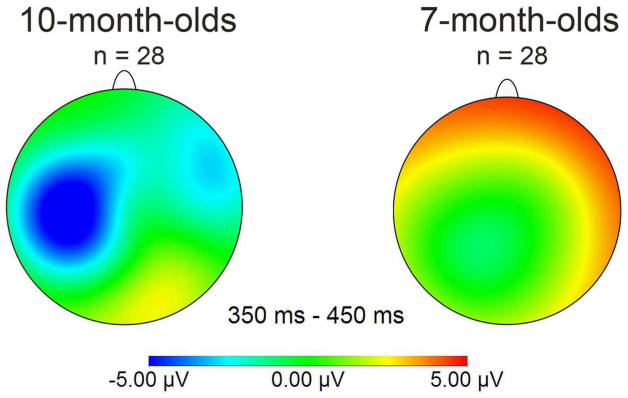


FIGURE 3 | Mean distribution plots for the ERP effect of familiarity (familiarized target – unfamiliar control words) in the 350–450 ms time window for overall group performance of 10-month-olds (left) and 7-month-olds (right).

Together, this suggests that our Dutch 7-month-old participants fall into two separate sub-groups, each showing evidence of being able to detect words previously heard in isolation when they re-occur in continuous speech. Note that word segmentation skill is here demonstrated in Dutch infants at an age at which behavioral evidence of segmentation is not available (Kuijpers et al., 1998)[1]. A majority of 7-month-olds demonstrated being able to segment words by showing a positive familiarity effect on right frontal electrodes. However, as **Figure 3** shows, this effect differs in polarity (positive instead of negative) as well as in distribution (on right frontal instead of on left electrodes), compared to other studies reporting word familiarity effects indexing word segmentation skill in 10-month-olds (Kooijman et al., 2005; Junge et al., 2012).

Nevertheless, the two age groups both show a negative familiarity effect for the familiarization phase, during which the infants were not required to segment words from speech. Moreover, one sub-group among the present 7-month-olds also showed a negative familiarity effect when speech segmentation skill was required. This makes it unlikely that brain maturation underlies this polarity difference observed between the 7- and 10-month-olds, which was only present for the continuous speech condition. We will return to this issue in the general discussion. In the following section we first examine whether the polarity differences in our participant population are related to later language development.

LANGUAGE SKILLS AT 3 YEARS

PARTICIPANTS

Of the 28 participants in the ERP experiment, two could no longer be reached and the parents of a further three declined to participate; 23 children (82%) thus returned for further testing. These children (all right-handed; 11 girls) were now on average 36.3 months of age (range 28.4–46.6 months).

We first examined whether this subset of 23 participants continued to show an overall negative familiarity effect for isolated words in the 200–500 ms time window (familiarization phase), yet an overall positive familiarity effect for the words within speech in the 350–450 ms time window (test phase). Analyses revealed again a significant negative familiarity effect for the familiarization phase ($F_{1,22} = 5.61$, $p = 0.027$), which was most pronounced over frontal electrodes (mean difference over frontal electrodes $-3.71\,\mu$V, SD $= 6.2$). For the test phase, which required infants to segment words from speech, there was again no main effect of Familiarity ($F_{1,22} < 1$), but the interaction between Familiarity and Quadrant was significant ($F_{3,66} = 5.17$, $p < 0.01$). The familiarity effect is significant over the whole right frontal quadrant ($F_{1,22} = 4.36$, $p < 0.05$) and has a positive polarity (mean $+ 2.49\,\mu$V, SD $= 5.7$). Hence, even with a smaller sample we see a negative familiarity effect for the familiarization phase yet a positive one for the test phase. The subset of 23 children is thus representative of the full sample.

We then looked for polarity differences in their 7-month-old ERP results concerning the speech segmentation condition. We focused on results from this phase, because it is here that we observe polarity differences, not only between seven- versus 10-month-olds, but also within the 7-month-olds. Note moreover that Junge et al. (2012) only observed links between infant ERP measures of word recognition and later language development when infants had to first segment words from speech, not when they heard them first in isolation. In particular we inspected the polarity of each participant's familiarity effect on left frontal electrodes, because it was on those electrodes that the familiarity effect was clearly present in 10-month-olds (Kooijman et al., 2005) and even turned out to be predictive of later vocabulary development in another sample of 10-month-olds (Junge et al., 2012). Moreover, as speculated in the previous section, a possible reason why we do not find any significant effect on left frontal electrodes for the 7-month-old overall analysis is that it is here that the two sub-groups overlap with their familiarity response (with reversed polarities), thereby canceling each other out. On this basis we identified two groups: nine "Negative responders" (three girls), with a negative-going ERP response resembling that found on average in both 10-month-old studies, and 14 "Positive responders," whose response was positive-going as in the grand average of the ERP study. When we re-examined the time window 350–450 ms for the 23 subjects in the test phase, with Group as between-subjects variable, we observed, besides the significant interaction of Familiarity × Quadrant ($F_{3,63} = 5.53$, $p = 0.003$), two interactions with the factor Group: a significant Familiarity × Group ($F_{1,21} = 24.3$, $p < 0.001$), and a near-significant three-way Familiarity × Quadrant × Group ($F_{3,63} = 2.67$, $p = 0.06$). This shows that the two groups not only differ in polarity of the familiarity response, but also in the distribution of the effect. For the negative responders the familiarity effect had a negative polarity and was only significant in the left frontal and left posterior quadrants ($F_{1,8} = 13.0$, $p < 0.01$; $F_{1,8} = 13.4$, $p < 0.01$), whereas for the positive responders the familiarity effect had a positive polarity and

[1]An ERP study uses more stimuli than an HPP study, and they are arranged differently (in ERP as in **Table 1**; in HPP, typically familiarization with two words, test with 6-sentence texts in which all sentences contain an instance of one of these words). Thus there were some differences between the present materials and those of Kuijpers et al.'s (1998) HPP study. When the present materials were adapted and tested in an HPP study with 7-month-olds, however, a null result was again observed (see Kooijman et al., 2008, for further detail).

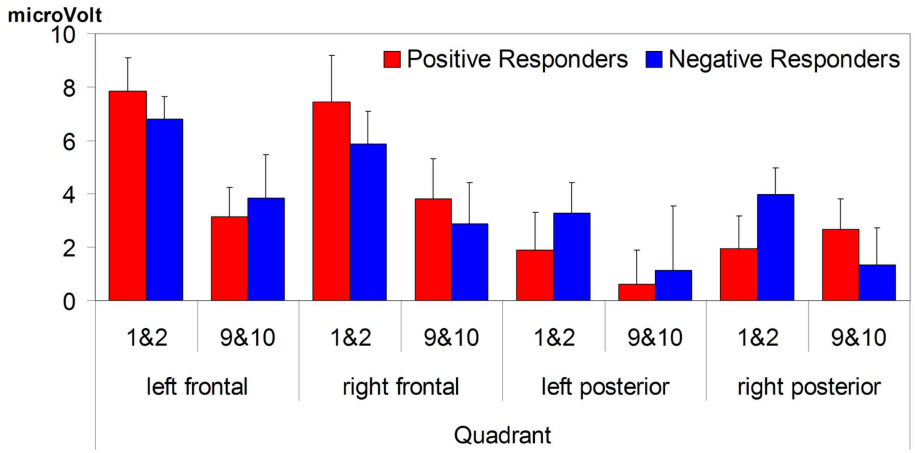


FIGURE 4 | Both groups show a similar decrease in positive amplitude for familiarized words in isolation (presented 9 & 10 times), compared to the first two times. For both groups, the decrease was most pronounced over frontal quadrants of the brain.

was significant at $p < 0.03$ in all quadrants except the left posterior quadrant.

We examined all other data available on the two groups. Two Positive responders (included in further analyses) reported having had speech therapy, and no Negative responders; but on no measure was there any significant difference between the two groups as a whole. Age did not differ (ERP experiment: positive responders mean age 217 days, Negative responders 218 days: $t_{21} = -0{,}213$, $p = 0.83$; follow-up testing: 37.6 and 34.4 months, respectively: $t_{21} = 1{,}307$, $p = 0.21$). Number of trials per condition in the ERP study did not differ: on average 21 trials per condition per Positive responder, and 20 trials per Negative responder ($t_{21} = 0.55$, $p = 0.59$; $t_{21} = 0.10$, $p = 0.92$ across familiar and unfamiliar words, respectively). Repetition effects in familiarization likewise did not differ: there were no significant interactions between Familiarity × Group for the first two versus the last two isolated word tokens in the familiarization phase ($F_{1,21} < 1$). Indeed, there were no polarity differences to be seen in the sub-groups' responses at this stage of the ERP experiment. **Figure 4** plots the response in μV for each sub-group to the first and the last pair of familiarization tokens, averaged for the 200–500 ms for each brain quadrant; it can be seen that there is a decrease in positivity (that is, a negative-going change) across familiarization that is virtually identical in average size for the two groups, and is further found for each group in each quadrant with only one exception (an insignificant shift in the opposite direction for Positive responders in the right posterior quadrant). This strongly suggests that our two sub-groups differ only in the abilities that are specifically needed for the ERP test phase but are not needed in familiarization.

Similarity in latency across the groups was also evident in onset analyses for the test phase: in Positive responders, the familiarity effect had an onset at 100 ms for right electrodes FT8 and RT, in Negative responders at 110 ms for left electrodes FT7 and LT. In short, the polarity and the distribution of the ERP response pattern for words presented in continuous speech were the sole significant differences that we could find between the two sub-groups. The mean distribution plot for each group is displayed in **Figure 5**;

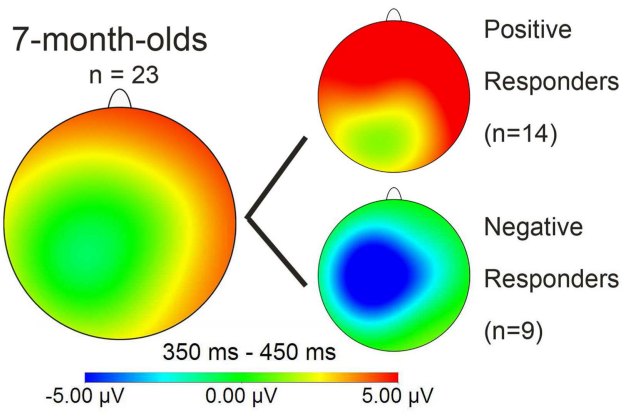


FIGURE 5 | Mean distribution plot for the ERP effect of familiarity (familiarized target – unfamiliar control words) in the 350–450 ms time window for those 7-month-olds who returned at 3 years for language testing; the two smaller plots divide the 7-month-olds into the sub-groups "Positive responders" and "Negative responders."

comparison with **Figure 3** makes clear that the Negative responder group deviates from the 28-participant seven-month average, and in fact closely resembles the pattern of negativity found with 10-month-old participants by Kooijman et al., 2005; see **Figure 3** above.

LANGUAGE SKILL TESTS

We administered two norm-referenced language tests to all children: the *Reynell Test voor Taalbegrip* "test of language comprehension" (Van Eldik et al., 1995), and the Schlichting et al. (1995) *Test voor Taalproductie* "test of language production." Together, the tests are a slightly modified Dutch translation of the Reynell (1985) Developmental Language Scales. They are the established scales used in the Netherlands for assessing language development problems, and are normed over 1,000 typically developing

children. The test results for each child are converted into language quotients (LQs), with a mean of 100 and a SD of 15 points, that depend on the child's age in months. An LQ below 85 is considered to indicate risk of language impairment. Both tests are graded in difficulty, allowing older children to start at a more advanced level, and both are suitable for children from 2 to 6 years.

The children were individually tested by the second author, unaware of their ERP profiles. In the first session they undertook the comprehension scale, in which they were asked to act out or point to requested objects. In the second session, scheduled on average 8 days (range 1–21 days) after the first session, they participated in two subtests of the production scale: one assessing sentence production, and one assessing expressive vocabulary. In the sentence subtest, children are required to make sentences of a similar structure to models given by the experimenter, to describe certain pictures, or arrays of toys. In the vocabulary subtest, children name objects or finish the experimenter's sentences describing pictures. In addition to both tests, parents were asked to complete a Dutch version of the "Speech and Language Assessment Scale" (Hadley and Rice, 1993), in which they rated their child's development on a variety of language skills compared to "other children of the same age," starting from 1 ("very poor") to 7 ("very good").

RESULTS

On the standardized language tests, all of these children achieved scores within or above the normal range. Overall, the children have high LQs for comprehension ($m = 115.4$, SD $= 11.8$), for sentence production ($m = 113.9$, SD $= 14.7$), and for word production ($m = 118.9$, SD $= 11.2$). Their parents rate their average language skills also as somewhat above those of their peers ($m = 4.7$, SD $= 0.9$). The scores are highly correlated (see **Table 2**).

Figure 6 shows that the Negative responders, with ERPs at 7 months resembling those of 10-month-olds, have significantly higher LQs than the Positive responders, whose ERPs at 7 months conformed to the overall seven-month group average. The Negative responders' scores fall on average at 1.5 SD above the LQ mean, and the inter-group difference is significant for both comprehension ($t_{21} = 2.37$, $p = 0.027$) and word production ($t_{21} = 5.85$, $p < 0.001$), and almost significant for sentence production ($t_{21} = 2.06$, $p = 0.052$).

Further, across all 23 subjects, the ERP effect indexing speech segmentation ability at 7 months (i.e., difference between familiarized test and unfamiliar control words over left frontal electrodes in the 350–450 ms time window) and the LQ for word production at 3 years were significantly correlated, as can be seen in **Figure 7**: the more negative the difference wave, the higher the LQ for word production at 3 years ($r_{\text{bivariate}} = -0.47$, $p = 0.02$; with LQs for comprehension and sentence production partialed out, $r_{\text{partial}} = -0.42$, $p = 0.06$).

Parents of Negative responders rated their children higher than parents of Positive responders did for their children ($t_{21} = 1.86$, $p = 0.077$). The average SLAS ratings, and separate group averages for each SLAS subscale, are shown in **Figure 8**; it can be seen that the Negative responders receive higher ratings in every case. The groups differ significantly on the syntax and talkativeness subscales ($t_{21} = 2.09$, $p < 0.05$, and $t_{21} = 2.58$, $p < 0.02$, respectively), and there is further a near-significant difference on the articulation subscale ($t_{21} = 1.82$, $p = 0.084$).

Together, these results show that ERPs for word recognition in continuous speech at 7 months are an indication of later language development. At 7 months the Negative responders delivered the brain response seen as a marker of segmentation in 10-month-olds. It is specifically in language processing that their brain responses differ from those of their age mates, and it is this specifically linguistic response that predicts their later vocabulary and sentence processing skills. Negative responders have higher language scores at 3 years than Positive responders, with the most marked difference being found for expressive vocabulary.

Table 2 | Correlation coefficients relating the language quotients and parental questionnaires at 3 years.

	Sentence production LQ	Word production LQ	SLAS average
Comprehension LQ	0.577**	0.515*	0.499*
Sentence production LQ	–	0.411	0.669***
Word production LQ	–	–	0.326

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$.

DISCUSSION

A 7-month-old's brain responses in a segmentation task provide advance evidence of the later course of language proficiency development. At 3 years, infants who at 7 months had shown a left-lateralized negative-going brain response to a familiarized word in a sentence context linguistically outperformed infants who had shown a distributed positive-going brain response to the same stimuli at 7 months. The infant language skill difference appeared across a wide range of measures collected at 3 years, involving language at both the word and the sentence level, and skills in both speech comprehension and speech production.

Recall that our comparisons across the infant sub-groups had found that isolated-token repetition effects, as evidenced by change in response to the last two in comparison to the first two tokens in familiarization, did not differ for the Negative versus the Positive responders. Repetition effects are evidence of memory abilities (Rugg, 1985), and thus it would appear that the difference between our sub-groups is not one of simple memory capacity, but one with a more sharply linguistic focus: the test phase requires segmentation of the familiarized word from surrounding speech, and it is in this skill in particular that the Negative responder group outstrips their Positive responder age mates.

The significant differences that motivated a split into two sub-groups concerned only the brain response that signaled segmentation: the response time-locked to onset of the word that had been familiarized, when it was heard embedded in a sentence context. This response differed across the two sub-groups in both polarity and distribution, and the two sub-groups that were identified in this manner turned out to have significant differences in linguistic performance nearly 2.5 years later. Recall that Junge et al. (2012) also observed that it was individual differences in word

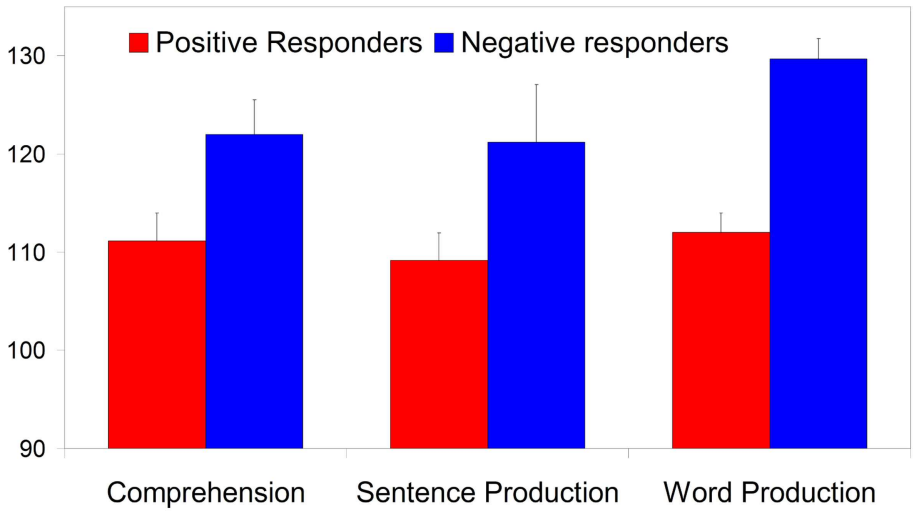


FIGURE 6 | The three language quotients at 3 years split by group performances at 7 months (error bars are one standard error from the mean). The group differences on comprehension and word production are significant (at $p < 0.05$ and $p < 0.001$ respectively), the sentence production difference just misses significance ($p = 0.052$).

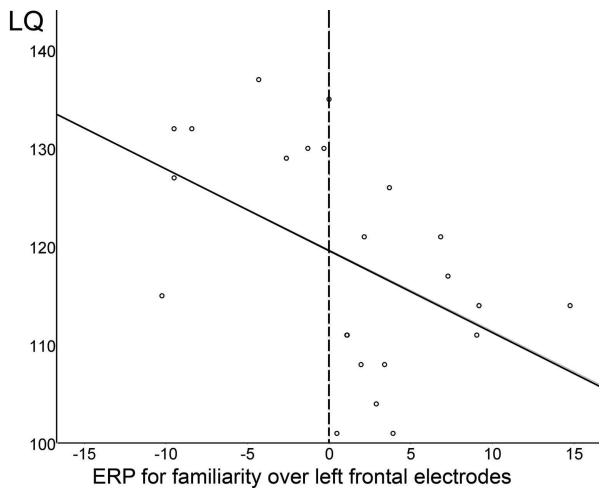


FIGURE 7 | The more negative the familiarity effect at 7 months (i.e. the more negative the difference wave between familiarized target and unfamiliar control words in the 350–450 ms time window over left frontal electrodes), the higher the quotient for word production at 3 years. The dotted line indicates the split between Negative and Positive responders.

recognition when words were presented in continuous speech, but not when presented in isolation, that were linked to future vocabulary. Together, these results strongly suggest that infant ERP responses of word recognition evidencing speech segmentation skill are predictive signals of linguistic development.

Mastery of segmentation is a fundamental skill indeed, because, as laid out in the introduction, most of the speech input an infant receives is in the form of multi-word utterances (Van de Weijer, 1999), and without being able to recognize words in these circumstances, the development of a substantial vocabulary cannot succeed. Segmenting speech into separate words forms part of the overall task of acquiring the phonology of the native language, in that the speech cues that inform lexical segmentation differ across languages. Evidence of segmentation in infant listening then appears earlier in some languages than in others, putatively for reasons of phonological salience and consistency of such segmentation cues. Mastering segmentation therefore rests on the construction of mental representations of language-specific phonology, prior to the availability of an extensive vocabulary from which such representations could have been abstracted.

It is perhaps little wonder that such a complex skill should vary in its rate of achievement across individuals. Such variation, and importantly, its relation to linguistic performance levels at 2 years, had already been demonstrated on the basis of behavioral measures both at a group level by Newman et al. (2006) and at an individual level by Singh et al. (2012). Moreover, related phonological skills of attunement to the native repertoire of phonetic contrasts have also been shown to vary across individuals and to be correlated with variation in later language skills (Kuhl et al., 2008); for instance, Tsao et al. (2004) measured the accuracy of vowel discrimination at 6 months, and also the speed with which a discrimination criterion could be reached, and found both measures to be predictive of vocabulary size in the second year of life. Although our results do not allow us to examine why it is that some infants displayed more mature speech segmentation skill than others, these findings further corroborate the proposition that such speech perception skills for the native language in infancy scaffold a child's future language development (Cristia et al., submitted).

In the present study we have shown that the dimensions of inter-individual variation in early segmentation performance can be captured in terms of patterns of ERPs in infants' brains. First, we have demonstrated that ERP evidence for segmentation is available earlier than behavioral evidence for the same skill. Although HPP

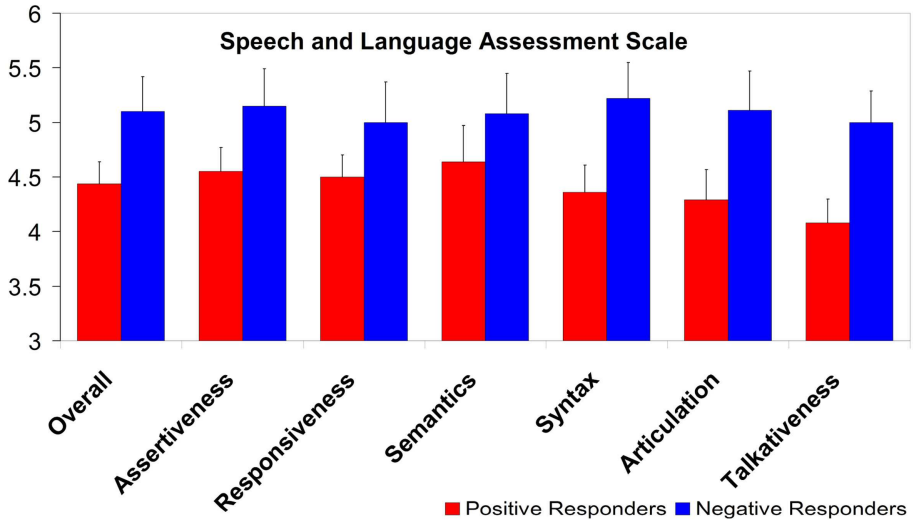


FIGURE 8 | Group ratings on the Speech and Language Assessment Scale, overall and per subscale. A score of "4" corresponds to parents rating their child's language performance as equal to their child's peers; higher scores reflect better language ratings. (Error bars are one standard error from the mean).

studies with 7-month-olds acquiring Dutch had shown no significant evidence that segmentation of continuous speech was in place at that age (Kuijpers et al., 1998), ERP measurement detected such evidence where behavioral techniques could not. The onset of a familiarized word in a continuously spoken sentence context produced a brain response that had a significantly more positive amplitude than the response induced in matched contexts by a matched word that had not been previously presented in familiarization.

Interestingly, the overall pattern observed in these 7-month-old brains was not the same as had been observed in the measurements made somewhat later in the first year of life by Kooijman et al. (2005, 2009), Goyet et al. (2010), and Junge et al. (2012). In all those studies, brain responses to the familiar words were on average more negative than the responses to the matched unfamiliar control words. Kooijman et al. (2005) report this pattern both for the familiarization phase (with responses to the last tokens in the 10-token list being more negative than responses to the first tokens) and for the test phase (where the same difference contrasted the familiarized word against its matched control in the test set). The result in the familiarization condition of the present study with 7-month-olds also showed the same negative-going effect. But in the test condition of the present study, the overall average difference brain response was opposite in direction, with familiarity being associated with a more positive brain signal than unfamiliarity.

A positive familiarity effect for words in a continuous speech situation has in fact been reported before, in infants younger than our sample of 7-month-olds (6-month-olds, Männel and Friederici, 2010). This may suggest that ERP effects of word recognition in infancy gradually change from a positive (up to 6 months) to a negative polarity (from 10 months onward). However, we reiterate that brain maturation alone cannot explain the variation in polarity in the present sample across conditions. As we saw, at a group level the same 7-month-olds show a familiarity effect with negative amplitude for the isolated word familiarization phase. It is only in the test phase (requiring segmentation skill) that a positive familiarity effect is seen. Moreover, other studies have reported differential ERP responses across conditions within the same set of children, an asymmetry that brain maturation alone obviously cannot explain. For instance, Conboy and Mills (2006) showed that the relative dominance of a language in bilingual children explained the distribution of language-relevant ERP components. Junge et al. (2012) showed that the distribution of the word familiarity effect also hinges on the relative difficulty of the task, with a more focal distribution for the easier task (words introduced in isolation) and a broader distribution for the harder task (when words were introduced within an utterance).

Both polarity and distribution of the ERP effects played a role in distinguishing the two sub-groups of these 7-month-olds in the current study, with language skills at 3 years differing along with these earlier ERPs. The present study indicates that this variation is an important indicator of how the individual brains are performing the present linguistic processing task. In our 7-month-old participant group, a minority produced the negative-going effect (consistently seen across familiarization and test phases in the earlier studies with 10- to 12-month-olds) in the test phase as well as in familiarization. When assessed at 3 years of age, this minority then proved to deliver better sentence production and sentence comprehension performance, to have larger expressive vocabularies, and to receive higher ratings of their language skills from their parents, than the remaining majority group from the same 7-month-old participant population. The question prompted by these results is then: why do some 7-month-olds show a positive amplitude and others a negative amplitude?

Although our data set is limited in sample size, and we cannot do any source localization to derive any explanation about the origin of this polarity differences, a possible answer to the

polarity issue can be found in other studies describing a similar phenomenon within the same age group. As described in the introduction, this is not the first occasion on which the same kind of significant ERP effect has been reported as negative under some conditions and as positive under others, even within the same age group. Both tone processing (Kudo et al., 2011) and phonetic discrimination (Garcia-Sierra et al., 2011) have been associated with such variation, and it has previously appeared in a segmentation-related task too (Männel and Friederici, 2010). Note further that ERP studies of early phonetic processing also used variation in polarity and distribution of responses to distinguish sub-groups within participant populations. Rivera-Gaxiola et al. (2005a) showed that differences in the patterns of 11-month-olds' responses to non-native versus native contrasts were related to later word production abilities. In an oddball task with a constant standard, all children produced much the same negativity in response to a deviant differing across a native phoneme boundary ("native deviant"). Two sub-groups differed, however, in their response to another deviant that differed from the standard to the same degree as the native deviant but across a non-native phoneme boundary ("non-native deviant"). One sub-group produced a negativity in response to the non-native deviant too, with a parietal localization. The other sub-group produced a right fronto-central positivity, instead (thus effectively distinguishing the non-native and native contrasts in kind; Rivera-Gaxiola et al. (2005b) had also observed such sub-groups forming when they tracked the gradual attunement to native contrasts across the second half of the first year). The latter sub-group then proved to have developed larger productive vocabularies by 18 and continuing to 30 months of age.

Rivera-Gaxiola et al. (2005b) hypothesized that the polarity differences denote differences in auditory processing, with a positivity reflecting acoustic processing and a negativity reflecting more mature processing, possibly due to increased experience with the native language (Kuhl et al., 2008). This would entail for our study that the Positive responders relied on acoustic salience (of, for instance, the stressed syllable), whereas the Negative responders achieved word recognition with a more mature mechanism (i.e., segmenting fluent speech into word-like units). It is in this light noteworthy that a similar left-going negative marker of word recognition later in infancy has also been observed in studies comparing familiar/known versus unfamiliar/unknown isolated word processing (Mills et al., 1993, 1997, 2004, 2005; Thierry et al., 2003). As Junge et al. (2012) hypothesized, it is likely that for young infants, with a very small vocabulary, this same recognition mechanism indexing word meaning has developed from one that at a younger age is mainly sensitive to word form repetitions.

Our results indicate that a familiarity effect in infancy with negative amplitude in a speech segmentation task is associated with a more mature response, which in turn is associated with better language development. It would be interesting to examine whether infants who exhibit different polarities indexing word recognition in different circumstances (in isolation versus in multi-word utterances) also differ in the neural generators they use for word recognition, or in the way they use the same generators to achieve this. The use of neural networks could in turn also be affected by individual differences in brain maturation, in closing of the fontanels or by listening strategies. However, more research is clearly necessary to uncover the origin of individual variation in polarity and in distributions; our sample size is too small to draw final conclusions. Future research should also address the development of the word familiarity effect, not only within infancy, but also from infancy to adulthood, since a broad positive effect is again seen in many adult studies (Rugg, 1985; Snijders et al., 2007; but see Cunillera et al., 2006).

Finally, an additional contribution of the present study is clear evidence that the inter-group differences are longer-lasting than previously known. We retested our participants at age three and found wide-ranging evidence of language skills advantages for the group that had shown the 10-month-like ERP effect at 7 months. Thus we have reconfirmed the relation of early segmentation ability to later linguistic proficiency, and have shown that it lasts at least into the fourth year of life. Most importantly, though, we have isolated an ERP marker associated with differences in early segmentation ability. Infants who at 7 months already show an advanced marker of segmentation skill continue to develop better language skill at least through their third birthday.

ACKNOWLEDGMENTS

This research was supported by the NWO-SPINOZA project "Native and non-native listening" (A. Cutler), and by a Max Planck Society doctoral fellowship to the second author. All authors were associated with the Max Planck Institute for Psycholinguistics at the time the research was carried out. The first two authors contributed equally to this research: the ERP study formed part of the first author's PhD dissertation, and the language skill testing and further analyses formed part of the second author's PhD dissertation. Partial reports of these results were presented to the workshop "Online methods in children's language processing," City University of New York, March 2006, and the 34th Annual Boston University Conference on Language Development, October 2009.

REFERENCES

Abla, D., Katahira, K., and Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.* 20, 952–964.

Baayen, R. H., Piepenbrock, R., and Van Rijn, H. (1993). *The CELEX Lexical Database [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bosch, L. (2011). "Precursors to language in preterm infants: speech perception abilities in the first year of life," in *Progress in Brain Research*, Vol. 189, eds O. Braddick, J. Atkinson, and G. M. Innocenti (Burlington: Academic Press), 239–257.

Conboy, B., and Mills, D. L. (2006). Two languages, one developing brain: event-related potentials to words in bilingual toddlers. *Dev. Sci.* 9, F1–F12.

Cunillera, T., Toro, J., Sebastián-Gallés, N., and Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res.* 1123, 168–178.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua* 92, 81–104.

DeCasper, A. J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C., and Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behav. Dev.* 17, 159–164.

Fear, B. D., Cutler, A., and Butterfield, S. (1995). The strong/weak syllable distinction in English. *J. Acoust. Soc. Am.* 97, 1893–1904.

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., et al. (2011). Bilingual language learning: an ERP study relating early brain responses to speech, language input, and later word production. *J. Phon.* 39, 546–557.

Goyet, L., de Schonen, S., and Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: an ERP study. *Brain Res.* 1332, 75–89.

Hadley, P., and Rice, M. L. (1993). Parental judgments of preschoolers' speech and language development: a resource for assessment and IEP planning. *Semin. Speech Lang.* 14, 278–288.

Junge, C. (2011). *The Relevance of Early Word Recognition: Insights from the Infant Brain.* Doctoral Dissertation, MPI Series in Psycholinguistics 67, Nijmegen.

Junge, C., Kooijman, V., Hagoort, P., and Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* 15, 463–473.

Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–25.

Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207.

Kooijman, V., Hagoort, P., and Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cogn. Brain Res.* 24, 109–116.

Kooijman, V., Hagoort, P., and Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy* 14, 591–612.

Kooijman, V., Johnson, E. K., and Cutler, A. (2008). "Reflections on reflections of infant word recognition," in *Early Language Development: Bridging Brain and Behaviour*, eds A. D. Friederici and G. Thierry (Amsterdam: John Benjamins), 91–114.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., and Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* 14, 1100–1106.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000.

Kuijpers, C., Coolen, R., Houston, D., and Cutler, A. (1998). "Using the head-turning technique to explore cross-linguistic performance differences," in *Advances in Infancy Research*, eds C. Rovee-Collier, L. Lipsitt, and H. Hayne (London: Ablex Publishing Corporation), 205–220.

Kushnerenko, E., Cepionene, R., Balan, P., Fellman, V., Huotilainen, M., and Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport* 13, 47–51.

Männel, C., and Friederici, A. D. (2010). Prosody is the key: ERP studies on word segmentation in 6- and 12-month-old children. *J. Cogn. Neurosci. Suppl.* 261–261.

Mills, D., Coffey-Carina, S. A., and Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Dev. Neuropsychol.* 13, 397–445.

Mills, D., Coffey-Corina, S. A., and Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *J. Cogn. Neurosci.* 5, 317–334.

Mills, D., Plunkett, K., Prat, C., and Schafer, G. (2005). Watching the infant brain learn words: effects of language and experience. *Cogn. Dev.* 20, 19–31.

Mills, D., Prat, C., Stager, C., Zangl, R., Neville, H., and Werker, J. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *J. Cogn. Neurosci.* 16, 1452–1464.

Narayan, C., Werker, J. F., and Beddor, P. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Dev. Sci.* 13, 407–420.

Newman, R., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P. W., and Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Dev. Psychol.* 42, 643–655.

Polka, L., and Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect, and cross-language comparisons. *Infancy* 17, 198–232.

Reynell, J. (1985). *Reynell Developmental Language Scales, 2nd revision.* Windsor: National Foundation for Educational Research.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., and Kuhl, P. K. (2005a). Neural patterns to speech and vocabulary growth in American infants. *Neuroreport* 16, 495–498.

Rivera-Gaxiola, M., Silva-Pereyra, J., and Kuhl, P. K. (2005b). Brain potentials to native and non-native speech contrasts in 7-and 11-month-old American infants. *Dev. Sci.* 8, 162–172.

Rugg, M. (1985). The effects of semantic priming and stimulus repetition on event-related potentials. *Psychophysiology* 22, 642–647.

Saffran, J. R., Werker, J., and Werner, L. (2006). "The infantss auditory world: hearing, speech, and the beginnings of language," in *Handbook of Child Development*, eds R. Siegler and D. Kuhn (New York: Wiley), 58–108.

Schlichting, L., van Eldik, M., Lutje Spelbroek, H., van der Meulen, S., and van der Meulen, B. (1995). *Schlichting Test voor Taalproductie.* Lisse: Swets & Zeitlinger.

Singh, L., Reznick, J. S., and Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Dev. Sci.* 15, 482–495.

Snijders, T., Kooijman, V., Hagoort, P., and Cutler, A. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Res.* 1178, 106–113.

Thierry, G., Vihman, M., and Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *Neuroreport* 14, 2307–2310.

Tsao, F.-M., Liu, H.-M., and Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Dev.* 75, 1067–1084.

Van de Weijer, J. (1999). *Language Input for Word Discovery.* Doctoral Dissertation, MPI Series in Psycholinguistics 9, Nijmegen.

Van der Hulst, H. (1984). *Syllable Structure and Stress in Dutch.* Dordrecht: Foris.

Van Eldik, M., Schlichting, J., Lutje Spelberg, H. van der Meulen, S., and van der Meulen, B. (1995). *Reynell Test voor Taalbegrip.* Lisse: Swets Test Services.

Weber, C., Hahne, A., Friedrich, M., and Friederici, A. D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cogn. Brain Res.* 18, 149–161.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 August 2012; accepted: 10 January 2013; published online: 08 February 2013.

Citation: Kooijman V, Junge C, Johnson EK, Hagoort P and Cutler A (2013) Predictive brain signals of linguistic development. Front. Psychology ***4****:25. doi: 10.3389/fpsyg.2013.00025*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 31 December 2012
doi: 10.3389/fpsyg.2012.00580

How each prosodic boundary cue matters: Evidence from German infants

Caroline Wellmann[1]*, Julia Holzgrefe[1], Hubert Truckenbrodt[2], Isabell Wartenburger[1] and Barbara Höhle[1]

[1] Department of Linguistics, University of Potsdam, Potsdam, Germany
[2] Centre for General Linguistics (ZAS), Berlin, Germany

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Kathy Hirsh-Pasek, Temple University, USA
Alejandrina Cristia, Max Planck Institute for Psycholinguistics, Netherlands

***Correspondence:**
Caroline Wellmann, Department of Linguistics, University of Potsdam, Karl-Liebknecht-Street 24-25, Potsdam 14476, Germany.
e-mail: caroline.wellmann@uni-potsdam.de

Previous studies have revealed that infants aged 6–10 months are able to use the acoustic correlates of major prosodic boundaries, that is, pitch change, preboundary lengthening, and pause, for the segmentation of the continuous speech signal. Moreover, investigations with American-English- and Dutch-learning infants suggest that processing prosodic boundary markings involves a weighting of these cues. This weighting seems to develop with increasing exposure to the native language and to underlie crosslinguistic variation. In the following, we report the results of four experiments using the headturn preference procedure to explore the perception of prosodic boundary cues in German infants. We presented 8-month-old infants with a sequence of names in two different prosodic groupings, with or without boundary markers. Infants discriminated both sequences when the boundary was marked by all three cues (Experiment 1) and when it was marked by a pitch change and preboundary lengthening in combination (Experiment 2). The presence of a pitch change (Experiment 3) or preboundary lengthening (Experiment 4) as single cues did not lead to a successful discrimination. Our results indicate that pause is not a necessary cue for German infants. Pitch change and preboundary lengthening in combination, but not as single cues, are sufficient. Hence, by 8 months infants only rely on a convergence of boundary markers. Comparisons with adults' performance on the same stimulus materials suggest that the pattern observed with the 8-month-olds is already consistent with that of adults. We discuss our findings with respect to crosslinguistic variation and the development of a language-specific prosodic cue weighting.

Keywords: infants, language acquisition, speech perception, prosodic bootstrapping, prosodic boundary cues, cue weighting, intonation phrase boundary, headturn preference procedure

INTRODUCTION

The system underlying the prosodic organization of language constitutes a complex linguistic subsystem with strong interfaces to other linguistic domains like the lexicon or the syntax. This paper deals with the correlation between prosodic phrasing and the syntactic structure of utterances which has already been the subject of numerous studies in the area of adult sentence processing as well as of infant language acquisition (e.g., Streeter, 1978; Scott, 1982; Hirsh-Pasek et al., 1987; Sanderman and Collier, 1997; Nazzi et al., 2000; Soderstrom et al., 2003; Peters, 2005). The question unifying these diverse areas of research is whether prosody provides information that can enter into the processing of the syntactic structure of utterances. In language acquisition research this approach is known as the prosodic bootstrapping account (Gleitman and Wanner, 1982), which assumes that infants can exploit acoustic information from their speech input to find solutions for several tasks they are faced with when accessing the grammatical system of their language. In this paper, we will have a closer look at German infants' sensitivity to the acoustic cues that mark a major prosodic boundary, that is, the intonation phrase boundary (IPB).

There are two properties that render IPBs especially useful within the prosody-syntax mapping. First, a rather clear-cut set of acoustic cues, namely pitch changes, lengthening of preboundary segments, and pauses, is associated with IPBs across different languages (e.g., Vaissière, 1983; Nespor and Vogel, 1986; Price et al., 1991; Wightman et al., 1992; Venditti et al., 1996; Hirst and Di Cristo, 1998; Peters et al., 2005; Féry et al., 2011). Secondly, again crosslinguistically, there exists a high coincidence of IPBs with major syntactic boundaries like sentence and clause boundaries (e.g., Cooper and Paccia-Cooper, 1980; Venditti et al., 1996; Vaissière and Michaud, 2006). Hence, sensitivity to the relevant acoustic cues would provide infants with a strong mechanism for chunking incoming speech into syntactically relevant units without requiring lexical or syntactic knowledge.

Indeed, numerous studies within the prosodic bootstrapping account have demonstrated that infants are equipped with a high sensitivity to prosodic information such as stress, rhythm, and intonation (for an overview, see Jusczyk, 1997). This also holds for the perception of acoustic information that is related to the marking of prosodic boundaries. Research in this area started with some landmark studies that tested infants' reactions to the presentation of natural speech in contrast to manipulated speech material in which pauses had been inserted at non-boundary positions (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989). These studies – using the headturn preference procedure (HPP) – showed that American infants as young as 7–10 months prefer to listen

to speech material showing a coincidence of the typical acoustic cues occurring at clausal boundaries compared to materials in which the coincidence of pauses with other prosodic cues had been disrupted. The fact that the same preference occurred with low-pass-filtered material strongly suggests that it is the disturbance of the prosodic organization of the utterances that causes the successful discrimination of both kinds of material. Studies with other languages using the same technique of pause insertion have provided evidence that this discrimination ability is not unique to English-learning infants: German as well as Japanese infants have been found to discriminate speech with pauses at clausal boundaries from speech with pauses inserted at non-boundary positions in their language (Hayashi and Mazuka, 2002; Schmitz, 2008).

Also using pause insertion, Jusczyk et al. (1992) investigated infants' sensitivity to boundaries of smaller units, namely clause-internal phrase boundaries. In their material, pauses were inserted either before the main verb, that is, at the boundary between the subject and the verb phrase, or after the main verb, that is, within the verb phrase. English-learning 9-month-olds preferred to listen to the materials in which the pause occurred at the phrasal boundary.

Gerken et al. (1994) compared sentences with lexical subjects (e.g., *The caterpillar ate . . .*) and sentences with pronominal subjects (e.g., *He ate . . .*) in which pauses had been inserted after either the subject or the verb. As lexical subjects form their own phonological phrase, there is a prosodic boundary between the subject and verb in the corresponding sentences while there is typically no prosodic boundary after a pronominal subject. Only in the lexical subject condition did 9-month-old infants prefer to listen to sentences with pauses after the subjects (e.g., *The caterpillar # ate . . ., He # ate . . .*) over those with pauses after verbs (e.g., *The caterpillar ate # . . ., He ate # . . .*). These results again suggest that the prosodic organization – and not the syntactic one – is relevant for infants' preference for natural material. Taken together, these studies provide evidence that by 9 months infants are sensitive to the acoustic markers at clausal as well as at phrasal boundaries.

More recent work has gone beyond the question of the perception of the acoustic correlates of major boundaries to the question as to whether the occurrence of prosodic boundaries affects the segmentation of continuous speech. Nazzi et al. (2000) were the first to test English-learning 6-month-olds' use of prosodic boundary cues to segment continuous speech. At the beginning of the experiment infants were familiarized with a sequence of words, once as a prosodically "well-formed" clause (e.g., *Leafy vegetables taste so good.*) and once as a prosodically "ill-formed," that is, non-clausal sequence that contained an internal clause boundary (e.g., *. . .leafy vegetables. Taste so good. . .*). These word sequences had been extracted from two different continuous passages. After familiarization infants were presented with two passages. One of them contained the familiarized prosodically well-formed sequence, the other the prosodically ill-formed sequence, which was now the end and the beginning of two adjacent sentences. This non-clausal unit contained a prosodic boundary that was marked by a pitch change, preboundary lengthening, and a pause. Infants listened significantly longer to the passage containing the clausal sequence than to the passage with the non-clausal sequence. These results suggest that word sequences that constitute a prosodic unit are better recognized than word sequences that span a prosodic boundary. Hence, prosodic boundary cues support the segmentation of clauses within a passage of sentences. These findings were replicated by Soderstrom et al. (2005) with a similar design, but more complex experimental materials. Specifically, it was demonstrated that prosodic boundary cues support English-learning infants' detection of familiar word sequences even across different passages of fluent speech.

Moreover, with a similar experimental design, Soderstrom et al. (2003) provided evidence that 6-month-old English-learning infants also use prosodic markers to detect syntactic units that are smaller than the clause, namely phrasal units such as noun and verb phrases. Interestingly, phrase boundaries were characterized by preboundary lengthening and pitch cues while there was no perceivable pause at the crucial position. This suggests that for the detection of phrase boundaries pause is not a necessary cue for 6-month-old English-learning infants.

The studies presented so far point to a crucial role of prosodic boundary information in infants' speech segmentation, especially during the first year of life. However, in a critical analysis of the prosodic bootstrapping account, Fernald and McRoberts (1996) doubt the reliability of acoustic correlates of prosodic boundaries as cues to syntactic units. The authors claim that none of the three markers is a reliable cue to syntactic boundaries as each cue also has non-linguistic functions (e.g., pitch changes for the regulation of affect) or linguistic functions other than syntax (e.g., vowel length as phonemic contrast). This would cause ambiguity of the acoustic correlates of boundaries whenever they occur at non-boundary positions. Fernald and McRoberts' argument may be weakened if a comprehensive analysis of a corpus of German adult-directed speech conducted by Peters et al. (2005) is taken into account. They found that IPBs were most frequently marked by pitch changes, followed by preboundary lengthening, while the occurrence of pause is rather rare. In addition, the analysis showed that each cue may occur individually, but that in a great majority of the cases boundaries are marked by a coalition of all three or two of the relevant cues. This convergence may decrease the ambiguity of prosodic boundary cues provided that the infant only considers a combination of cues to be a boundary marker.

In fact a detailed study by Seidl (2007) that tested the perceptual impact of each of the prosodic cues provided evidence that English-learning 6-month-old infants rely on a combination of cues in their boundary processing. The investigations were based on the materials and the experimental design used by Nazzi et al. (2000). Seidl successively neutralized each acoustic correlate of the prosodic boundaries in the familiarization sequences. Thereby, the acoustic realization of the cue under investigation no longer differed between the two sequences. The question was whether infants, on the basis of the remaining prosodic cues, would still differentiate the clausal and the non-clausal familiarization sequences and recognize the clausal sequence in the passage during testing.

Infants' detection of the clausal sequence was not disturbed by the neutralization of the pause cue. This indicates that pitch change and preboundary lengthening were sufficient cues for the 6-month-old English-learning infants, whereas the pause was not necessary. Furthermore, preboundary lengthening also proved not to be a necessary cue, because infants still recognized the clausal

sequence when preboundary lengthening was neutralized. However, when the pitch cue was neutralized the infants no longer detected the clausal sequence in the passage. Hence, pitch change proved to be a necessary boundary cue for American infants' clause segmentation. A further experiment investigated whether pitch change as a single cue would suffice, that is, both preboundary lengthening and pause were neutralized. This kind of acoustic manipulation disturbed infants' detection of the clausal sequence, indicating that a pitch change alone is not sufficient. In conclusion, a combination of pitch change and preboundary lengthening or pitch change and pause was necessary to trigger clause segmentation in 6-month-old English-learning infants. Seidl (2007) argued that by 6 months English-learning infants do not treat prosodic cues equally, but have, at least partially, become attuned to adults' weighting of prosodic cues in their native language (Streeter, 1978; Scott, 1982; Aasland and Baum, 2003).

Seidl and Cristià (2008) expanded these investigations by testing 4-month-old English-learning infants with the same materials. In contrast to the 6-month-olds, this younger group was successful in clause segmentation only when pitch change, lengthening, and pause in combination signaled the boundary. Neutralization of one of the prosodic cues led to failure in segmentation. Seidl and Cristià (2008) concluded that 4-month-old English-learning infants segment clauses by considering all prosodic boundary cues.

In a following study, Johnson and Seidl (2008) explored whether infants' weighting of prosodic boundary cues varies across languages. The experimental design of Seidl (2007) was applied with Dutch material to Dutch 6-month-olds. Like the English-learning infants, the Dutch learners segmented the clausal sequence from the text passage. However, when the pause was neutralized in the familiarization sequences Dutch-learning infants failed to segment the clausal sequence from the text passage. Johnson and Seidl (2008) considered two interpretations. One is related to the strength of the prosodic cues. The magnitude of pitch change and preboundary lengthening might not have been salient enough to trigger the clause segmentation. Acoustic analyses of the stimuli had revealed that the saliency of the pitch reset and the pause duration at the clausal boundary differed in the materials used across the two languages. Compared to the English stimuli the pitch reset in the Dutch stimuli was only half the magnitude, whereas the pause was more than twice as long. However, the qualitative difference in the prosodic cues in the Dutch versus English stimuli might reflect language-specific boundary markings as Dutch compared to English generally tends to have a smaller pitch range (Collins and Mees, 1981; Willems, 1982). Therefore, Johnson and Seidl argued for a different interpretation: by 6 months, with increasing exposure to the native language, Dutch-learning and English-learning infants have developed a language-specific prosodic cue weighting that influences infants' clause segmentation procedures.

Taken together, these findings indicate that infants' sensitivity to acoustic cues as prosodic boundary markers is subject to a developmental change during early infancy – perhaps a change from a more general perceptually driven mechanism that relies on a broad set of acoustic cues to a mechanism that is attuned to the specific properties of the target language.

To further investigate the question of an early weighting of prosodic boundary cues, the present study set out to test infants learning German, a language in which we have – at least for adult-directed spontaneous speech – specific knowledge about the frequency of occurrence of prosodic cues at IPBs (Peters et al., 2005), the prosodic unit under investigation in this study. Moreover, from a study with German listeners, findings on adults' weighting of the relevant acoustic cues are available: in a prosodic judgment task Holzgrefe et al. (2012) tested whether the presence of the cues pitch change and preboundary lengthening in the absence of the pause cue would suffice to signal a boundary. Listeners were presented with coordinated sequences of three names in different prosodic groupings. Their task was to judge the heard sequence as to whether or not it had an internal boundary. The German adult listeners identified the internal boundary when both, a pitch change as well as preboundary lengthening, but no pause, were present in the sequence; however, pitch change alone or lengthening alone was not sufficient. In the present study the same linguistic materials were used to test whether German infants' processing of prosodic boundary cues is similar to that shown for German adults.

Hence, in contrast to previous studies, we did not present complex clauses (Nazzi et al., 2000; Seidl, 2007; Johnson and Seidl, 2008; Seidl and Cristià, 2008), but well-formed sequences that allowed for a precise acoustic characterization of the phonetic instantiation of the crucial prosodic boundaries which we considered to be the basis for a controlled acoustic manipulation of the stimuli. Thus, going beyond the previous studies with English- and Dutch-learning infants, the results of the infants tested in the current study could be related to findings from adults, allowing a direct comparison of German adults' and infants' cue weighting.

Again in contrast to previous studies, we did not test infants' segmentation, but their discrimination ability. We suggest that infants' attunement to specific properties of their native language is not only displayed in segmentation tasks as revealed by the work of Johnson and Seidl (2008), Seidl (2007), and Seidl and Cristià (2008). Instead, perceptual reorganization with respect to cue weighting should also be reflected in discrimination performance as has been shown for tone and phonemic contrasts in previous research (Werker and Tees, 1984; Polka and Werker, 1994; Mattock and Burnham, 2006; Mattock et al., 2008). If prosodic boundary cues are perceptually weighted individually, we assume that the less weighted information will contribute less to both discrimination and segmentation.

Experiment 1 served as a baseline to ensure that in our experimental design German-learning infants perceive a boundary signaled by all three prosodic cues. In Experiment 2 we investigated whether the specific combination of a pitch change and preboundary lengthening is sufficient for boundary detection. Hereby, the question whether pause is a necessary cue would be examined. We did not test a combination of two cues that included the pause cue, because we expected that 8-month-olds would discriminate between stimuli with and without a pause easily given that the pause is a rather strong acoustic cue, especially in a mere discrimination task. In fact, in a similar study with younger German infants (Wellmann et al., in preparation) we found that even 6-month-olds are able to use the pause cue. More precisely, a pitch change together with preboundary lengthening was not sufficient for 6-month-olds, but the combination of pause and lengthening

was. Thus, a pause, but not a pitch change was a necessary cue for 6-month-olds. This finding moreover suggests that successful boundary detection depends on the specific cue constellation, rather than on the number of boundary cues provided.

After testing the combination of pitch change and preboundary lengthening, we examined the impact of each of the two as single cues: Experiment 3 tested pitch change and Experiment 4 preboundary lengthening.

EXPERIMENT 1: A BASELINE STUDY ON INFANTS' SENSITIVITY TO PITCH CHANGE, PREBOUNDARY LENGTHENING, AND PAUSE

In Experiment 1, we sought to ensure that 8-month-old German-learning infants are able to perceive a prosodic boundary that is signaled by the three main prosodic cues pitch change, preboundary lengthening, and pause. This would provide a verification of the experimental design and material as suitable for studying the perception of single prosodic boundary cues. As previous research has revealed that infants are sensitive to prosodic boundary information (e.g., Hirsh-Pasek et al., 1987; Nazzi et al., 2000), infants tested in Experiment 1 should be able to perceive a prosodic boundary. Experiment 1 aimed at creating a baseline for the subsequent experiments, in which the constellation of prosodic cues would be systematically varied.

MATERIALS AND METHODS

Participants

Twenty-four 8-month-old infants (12 girls) were tested. The mean age was 8 months, 16 days (range: 8 months, 3 days–8 months, 30 days). All infants who participated in this and the following experiments were from monolingual German-speaking families, born full-term and normal-hearing. Eleven additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (2), crying or fussiness (3), mean listening times of less than 3 s per condition (3), technical problems (2), and experimenter error (1).

Stimuli

The stimuli consisted of a sequence of three German names that were coordinated by *und* ("and"). The advantage of using coordinated structures instead of clauses lies in the better control of phonological and consequently prosodic parameters. Thus, we used the following three names, which only contained sonorant sounds: *Moni, Lilli, Manu*. This allowed for a reliable measurement of the fundamental frequency – the acoustic correlate of the pitch contour.

Several recordings of the same sequences of names were made in an anechoic chamber equipped with an AT4033a audio-technical studio microphone, using a C-Media Wave soundcard at a sampling rate of 22050 Hz with 16 bit resolution. A young female German native speaker from the Brandenburg area was instructed to read the sequence in two different prosodic groupings, as indicated by different bracketing as in (1).

(1) a. (Moni und Lilli und Manu)
b. (Moni und Lilli) (und Manu)

Each name is a syntactic XP and is correspondingly set off by a phonological phrase boundary from the other names (Gussenhoven, 1992; Truckenbrodt, 1999, 2007). Both sequences contain the same string and are disambiguated either by grouping all three names together as shown in (1a) or by grouping the first two names together and the final one apart as shown in (1b). This disambiguation employs the next higher level of the prosodic hierarchy, that is, the intonation phrase (IP). Thus, sequences of type (1a) are produced as a single IP, that is, without an internal boundary. In contrast, sequences of type (1b) are produced with an IPB after the second name, and consequently consist of two IPs. For each type of prosodic phrasing, the speaker produced six different acoustic realizations (tokens). The intended prosodic grouping was confirmed by two independent listeners who were naïve to the given bracketing.

The presence of the characteristics of an IPB in the sequences of names were confirmed by a detailed acoustic analysis of the recordings using PRAAT software (Boersma and Weenink, 2011). Measurements were carried out at the critical boundary position, namely on and after the second name. The analysis concentrated on the three acoustic correlates of prosodic boundary cues – fundamental frequency (F_0), the duration of the final vowel, and the pause. Examples of the oscillogram and the fundamental frequency aligned with the segments for sequences without an IPB are shown in **Figure 1A**, and for sequences with an IPB in **Figure 1B**. Details of the acoustic analysis are presented in **Table 1**.

The target word for the analysis was decomposed into four intervals corresponding to the phonetic segments, that is, the single consonantal and vocalic parts of the signal. F_0 was measured at the midpoint of the first segment and at the position of the maximum F_0 on the final vowel. The difference between these values was used to calculate the pitch change preceding the boundary. In sequences with an IPB, a pitch rise occurred, starting at the second syllable of the word and leading to a high boundary tone at the final vowel. This pitch change was 2.5 times greater in sequences with an IPB compared to sequences without an IPB (see **Table 1** and **Figure 1A** vs. **1B**). A mean pitch reset of 25 Hz from the high boundary tone to the midpoint of the following conjunction *und* ("and") was measured in sequences with an IPB, whereas the pitch change was only 3 Hz at the same location in sequences without an IPB. Thus, the pitch reset was greater in sequences with a boundary, but compared to Seidl's (2007) stimuli the overall extent of the reset was rather small, as the conjunction *und* was also uttered on a high pitch level (see the pitch contour in **Figure 1B**). First and foremost, in our stimuli the pitch cue in sequences with an IPB was provided by the pitch rise on the target name.

Preboundary lengthening was calculated by measuring the length of the final vowel in both prosodic types. Transitions between the final vowel and the onset of the conjunction *und* were not included. The vowel duration was about 1.8 times longer before a boundary compared to the same vowel in the sequence without an internal IPB.

The duration of the pause after the target name had a mean of 506 ms in sequences with an internal IPB. In contrast, no pause was present at this position in sequences without an internal IPB.

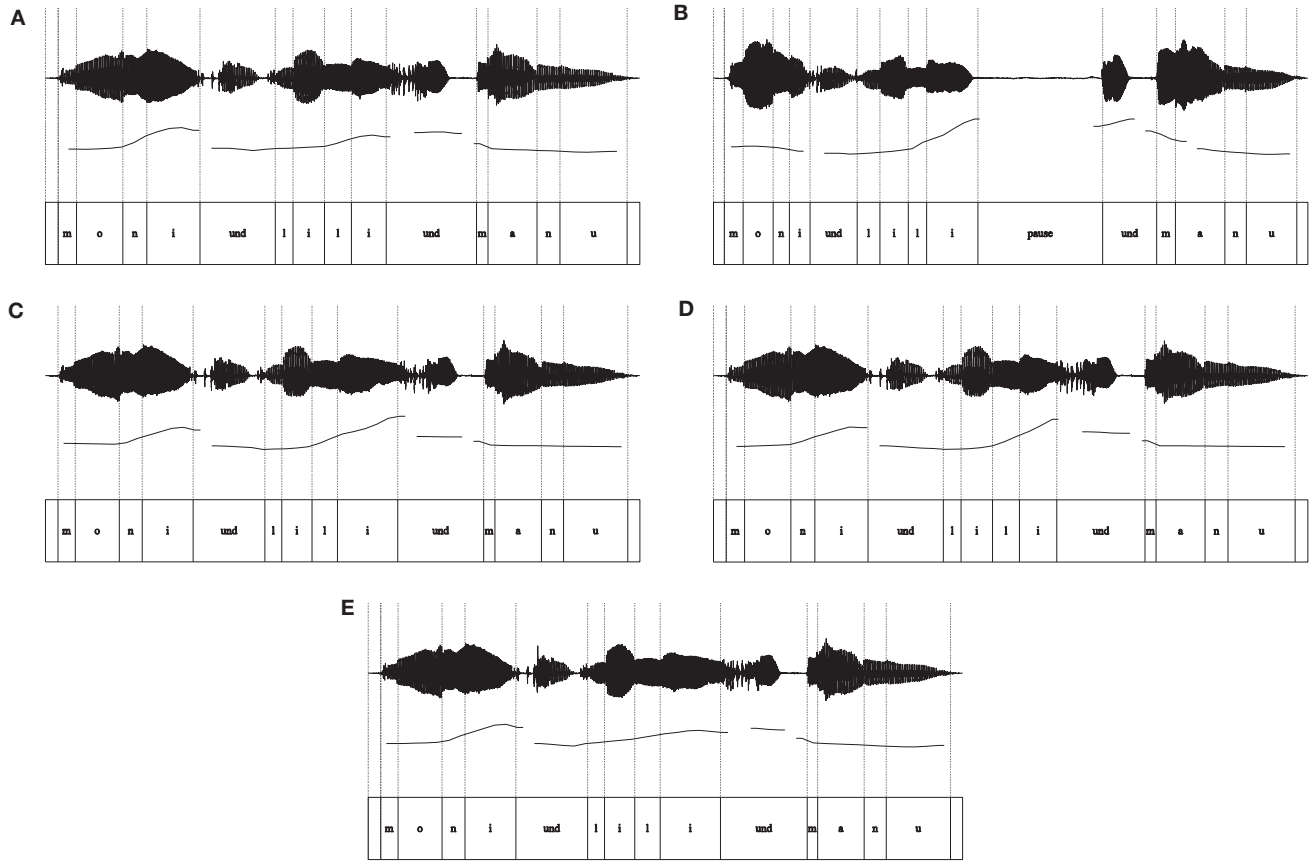


FIGURE 1 | Oscillograms and pitch contours aligned with the text. **(A)** Sequence without an internal IPB used in Exp. 1, **(B)** Sequence with an internal IPB used in Exp. 1, **(C)** Sequence with pitch change and preboundary lengthening used in Exp. 2, **(D)** Sequence with pitch change used in Exp. 3, **(E)** Sequence with preboundary lengthening used in Exp. 4.

Table 1 | Mean values and range of the acoustic correlates of prosodic boundary cues in the experimental stimuli.

Acoustic correlate	Without an internal IPB	With an internal IPB
	*[Moni und **Lilli** und Manu]*	*[Moni und **Lilli**] [und Manu]*
Pitch rise in Hz	88 (77–110)	220 (197–240)
Pitch rise in semitones	6.7 (5.8–8.2)	14.0 (12.8–14.6)
Maximum pitch in Hz	277 (264–293)	397 (371–422)
	*[Moni und Lill**i** und Manu]*	*[Moni und Lill**i**] [und Manu]*
Final vowel duration in ms	99 (91–110)	175 (162–186)
	*[Moni und Lilli **#** und Manu]*	*[Moni und Lilli]**#** [und Manu]*
Pause duration in ms	0	506 (452–556)

To summarize, sequences with an internal IPB clearly revealed the acoustic correlates of the three main prosodic boundary cues similar to IPBs in German spontaneous speech (Peters et al., 2005). A pitch rise occurred on the target name followed by a pitch reset after a pause. Preboundary lengthening was observed at the final vowel of the target name.

Following the acoustic analyses the different recordings (tokens) were used to create sound files for presentation as trials during the experiment. For each prosodic type, the six tokens were randomly concatenated with a silent interval of 1 s inserted between them. In this way, six sound files per prosodic grouping were created such that each file consisted of a different order of tokens. The average duration of tokens without an IPB was 1.76 s (range: 1.71–1.87 s), while it was 2.16 s (range: 2.13–2.2 s) for tokens with an IPB.

To match the sound files of the two prosodic types with respect to length the number of tokens within each file was varied. Files of the grouping with an IPB contained six tokens and had an average duration of 18.97 s. However, files of the condition without an IPB contained seven tokens (i.e., one random token was repeated), leading to an average duration of 19.32 s (range: 19.16–19.43 s).

Procedure

The HPP including a familiarization phase (Hirsh-Pasek et al., 1987; Jusczyk and Aslin, 1995) was used in this and all subsequent

experiments. During the experimental session, the infant was seated on the lap of a caregiver in the center of a test booth. The caregiver listened to music over headphones to prevent influences on the infant's behavior. Furthermore, she was instructed not to interfere with the infant's behavior during the experiment. The experimenter sat in an adjacent room, where she observed the infant's behavior on a mute video monitor and controlled the presentation of the visual and the acoustic signals by a button box.

Three lamps were fixed inside the booth: a green one on the center wall, and red ones on each of the side walls. Directly above the green lamp on the center wall was an opening for the lens of a video camera. Behind each of the red lights a JBL Control One loudspeaker was mounted. Each experimental trial started with the blinking of the green center lamp. When the infant oriented to the green lamp, it was turned off and one of the red lamps on a side wall started to blink. When the infant turned her head toward the red lamp, the speech stimulus was started, delivered via a Sony TA-F261R audio amplifier to the loudspeaker at the same side. The trial ended when the infant turned her head away for more than 2 s, or when the end of the speech file was reached. If the infant turned away for less than 2 s, the presentation of the speech file continued but the time spent looking away was not included in the total listening time. The whole session was digitally videotaped. The experimenter's coding was recorded and served for the calculation of the duration of the infant's headturns during the experimental trials (for comparable experimental setups, see Höhle et al., 2006; Höhle et al., 2009).

Half of the infants were familiarized to the sequences without an IPB (Group 1), while the other half were familiarized to the sequences including an IPB (Group 2). The familiarization was set such that at least 20 tokens in each familiarization condition were presented, that is, when familiarized to sequences without an IPB the familiarization lasted until the infant had accumulated 55 s of listening time. For the familiarization with an IPB the criterion was 63 s of accumulated listening time. This requirement was chosen to match the familiarization duration used in Nazzi et al. (2000).

Two different kinds of familiarization were chosen to control for a possible effect of the prosodic structure of the sequences presented. One might hypothesize that a familiarization to sequences without an internal IPB might be more effective. This is supported by Nazzi et al.'s (2000) findings that infants recognize word sequences that constitute a prosodic unit better than sequences that are a non-unit like our sequences with an IPB. Therefore, we planned to compare the data of both familiarization groups.

The familiarization was followed by a test phase that comprised 12 test trials. In six trials, the sound files without an internal IPB were presented, in the other six trials the sound files of the sequences with an IPB. Thus, half of the test trials contained exactly the same sound files that the infants had previously heard during familiarization, whereas the other half consisted of sound files with the type of prosodic grouping that had not been presented during familiarization. The test trials were grouped in three blocks of four trials each (two with and two without an internal IPB in a random order). Additionally, within each block the side of presentation of the sequences of the two prosodic types was counterbalanced so that the prosodic condition and the side of presentation were not associated. The duration of each experimental session depended on the infant's behavior and varied between 4 and 6 min.

RESULTS AND DISCUSSION

Mean listening times to the test trials with and without an IPB were calculated for each infant. Because all listening times were shorter than 18.97 s (the maximum trial length in the condition with an IPB), an adjustment of the listening times to the longer duration of the trials without an IPB was not necessary.

On average, infants listened for 6.32 s (SD = 2.39) to the familiarized prosodic grouping, and for 7.13 s (SD = 2.12) to the novel prosodic grouping (see **Figure 2**). This difference was significant, $t(23) = 2.30$, $p = 0.031$, two-tailed. Eighteen out of 24 infants had longer listening times to the novel test items. A repeated-measures ANOVA with the within-subject factor familiarity (familiarized versus new prosodic pattern) and the between-subject factor prosodic type (familiarization with versus without an internal IPB) showed a main effect of familiarity, $F(1,22) = 5.36$, $p = 0.030$, and a main effect of prosodic type, $F(1,22) = 4.44$, $p = 0.047$, but no significant interaction between prosodic type and familiarity, $F(1,22) = 1.237$, $p = 0.278$.

A further analysis of the data separated by prosodic type heard during familiarization was conducted. This analysis revealed a significant preference for novel test items in the group familiarized with the sequences without an IPB, $t(11) = 2.40$, $p = 0.035$. The mean listening time to the novel prosodic pattern was 6.48 s (SD = 1.23) and to the familiarized prosodic pattern 5.29 s (SD = 1.46). No such preference was present in the group familiarized with sequences including an IPB, $t(11) = 0.860$, $p = 0.408$. Infants in this group listened to the novel test trials on average for 7.77 s (SD = 2.64) and to the familiar test trials for 7.36 s (SD = 2.73). Mean listening times separated by familiarization group are depicted in **Figure 3**.

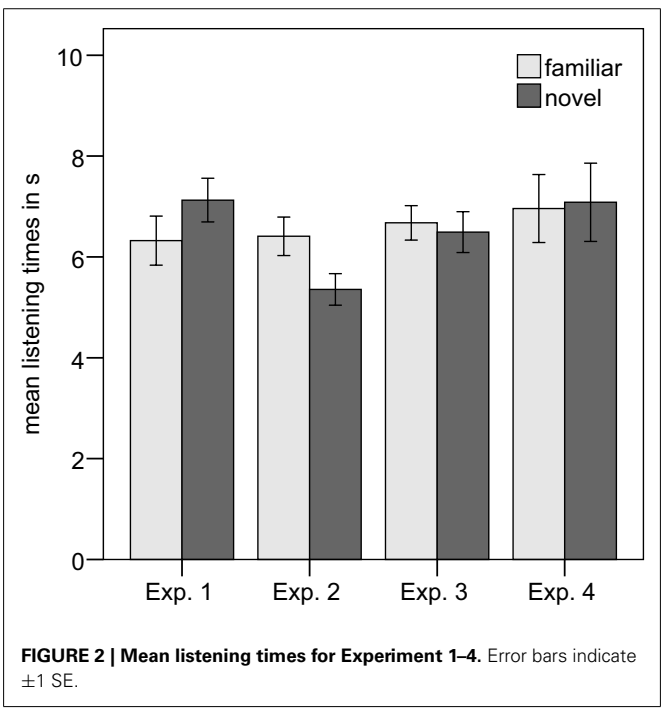


FIGURE 2 | Mean listening times for Experiment 1–4. Error bars indicate ±1 SE.

Experiment 1 served as a baseline study to ensure that the stimuli – sequences of names that have two different prosodic groupings – and our experimental design are suitable for studying the perception of prosodic boundary cues in German-learning infants. After being familiarized with one of the two prosodic phrasings, 8-month-old infants showed an overall preference for the novel prosodic grouping. Thus, German-learning infants are able to discriminate the two prosodic groupings. Even though we found no significant interaction between prosodic type and familiarity, a separate analysis of the two familiarization groups revealed that the difference in listening times was significant only when the familiarization strings did not have an internal IPB. Thus, discrimination of sequences with versus without an IPB was affected by the prosodic type heard during familiarization. How can we explain this effect? During familiarization the infants' task is to build up a representation of the auditory stimulus, to which they will compare the test stimuli. Presumably, infants can more easily build up representations of sequences without an IPB because these are easier to process and memorize, as Nazzi et al.'s (2000) study demonstrated. Secondly, both familiarization conditions differ in the number of IPs: stimuli played to Group 1 do not contain any prosodic boundary cue and, hence, only consist of a single IP, that is, one prosodic unit. In contrast, stimuli presented to Group 2 were sequences including an IPB, which splits the sequences into two separate IPs, that is, two prosodic units.

A study by Mandel et al. (1994) suggests that infants at the age of 2 months already perceive prosodic units as an organizational unit in the speech stream. Infants detected phonetic changes in word sequences when the words were prosodically grouped into a major linguistic unit, but not when the words were presented as isolated words in a list or as a fragment of two adjoining clauses. Mandel et al. argued that the organization of words in a prosodic unit helps infants to process and memorize the speech signal. For our experiment this implies that the representation of the familiarization sequence is built up more easily when the sequence consists of a single prosodic unit, like our sequences without an internal IPB. These are – compared to the sequences with an IPB – easier to process during the familiarization phase and thus can be better remembered during the test phase.

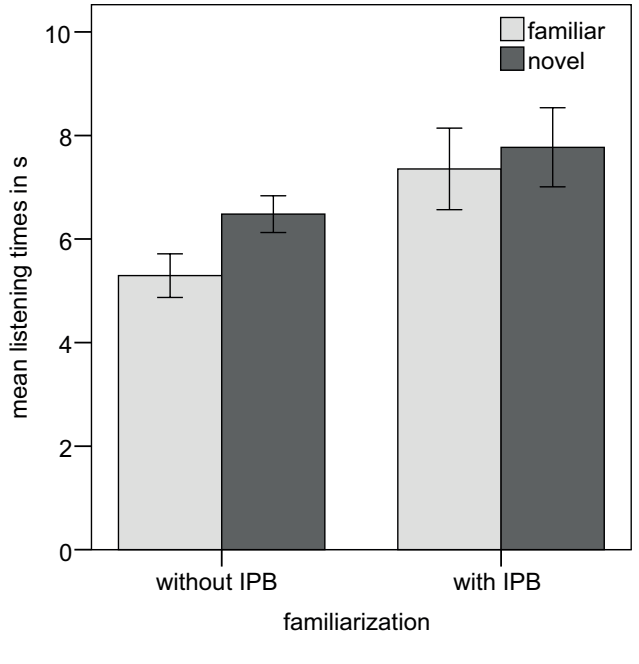


FIGURE 3 | Mean listening times for Experiment 1, separated by familiarization group.

The difference found in the two familiarization groups motivated a modification of the experimental design implemented in the subsequent experiments. As a full design with two separate familiarization conditions was not relevant to our research question, we decided to only use strings without an internal IPB as familiarization stimuli. In doing so, we chose the condition that yielded the most robust results.

In sum, Experiment 1 showed that 8-month-old German-learning infants are sensitive to the presence of an IPB in short coordinated sequences of names when the IPB is marked by the acoustic correlates of the main prosodic boundary cues pitch change, preboundary lengthening, and pause. Hence, not only clauses – like those that were used in previous studies (e.g., Hirsh-Pasek et al., 1987; Nazzi et al., 2000; Seidl, 2007; Schmitz, 2008) – are suitable for investigating infants' sensitivity to prosodic boundaries. Rather, coordinate structures, which can be carefully controlled for phonological parameters, may serve as stimuli to characterize the impact of each prosodic boundary cue in a discrimination task.

The subsequent experiments contain only one kind of familiarization, namely the familiarization to sequences without an IPB. In these experiments, the number of prosodic boundary cues in the stimuli is reduced stepwise. This is done to determine whether infants' discrimination ability remains or is disturbed when different constellations of prosodic boundary cues are given.

EXPERIMENT 2: SENSITIVITY TO PITCH CHANGE AND PREBOUNDARY LENGTHENING

In Experiment 2, we investigated infants' sensitivity to two of three prosodic boundary cues, namely pitch change in combination with preboundary lengthening. Specifically, we asked whether the pitch change and the lengthening of the preboundary vowel suffice as boundary cues or whether the pause is a necessary prosodic boundary cue. If pause is a necessary cue for the discrimination of two prosodic groupings, infants were not expected to show significantly different listening times to novel versus familiar test items. In contrast, if pitch change and preboundary lengthening are sufficient cues, we expected a significant listening preference.

MATERIALS AND METHODS

Participants

Sixteen 8-month-old infants (eight girls) were tested. The mean age was 8 months, 11 days (range: 8 months, 1 day–9 months, 8 days). Ten additional infants were tested but their data were not included in the analysis for the following reasons: crying or fussiness (6), mean listening times of less than 3 s per condition (2), and noise (2).

Stimuli

In Experiment 2, again sequences with and without an internal prosodic boundary cues were presented. The stimuli without any boundary cues were the same as the stimuli used in Experiment 1. The sequences containing a pitch change and preboundary lengthening were construed from the sequences without an IPB by acoustic manipulation – according to the values that had been measured in the sequences with an IPB recorded for Experiment 1. Hereby, we created two types of stimuli that only differed in fundamental frequency and duration at the critical boundary position, that is, on the second name. Apart from that, the sequences of both prosodic types were acoustically identical.

The manipulation was carried out with the PRAAT software. For duration, the final vowel of the target name was lengthened to 180%. This factor was chosen because in Experiment 1 the crucial vowel was on average 1.8 times longer in sequences with an IPB than in sequences without an IPB.

For the manipulation of the pitch contour, first the sequences without an IPB were stylized (two semitones), that is, the number of pitch points was reduced. The reference values of the fundamental frequency were measured on the target name in the sequences with an internal IPB from Experiment 1 – at the midpoints of the four segments [l], [l], [l], and [i]and at the position of the maximum pitch present on the preboundary vowel. Then, pitch points with the mean values at these time points were inserted at the same positions into the sequences without an IPB. We obtained new stimuli for the prosodic type with pitch change and preboundary lengthening. They contained a natural sounding pitch rise of 212 Hz (13.65 semitones) and a preboundary lengthening with a factor of 1.8. The pitch contour and wave form of a sequence with manipulated pitch and lengthening are depicted in **Figure 1C**.

To avoid comparing natural with acoustically manipulated stimuli we carried out a slight acoustic manipulation of the sequences without an IPB as well: a stylization of the pitch contour (two semitones). After acoustic manipulation, all sequences were resynthesized using the PSOLA function in PRAAT.

Six differently ordered speech files with the same set of tokens in each prosodic condition were created from the acoustically manipulated sequences. The speech files of the condition without an IPB contained seven tokens (i.e., one random token was repeated) and had an average duration of 18.33 s (range: 18.23–18.43 s). The files of the condition with added pitch and lengthening cues also contained seven tokens (again one random token was repeated) and had an average duration of 18.81 s (range: 18.79–19.01 s).

Procedure

The procedure was the same as in Experiment 1 with a modification concerning the familiarization phase. Infants in Experiment 2 were only familiarized to sequences without an IPB, but not to sequences with boundary cues. The familiarization lasted until at least 20 sequences had been presented leading to a minimum duration of 52 s.

RESULTS AND DISCUSSION

Infants oriented on average for 6.41 s (SD = 1.53) to the familiarized prosodic grouping, and for 5.36 s (SD = 1.25) to the novel prosodic grouping (see **Figure 2**). This difference was significant, $t(15) = -3.59$, $p = 0.003$, two-tailed. Thirteen of 16 infants had longer listening times to the familiar test items.

Experiment 2 tested whether German-learning infants still perceive an IPB when only a subset of prosodic cues, pitch change, and preboundary lengthening, is present. A significant familiarity effect was displayed indicating that the infants were able to discriminate the stimuli of the two prosodic patterns in Experiment 2. Interestingly, the direction of preference reversed from Experiment 1 to Experiment 2. While infants in Experiment 1 preferred to listen to the novel prosodic pattern, in Experiment 2 the familiar pattern was preferred. According to the model by Hunter and Ames (1988), this shift in preference can be explained by higher task demands in Experiment 2. Hunter and Ames claimed that the direction of preference is affected by three factors: age, duration of familiarization, and task difficulty. As we held the first two factors constant, we assume that the shift in preference from Experiment 1 to Experiment 2 is caused by increased task difficulty: if only two instead of three prosodic cues mark the difference between the stimuli, it becomes harder to distinguish both conditions as less information is available. In turn, the task of discriminating the two prosodic patterns is more difficult and leads infants to a preference for the familiar sequences. Hence, for German 8-month-olds pitch change and preboundary lengthening in combination are sufficient. Pause is not a necessary boundary cue, however, processing different prosodic groupings without the information provided by the pause cue seems to be more demanding.

EXPERIMENT 3: SENSITIVITY TO PITCH CHANGE

Experiment 2 showed that German infants are able to discriminate the two prosodic groupings when a boundary is signaled by a pitch change and preboundary lengthening in combination. In Experiment 3 we asked whether only one cue, the pitch change, is sufficient for German 8-month-olds to perceive a boundary.

MATERIALS AND METHODS

Participants

Seventeen infants (seven girls) were tested. The mean age was 8 months, 13 days (range: 8 months, 4 days–8 months, 29 days). Six additional infants were tested but their data were not included in the analysis for the following reasons: crying or fussiness (4), and mean listening times of less than 3 s per condition (2).

Stimuli

In Experiment 3, sequences without an IPB and sequences with an inserted pitch rise were contrasted. For the condition without an IPB the same sequences as in Experiment 2 were used. For the condition with added pitch cue a manipulation of the pitch contour was carried out similar to that in Experiment 2: a pitch rise was inserted on the second name of the six sequences without an IPB. In contrast to the stimuli in Experiment 2 no duration manipulation was conducted. Thus, the pitch change with the high boundary tone was the only signal of an IPB (see **Figure 1D**). From these pitch-manipulated sequences six differently ordered speech files were created with seven tokens per file (i.e., one of the six exemplars was randomly repeated). The speech files of the condition without an IPB were the same as in Experiment 2. The average duration of the speech files was the same in both prosodic

conditions as there was no duration manipulation ($M = 18.33$ s; range: 18.23–18.43 s).

Procedure

The procedure was the same as in Experiment 2. Infants were familiarized to sequences without an IPB until at least 20 sequences had been presented. This led to a minimum duration of 52 s.

RESULTS AND DISCUSSION

Infants listened on average for 6.68 s (SD = 1.41) to the familiarized prosodic grouping and for 6.49 s (SD = 1.67) to the novel prosodic grouping (see **Figure 2**). This difference was not significant, $t(16) = 0.522$, $p = 0.609$. Ten of 17 infants had longer listening times to the familiar test items.

In Experiment 3 only a pitch rise indicated a different prosodic grouping. Neither a pause nor lengthening of the preboundary vowel was present. The infants did not differentiate between sequences with added pitch cue and sequences without an IPB. Hence, the presence of a pitch change alone is not sufficient for German infants to perceive a prosodic boundary.

Apart from the specific cue constellation presented, Experiment 3 generally differs from Experiment 2 with regard to the number of IPB cues provided in the stimuli, that is, whereas in Experiment 2 two boundary cues were available, in Experiment 3 we only inserted one cue. Hereby, the boundary is generally less marked in Experiment 3. The failure to discriminate the two conditions could hence be due to the mere number of cues being relevant for boundary detection, instead of the specific kind of cue or cue constellation (but see General Discussion).

EXPERIMENT 4: SENSITIVITY TO PREBOUNDARY LENGTHENING

German 8-month-olds are able to perceive an IPB when a pitch change and preboundary lengthening occur together (Experiment 2) but not when only a pitch change is present (Experiment 3). Experiment 4 tested whether preboundary lengthening as a single boundary cue is sufficient.

MATERIALS AND METHODS

Participants

Sixteen infants (eight girls) were tested. The mean age was 8 months, 10 days (range: 7 months, 30 days–8 months, 29 days). Six additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (1), crying or fussiness (2), and mean listening times of less than 3 s per condition (3).

Stimuli

In Experiment 4, sequences without an IPB and sequences with inserted preboundary lengthening were contrasted. For the condition without an IPB the same sequences as in Experiment 2 were used. For the condition with inserted preboundary lengthening a manipulation of the duration of the final vowel was carried out similar to that in Experiment 2: in six exemplars of the sequences without an IPB the final vowel was lengthened to 180% (see **Figure 1E** for an example). The sequences were concatenated in a random order to speech files.

The speech files of the condition without an IPB were the same as in Experiment 2. They contained six different tokens and had an average duration of 18.33 s (range: 18.23–18.43 s). The speech files of the condition with preboundary lengthening also contained six tokens and lasted for 18.89 s on average (range: 18.79–19.01 s).

Procedure

The procedure was the same as in Experiment 2. Infants were familiarized to sequences without an IPB until at least 20 sequences had been presented. This led to a minimum duration of 52 s.

RESULTS AND DISCUSSION

The listening time in one individual trial of the condition with the lengthening cue exceeded the duration of the longest speech file in the condition without an IPB. Therefore, the listening time in this trial was reduced to the maximum trial length of sequences without an IPB, which was 18.43 s.

The mean listening time to the familiarized prosodic grouping was 6.96 s (SD = 2.7) and to the novel pattern 7.08 s (SD = 3.1; see **Figure 2**). This difference was not significant, $t(15) = -0.221$, $p = 0.828$. Nine of 16 infants had longer listening times to the familiar test trials.

Experiment 4 suggests that preboundary lengthening as a single cue is not sufficient to trigger the perception of a prosodic boundary in German 8-month-old infants. However, in combination with a pitch cue, as tested in Experiment 2, it becomes an effective boundary marker. As for Experiment 3, we also have to consider that the insufficiency of preboundary lengthening alone compared to its effectiveness in combination with a pitch change could also be explained by the number of cues (but see General Discussion).

GENERAL DISCUSSION

The aim of the present study was to specify the relevance of pitch change and preboundary lengthening as combined and as single prosodic cues in German-learning infants' perception of major prosodic boundaries. Experiment 1 showed that 8-month-olds are able to discriminate different prosodic groupings – specifically, familiar sequences without a prosodic boundary from unfamiliar sequences with a prosodic boundary – when the boundary is clearly marked by all three boundary markers.

In further experiments stimuli were acoustically manipulated with respect to pitch and preboundary lengthening. We focused on investigating infants' processing of boundaries in the absence of the pause cue. Pauses are perceptually highly salient and we assumed that in a discrimination task like ours, infants would easily detect the presence of a pause. Especially in short coordinated structures as used in this study pauses are easy to notice as they constitute approximately a fourth of the overall duration of the sequence. Furthermore, we know from other studies (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Schmitz, 2008; Wellmann et al., in preparation) that infants by the age of 6–10 months are highly sensitive to pauses.

When we manipulated the stimuli such that only a pitch change and preboundary lengthening indicated the presence of an IPB (Experiment 2), infants still detected the boundary. We concluded that pause is not necessary, but it seems to ease infants' processing. This was indicated by a shift in preference from a novelty

effect in Experiment 1 to a familiarity effect in Experiment 2. We argued that higher task demands in Experiment 2 are responsible for the preference for familiar stimuli (see Hunter and Ames, 1988). In Experiments 3 and 4 the impact of the single prosodic cues pitch change and preboundary lengthening were tested. Sequences with pitch as a single cue (Experiment 3) were not differentiated from sequences without any boundary cue. Nor was preboundary lengthening alone (Experiment 4) sufficient to trigger the perception of a boundary. This might indicate that infants do not take single cues into account, as cue combinations are very frequent whereas the occurrence of single cues is rather rare (Peters et al., 2005). However, the weighting of prosodic boundary cues might depend on the strength of the specific cue, that is, its phonetic magnitude. When implementing the cues in Experiments 2–4 we used the acoustic values measured in natural sequences that contained all three cues. It is conceivable that the specific strength of each cue in production depends on the constellation of cues, that is, when a cue occurs alone or in a subset its magnitude might be larger than when it occurs together with all main cues. Thus, it remains possible that a larger pitch rise in Experiment 3 or longer preboundary lengthening in Experiment 4 might have been sufficient to trigger boundary perception by a single cue. We also considered the reduced number of boundary cues as an explanation for the insufficiency of the single cues compared to their occurrence in combination. However, in a study with 6-month-olds (Wellmann et al., in preparation) we found that pause, but not a pitch change, was sufficient though the number of cues was kept constant. Therefore, we argue that the specific cue constellation, and not the number of cues, is decisive for the detection of a boundary.

Another restriction when interpreting the data concerns the fact that the stimuli presented during the test phase differed across experiments in the presence or absence of boundary cues, but potentially also with respect to their naturalness. Thus, infants' different performance patterns could be due to infants' disliking of one kind of stimuli in one but not the other experiment. Pitch change or preboundary lengthening might be effective as single cues when produced naturally, but infants could find stimuli with a single inserted cue odd, thus, would not pay attention and consequently fail to discriminate test stimuli. Hereby, infants' cue weighting and their liking of stimuli might be confounded. However, when editing the stimuli with inserted cues, we took special care to create stimuli that are perceptually distinguishable, but comparably natural sounding in all experiments. Hence, we rather argue that the different performance patterns suggest that perception depends on the specific cue constellation: pitch change and preboundary lengthening in combination are sufficient to trigger boundary perception in German 8-month-old infants and hence, pause is not a necessary cue. Whether pitch change or preboundary lengthening is a necessary cue cannot be answered from these experiments. Still, both of them are not sufficient as single boundary cues: when they occur individually, stimuli are not differentiated from sequences without prosodic boundary marking – at least if the single cues are presented with the same acoustic parameters as when they occur combined.

In summary, two parallels of these findings to previous research are obvious: first, they resemble findings on the processing of these cues in German adults (Holzgrefe et al., 2012), and secondly, they show a strong overlap with the findings by Seidl (2007) for English-learning infants. Both parallels will be discussed separately in the following section.

To our knowledge, the present study is the first that has used the same material with infants that had previously been used with adults in a prosodic judgment task (Holzgrefe et al., 2012). In this study, adults were asked to interpret the aurally presented sequences as having no internal boundary [a and b and c], or as having an internal boundary after the second name [a and b] [and c]. The effects that the specific prosodic cues had on these decisions mirror the pattern we found with the German-learning infants: sequences that provided pitch change or preboundary lengthening as single cues either were judged as having no boundary or listeners performed at chance level. However, when a combination of pitch change and preboundary lengthening occurred in the sequence, they were clearly identified as consisting of two prosodic units. Moreover, infants' behavior in our study is in line with the distribution of prosodic boundary cues found in spontaneous speech of German adults (Peters et al., 2005): first, the majority of IPBs are marked by a coalition of cues. Secondly, compared to pitch change and preboundary lengthening, pause is a rather rare marker of IPBs. This suggests that pause is not reliable and listeners should be able to cope without it.

It is rather surprising that the experiments with the adults and the infants show exactly the same pattern of results with respect to cue effectiveness even though the tasks that had to be performed by the participants were clearly different: while the adults had to exploit the acoustic information to assign a prosodic phrasing to the utterances, the children only had to discriminate between the different prosodic contours. If we consider these findings in the light of Johnson and Seidl's (2008) assumption that a language-specific weighting of prosodic boundary cues takes place, our results suggest that the German 8-month-olds have already attuned to the German system as they show a parallel pattern of responding to the cues to that of adults. Furthermore, our results indicate that cue weighting leads to a perceptual reorganization that has an effect on the ability to discriminate verbal materials containing the relevant phonetic information.

Additional empirical support for this conclusion is required and may come from crosslinguistic studies that compare children learning languages that exhibit relevant differences in the acoustic instantiation of prosodic boundary cues. In addition, one may compare the current findings to the performance of younger infants. This would allow a developmental trajectory to be followed from a language-general perceptual system that is not yet fully adapted to the properties of the phonological system of the ambient language to a language-specific perceptual system that is attuned to these properties.

Crosslinguistic research in the area of the processing of prosodic boundaries is still sparse. Additionally, a crosslinguistic comparison may be impeded because of differences in the experimental material of our and previous studies: we used coordinated noun phrases, whereas previous studies on English and Dutch (Seidl, 2007; Johnson and Seidl, 2008) presented clauses. Even though both kinds of material have a different syntactic structure, the prosodic structure is similar. Clause boundaries in Seidl's (2007)

and Johnson and Seidl's (2008) studies coincide with IPBs. In our sequences of names each name forms a phonological phrase. To convey the intended internal grouping, that is, separating the first two names from the third, our speaker needed to group the first two names into a larger prosodic unit by producing a larger prosodic boundary after the second name. In line with current models of prosodic phrasing (Gussenhoven, 1992; Truckenbrodt, 1999, 2007) we argue that therefore the first two names of the internally grouped sequences constitute an intonation phrase. This account is supported by the acoustic analysis we carried out on the respective IPB cues. Hence, even though the stimuli differ across studies, the prosodic level under investigation is comparable allowing us to compare ours and previous findings crosslinguistically. German infants' behavior compared to American 6-month-olds' (Seidl, 2007) shows no indications of crosslinguistic variation. Like the German infants in our study, the 6-month-old American infants did not provide any evidence of detecting a boundary when it was solely cued by pitch change or preboundary lengthening, but only if a combination of these cues occurred in the stimuli. However, given the high overlap in the prosodic systems of English and German, the missing crosslinguistic variation could simply reflect the fact that the two languages do not differ crucially in the area under investigation.

However, a comparison of the results of the experiments with German- and English-learning infants on the one hand and Dutch-learning infants on the other gives some indications of crosslinguistic variation. While the 6-month-old Dutch infants tested by Johnson and Seidl (2008) needed a pause to detect the prosodic boundary, the German and American infants were able to perceive a boundary with pitch change and preboundary lengthening only. This might indicate a true crosslinguistic variation between German and Dutch and English and Dutch.

Regarding the difference observed between the German and Dutch infants' reliance on the prosodic cues, we have to take into account that it may arise from a purely developmental change. The Dutch infants were 2 months younger than the German ones. It is thus possible that older Dutch babies will be able to detect prosodic boundaries that are not marked by a pause. In addition, it is feasible that German 6-month-olds will not detect a prosodic boundary when no pause is present. This would suggest a developmental change in prosodic cue perception from 6 to 8 months in Dutch and German infants. Future studies comparing German and Dutch infants of the same age will have to disentangle whether the observed difference is due to crosslinguistic variation or is caused by developmental aspects.

Regarding the difference between English- and Dutch-learning infants' sensitivity to prosodic boundary markers, Johnson and Seidl (2008) took this as an indication of the emergence of a language-specific cue weighting, as the results reflected differences in the way that the prosodic boundaries were marked in the Dutch material and the English material, with a longer pause but smaller pitch reset in Dutch as compared to English. Additional evidence for this view comes from the study by Seidl and Cristià (2008), which revealed that younger, 4-month-old English-learning infants only rely on a combination of all three cues. The authors argued that younger infants' perception reflects holistic mechanisms that do not depend on language-specific factors. Later in development, infants follow an analytical segmentation strategy that implies language-specific processing (Seidl, 2007). This indicates a developmental shift from 4 to 6 months of age. Based on this reasoning, a further study with German-learning infants younger than the age tested in our study would be necessary to provide more evidence for this kind of developmental change.

Furthermore, it would be highly interesting to look at languages in which the way prosodic boundaries are marked is more different than in the closely related languages English, German, and Dutch. The advantage of the linguistic material used in this study is that it can easily be adapted to other languages. One relevant language to look at would be French. Two features might lead to a greater saliency of preboundary lengthening. First, French does not have lexical stress and thus has no pitch accents. In languages without pitch accents syllable duration is much less varied within phrases. Secondly, French is a syllable-timed language. The inventory of syllable types is smaller in syllable-timed than in stress-timed languages. Smaller syllable inventories comprise simpler syllables, whereas languages with more syllable types tend to have heavier syllables (Ramus et al., 2000). Consequently, syllable duration is less varied in syllable-timed than in stress-timed languages. Both aspects, no lexical stress and a smaller syllable inventory, lead to the assumption that whenever syllables are lengthened, namely phrase-finally, this provides a clear acoustic contrast to phrase-internal syllable durations. Empirical evidence for a greater phonetic extent of preboundary lengthening comes from a production study with German and French adults by Féry et al. (2011). They found that the difference in duration between phrase-internal and phrase-final words was significantly higher in French speakers than in German speakers, who used preboundary lengthening to a smaller degree. Thus, preboundary lengthening might be a more important cue for the perception of prosodic boundaries in French adults and infants compared to the speakers and learners of the languages looked at so far. Again, this question is left open for further research.

Also, tone languages that deploy lexical tones on each syllable should be studied (e.g., Chinese). Where pitch is used to encode lexical distinctions, its role in encoding boundaries is reduced (Fernald and McRoberts, 1996). Therefore, one can hypothesize that infants acquiring such a tone language focus more on other boundary cues, like pause and preboundary lengthening. Pitch would then be perceptually weighted less in this kind of tone language than in non-tone languages.

The results of our study contribute in an important way to our understanding of how prosodic information may support children's early phrasing of incoming linguistic material and hence provide further evidence for the prosodic bootstrapping account. Fernald and McRoberts (1996) outlined the unreliability of prosodic cues due to their multiple functions. Our results as well as Seidl's (2007) data show that infants only consider a combination of at least two cues as a marker for a prosodic boundary – and even younger infants rely on the convergence of all cues that serve as prosodic boundary markers (Seidl and Cristià, 2008). With these constraints infants have a powerful mechanism to make specific use of these correlations of cues as boundary markers and to ignore the same acoustic information when it is not accompanied by correlating cues.

ACKNOWLEDGMENTS

This research was funded by a grant from the German Science Foundation, priority program 1234, to Barbara Höhle and Isabell Wartenburger (HO 1960/13-1; FR 2865/2-1). Isabell Wartenburger is supported by the Stifterverband für die Deutsche Wissenschaft (Claussen-Simon-Stiftung). We thank Tom Fritzsche for technical assistance in the Potsdam BabyLab and Anne Beyer, Sophia Fischer, Babette Graf, Mareike Orschinsky, Teresa Leitner, and Marie Zielina for their help in recruiting and testing the infants. We are grateful to Romy Räling for stimuli production. Thanks to all parents and their children who participated in this study.

REFERENCES

Aasland, W. A., and Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: a comparison of processing by left- and right-hemisphere brain-damaged individuals. *Brain Lang.* 87, 385–399.

Boersma, P., and Weenink, D. (2011). *Praat: Doing Phonetics by Computer [Computer Program]*, Version 5.3.03. Available at: http://www.praat.org/ (accessed November 21, 2011).

Collins, B., and Mees, I. M. (1981). *The Phonetics of English and Dutch*. Leiden: Brill.

Cooper, W., and Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Fernald, A., and McRoberts, G. (1996). "Prosodic bootstrapping: a critical analysis of the argument and the evidence," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Mahwah, NJ: Lawrence Erlbaum Associates), 365–388.

Féry, C., Hörnig, R., and Pahaut, S. (2011). "Correlates of Phrasing in French and German from an Experiment with Semi-Spontaneous Speech," in *Intonational Phrasing in Romance and Germanic*, eds C. Gabriel and C. Lleó (Amsterdam: John Benjamins), 11–41.

Gerken, L., Jusczyk, P. W., and Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition* 51, 237–265.

Gleitman, L. R., and Wanner, E. (1982). "Language acquisition: the state of the state of the art," in *Language Acquisition: The State of the Art*, eds E. Wanner and L. R. Gleitman (Cambridge: Cambridge University Press), 3–48.

Gussenhoven, C. (1992). "Sentence accents and argument structure," in *Thematic Structure: Its Role in Grammar*, ed. I. Roca (Berlin: Foris), 79–106.

Hayashi, A., and Mazuka, R. (2002). Developmental change in infants' preferences for continuousness in Japanese speech. *Paper presented at the 13th Biennial International Conference on Infant Studies*, Toronto.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P., Wright Cassidy, K., Druss, B., and Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition* 26, 269–286.

Hirst, D., and Di Cristo, A. (1998). "A survey of intonation systems," in *Intonation systems. A Survey of Twenty Languages*, eds D. Hirst and A. Di Cristo (Cambridge: Cambridge University Press), 1–44.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Develop.* 32, 262–274.

Höhle, B., Schmitz, M., Santelmann, L. M., and Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: evidence for lexical and structural influences on children's early processing capacities. *Lang. Learn. Dev.* 2, 277–300.

Holzgrefe, J., Schröder, C., Höhle, B., and Wartenburger, I. (2012). "Processing of prosodic boundary cues as revealed by event-related brain potentials," in *Poster Presented at the 13th Conference on Laboratory Phonology (LabPhon 13)*, Stuttgart.

Hunter, M. A., and Ames, E. W. (1988). "A multifactoral model of infant preferences for novel and familiar stimuli," in *Advances in Infancy Research*, eds L. P. Lipsitt and C. Rovee-Collier (Norwood, NJ: Ablex), 69–95.

Johnson, E. K., and Seidl, A. (2008). Clause segmentation by 6-month-old infants: a crosslinguistic perspective. *Infancy* 13, 440–455.

Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–23.

Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G., Kennedy, L. J., Woodward, A., and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cogn. Psychol.* 24, 252–293.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., and Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *J. Child Lang.* 16, 55–68.

Mandel, D. R., Jusczyk, P. W., and Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition* 53, 155–180.

Mattock, K., and Burnham, D. (2006). Chinese and English infants' tone perception: evidence for perceptual reorganization. *Infancy* 10, 241–265.

Mattock, K., Molnar, M., Polka, L., and Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition* 106, 1367–1381.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., and Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: effects of prosodic well-formedness. *Infancy* 1, 123–147.

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Peters, B. (2005). "Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache," in *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)*, eds K. J. Kohler, F. Kleber, and B. Peters (Kiel: IPDS), 203–345.

Peters, B., Kohler, K. J., and Wesener, T. (2005). "Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache," in *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)*, eds K. J. Kohler, F. Kleber, and B. Peters (Kiel: IPDS), 143–184.

Polka, L., and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 421–435.

Price, P. J, Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* 90, 2956–2970.

Ramus, F., Nespor, M., and Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition* 75, 265–292.

Sanderman, A., and Collier, R. (1997). Prosodic phrasing and comprehension. *Lang. Speech* 40, 391–409.

Schmitz, M. (2008). *The Perception of Clauses in 6- and 8-month-old German-Learning Infants: Influence of Pause Duration and the Natural Pause Hierarchy*. Potsdam: Universitätsverlag.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *J. Acoust. Soc. Am.* 71, 996–1007.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *J. Mem. Lang.* 57, 24–48.

Seidl, A., and Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Dev. Sci.* 11, 596–606.

Soderstrom, M., Kemler Nelson, D. G., and Jusczyk, P. W. (2005). Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behav. Dev.* 28, 87–94.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., and Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Mem. Lang.* 49, 249–267.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *J. Acoust. Soc. Am.* 64, 1582–1592.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguist. Inquiry* 30, 219–255.

Truckenbrodt, H. (2007). "The syntax-phonology interface," in *The Cambridge Handbook of Phonology*, ed. Paul de Lacy (Cambridge: Cambridge University Press), 435–456.

Vaissière, J. (1983). "Language-independent prosodic features," in *Prosody: Models and Measurements*, eds A. Cutler and D. R. Ladd (Berlin: Springer), 53–66.

Vaissière, J., and Michaud, A. (2006). "Prosodic constituents in French: a data-driven approach," in *Prosody and Syntax*, eds I. Fónagy, Y. Kawaguchi, and T. Moriguchi (Amsterdam: John Benjamins), 47–64.

Venditti, J. J., Jun, S.-A., and Beckman, M. E. (1996). "Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Mahwah, NJ: Lawrence Erlbaum Associates), 287–311.

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Develop.* 7, 49–63.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717.

Willems, N. (1982). *English Intonation from a Dutch Point of View*. Dordrecht: Foris.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 July 2012; accepted: 10 December 2012; published online: 31 December 2012.

Citation: Wellmann C, Holzgrefe J, Truckenbrodt H, Wartenburger I and Höhle B (2012) How each prosodic boundary cue matters: Evidence from German infants. Front. Psychology ***3****:580. doi: 10.3389/fpsyg.2012.00580*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 30 October 2012
doi: 10.3389/fpsyg.2012.00451

Prosodic cues to word order: what level of representation?

Carline Bernard [1] ***and Judit Gervain*** [1,2] ***

[1] *Laboratoire Psychologie de la Perception (UMR 8158), Université Paris Descartes, Sorbonne Paris Cité, Paris, France*
[2] *Laboratoire Psychologie de la Perception (UMR 8158), CNRS, Paris, France*

Edited by:
Claudia Männel, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Mohinish Shukla, University of Massachusetts Boston, USA
Jean-Remy Hochmann, Harvard University, USA

***Correspondence:**
Judit Gervain, Laboratoire Psychologie de la Perception (UMR 8158), CNRS & Université Paris Descartes, 45 rue des Saints Pères, 75006 Paris, France.
e-mail: judit.gervain@parisdescartes.fr

Within language, systematic correlations exist between syntactic structure and prosody. Prosodic prominence, for instance, falls on the complement and not the head of syntactic phrases, and its realization depends on the phrasal position of the prominent element. Thus, in Japanese, a functor-final language, prominence is phrase-initial, and realized as increased pitch (^ ***Tōkyō*** *ni* "Tokyo to"), whereas in French, English, or Italian, functor-initial languages, it manifests itself as phrase-final lengthening (*to* ***Rome***). Prosody is readily available in the linguistic signal even to the youngest infants. It has, therefore, been proposed that young learners might be able to exploit its correlations with syntax to bootstrap language structure. In this study, we tested this hypothesis, investigating how 8-month-old monolingual French infants processed an artificial grammar manipulating the relative position of prosodic prominence and word frequency. In Condition 1, we created a speech stream in which the two cues, prosody and frequency, were aligned, frequent words being prosodically non-prominent and infrequent ones being prominent, as is the case in natural language (functors are prosodically minimal compared to content words). In Condition 2, the two cues were misaligned, with frequent words carrying prosodic prominence, unlike in natural language. After familiarization with the aligned or the misaligned stream in a headturn preference procedure, we tested infants' preference for test items having a frequent word initial or a frequent word final word order. We found that infants' familiarized with the aligned stream showed the expected preference for the frequent word initial test items, mimicking the functor-initial word order of French. Infants in the misaligned condition showed no preference. These results suggest that infants are able to use word frequency and prosody as early cues to word order and they integrate them into a coherent representation.

Keywords: prosodic bootstrapping, word order, French, language acquisition

INTRODUCTION

The languages of the world show considerable variation in word order. In Japanese, for instance, the object precedes the verb [*ringo-wo taberu* (apple.acc[1] eat) "eat an apple"] and postpositions follow their nouns [*Tokyo kara* (Tokyo from) "from Tokyo"] etc. In French, by contrast, the object follows the verb [*manger une pomme* (eat.inf[2] an apple) "eat an apple"] and prepositions precede their nouns [*de Paris* (from Paris) "from Paris"]. As the examples suggest, this variation is not random: most languages conform to a basic word order type, which is usually characterized by the relative order of the object and the verb or by the typical position of function words within phrases (Greenberg, 1978; Dryer, 1992). Thus, Japanese is an OV or functor-final language, while French is VO or functor-initial. Crucially, the order of words in several phrase types correlates with that of the object and the verb. In OV languages, adpositions follow nouns, subordinate clauses precede the main verb and possessors precede the possessed. The opposite orders are observed in VO languages.

This knowledge is fundamental to language use, as it allows the efficient production and comprehension of multiword utterances. Indeed, infants know the basic word order of their mother tongue from their earliest multiword productions (Brown, 1973) and perceptually recognize word orders typical of their native language even earlier (e.g., Weissenborn et al., 1996; Höhle et al., 2001; Gervain et al., 2008). Importantly, the early mastery of word order might have a facilitatory effect on language acquisition, allowing young infants to correctly assign a grammatical function to novel structures or words they encounter.

How is word order learned? The purpose of the current paper is to contribute to a growing literature on the bootstrapping account of word order acquisition (Mazuka, 1996; Morgan and Demuth, 1996; Weissenborn et al., 1996; Gervain et al., 2008; Shukla and Nespor, 2010). Bootstrapping is a learning mechanism whereby the learner infers abstract, structural, perceptually unavailable properties of the target language on the basis of perceptually available cues in the input, which are correlated with the former (Morgan and Demuth, 1996). Under this view, the acquisition of a rudimentary, but already abstract representation of basic word order starts very early on, even before, and independently of the

[1] acc: accusative case
[2] inf: infinitive

acquisition of a sizeable lexicon, on the basis of perceptually available cues such as word frequency and prosody, which correlate with word order. This bootstrapping account belongs to a larger family of theories on language development that assume language acquisition to rely on abstract structural representations from early on (Pinker, 1984; Gleitman et al., 1988; Fisher et al., 1991). These accounts contrast with the lexicalist view (Akhtar and Tomasello, 1997; Tomasello, 2000), according to which the knowledge of word order is initially linked to specific lexical items and becomes abstract only later, possibly only in the mature grammar.

Several recent studies have provided evidence that prelexical infants possess at least a simple representation of the basic word order of their native language. Specifically, two cues have been identified that infants might be able to exploit as indicators of the word order type of their mother tongue: word frequency and phrasal prosody.

Frequency-based word order bootstrapping relies on the observation that natural languages have two general word classes (Fukui, 1986; Abney, 1987): function words (articles: *the*, *a*, adpositions: *in*, *on*, *to*, pronouns: *he*, *she*, *they* etc.), indicating the morphosyntactic structure of sentences, and content words, carrying lexical meaning. Function words are typically more frequent than content words. Indeed, the 30–50 most frequent words are usually functors in all of the languages that have been studied in both adult- and child-directed speech (Kucera and Francis, 1967; Morgan et al., 1996; Gervain et al., 2008). Further, these frequent words often occupy utterance-initial and utterance-final positions, known to be perceptually salient and recognized even by young infants (Aslin et al., 1996). Importantly, the specific position they occupy correlates with word order: in OV languages, functors tend to appear phrase-finally, whereas they are phrase-initial in VO languages (Gervain et al., 2008). Thus tracking the most frequent words and their positions relative to salient utterance boundaries provides a cue to word order. It has been shown that 8-month-old monolingual Japanese and Italian infants are able to use this cue in an artificial grammar learning task to bootstrap the opposite word orders that characterizes their native languages (OV for Japanese, VO for Italian). In this study, infants were familiarized with an artificial grammar consisting of strictly alternating frequent and infrequent nonce words. As no phase-information is given (the beginning and the end of the stream are ramped in amplitude), the structure of this grammar is ambiguous between a frequent word initial (FI) and a frequent word final (FF) parse. In the test phase, infants are tested on their preference for FI and FF sequences. As predicted, Italian infants preferred the FI items, while Japanese babies looked longer at the FF items, reflecting the typical word order of these two languages. It is important to note that both FI and FF sequences were taken from the familiarization stream, so they were both familiar to infants. The only difference between the two groups that could explain the observed differences in their preferences during test was the opposite word orders of their mother tongues. This study thus shows that 8-month-old infants already have an expectation about the word order of their native language in terms of the relative position of frequent and infrequent words, and use it to parse a novel stream.

However, word frequency is not the only cue to word order (Morgan et al., 1996) and under some circumstances, it might not even be sufficient on its own. If an infant is exposed to a mixed language like German or Dutch, in which both OV and VO structures appear (German: (*weil ich) Papa sehe* because I Daddy see "because I see Daddy" and (*denn ich) sehe Papa* because I Daddy see "because I see Daddy," Dutch: *op de trap* up the stairs "up the stairs" & *de trap op*), or to two languages with opposite orders, e.g., Japanese and Italian, then both FI and FF orders are found in the input she receives. Another well-established cue to word order, which can be used in combination with word frequency, is phrasal prosody (for a recent formulation of the proposal, see Shukla and Nespor, 2010). The prominence typically falls on the content word, i.e., the infrequent element, in prosodic phrases, hence its position correlates with word order. It is usually phrase-initial in OV or functor-final languages and phrase-final in VO or functor-initial languages (Nespor and Vogel, 1986). Even more importantly, the acoustic realization of phrasal prominence differs in these two positions, i.e., it correlates with word order. In OV languages, phrasal prominence is typically realized as increased pitch and/or intensity on the stressed vowel of the prominent word, so phrases tend to have a high-low or strong-weak pattern, whereas in VO languages, prominence is realized as increased duration on the stressed vowel of the prominent element, so phrases shown a short-long pattern (Nespor et al., 2008). Interestingly, this has been shown to hold true not only across languages, but also within a language, e.g., in the OV and VO phrases of German (Nespor et al., 2008). This differential acoustic realization means that there is a low-level, perceptually available cue in the input signal that correlates with word order. Further, it has been argued that these different acoustic features, i.e., pitch/intensity vs. duration, trigger different perceptual groupings. Known as the iambic-trochaic law (ITL, Hayes, 1995) and originally described for non-linguistic auditory stimuli (Bolton, 1894; Woodrow, 1951), this principle argues that elements contrasting in intensity or pitch are naturally perceived as having initial prominence, i.e., trochaic grouping, while elements contrasting in duration are perceived as prominence-final, i.e., iambic. This principle together with the different acoustic realization of prominence in OV vs. VO languages provides an automatic bootstrapping mechanism to cue word order (Mazuka, 1996; Nespor et al., 1996, 2008; Höhle et al., 2001; Shukla and Nespor, 2010).

Are infants able to exploit this cue? Sensitivity to prosody appears very early in development. Newborns' communicative cries already show similarities with the prosodic patterns of the languages heard *in utero*, evidencing prenatal learning of prosody (Mampe et al., 2009). By 2 months of age, infants are able to discriminate the typical OV and VO prosodies described above (derived from Turkish and French, respectively), even when the stimuli are resynthesized to suppress all other distinctive features, e.g., segmental information (Christophe et al., 2003). Prosodic grouping preferences following the ITL have been documented as early as 6–8 months of age. Specifically, monolingual Japanese (OV) and monolingual English (VO) infants show language-specific prosodic grouping at 7–8 months, but not yet at 5–6 months (Yoshida et al., 2010) for the durational contrast with pure tone, i.e., non-linguistic, stimuli. Pitch and intensity were not tested in this study. For speech sequences, prosodic grouping was observed in monolingual Italian (VO) infants at 7 months with the pitch/intensity contrast, but not with duration (Bion et al., 2011).

Differences in the nature and complexity of the stimuli used in the two studies might explain why a duration-based grouping preference was found in one VO-exposed population (English infants in the Yoshida et al., 2010 study), but not in the other (Italian infants in the Bion et al., 2011; study). Taken together, these studies suggest that prosodic grouping preferences start to emerge at around 7–8 months of age in the monolingual populations tested. Similar results were obtained when prosodic cues were combined with statistical information in a word segmentation task: 9-month-old infants were able to use intensity as a cue to word onset and duration as a cue to word offset with both pure tones and speech stimuli, while 6.5-month-old infants could only use the intensity cue, but not duration (Hay and Saffran, 2011).

Recently, infants' ability to use prosody, and more specifically the ITL as a cue to word order has been tested directly (Gervain and Werker, under review). Seven-month-old OV (one of Japanese, Korean, Hindi/Punjabi, Farsi, or Turkish) – VO (English) bilinguals were exposed to a structurally ambiguous artificial grammar similar to the one used in Gervain et al. (2008). Importantly, prosody was added to the stream: half of the infants were exposed to the stream with OV prosody (pitch contrast), the other half to VO prosody (durational contrast). The test items were the same FI and IF sequences as in Gervain et al. (2008) with no prosodic cues (flat pitch and constant duration). Infants exposed to OV prosody showed a preference for the IF items, while infants in the VO prosody condition looked longer at the FI items. This suggests that OV–VO bilinguals are able to use phrasal prosody, in combination with word frequency, as a cue to select between the opposite word orders of their native languages. Interestingly, VO (English) monolinguals tested with the unfamiliar OV prosody did not show any preference, although they did prefer FI items when tested with no prosody, i.e., with only word frequency as a cue, replicating the monolingual Japanese and Italian findings (Gervain et al., 2008). This might indicate that by 7 months of age, monolinguals possess a stable representation of word order in terms of the distribution of frequent functors, which cannot be overridden by prosody when there is a conflict between the two cues (as was the case for the English monolinguals). An alternative explanation is that monolinguals may be less efficient at processing multiple cues, i.e., prosody and frequency, than bilinguals (Kovacs and Mehler, 2009a,b) and showed no preference in this task as a result of cognitive overload.

The current study, therefore, addresses two questions. First, we ask whether monolinguals are able to process word frequency and phrasal prosody simultaneously as cues to word order. Second, if they are, how do they integrate the two cues? To address these issues, we ran two studies (**Figure 1**), adapting the VO prosody condition from Gervain and Werker (under review). In Condition 1, the stimuli were identical to the VO prosody condition of Gervain and Werker (under review), with prosody and frequency perfectly aligned, i.e., with lengthening on the infrequent words as in natural language. We reasoned that for the monolingual French (VO) infants we tested, there is no conflict between prosody and frequency in this condition, so if they are able to process the two cues simultaneously, they should show a FI (VO) preference during test. If, however, the reason for their null preference in the Gervain and Werker (under review) study was the simultaneous presence of two cues, then they should also fail to show a preference in the present study. In Condition 2, we also used VO prosody and word frequency as cues, but now they were misaligned: prosodic prominence was shifted by one word, rendering the frequent words longer. This pattern, i.e., prosodic prominence on function words, is unusual in natural languages. Therefore, if infants integrate the two cues at the level of individual lexical items, then an ill-formed, misaligned representation arises, possibly disrupting infants' preference for the FI (VO) pattern. If, however, prosody and frequency are processed separately, infants might still show a FI preference, because when considered independently, both cues are well-formed, native-like indicators of the functor-initial order of French.

MATERIALS AND METHODS

PARTICIPANTS

Thirty (13 girls and 17 boys) 8-month-old (mean age: 8 months and 6 days, range: 6 months and 24 days to 8 months and 25 days) infants participated in Condition 1. Among these 30 children, five had one parent who spoke a language other than French: Arabic (2), Antillean Creole (1), Hungarian (1), Italian (1). Only the Italian-exposed infant was retained for analysis. Six other children did not complete the experiment because of fussiness and crying. Thus, 20 infants entered the analysis of Condition 1.

Another 36 (18 girls and 18 boys) 8-month-old (mean age: 8 months and 3 days, range: 6 months and 22 days to 8 months and 27 days) infants participated in Condition 2. Among these 36 children, seven had one parent who spoke a language other than French: English (1), Russian (1), Spanish (3), and Turkish (2). The Turkish and Russian-exposed infants were not retained for analysis. However, the English- and Spanish-exposed infants were, as both languages are VO with phrasal prosodies that are sufficiently similar to that French. In addition, 11 children did not complete the experiment because of technical problems (3), fussiness and crying (6), and too short or too long looking times (2). Since the duration of a test item was 960 ms and the maximum duration of a trial test was 21.84 s, we kept only the trials with fixation times strictly between these two values. Also, babies with more than two test trials rejected were not included in the final data analysis. Thus, 22 infants entered the analysis of Condition 2.

All parents gave informed consent before participation, and completed an information sheet.

MATERIAL

An artificial grammar with ambiguous underlying structure was created for Conditions 1 and 2 (**Figure 1**), following Gervain and Werker (under review): a four-syllable-long basic unit AXBY was concatenated repeatedly. The A and B categories had one token each, while the X and the Y categories contained nine tokens, making individual X and Y tokens nine time less frequent than the A and B tokens. The lexicon of the artificial grammar consisted of the following words: A: *fi*, B: *ge*, X: *ru, pe, du, ba, fo, de, pa, ra, to,* Y: *mu, ri, ku, bo, bi, do, ka, na, ro.* This basic structure gave rise to a continuous stream of strictly alternating frequent (A and B) and infrequent (X and Y) words, mimicking function words and content words, respectively. The initial and final 15 s of the stream were ramped in amplitude in order to mask any

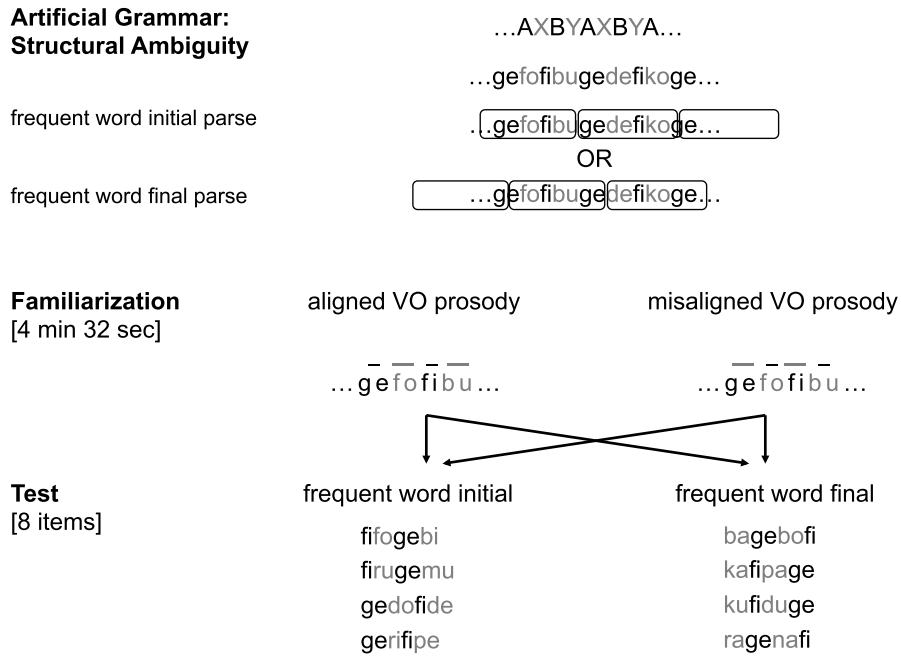


FIGURE 1 | The material used in Conditions 1 and 2.

phase-information. The familiarization stream was thus ambiguous between a frequent word initial or frequent-infrequent (e.g., AXBY) and a frequent word final or infrequent-frequent (IF; e.g., XBYA) parse.

The familiarization stream was synthesized using the fr4 female diphone database of MBROLA (Dutoit, 1997). In the two conditions, we used the same pitch (200 Hz) for all syllables (both frequent and infrequent words). We added native prosody (VO prosody) to the stream. We manipulated the relative position of prosodic prominence and word frequency. In Condition 1, the two types of cues were congruent: the non-prominent frequent words were short (240 ms) and the prominent infrequent words were long (320 ms). In Condition 2, we misaligned word frequency and word length so that frequent words were long (320 ms) and infrequent words were short (240 ms). The total duration of the two types of familiarization streams was 4 min 32 s.

The test items were eight four-syllabic chunks from the stream. Four of them instantiated the frequent-infrequent (FI) order (corresponding to a VO language; *fifogebi/firugemu/gedofipe/gerifipe*), the other four the IF order (corresponding to an OV language; *kafipage/kufiduge/bagebofi/ragenafi*). The prosody was flat for all the test items: with a constant 240 ms syllable duration, resulting in 960 ms long test items.

PROCEDURE

Participants were tested individually in a sound-attenuated room, with a low light intensity. The Headturn Preference Procedure (HPP, KemlerNelson et al., 1995) was used. Babies were seated on their caregiver's lap in front of a central attention-getter light. Each experimental session consisted of a familiarization phase (with one of the two streams: word length and word frequency aligned or misaligned) immediately followed by a test phase. During the familiarization phase, a continuous stream, which lasted 4 min 32 s, was presented to the participants from two side speakers, associated with two attention-getter lights. During the familiarization phase, the lights were contingent upon the infants' looking behavior, but were independent of the sound stimuli. During test trials, babies heard one of the eight four-syllabic chunks from the stream (four per condition). Before each test trial, infants' attention was drawn to the central attention-getter light. Once this was achieved, the central light was turned off, and one of the sidelights was turned on. A test trial began when infants turned away from the central light and attended to the flashing sidelight. The test item was then presented at the same side. When babies looked for the maximum duration of the trial or if they looked away for more than 2 s, the trial ended, the sidelight was turned off, and the central attention-getter light started blinking again.

Each child heard eight test items: four in each condition (FI or IF). Stimuli were pseudo-randomized for each participant: there could not be more than two consecutive test items in the same condition. They were also counterbalanced between participants.

An experimenter observed infants' behavior on a video monitor placed outside the experimental booth and controlled the lights and the stimuli. She listened to masking music and was blind to the stimuli being presented. Infants' looking behavior was coded offline using the video recording made during the experiment.

RESULTS

The average looking times to FI and IF items in the two conditions are shown in **Figure 2**. We conducted an ANOVA with Familiarization Condition (Cond 1 aligned/Cond 2 misaligned) as

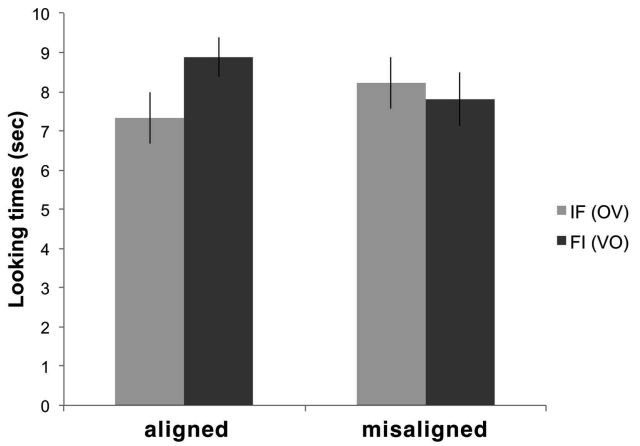


FIGURE 2 | The average looking times for FI and IF items in Conditions 1 and 2.

a between-subjects factor and Test Item Type (FI/IF) as a within-subjects factor. We obtained a significant Familiarization Condition × Test Item Type interaction [$F(1,40) = 5.3983$, $p = 0.026$]. This was due to significantly longer looking times (Scheffe *post hoc* test $p = 0.010$) to *F*I test items than to IF ones in Exp 1 (aligned familiarization), but not in Exp 2 (misaligned familiarization). No other effect was significant.

DISCUSSION

In this study, we tested whether monolingual French-exposed 8-month-old infants are able to use word frequency and prosody as simultaneous cues to a rudimentary representation of the word order type of their native language. In an artificial grammar learning task, we found that they indeed showed the predicted preference for frequent word initial test items, mimicking the functor-initial word order of French, when the two cues were aligned at the level of lexical items, i.e., frequency words were non-prominent, but not when they were misaligned.

A possible alternative interpretation could be that infants in Condition 1 simply did not use prosody as a cue and succeeded on the basis of the frequency cue alone, as did monolingual Japanese and Italian infants in the Gervain et al. (2008) study. However, this interpretation is not probable, because if infants ignored prosody altogether in Condition 1, we would expect them to do the same in Condition 2, showing the same FI preference, contrary to fact.

Our results, therefore, suggest that monolinguals are not hindered by the presence of simultaneous cues as long as the prosodic cue is coherent with the frequency cue. This coherence is required at least at two levels. First, frequency and prosody cannot be in conflict: the OV prosody used with English-exposed infants in the Gervain and Werker (under review) study gives rise to a null preference, as neither cue overrides the other, i.e., they carry equal weight. Second, the prosodic cue and the word frequency cue need to be aligned at the lexical level, suggesting that the two cues are processed in an integrated manner.

What representations are formed through this integrative process? Further research is needed to explore the full details of how word order is acquired. It is not clear, for instance, whether both the frequent and the infrequent words are learned, or only the frequent ones. What the present study shows, however, is that infants expect lexical categories that follow the characteristics of those found in natural languages. Thus, they expect frequent words to occupy the typical positions of functors and to be prosodically less prominent than infrequent words, reflecting their knowledge of the typical features of functors, and content words. This is in accordance with previous results showing that infants as young as newborns are able to discriminate functors and content words on the basis of their different perceptual properties, and have expectations about their function and sentential position at an early age (Gerken et al., 1990; Gerken and McIntosh, 1993; Morgan et al., 1996; Shi et al., 1999, 2006; Shi and Werker, 2001, 2003; Höhle and Weissenborn, 2003; Hochmann et al., 2010). Further, this knowledge is abstract enough to allow generalization to a novel language, reflecting the existence of a representation of word order in terms of functor positions.

This simple representation of basic word order type in terms of function word position might be a first step in bootstrapping more complex word order phenomena and grammatical structure in general. During subsequent language development, infants might enrich this representation relying on several sources. They might be able to exploit the correlations that exist between the position of functors and other word order phenomena, such as the relative order of Verbs and their Objects, main and subordinate clauses etc (Kucera and Francis, 1967; Gervain et al., 2008). They might rely on their emerging vocabulary of object and action labels (Bergelson and Swingley, 2012) or their increasing understanding of intentionality (Csibra and Gergely, 2009) to determine the syntactic and semantic patterns of simple utterances in their input and generalize them to understand and produce more complex structures, as suggested by the semantic (Pinker, 1984) and syntactic bootstrapping hypotheses (Gleitman et al., 1988; Fisher et al., 1991).

If infants integrate word frequency and phrasal prosody at the level of lexical categories, as argued above, can we really conclude that this bootstrapping mechanism is prelexical and independent of vocabulary learning, as claimed before? In our view, this conclusion is justified for at least two reasons. First, infants' knowledge appears to be category- and not item-based. There is nothing about the specific words used as frequent and infrequent items in our study that requires them to be prosodically weak or strong, respectively. It is infants' knowledge about the lexical category of functors and content words in natural language that allows them to process the aligned grammar as well-formed and the misaligned one as ill-formed. Second, although recent results suggest that infants show evidence of word learning between 6–9 months of age (Bergelson and Swingley, 2012), at 8 months, the age tested in this study, they certainly do not yet have a sizeable lexicon. Therefore, they have no item-based knowledge in the sense of Tomasello (2000) that could support the word order representations we have uncovered in this study.

Taken together, our findings suggest that a first representation of a fundamental property of the native language,

word order, is bootstrapped very early in development on the basis of perceptual cues such as word frequency and phrasal prosody. This early acquisition might have a cascading effect on the subsequent development of the native grammar and the lexicon.

ACKNOWLEDGMENTS

This work was supported by an ANR Jeunes Chercheurs Jeunes Chercheuses Grant nr. 21373, a Fyssen Startup Grant and "The Empirical Foundations of Linguistics" Labex project to Judit Gervain.

REFERENCES

Abney, S. (1987). *The English Noun Phrase in its Sentential Aspect*. Cambridge: MIT.

Akhtar, N., and Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Dev. Psychol.* 33, 952.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., and Bever, T. G. (1996). "Models of word segmentation in fluent maternal speech to infants," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Mahwah: Erlbaum), 117–134.

Bergelson, E., and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258.

Bion, R. A. H., Benavides-Varela, S., and Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Lang. Speech* 54, 123.

Bolton, T. L. (1894). Rhythm. *Am. J. Psychol.* 6, 145–238.

Brown, R. W. (1973). *A First Language; The Early Stages*. Cambridge, MA: Harvard University Press.

Christophe, A., Marina, N., Maria, T. G., and Brit, V. O. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Dev. Sci.* 6, 211–220.

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci. (Regul. Ed.)* 13, 148–153.

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language* 68, 81–138.

Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*, Vol. 3. Dordrecht: Kluwer Academic Publishers.

Fisher, C., Gleitman, H., and Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cogn. Psychol.* 23, 331–392.

Fukui, N. (1986). *A Theory of Category Projection and its Applications*. Cambridge: MIT.

Gerken, L., Landau, B., and Remez, R. (1990). Function morphemes in young children's speech perception and production. *Dev. Psychol.* 26, 204–216.

Gerken, L., and McIntosh, B. (1993). Interplay of function morphemes and prosody in early language. *Dev. Psychol.* 29, 448–457.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., and Mehler, J. (2008). Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cogn. Psychol.* 57, 56–74.

Gleitman, L. R., Gleitman, H., Landau, B., and Wanner, E. (1988). "Where learning begins: initial representations for language learning," in *Linguistics: The Cambridge Survey: Vol. 3. Language: Psychological and Biological Processes* (Cambridge: Cambridge University Press), 150–193.

Greenberg, J. H. (1978). *Universals of Human Language*. Stanford: Stanford University Press.

Hay, J. F., and Saffran, J. R. (2011). Rhythmic grouping biases constrain infant statistical learning. *Infancy* 17, 610–641.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hochmann, J. R., Endress, A. D., and Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition* 115, 444–457.

Höhle, B., and Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Dev. Sci.* 6, 122–127.

Höhle, B., Weissenborn, J., Schmitz, M., and Ischebeck, A. (2001). "Discovering word order regularities: the role of prosodic information for early parameter setting," in *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition*, eds J. Weissenborn and B. Höhle (Amsterdam: John Benjamins), 249–265.

KemlerNelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A. E., and Gerken, L. (1995). The headturn preference procedure for testing auditory perception. *Infant Behav. Dev.* 18, 111–116.

Kovacs, A. M., and Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6556–6560.

Kovacs, A. M., and Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science* 325, 611–612.

Kucera, H., and Francis, W. N. (1967). *Computational Analysis of Present-day American English*. Providence: Brown University Press.

Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Curr. Biol.* 19, 1994–1997.

Mazuka, R. (1996). *Can a Grammatical Parameter be Set Before the First Word? Prosodic Contributions to Early Setting of a Grammatical Parameter. Signal to Syntax*. Hillsdale, NJ: Lawrence Erlbaum Associates, 313–330.

Morgan, J. L., and Demuth, K. (1996). *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Morgan, J. L., Shi, R., and Allopenna, P. (1996). "Perceptual bases of rudimentary grammatical categories: toward a broader conceptualization of bootstrapping," in *Signal to Syntax* (Hillsdale, NJ: Lawrence Erlbaum Associates), 263–283.

Nespor, M., Guasti, M. T., and Christophe, A. (1996). "Selecting word order: the rhythmic activation principle," in *Interfaces in Phonology*, ed. U. Kleinhenz (Berlin: Akademie Verlag), 1–26.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., and Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio* 7, 1–28.

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*, Vol. 28. Dordrecht: Foris.

Pinker, S. (1984). *Language Learnability and Language Development*, Vol. 7. Cambridge, MA: Harvard University Press.

Shi, R., Cutler, A., Werker, J., and Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *J. Acoust. Soc. Am.* 119, EL61–EL67.

Shi, R., and Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychol. Sci.* 12, 71–76.

Shi, R., and Werker, J. F. (2003). The basis of preference for lexical words in 6-month-old infants. *Dev. Sci.* 6, 484–488.

Shi, R., Werker, J. F., and Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72, B11–B21.

Shukla, M., and Nespor, M. (2010). "Rhythmic patterns cue word order," in *The Sound Patterns of Syntax*, ed. N. Erteschick-Shir (Oxford: Oxford University Press), 174–188.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* 74, 209–253.

Weissenborn, J., Höhle, B., Kiefer, D., and Cavar, D. (1996). "Children's sensitivity to word-order violations in german: evidence for very early parameter-setting," in *Proceedings of the Boston University Conference on Language Development*, Vol. 22. Somerville: Cascadilla Press.

Woodrow, H. (1951). "Time Perception," in *Handbook of Experimental Psychology*, ed. S. S. Stevens (New York: Wiley), 1224–1236.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., et al. (2010). The development of perceptual grouping biases in infancy: a Japanese-English cross-linguistic study. *Cognition* 115, 356–361.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 August 2012; accepted: 08 October 2012; published online: 30 October 2012.

Citation: Bernard C and Gervain J (2012) Prosodic cues to word order: what level of representation? Front. Psychology ***3****:451. doi: 10.3389/fpsyg.2012.00451*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 05 March 2013
doi: 10.3389/fpsyg.2013.00106

Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages

Laura Bosch[1,2]*, Melània Figueras[1], Maria Teixidó[1] and Marta Ramon-Casas[1]

[1] Department of Basic Psychology, University of Barcelona, Barcelona, Spain
[2] Institute for Research in Brain, Behavior and Cognition (IR3C), University of Barcelona, Barcelona, Spain

Edited by:
Jutta L. Mueller, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Toben H. Mintz, University of Southern California, USA
Amanda Seidl, Purdue University, USA

***Correspondence:**
Laura Bosch, Department of Basic Psychology, University of Barcelona, Passeig Vall d'Hebron, 171, 08035 Barcelona, Spain.
e-mail: laurabosch@ub.edu

The ability to extract word-forms from sentential contexts represents an initial step in infants' process toward lexical acquisition. By age 6 months the ability is just emerging and evidence of it is restricted to certain testing conditions. Most research has been developed with infants acquiring stress-timed languages (English, but also German and Dutch) whose rhythmic unit is not the syllable. Data from infants acquiring syllable-timed languages are still scarce and limited to French (European and Canadian), partially revealing some discrepancies with English regarding the age at which word segmentation ability emerges. Research reported here aims at broadening this cross-linguistic perspective by presenting first data on the early ability to segment monosyllabic word-forms by infants acquiring Spanish and Catalan. Three different language groups (two monolingual and one bilingual) and two different age groups (8- and 6-month-old infants) were tested using natural language and a modified version of the HPP with familiarization to passages and testing on words. Results revealed positive evidence of word segmentation in all groups at both ages, but critically, the pattern of preference differed by age. A novelty preference was obtained in the older groups, while the expected familiarity preference was only found at the younger age tested, suggesting more advanced segmentation ability with an increase in age. These results offer first evidence of an early ability for monosyllabic word segmentation in infants acquiring syllable-timed languages such as Spanish or Catalan, not previously described in the literature. Data show no impact of bilingual exposure in the emergence of this ability and results suggest rapid gains in early segmentation for words that match the rhythm unit of the native language.

Keywords: word segmentation, syllable-timed languages, natural speech, rhythmic unit, preference pattern, infants

INTRODUCTION

The identification of possible word-forms within sentential contexts represents an initial step in infants' process toward lexical acquisition. Extracting word units from the input and detecting repetitions of these units in different contexts is considered a basic skill related to early vocabulary construction. Research has already shown an associative link between these early skills and later language outcomes (Newman et al., 2006; Junge et al., 2012; Singh et al., 2012). Characterizing the emergence of word segmentation ability is, thus, important in relation to the early building and growing of lexical knowledge. More specifically, exploring this emergent capacity in infants exposed to languages with different rhythmic structure offers the opportunity to identify differential features in the segmentation strategies used by infants, as well as possible variation in its developmental time-course. Finding evidence of variation in the time course for word segmentation might ultimately be useful to account for possible differences in early lexical acquisition processes from a cross-linguistic perspective. The present research addresses this issue by exploring early word segmentation abilities in infants exposed to Spanish and Catalan, two Romance languages whose rhythmic properties differ from the properties of languages that have already been analyzed in previous word segmentation studies.

The ability to segment and recognize unfamiliar words from fluent speech was first explored in the pioneering research developed by P. W. Jusczyk and R. N. Aslin in 1995. In their seminal paper they showed that 7½ -month-old, but not 6-month-old English-learning infants, were able to extract short, monosyllabic word-forms from natural speech passages containing repetitions of two different target words (Jusczyk and Aslin, 1995). Whether familiarized to lists of words and then tested with passages, or familiarized to passages and then tested on words, infants in both testing conditions showed the capacity to extract and recognize possible "lexical" units (word-forms) and they did so by retaining rather detailed information about the phonetic form of these word candidates. Even though words in that experiment were short, simple monosyllabic items (*bike*, *dog*, *cup*, *feet*), infants younger than 7 months of age did not succeed in the task. Follow-up work explored the ability to segment bi-syllabic words and it was shown that for words following the predominant stress

pattern in the language (i.e., the trochaic or strong/weak -SW- stress pattern in the case of English), this ability was also present by 7½ months of age (Jusczyk et al., 1999a). Taken together these results were interpreted as an indication that prosodic information, here based on the predominant stress pattern of content words in English (around 90% of content words begin with a stressed syllable, according to Cutler and Carter, 1987), could be used by infants to successfully find word-form units in connected speech. This prosodic hypothesis (defined as the Metrical Segmentation Strategy -MSS- in Jusczyk, 1999) could explain both the results from the monosyllabic and the trochaic word segmentation experiments, as items in the monosyllabic study were strong syllables with full vowels. The importance of prosodic information in early word segmentation was first described in these early studies and subsequent work contributed to give support to the relevant role of prosody in infants' dealing with the word segmentation problem (Johnson and Jusczyk, 2001; Johnson and Seidl, 2008, but see Thiessen and Saffran, 2003; Pelucchi et al., 2009 for an alternative position to the prosodic bootstrapping approach).

If we admit that segmentation strategies based on prosodic information derive from the specific rhythmic properties of the native language, then these strategies might differ in populations acquiring languages with different rhythmic structure. Research with young infants has shown that they are sensitive to global prosodic features contained in the linguistic input (Bosch and Sebastián-Gallés, 1997; Nazzi et al., 2000a). These prosodic features may offer first cues to segment the input into linguistically relevant units such as clauses and phrases within which word-form units can eventually be extracted (Hirsh-Pasek et al., 1987; Nazzi et al., 2000b; Soderstrom et al., 2003; Seidl, 2007; Seidl and Cristià, 2008). But beyond these global prosodic cues, attention to the specific rhythmic properties of the native language and detection of the specific rhythmic unit operating in that language can lead to the emergence of segmentation strategies most adequate to extract words from fluent speech. This is actually the hypothesis behind the so-called *early rhythmic segmentation proposal* developed by Nazzi et al. (2006). Cross-linguistic differences regarding the type of rhythmic strategy and rhythmic unit used for segmentation can then be expected for languages differing in their rhythmic properties.

A gross partition of the languages based on linguistic rhythm has traditionally identified three broad rhythmic types, i.e., stress-timed, syllable-timed and mora-timed languages, each of them associated to a different underlying rhythmic unit (Abercrombie, 1967; Ladefoged, 1975). Germanic languages such as English, Dutch, or German would belong to the first type, having the trochaic stress unit at the basis of their rhythmic structure; Romance languages such as French, Italian or Spanish would be examples of the second type, having the syllable as the basic rhythmic unit and, finally, languages like Japanese would belong to the third type relying on the sub-syllabic *mora* as the basic unit of rhythm. Initially, these typologies were considered to derive from the notion of isochrony between successive units (syllables, feet or morae depending on the type of language), however, subsequent measurements obtained from different languages questioned this idea. Linguistic rhythm is more accurately described as an alternation of elements: vowels and consonants at the most basic level and syllables (stressed and unstressed) and feet at subsequent levels (Nespor et al., 2011). Factors such as variability in syllable structure complexity and the degree of vowel reduction are considered key elements in accounting for language rhythm differences. The study of durational correlates of such phonological phenomena has become the focus of research aimed at identifying the specific properties underlying rhythmic differences between languages.

Different rhythm metrics have been used in studies analyzing limited sets of cross-linguistic material. Ramus et al. (1999) measured duration of vocalic and consonantal intervals in the speech signal. By plotting the percentage of total utterance duration comprising vocalic intervals (%V) against the standard deviation of consonantal intervals (ΔC), they succeeded at adequately grouping the eight languages under study (English, Dutch, Polish, Spanish, Italian, French, Catalan, and Japanese) into the three traditional rhythm typologies. Low et al. (2000) proposed pairwise variability indices (nPVI and rPVI, normalized and raw, respectively) in an attempt to better capture the durational differences between successive vocalic and consonantal intervals. However, only measurements from the nPVI-V scores could group separately English, German, and Dutch on the one hand, and Spanish and French on the other, failing to place Japanese in a different area. Interestingly, languages considered more difficult to classify in terms of rhythm structure, such as Catalan and Polish (Nespor, 1990), showed intermediate positions in the PVI space. More recently, White and Mattys (2007) using rate-normalized metrics of vocalic interval variation (VarcoV) plotted against %V measurements, showed again that "stress-timed" Dutch and English, and "syllable-timed" French and Spanish could be distinguished, but at the same time their analysis revealed that the notion of a strictly categorical distinction between these two rhythmic typologies was far from perfect, with Dutch and French placed in a more intermediate position between stress-timed English and syllable-timed Spanish. In general, results from these metrical studies offer empirical support for the existence of broad rhythmic distinctions between languages, but critically, they also provide a more nuanced perspective on the nature of rhythmic differences that goes beyond the initial notion of three distinct language typologies (see White et al., 2012). From this perspective, differences in the emergence of the word segmentation ability may be found not only when comparing languages traditionally ascribed to a different rhythmic typology (e.g., stress-timed English and syllable-timed French), but also for languages traditionally grouped under the same typology (e.g., syllable-timed French and Spanish, or stress-timed English and Dutch).

What evidence can be found about cross-linguistic differences in the skill to segment words from fluent speech early in development? A review of the early word segmentation literature immediately reveals that research has been developed mostly in English and cross-linguistic data are still scarce. As already mentioned first evidence of word segmentation with natural language material came from English-learning 7½ -month-old infants and restricted to specific types of words such as monosyllabic and trochaic items (Jusczyk and Aslin, 1995; Jusczyk et al., 1999a). Evidence for this ability at an earlier age (6 months) was later

attested by using a slightly different methodological approach, in which highly familiar words (infants' own names) preceded the target monosyllabic units (those used in Jusczyk and Aslin, 1995 study) in familiarization passages (Bortfeld et al., 2005). In this situation, familiar words were probably acting as anchors and facilitated segmentation of the adjacent elements, which could not otherwise be easily extracted. Without additional cues to segmentation, English-learning infants are just beginning to segment simple word forms from fluent speech around 7 months of age. It is interesting to note that evidence of segmentation is shown by a familiarity preference pattern, whether familiarization be based on passages or word lists. The direction of the preference has been linked to task demands (Hunter and Ames, 1988). The specific direction of the preference pattern in word segmentation tasks (novelty versus familiarity) and its changes during development can be explained from factors such as the duration of the familiarization, stimulus complexity, degree of similarity between familiarization and test stimuli, and more generally, from expertise acquired with age (Thiessen et al., 2005). Thus at younger ages, when segmentation ability is just emerging, a familiarity preference is to be expected, as found by Jusczyk and Aslin (1995) in 7½ month-old infants.

For studies examining the early emergence of segmentation ability in stress-timed languages other than English, only some data from German and Dutch are available. Segmentation of unstressed closed-class elements has been shown in German-learning infants from 7½ months on, but not before (Höhle and Weissenborn, 2003). The procedure involved familiarization with isolated words and test on passages and, as expected, a familiarity preference was found paralleling Jusczyk and Aslin's results, but this time on unstressed material (although from an acoustical perspective, closed-class grammatical morphemes experience less vowel reduction in German than in English). According to the rhythmic segmentation hypothesis these unstressed monosyllabic elements should have been difficult to segment at that age. It was argued, however, that their special status and potential role in the acquisition of morpho-syntactic knowledge might have favored successful segmentation at an early age.

Evidence from Dutch-learning-infants revealed a slightly later emergence of the segmentation ability (at 9, but not at 7 months of age) using HPP and trochaic words (Kuijpers et al., 1998). Further research replicated 9-month-olds' segmentation of trochees in this language and confirmed that the ability to extract words from fluent speech is not dependent on familiarity with the phonetic structure of the input, as English-learning infants also succeeded in the task with Dutch material (Houston et al., 2000). The same strategy could be exported to successfully extract word units in another stress-timed language with similar rhythmic properties. In spite of the slightly older age of the Dutch participants, segmentation evidence resulted from a familiarity preference. Could rhythmic differences between English and Dutch, as described by White and Mattys (2007) using VarcoV and %V measurements, have impacted speed of segmentation? This remains an open question that deserves further analysis. Unfortunately, no data from monosyllabic word segmentation in Dutch are available, which might have revealed successful segmentation at an earlier age than that obtained for trochees.

The above mentioned studies involve languages traditionally grouped under the stress-timed category, whose rhythmic unit is not the syllable. Will monosyllabic word segmentation be facilitated early in development if the rhythmic unit of the ambient language is the syllable? And will segmentation of bi-syllabic words initially be delayed, being first segmented as two independent syllabic units and only later as whole units? Infant segmentation data from syllable-timed languages are actually limited to French, although evidence obtained from two different French dialects (European and Canadian) is available.

Monosyllabic word segmentation in French has not been extensively explored and only data available from a dissertation indicate that Parisian 7½ month-olds could successfully segment monosyllabic CVC items, using HPP with familiarization to words and test on passages (Gout, 2001). A familiarity preference was also obtained there[1]. No data from infants tested at a younger age were gathered, so we do not know if monosyllabic words in a syllable-timed language are actually easier to extract from fluent speech than similar words in stress-timed languages. What we actually know, however, is that bi-syllabic word segmentation in French is not easily attained, at least according to data from infants exposed to the European French dialect who could only succeed at successfully segmenting iambs by 16 months of age (Nazzi et al., 2006). Data from French-learning infants exposed to the Canadian dialect did not replicate European French results, however. No "delayed" segmentation ability was identified in Canadian French-learning infants compared to a group of Canadian English young learners tested at 8 months on two-syllable word segmentation (iambic and trochaic patterns, respectively): both groups succeeded, although segmentation strategies certainly differed and were adjusted to the properties of the native language, so no group was able to segment the items in the other language (Polka and Sundara, 2012).

Because Nazzi et al.'s (2006) and Polka and Sundara's (2012) work involved a considerable amount of experiments to more thoroughly explore segmentation abilities in the populations under study, some relevant findings about the segmentation of the syllabic components of the iambic items could be identified. Clear evidence of final syllable segmentation was obtained at 12 months and some evidence of initial syllable segmentation could also be found at the same age in European French infants suggesting that a syllable-based segmentation procedure is applied before bi-syllabic words can be successfully segmented as whole units (Nazzi et al., 2006). Similarly, although at an earlier age, results in Canadian French also revealed some ability to segment each isolated syllable of the iambic target words, although the transition from an initial syllable-based segmentation to a successful whole bi-syllabic word segmentation could not be established in that research as only groups of 8-month-olds' were tested (Polka and Sundara, 2012). Relevant for our own research on monosyllabic word segmentation, the Canadian study found opposite response patterns when familiarization involved whole iambic

[1] In a different study focused on the segmentation of monosyllabic verb forms by Canadian-French learning infants, positive evidence of segmentation, also based on a familiarity preference, was obtained at 11 months of age, but not earlier (Marquis and Shi, 2008).

words or only their syllabic components. The novelty preference pattern obtained when syllables instead of whole words were presented in the familiarization phase suggests that syllables might be more easily identified because they match the rhythmic unit in this language.

Taken together, and compared to data from segmentation in stress-timed languages, research done in French reveals important cross-linguistic differences, not only in the emergence of segmentation abilities, but also in the strategies used, which reflect the rhythmic nature of the language of exposure. However, French results are not clear-cut especially due to the non-trivial timing difference in the emergence of segmentation abilities found when both dialects are compared. Even if differences can be attributed to factors derived from specific properties of these dialects or the testing material, the fact is that behavioral results so far have only partially confirmed a hypothetic ease to segment monosyllabic words or track syllabic elements in fluent speech, as it could be expected if the syllable is the rhythmic unit for segmentation in syllable-based languages (but see Goyet et al., 2010 for a re-assessment of syllabic segmentation using ERP measures). Studying early segmentation abilities in infants acquiring other syllable-timed languages could shed more light on the early rhythmic segmentation hypothesis and help clarify results obtained so far.

Spanish and Catalan have also been traditionally grouped under the syllable-timed typology, although some metric distinctions have been described in studies comparing the rhythmic properties of these two languages. Some authors consider Catalan a rhythmically-intermediate language between the stress-timed and syllable-timed typologies (Nespor, 1990). Catalan, but not Spanish, has vowel reduction, a property that can affect syllabic rhythm and determines differences in the type of vowels that can appear in unstressed syllable positions (Prieto et al., 2012). Catalan allows for more complex consonant clusters in coda position, while syllabic structures are simpler in Spanish. As a consequence, %V metrics have been found to be significantly lower in Catalan than in Spanish. However, higher variability in vocalic interval duration (i.e., higher VarcoV scores that characterize languages with vowel reduction) has not been confirmed, with Catalan even showing lower variability scores than Spanish according to Payne et al.'s (2009) work. More recent research has corroborated that vowel reduction in Catalan does not seem to substantially increase variability in vowel interval duration (Prieto et al., 2012). In sum, while some rhythmic differences between Catalan and Spanish exist, the classification of Catalan as a rhythmically-intermediate language between syllable-timed and stress-timed typologies remains controversial. Although an in-depth and systematic comparison between Spanish, Catalan, and French rhythm metrics is not available, measures from different studies involving different sets of material would suggest a non-overlapping distribution of these three "syllable-timed" languages over the %V and VarcoV rhythmic plane (White and Mattys, 2007; Payne et al., 2009). Among these three languages, Spanish would show the highest %V and the lowest VarcoV scores, while French would show the opposite tendency (i.e., higher VarcoV and lower %V scores), and Catalan would be placed in an intermediate position, probably more similar to Spanish in terms of vocalic interval variability (VarcoV), as the above mentioned studies have revealed. Given these differential metrical characteristics, Spanish and Catalan are good language candidates to extend word segmentation studies in syllable-timed languages other than French and explore infants' early use of a syllabic segmentation strategy. In particular, the comparison between Catalan-learning and Spanish-learning groups can reveal if the (minor) rhythmic differences between these two languages have an impact on the emergence of the segmentation ability.

To sum up, the present research was designed to explore the emergent ability to segment monosyllabic word-forms by infants acquiring Spanish, Catalan, but also both languages simultaneously from birth. To our knowledge, word segmentation abilities in bilingual infants have begun to be explored only in English-French environments, with preliminary data available so far showing bi-syllabic word segmentation ability in both languages by 8 months of age (Polka and Sundara, 2003). The inclusion of bilingual participants in this research, exposed to languages traditionally grouped into the same rhythmic class, but nonetheless showing some minor differential rhythmic properties, can contribute to clarify the actual impact that bilingual exposure can have on the emergent ability to extract words from connected speech, when segmentation strategies derived from each of the ambient languages are likely to converge.

In the present research, evidence of an emergent segmentation ability will be explored using the HPP technique, in line with the work just reviewed coming from both stress-timed and syllable-timed languages. However, we have selected the less frequent order in this type of experiments, involving passages first and test on lists of isolated words. Because similar segmentation effects were obtained independently of the testing order in the original Jusczyk and Aslin's (1995) study, we opted for the passages-first order to promote segmentation spontaneously arising from a more natural context and to avoid initially biasing participants to attend to a specific word or syllabic unit.

In our first experiment we analyzed 8-month-olds' ability to segment words that match the rhythmic unit of their native language. No great difficulties were expected for monosyllabic word segmentation in our Catalan and Spanish participants, but given the limited data available in French and the slightly delayed emergence of the segmentation ability, even for the syllabic components of the bi-syllabic words, found by Nazzi et al. (2006), evidence from the three groups tested at 8 months would be most informative about the timing of this emergent ability in languages different from French but having syllables as the basic rhythmic units.

In a second experiment we wanted to further explore if evidence of monosyllabic word segmentation could be found at an earlier age (6 months) in syllable-timed languages compared to stress-timed ones, due to the direct match between the target elements (monosyllabic words) and the rhythmic unit for segmentation (the syllable). If confirmed, results would not only give support to the early rhythmic segmentation hypothesis, but they would also suggest the need to take into account additional differences in the rhythmic properties of languages traditionally

grouped into the same rhythmic typology, as these properties might lead to differences in the timing of the emergence of the segmentation ability. Recall that the earliest evidence for monosyllabic word segmentation in French comes from a single experiment with 7½-month-olds showing a familiarity preference response pattern (Gout, 2001; Gout, unpublished dissertation).

The ultimate aim of the present study is to set the groundwork for future research exploring the emergence of the ability to segment multi-syllabic word-forms both in Spanish- and in Catalan-learning infants. Knowledge about the ability and the segmentation strategies used to extract short, simple monosyllabic units from connected speech can offer valuable information to better understand the specific problems that segmenting bi- and tri-syllabic words in syllable-timed languages with variable stress can pose to the infant learner.

EXPERIMENT 1: WORD SEGMENTATION AT 8 MONTHS

PARTICIPANTS

A total of 54 healthy full-term infants with no history of hearing or vision problems were included in the sample divided into three groups ($N = 18$ in each group) according to the language/s spoken in their environment (Catalan only, Spanish only or both languages on a daily basis). Mean age of the infants in the Catalan monolingual group was 8 months 4 days (range: 7 months, 15 days–8 months, 22 days); in the Spanish monolingual group was 8 months 6 days (range: 7 months, 19 days–8 months, 25 days) and in the bilingual group was 8 months 6 days (range: 7 months, 13 days–8 months, 15 days). No significant between-group age differences were found ($F < 1$). Participants were assigned to different language groups based on the information obtained through a questionnaire to the parents that offered an estimate of the daily and weekly amount of exposure to the languages in their environment (Bosch and Sebastián-Gallés, 2001). To be included in a monolingual group, participants had at least 75% of regular exposure to either Catalan or Spanish, while a more balanced distribution between these two languages was required for inclusion in the bilingual group. Mean percentage of exposure to Catalan in the Catalan monolingual group was 92% (range: 75–100%) and to Spanish in the Spanish monolingual group was 93% (range: 80–100%). From the 18 infants in the bilingual group, seven had a higher amount of exposure to Spanish than to Catalan (66–34%) and they were tested on Spanish material. The remaining infants had a higher exposure to Catalan than to Spanish (63–37%) and were tested on Catalan material. Fourteen additional infants were also tested but excluded from the final sample due to fussiness or crying leading to incomplete testing (4), very short looking time—below 1 s—to trials in the test phase (6), preterm birth (1) and experimental error (3).

STIMULI

Target Spanish and Catalan monosyllabic words with full vowels and a CVC (*bus*, *mar*, *gol* –"bus," "sea," and "goal"-) or CCVC (*tren* –"train") structure were selected because of their cognate status in the languages under study (for simplicity, from now on we will refer to all target words as having a monosyllabic CVC structure). Target words were nouns that are infrequent in the first receptive and expressive vocabularies of 1-year-olds acquiring Spanish, Catalan or both (Águila et al., 2005).

Four passages were created, formed by six different sentences each with the target word appearing once per sentence in different positions (twice in initial, twice in medial and twice in final sentence positions). Because the experimental design involved two different conditions (half of the participants were familiarized with "train-bus" passages -TB-, and the other half with "gol-mar" passages -GM-), parallel sentences were used to make conditions equivalent (see **Table 1**). Adjacent syllables to the target words (from words preceding or following the target nouns) were controlled so that no specific syllabic sequences appeared repeatedly within the passage. Mean duration of the sentences in the passages was 2.3 s and total length of the passages was adjusted to 18 s by inserting short pauses of about 700 ms between sentences. Passages had 45–46 syllables each and especial care was taken to build equivalent passages for the Spanish and Catalan versions of the material. Sentences were not always perfect translations because length of the words tends to be shorter in Catalan

Table 1 | Catalan and Spanish sentences forming the passages used in the familiarization phase.

"Tren" passage (*train*)

Catalan: Un tren té sis o set vagons. Veig un gran tren des d'aquí. El tren no s'atura mai. A la foto hi ha aquell tren. Mira aquest cotxe a prop del tren. Arriben en tren molt d'hora

Spanish: Un tren tiene seis vagones. Veo un gran tren desde aquí. El tren nunca está parado. En la foto está aquel tren. Mira este coche junto al tren. Llegan en tren mañana

"Bus" passage (*bus*)

Catalan: Un bus va venir de sobte. Esperava el primer bus. Recordo aquest bus cada dia. El bus no era massa bo. M'encanta el seu bus de cartró. Somiaré amb el meu bus

Spanish: Un bus llega de repente. Esperan otro bus. Recuerdo aquel bus cada día. El bus no era largo. Me encanta su bus de cartón. Soñaré con este bus

"Mar" passage (*sea*)

Catalan: Un mar té milers de peixos. Veig un gran mar des d'aquí. El mar no s'atura mai. A la foto hi ha aquell mar. Mira aquest cotxe a prop del mar. Arriben per mar molt d'hora

Spanish: Un mar tiene muchos peces. Veo un gran mar desde aquí. El mar nunca está calmado. En la foto está aquel mar. Mira este coche junto al mar. Llegan por mar mañana

"Gol" passage (*goal*)

Catalan: Un gol va venir de sobte. Esperava el primer gol. Recordo aquest gol cada dia. El gol no era massa bo. M'encanta el seu gol de taló. Somiaré amb el meu gol

Spanish: Un gol llega de repente. Esperan otro gol. Recuerdo aquel gol cada día. El gol no era bueno. Me encanta su gol de tacón. Soñaré con este gol

In the experimental design half of the participants were familiarized to "Tren-Bus" (train-bus) passages and the other half to "Mar-Gol" (sea-goal) passages.

and we wanted to keep with the same number of syllables per sentence. In spite of minor meaning differences between Spanish and Catalan sentences (irrelevant to study word segmentation in early infancy), the final passages represent equivalent versions of the material in terms of number of syllables and total length.

Word lists involving 12 isolated productions of each of the four target words were also needed for use in the test phase of the experiments. Six different tokens of the same noun repeated twice in a randomized order formed each of the four experimental word lists in this study. Total length of the word lists was18 s as lists were built by adding silence to the end of the stimulus to reach a 1.5 s duration (mean length of the words in the lists in each language is reported in **Table 2**).

Passages and words were produced by a highly proficient Spanish-Catalan bilingual female speaker and she was instructed to use infant direct speech, as if speaking to a young child. The stimuli were recorded in a single session in a comfortable, sound attenuated booth equipped with an omni-directional microphone. Utterances were recorded directly onto a Pentium-III PC using Sound Edit (version 2.99) software. The operating system was Windows XP. Online monitoring ensured optimal sound quality recording.

Finally, to ensure similarity between the materials for each language, acoustic analyses on target words extracted from the passages and words in the lists were conducted using Praat software (version 5.3.22). Mean values of word duration and amplitude (for the entire word) and pitch (calculated on the vocalic portion of the word) were obtained and are reported in **Table 2**, both for Catalan and Spanish material. As expected, statistical analyses only revealed significant differences in duration between words extracted from the passages and words produced in isolation. Amplitude and pitch measurements were found equivalent both within each language and also between languages (see details in **Table 2**).

Table 2 | Acoustic measures of target words in passages (familiarization) and lists (test) for Catalan and Spanish material.

	Passage words	List words	
	Mean (*SD*); range	**Mean (*SD*); range**	***p***
DURATION (ms)			
Catalan	380 (43.4); 331–488	596 (78.2); 474–763	***
Spanish	378 (59.5); 290–536	599 (120); 422–863	***
p	n.s.	n.s.	
AMPLITUDE (dB)			
Catalan	71.9 (4.2); 65.4–80	70.5 (1.2); 68.2–72.5	n.s.
Spanish	72.9 (3.1); 68.4–78.4	71.4 (3.3); 61.2–77.2	n.s.
p	n.s.	n.s.	
PITCH (Hz)			
Catalan	256 (61); 174–388	267 (40); 201–372	n.s.
Spanish	274 (57); 194–396	283 (56); 201–406	n.s.
p	n.s.	n.s.	

Measurements include whole word duration (ms), followed by amplitude (dB) and pitch (Hz) calculated on the vocalic portion of the words. Significant differences are also indicated.

Results of $t_{(23)}$ tests: $^{***}p < 0.001$; n.s. $p > 0.05$.

PROCEDURE

The familiarization-preference procedure with familiarization to passages and test on lists of words [as in Experiment 4, by Jusczyk and Aslin (1995)], was implemented in this research. The testing took place in a three-sided test booth, but instead of a frontal and two lateral lights typically used in the HPP set-up, a frontal display involving three computer screens and two concealed loudspeakers below the left and right monitor screens was used [this set-up had already been satisfactorily used by Bosch and Sebastián-Gallés (2001), to test for language discrimination in young infants]. Babies were seated on their parent's lap facing these three frontal monitor screens. Parents were listening to music through headphones throughout the whole experimental session. An experimenter, inside the testing room but out of the view of the infant, watched infants' looking behavior through a TV monitor, controlled trial presentation and recorded online infants' attention. By pressing and releasing the mouse button the experimenter could register the direction and duration of the infant look fixation toward the side screen involved in the presentation of the audio files in each trial. Online information about total attention time in each trial for each participant was stored and could later be checked against the results from off-line coding of the recordings to assess reliability of the measures and detect experimenter errors.

The experimental session began with a familiarization phase in which TB or GM passages were presented on alternating trials until the infant accumulated 45 s of attention time to each passage. Because of this criterion, infants could hear the target words in sentential contexts about 18 times each. Immediately after completing familiarization, the test phase began. It involved 16 test trials (four target word lists presented in four blocks). Words within each list were randomly presented and the order varied for each participant. At the beginning of each trial the central monitor displayed a flashing green circle to direct infants' attention toward the center. Immediately afterwards, one of the two lateral monitors displayed a flashing red circle to capture infant's attention and as soon as the infant oriented toward that side screen the audio files were presented. Auditory material was played until trial completion (18 s) or until the infant ceased to look in that direction for more than two consecutive seconds. In case of trial interruption, passage presentation was not resumed in the next trial, but started again from the beginning. Looks away below 2 s duration did not interrupt trial presentation but time away was not included in the final amount of fixation for that specific trial.

DESIGN

Half of the infants were familiarized to passages containing the target nouns "tren-bus" (TB condition) and the other half to passages containing the target nouns "gol-mar" (GM condition). In the test phase all participants were presented with the four target word lists.

RESULTS

Separate analyses were run, one on changes in attention time from the first to the last trial in the familiarization phase, and the other on attention time to familiar vs. novel words in the test phase. Regarding changes in attention during familiarization to the passages, a repeated-measures ANOVA with looking time as dependent measure, language group (Catalan, Spanish, bilingual) and familiarization condition (TB or GM) as a between-group factors and trial (first, last) as repeated measures revealed a highly significant effect of familiarization trial [$F_{(1,\ 48)} = 82.9$; $p = 0.0001$; $\eta^2 = 0.63$], but no effect of language group or condition and no significant interactions (all F's < 1). All three groups thus showed similar decays in attention from the first to the last trial in the familiarization phase when they were presented with the passages containing repetitions of two target words (mean attention time to first and last trial was, respectively, 15.1 s and 9.9 s in the Catalan monolingual group, 16 s and 8.2 s in the Spanish monolingual group and 15.2 s and 10.5 s in the bilingual group). Paired t tests conducted separately for each group on attention time to first and last familiarization trial confirmed the similarity in behavior [Spanish monolingual: $t_{(17)} = 6.1$, $p = 0.0001$; Cohen's $d = 2.0$]; [Catalan monolingual: $t_{(17)} = 4.7$, $p = 0.0001$; Cohen's $d = 1.41$]; [bilingual: $t_{(17)} = 4.9$, $p = 0.0001$; Cohen's $d = 1.21$].

We also analyzed if groups differed in the number of trials to reach criterion. A one-way ANOVA on number of trials in the familiarization phase as dependent measure and condition (TB vs. GM) and language group (Spanish, Catalan and bilingual) as between-subjects factors revealed no significant effects (F's < 1) or interaction [$F_{(2,\ 48)} = 1.77$, $p = 0.18$; $\eta^2 = 0.69$]. These results suggest that duration of the familiarization and participants' looking behavior in this phase can be considered equivalent.

To assess word segmentation, mean attention time to familiar vs. novel words in the test phase was computed for each participant (see **Figure 1**). A repeated-measures ANOVA with mean looking time as dependent measure, language group (Catalan, Spanish, bilingual) and familiarization condition (TB or GM) as a between-group factors and type of word (familiar, novel) as repeated measures was run. Results only revealed a highly significant main effect of type of word [$F_{(1,\ 48)} = 21.6$; $p = 0.0001$; $\eta^2 = 0.31$], with no language group or condition effects (both F's < 1) and no interactions. Paired t tests conducted separately for each group on mean attention time to familiar vs. novel word lists confirmed the presence of significant differences in attention to the two types of words, thus indicating that segmentation of monosyllabic words had been reached [Spanish monolingual: familiar words $M = 6.6$ s ($SD = 3.2$) and novel words $M = 7.9$ s ($SD = 2.7$); $t_{(17)} = -2.7$, $p = 0.015$; Cohen's $d = 0.41$]; [Catalan monolingual: familiar words $M = 6.1$ s ($SD = 2.8$) and novel words $M = 7.5$ s ($SD = 3$); $t_{(17)} = -2.6$, $p = 0.019$; Cohen's $d = 0.48$]; [bilingual: familiar words $M = 5.7$ s ($SD = 2.3$) and novel words $M = 7$ s ($SD = 2.7$); $t_{(17)} = -2.8$, $p = 0.011$; Cohen's $d = 0.51$]. Interestingly, however, the pattern of preference that was obtained at 8 months across all three groups was not the expected one, as a familiarity preference rather than novelty is typically observed in segmentation tasks using natural language. The monosyllabic nature of the target items, the fact that they were presented twice in sentence-final position in the familiarization passages, together with the use of IDS and a sufficiently long familiarization phase are possible factors that might explain this unexpected novelty preference, which is more likely to be found when the task is relatively easy and can be completed within the temporal limits established by the procedure.

EXPERIMENT 2: WORD SEGMENTATION AT 6 MONTHS

PARTICIPANTS

As in Experiment 1, a total of 54 healthy full-term infants with no history of hearing or vision problems were included in the sample divided into three groups ($N = 18$ in each group) according to the language/s spoken in their environment (Catalan only,

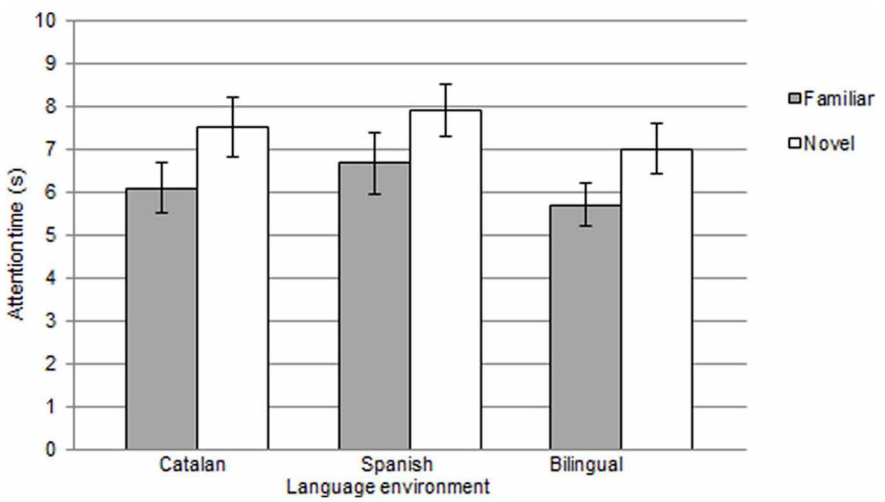


FIGURE 1 | Mean attention time (s) and standard error to familiar and novel words presented in the test phase, for the 8-month-old infants, grouped by language environment (monolingual Catalan, monolingual Spanish and bilingual).

Spanish only or both languages on a daily basis). Mean age of the infants in the Catalan groups was 6 months 6 days (range: 5 months, 22 days–6 months, 29 days); in the Spanish monolingual group was 6 months 4 days (range: 5 months, 19 days–6 months, 27 days) and in the bilingual group was 6 months 7 days (range: 5 months, 19 days–6 months, 27 days). No significant between-group age differences were found ($F < 1$). Following the information from the initial language questionnaire to parents, participants were assigned to different language groups. Inclusion criteria were the same as in Experiment 1. Mean percentage of exposure to Catalan in the Catalan monolingual group was 91% (range: 80–100%) and to Spanish in the Spanish monolingual group was 95% (range: 75–100%). From the eighteen infants in the bilingual group, 12 had a higher amount of exposure to Spanish than to Catalan (65–35%) and they were tested on Spanish material. The remaining six had a higher exposure to Catalan than to Spanish (64–36%) and they were tested on Catalan material. Twenty-nine additional infants were also tested but excluded from the final sample due to fussiness or crying leading to incomplete testing (22), very short looking time—below 1 s—to trials in the test phase (5) and experimental error (2).

STIMULI, PROCEDURE, AND DESIGN

Same as in Experiment 1.

RESULTS

Separate analyses were also run on data from these younger-age groups to explore attention behavior in the familiarization phase (expected decay of looking time) and possible differences in attention time to familiar vs. novel words in the test phase, as indicative of successful word segmentation.

Concerning attention behavior during familiarization, a repeated-measures ANOVA with looking time as dependent measure, language group (Catalan, Spanish, bilingual) and familiarization condition (TB or GM) as a between-group factors and trial (first, last) as repeated measures revealed a highly significant effect of familiarization trial [$F_{(1,\ 48)} = 42.9$; $p = 0.0001$; $\eta^2 = 0.47$], and no effect of language group or condition and no significant interactions [familiarization trial × language group: $F_{(2,\ 48)} = 1.15$; $p = 0.32$; $\eta^2 = 0.04$; familiarization trial × condition: $F_{(1,\ 48)} = 1.7$; $p = 0.19$; $\eta^2 = 0.03$; familiarization trial × language group × condition: $F < 1$]. All three groups showed a decrement in their attention time during familiarization to passages containing repetitions of target words (mean attention time to first and last trial was, respectively, 15.5 s and 12.3 s in the Catalan monolingual group, 15.5 s and 10.6 s in the Spanish monolingual group and 14.9 s and 9.1 s in the bilingual group). Paired t tests conducted separately for each group on attention time to first and last familiarization trial confirmed the similarity in this behavior [Spanish monolingual: $t_{(17)} = 4.1$, $p = 0.001$; Cohen's $d = 1.24$]; [Catalan monolingual: $t_{(17)} = 2.9$, $p = 0.009$; Cohen's $d = 0.85$]; [bilingual: $t_{(17)} = 4.2$, $p = 0.001$; Cohen's $d = 1.29$]. We also analyzed if groups differed in the number of trials to reach criterion. A One-Way ANOVA on number of trials in the familiarization phase as dependent measure and condition (TB vs. GM) and language group (Spanish, Catalan, and bilingual) as between-subjects factors revealed no significant effect of language group ($F < 1$), but a significant effect of condition [$F_{(1,\ 48)} = 5.59$, $p = 0.02$; $\eta^2 = 0.1$], with no significant group × condition interaction [$F_{(2,\ 48)} = 1.11$, $p = 0.33$; $\eta^2 = 0.04$]. Follow-up t tests revealed that mean number of trials to reach criterion in the TB condition (9.1) was significantly higher than in the GM condition (7.5) [$t_{(26)} = -2.56$, $p = 0.017$; Cohen's $d = 0.67$]. Further t tests by language groups indicated that only in the monolingual Catalan group the number of trials to reach criterion was significantly different by condition [$t_{(8)} = -2.8$, $p = 0.02$; Cohen's $d = 1.5$]. Differences by condition did not reach significance in the other two groups [monolingual Spanish: $t_{(8)} = -1.04$, $p = 0.32$; Cohen's $d = 0.26$; bilingual group $t < 1$].

To analyze word segmentation ability in the younger groups, mean attention time to familiar vs. novel words in the test phase was computed for each participant (see **Figure 2**). A repeated-measures ANOVA with mean looking time as dependent measure, language group (Catalan, Spanish, bilingual) and familiarization condition (TB or GM) as a between-group factors and type of word (familiar, novel) as repeated measures was run. Results only revealed a highly significant main effect of type of word [$F_{(1,\ 48)} = 18.9$; $p = 0.0001$; $\eta^2 = 0.87$], with no language group or condition effects [language group: $F_{(2,\ 48)} = 1.9$; $p = 0.15$; $\eta^2 = 0.07$; and condition $F < 1$] and no significant interactions (F's < 1). Paired t tests conducted separately for each group on mean attention time to familiar vs. novel word lists confirmed the presence of significant differences in attention to the two types of words, thus indicating that segmentation of monosyllabic words had successfully been reached at this early age [Spanish monolingual: familiar words $M = 8.6$ s ($SD = 3.1$) and novel words $M = 7.1$ s ($SD = 2.7$); $t_{(17)} = 2.2$, $p = 0.035$; Cohen's $d = 0.48$]; [Catalan monolingual: familiar words $M = 7.1$ s ($SD = 2.8$) and novel words $M = 6.1$ s ($SD = 2.7$); $t_{(17)} = 4.3$, $p = 0.0001$; Cohen's $d = 0.39$]; [bilingual: familiar words $M = 6.7$ s ($SD = 2.8$) and novel words $M = 5.5$ s ($SD = 2.9$); $t_{(17)} = 2.2$, $p = 0.038$; Cohen's $d = 0.41$]. Overall, results indicate that 6 month olds (monolinguals and bilinguals) can segment monosyllabic words from sentential contexts and evidence for this ability is reflected in the familiarity preference response pattern observed for words in the test phase. This is actually the usual preference pattern obtained in this type of task and it differs from the pattern found with the older groups tested in this research using exactly the same material and procedure.

A final analysis involving data from the two age groups was undertaken and only the age (6 vs. 8 months) per type of list (familiar vs. novel) interaction was deemed significant [$F_{(1,\ 102)} = 40.4$; $p = 0.0001$; $\eta^2 = 0.28$], confirming the radical change in the direction of preference that had taken place between the two ages under analysis. No other effects or interactions were found significant in this global analysis. We also extended the analysis to the attention time measures in the familiarization phase to check for any between-age differences in attention behavior to trials in the familiarization that could be related to the word preferences observed in the test phase. Results yielded no evidence of significant differences by age related to the attention time measures in the familiarization phase [$F_{(1,\ 102)} = 1.55$, $p = 0.21$; $\eta^2 = 0.01$], nor in the number of

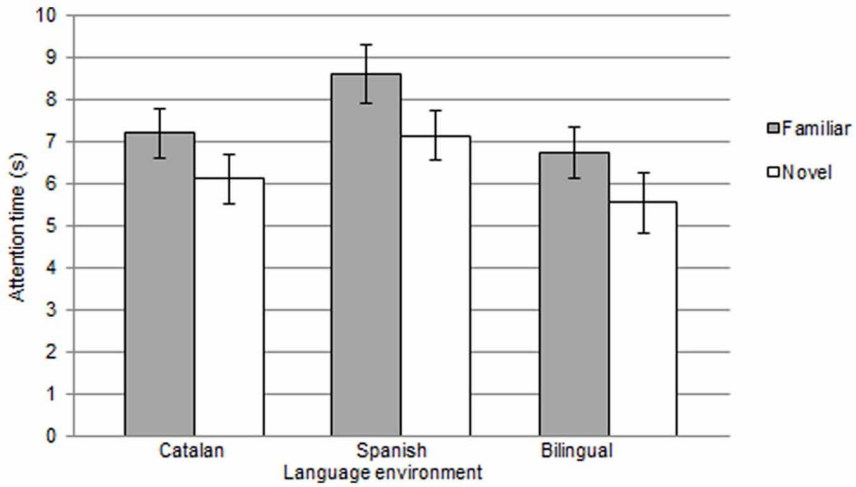


FIGURE 2 | Mean attention time (s) and standard error to familiar and novel words presented in the test phase, for the 6-month-old infants, grouped by language environment (monolingual Catalan, monolingual Spanish and bilingual).

trials to reach criterion [$F_{(1,\ 107)} = 1.34$, $p = 0.24$; $\eta^2 = 0.14$]. To sum up, although age differences did not seem to affect the behavior in the familiarization phase, they were determinant in the preference pattern observed in the test.

DISCUSSION

This research has explored young infants' emerging ability to segment simple, monosyllabic word-forms from fluent speech (natural language) in syllable-timed languages other than French. Six and eight-month-old participants growing up in Catalan monolingual, Spanish monolingual and Spanish-Catalan bilingual families were tested on a version of HPP, with familiarization to passages containing repetitions of two different target words in sentential contexts, and tested on words. Results revealed that all groups at both ages were able to successfully segment words from the passages and recognize them in the test phase. Critically, however, the predominant response pattern obtained differed with age. While younger infants showed a familiarity preference, by far the most frequent pattern found in segmentation studies using natural language paradigms (Jusczyk and Aslin, 1995; Jusczyk et al., 1999a,b; Mattys et al., 1999; Mattys and Jusczyk, 2001a,b; Houston et al., 2004), older groups showed a novelty preference, i.e., a preference for the lists involving words not included in the familiarization passages.

Familiarity or novelty preference patterns are usually attributed to the ease or difficulty to solve the task at hand (Hunter and Ames, 1988). As an example, in segmentation studies using artificial language paradigms evidence of segmentation is usually linked to a novelty preference for part-words over words in the familiarization material. This is no surprise as artificial languages in these studies are considered more simplistic than natural languages, syllabic sequences lacking the higher levels of variability found in any of the relevant dimensions in natural speech material (Pelucchi et al., 2009). Based on language complexity factors, word segmentation experiments run on natural speech material are thus likely to yield results showing a familiarity rather than a novelty preference and this is the pattern that a priori could be expected in our research. However, our study focused on simple elements (monosyllabic words) to be segmented from passages recorded in IDS style and participants were acquiring a language in which the rhythmic unit is the syllable (Nazzi et al., 2006), so even if the experiment was run on natural language material, the sum of all these factors may have contributed to simplify the task, thus leading to a novelty preference pattern in infants' responses. This can especially be true at older ages, when greater ability in word segmentation might have already been acquired.

There are thus a number of factors that may have facilitated word segmentation in the populations under study. The use of IDS in the recording of the material is one of them. IDS has been described as having a slower rate of speech, longer pauses and greater pitch excursions favoring infant's attention to it (Fernald and Kuhl, 1987). This style also uses simplified sentence structures that together with prosodic exaggeration can facilitate speech processing and the extraction of units from the segmentation perspective. Support for this interpretation comes from a segmentation study in which nonsense sentences either with in ADS or IDS style were used to test segmentation ability in 7- and 8-month-old infants (Thiessen et al., 2005). Results indicated that only in the IDS condition segmentation could be reached, thus the prosodic characteristics of IDS seem to have facilitated the extraction of word-form units from material that was otherwise equivalent in terms of the statistical cues that could be used for segmentation. In our material, where sentences in the passages were about 7–8 syllables long and target words where often aligned to phrase boundaries, clearly demarcated by pauses, the extraction of the target elements from the passages was certainly facilitated (Seidl and Johnson, 2006). It is worth mentioning here that words in the passages and words in the lists differed in duration, as reported in **Table 2**. Variability did not preclude recognition of the target items: infants in any of the two age groups in this study did not fail to notice the correspondence

between the target words placed in sentential contexts and the words presented in isolation in the test phase, when duration was longer than when they were produced in sentential contexts. This result is similar to what has already been found in previous research using natural speech (Jusczyk and Aslin, 1995; Jusczyk et al., 1999a; Mattys and Jusczyk, 2001a,b), but it is reported here for infants tested at a younger age (6 months).

Another factor favoring word segmentation in our research is related to the length of the words (monosyllabic CVC items) and the match with the rhythmic unit of the languages under study. This is actually a key issue in our research. From the early rhythmic segmentation hypothesis, syllabic units would play a determinant role at the onset of word segmentation for infants acquiring languages with a syllable-timed rhythm (Nazzi et al., 2006). Thus, segmentation of monosyllabic word-forms should be easier in these languages than in languages belonging to a different rhythmic typology in which the rhythmic unit might not be the syllable. The fact that positive evidence for monosyllabic word segmentation has been obtained at 6 months of age in either Catalan-learning and Spanish-learning infants suggests that the match between the rhythmic unit and the length of the target words in our study may have favored an early onset of the segmentation abilities in our populations. Recall here that English-learning infants succeeded at monosyllabic word segmentation with natural language material at 7½ months of age but not earlier (Jusczyk and Aslin, 1995), unless highly familiar words such as the infants' own names preceded the target monosyllabic units facilitating the extraction of adjacent elements (Bortfeld et al., 2005). Without these or other additional cues to segmentation, English-learning infants seem to start segmenting simple word forms from fluent speech around 7 months of age.

It is interesting to note that in spite of the presence of some differential features between Catalan and Spanish possibly affecting their rhythmic properties (vowel reduction and more complex consonantal codas in the former), these differences have not had any clear impact on the emergence of the segmentation ability for monosyllabic word-forms. Neither the monolingual, nor the bilingual groups in this research have shown significant differences in their behavior in the segmentation task.

Data from other syllable-timed languages that could support the early rhythmic segmentation hypothesis are limited to French and mostly focused on bi-syllabic word segmentation (Nazzi et al., 2006; Goyet et al., 2010 for European French, and Polka and Sundara, 2012, for Canadian French) so no data are available regarding 6-month-old French-learning infants solving a word segmentation task. However, as mentioned in the introduction, these studies have reported a certain ease for syllabic segmentation compared to the segmentation of bi-syllabic words, so in spite of the differences and controversies when European and Canadian French segmentation studies are compared, there is some converging evidence about the facilitative role of the syllable as a unit for segmentation in these syllable-timed languages. But, finding differences between Spanish and Catalan on the one hand, and French on the other, on the early onset of word segmentation for monosyllabic units is also a possibility to be taken into account. Spanish and Catalan have contrastive and variable stress, a property not shared with French. Stress in French falls invariably on the last syllable of each word of phrase but it is actually mostly reduced in fluent speech; prominence of stressed syllables is very similar to their unstressed neighbors and only words in utterance-final position get some prosodic marking in the form of vowel lengthening (Tranel, 1987). The presence of variable stress in languages such as Catalan or Spanish can be a factor that enhances the perception of syllabic units, thus leading to an earlier onset of the segmentation ability at least for short monosyllabic items. This remains an open question requiring further analysis, but the positive effect of variability in the input to the young learner has already been pointed out in research addressing different aspects of language acquisition. For instance, high levels of acoustic/phonetic variability deriving from the use of multiple exemplars (several tokens from multiple speakers) in a word learning task involving phonologically similar words led participants to successful learning while they failed in a more simple, single-exemplar condition (Rost and McMurray, 2009). Another example can be found in research showing that the learning of non-adjacent dependencies is facilitated with decreasing predictability between adjacent elements, that is, the extraction of the invariant structure (the stable elements of a stimulus set) is actually easier with increasing variability of the irrelevant intervening elements (Gómez, 2002). Back to word segmentation in syllable-timed languages, it is possible to hypothesize that the presence of variable stress in the input may have enhanced the detection and extraction of monosyllabic word units. This is an issue to be further analyzed in future studies, where the facilitation effects of the syllable as the rhythmic unit for segmentation could be more carefully analyzed after controlling for other facilitation effects derived from the paradigm, task demands, or the specific properties of the speech material in the test.

The developmental change in the preference pattern obtained in our data, suggests rapid gains in segmentation ability for these short, monosyllabic units that match the rhythmic unit of the ambient language. Because the paradigm and material used in our experiments were exactly the same at both ages, the reversal of the preference pattern seems to confirm the ease to extract these short units from sentential contexts with increasing age. A reversal of the preference pattern had also been described in the literature (Thiessen et al., 2005), but in that case not only age but an extended familiarization phase were both factors that modified the pattern obtained at an earlier age. This is not the case in our study as no manipulation of the paradigm was done. A simpler interpretation is thus that infants have gained expertise in segmenting fluent speech, especially regarding monosyllabic elements. The question remains whether similar results would be obtained for CV items, instead of CVC, and whether segmentation of function CVC or CV words in these languages, involving unstressed vowels, would also be successfully solved at an early age and by all language groups (the presence of vowel reduction in Catalan but not in Spanish may also contribute to differential results). This is clearly a topic to be explored in future research.

The present paper has included participants growing up bilingual and their results deserve some comments. An early onset of monosyllabic word segmentation abilities has also been found in our Spanish-Catalan infant participants. The timing and

characteristics of their segmentation ability do not seem to differ from results obtained in monolingual infants in our study, at least from a behavioral perspective. The bilingual results are relevant in that they are the first evidence of segmentation abilities in bilinguals acquiring languages with rather similar rhythmic properties [so far, only preliminary data exist from French-English bilinguals showing bi-syllabic word segmentation ability in both languages by 8 months of age, as reported by Polka and Sundara (2003)]. Although bilinguals in our research are exposed to languages that do not greatly differ in their rhythmic properties and, from this perspective, it could be predicted that no delays or differences in solving the segmentation task would be found, bilingual exposure might nevertheless lead to small differences in the developmental time-course of certain speech and language abilities, as for instance those found in the phonetic categorization domain (Bosch and Sebastián-Gallés, 2003; Sebastián-Gallés and Bosch, 2009). Even if the ambient languages do not show great differences in their rhythmic characteristics, segmentation abilities might have been slightly delayed in this population, just as a consequence of adaptive processes to cope with the more complex input the bilingual is exposed to. This was not the case in our data, with bilinguals showing parallel results to their monolingual counterparts. These data extend to the word segmentation domain the notion that bilingual exposure does not alter the pattern of acquisition as observed in monolingual populations (Werker and Byers-Heinlein, 2008).

To sum up, results from this research (a) offer evidence of an early ability for monosyllabic word segmentation in syllable-timed languages such as Spanish and Catalan, not previously described in the literature; (b) reveal no differences between monolingual and bilingual participants in this task, probably because both languages in the bilingual environment share the same rhythmic properties; and (c) show a specific developmental pattern that is compatible with an interpretation based on the facilitation effect that can be observed when rhythmic properties of the language match with the units to be extracted from fluent speech. These results should be the basis for further research exploring disyllabic word segmentation in the same linguistic population. They can also offer relevant information for future cross-linguistic research and they should be useful in studies comparing normally developing infants and clinical groups at risk for language delays in speech segmentation tasks.

ACKNOWLEDGMENTS

Research supported by grant PSI2011-25376 from the Spanish *Ministerio de Ciencia y Economía*. We thank Jorgina Solé for help in recruiting and testing infants and all the families and infants that took part in this research.

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: University of Edinburgh Press.

Águila, E., Ramon, M., Pons, F., and Bosch, L. (2005). "Efecto de la exposición bilingüe sobre el desarrollo léxico inicial [Effect of bilingual exposure on early lexical development]," in *Estudios Sobre la Adquisición del Lenguaje*, eds M. A. Mayor Cinca, B. Zubiauz de Pedro, and E. Díez-Villoria (Salamanca, Spain: Ediciones Universidad de Salamanca), 676–692.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech stream segmentation. *Psychol. Sci.* 16, 298–304.

Bosch, L., and Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition* 65, 33–69.

Bosch, L., and Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy* 2, 29–49.

Bosch, L., and Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Lang. Speech* 46, 217–243.

Cutler, A., and Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Comput. Speech Lang.* 2, 133–142.

Fernald, A., and Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant. Behav. Dev.* 10, 279–293.

Gómez, R. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436.

Gout, A. (2001). *Etapes Précoces de l'acquisition du Lexique*. Unpublished dissertation. Ecole des Hautes Etudes en Sciences Sociales, Paris, France.

Goyet, L., de Schonen, S., and Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: an ERP study. *Brain Res.* 1332, 75–89.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., and Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition* 26, 269–286.

Höhle, B., and Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed class elements in continuous speech. *Dev. Sci.* 6, 122–127.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., and Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychon. Bull. Rev.* 7, 504–509.

Houston, D. M., Santelmann, L. M., and Jusczyk, P. W. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Lang. Cogn. Proc.* 19, 97–136.

Hunter, M. A., and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Adv. Infancy Res.* 5, 69–95.

Johnson, E., and Seidl, A. (2008). At eleven months, prosody still outranks statistics. *Dev. Sci.* 11, 1–11.

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 548–567.

Junge, C., Kooijman, V., Hagoort, P., and Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* 15, 463–473.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci.* 3, 323–328.

Jusczyk, P. W., and Aslin, R. (1995). Infant's detection of the sound patterns words in fluent speech. *Cogn. Psychol.* 29, 1–23.

Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999a). The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207.

Jusczyk, P. W., Hohne, E. A., and Bauman, A. (1999b). Infants' sensitivity to allophonic cues for word segmentation. *Percept. Psychophys.* 61, 1465–1476.

Kuijpers, C., Coolen, R., Houston, D., and Cutler, A. (1998). "Using the head-turning technique to explore cross-linguistic performance differences," in *Advances in Infancy Research, Vol. 12*, eds C. Rovee-Collier, L. Lippsitt, and H. Hyane (London: Ablex), 205–220.

Ladefoged, P. (1975). *A Course in Phonetics*. New York, NY: Harcourt Brace Jovanovich.

Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterisations of speech rhythm: syllable-timing in Singapore English. *Lang. Speech* 43, 377–401.

Marquis, A., and Shi, R. (2008). Segmentation of verb forms in preverbal infants. *J. Acoust. Soc. Am.* 123, EL105–EL110.

Mattys, S. L., and Jusczyk, P. W. (2001a). Do infants segment words or recurring contiguous patterns? *J. Exp. Psychol. Hum. Percept. Perform.* 27, 644–655.

Mattys, S. L., and Jusczyk, P. W. (2001b). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cogn. Psychol.* 38, 465–494.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., and Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: emerging evidence for crosslinguistic differences. *J. Mem. Lang.* 54, 283–299.

Nazzi, T., Jusczyk, P. W., and Johnson, E. K. (2000a). Language discrimination by English learning 5-month olds: effects of rhythm and familiarity. *J. Mem. Lang.* 43, 1–19.

Nazzi, T., Kemler Nelson, D., Jusczyk, P., and Jusczyk, A. M. (2000b). Six-month-olds' detection of clauses embedded in continuous speech: effects of prosodic well-formedness. *Infancy* 1, 123–147.

Nespor, M. (1990). "On the rhythm parameter in phonology," in *Logical Issues in Language Acquisition,* ed I. M. Roca (Dordrecht: Foris), 157–175.

Nespor, M., Shukla, M., and Mehler, J. (2011). "Stress-timed vs. syllable-timed languages," in *The Blackwell Companion to Phonology, Vol. II*, eds M. Van Oostendorp, C. J. Ewen, E. V. Hume, and K. Rice (Chichester, UK: Blackwell Publication Inc.), 1147–1157.

Newman, R. S., Ratner, N. B., Jusckzyk, A. M., Jusckzyk, P. W., and Dow, K. A. (2006). Infant's early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Dev. Psychol.* 42, 643–655.

Payne, E., Post, B., Astruc, L., Prieto, P., and Vanrell, M. (2009). Rhythmic modification in child directed speech. *Oxf. Univ. Work. Pap. Ling. Philol. Phon.* 12, 123–144.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev.* 80, 674–685.

Polka, L., and Sundara, M. (2003). "Word segmentation in monolingual and bilingual infant learners of English and French," in *Proceedings of the 15th International Congress of Phonetic Sciences,* eds M. J. Sole, D. Recasens, and J. Romero (Barcelona: Causal Productions), 1021–1024.

Polka, L., and Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect, and cross-language comparisons. *Infancy* 17, 198–232.

Prieto, P., Vanrell, M., Astruc, L., Payne, E., and Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Commun.* 54, 681–702.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292.

Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Dev. Sci.* 12, 339–349.

Sebastián-Gallés, N., and Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: is the distributional account all there is to it? *Dev. Sci.* 12, 874–887.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *J. Mem. Lang.* 57, 24–48.

Seidl, A., and Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Dev. Sci.* 11, 596–606.

Seidl, A., and Johnson, E. K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Dev. Sci.* 9, 565–573.

Singh, L., Reznick, J. S., and Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Dev. Sci.* 15, 482–495.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., and Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Mem. Lang.* 49, 249–267.

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71.

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries in 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.

Tranel, B. (1987). *The Sounds of French: An Introduction*. Cambridge: Cambridge University Press.

Werker, J. F., and Byers-Heinlein, K. (2008). Bilingualism in infancy: first steps in perception and comprehension. *Trends Cogn. Sci.* 12, 144–151.

White, L., and Mattys, S. L. (2007). Calibrating rhythm: first language and second language studies. *J. Phon.* 35, 501–522.

White, L., Mattys, S. L., and Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *J. Mem. Lang.* 66, 665–679.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2012; accepted: 14 February 2013; published online: 05 March 2013.

Citation: Bosch L, Figueras M, Teixidó M and Ramon-Casas M (2013) Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. Front. Psychol. ***4****:106. doi: 10.3389/fpsyg.2013.00106*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 17 January 2013
doi: 10.3389/fpsyg.2012.00590

Discovering words in fluent speech: the contribution of two kinds of statistical information

Erik D. Thiessen* and Lucy C. Erickson

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Maren Schmidt-Kassow, Goethe University, Germany
Juan M. Toro, University Pompeu Fabra, Spain
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

****Correspondence:***
Erik D. Thiessen, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
e-mail: thiessen@andrew.cmu.edu

To efficiently segment fluent speech, infants must discover the predominant phonological form of words in the native language. In English, for example, content words typically begin with a stressed syllable. To discover this regularity, infants need to identify a set of words. We propose that statistical learning plays two roles in this process. First, it provides a cue that allows infants to segment words from fluent speech, even without language-specific phonological knowledge. Second, once infants have identified a set of lexical forms, they can learn from the distribution of acoustic features across those word forms. The current experiments demonstrate both processes are available to 5-month-old infants. This demonstration of sensitivity to statistical structure in speech, weighted more heavily than phonological cues to segmentation at an early age, is consistent with theoretical accounts that claim statistical learning plays a role in helping infants to adapt to the structure of their native language from very early in life.

Keywords: statistical learning, word segmentation, lexical stress, infant language, phonology

INTRODUCTION

The ability to segment words from fluent speech is taken for granted by adults, but it represents a major accomplishment for infants. Unlike the white spaces between words on the written page, pauses do not consistently mark word boundaries in fluent speech. This is not troublesome for adults, who can identify word boundaries in large part due to their familiarity with the word forms in their native language (e.g., Nazzi et al., 2005; Norris and McQueen, 2008). Infants, though, begin the task of word segmentation unable to take advantage of familiar word forms. The challenge faced by infants is comparable to the task faced by adults attempting to identify words spoken in a foreign language. Nevertheless, infants succeed in this task before they have amassed a large lexicon of familiar word forms (e.g., Jusczyk and Aslin, 1995; Bortfeld et al., 2005). Two cues have been suggested to play a role in infants' earliest ability to segment words from fluent speech: conditional statistical information, and information about the prosodic structure of words (Thiessen and Saffran, 2003). These cues are likely to work together in natural languages, but an open developmental question is which is available to infants earlier in development. In this series of experiments, we will examine the hypothesis that sensitivity to conditional structure is available from an earlier age, and that statistical learning helps infants discover the predominant prosodic structure of words in their native language.

There is no doubt that information about the prosodic structure of words plays a role in infants' and adults' word segmentation. The difference between stressed and unstressed syllables is perceptually available to infants from a young age (e.g., Jusczyk and Thompson, 1978; Weber et al., 2005). To the extent that stressed and unstressed syllable systematically occur in particular word positions, this distinction can serve as a cue to word boundaries. In English, for example, most bisyllabic content words follow a trochaic pattern: they begin with a stressed syllable, and are followed by an unstressed syllable (Cutler and Carter, 1987). English-learning infants prefer to listen to trochaic words over words with a weak-strong (iambic) pattern (Jusczyk et al., 1993). When exposed to a stream of syllables, English-learning infants and English-speaking adults treat the stressed syllables as word onsets (e.g., Cutler and Norris, 1988; Echols et al., 1997; Jusczyk et al., 1999). Importantly, though, not all languages show this trochaic predominance; lexical items in other languages may be predominantly iambic. Therefore, English-learners trochaic bias is likely acquired from experience with the language (Thiessen and Saffran, 2007).

By contrast, sensitivity to conditional statistical information does not require language-specific knowledge; it is a cue to word segmentation that is available cross-linguistically. This cue is relevant to word segmentation because sounds within a word are more likely to co-occur than sounds across word boundaries (Hayes and Clark, 1970). For example, *copter* is very likely to occur after *heli*; but many words could potentially occur after *helicopter*. Conditional statistics – such as transitional probability (e.g., Saffran et al., 1996) – reflect the likelihood of co-occurrence among elements of the input. A body of prior research indicates that both infants and adults are able to segment words from fluent speech on the basis of conditional statistical information. For example, artificial language experiments demonstrate that after exposure to a sequence of syllables, both infants and adults are able to distinguish between syllable groups with high conditional relations (i.e., words), and syllable groups with low conditional relations, such as groupings that occur across word boundaries (e.g., Aslin et al., 1998; Thiessen and Saffran, 2004).

A variety of different computational accounts have been proposed to explain sensitivity to conditional statistical information (for discussion, see Frank et al., 2010). The most successful of these models – clustering models – search for and store clusters of statistically coherent elements (e.g., Perruchet and Vinter, 1998; Orban et al., 2008). These models predict that after exposure to speech, participants should have extracted a set of candidate lexical items (e.g., Giroux and Rey, 2009). Research with both infants and adults is consistent with this prediction. For example, infants accept words from the synthesized speech in English utterances after exposure to a stream of synthesized speech (Saffran, 2001). Similarly, infants and adults learn labels for novel objects more easily when provided the opportunity to segment the labels from fluent speech (Graf Estes et al., 2007; Mirman et al., 2008).

In word segmentation tasks, for example, this means that exposure to fluent speech leads to learners extracting a set of candidate lexical items. Evidence that learners are extracting clusters of statistically coherent elements can be seen even for non-linguistic stimuli (e.g., Fiser and Aslin, 2005), suggesting that this extraction is a domain-general aspect of conditional statistical learning.

The fact that infants are capable of extracting and storing word forms is consistent with a statistical bootstrapping account of the development of word segmentation (Thiessen and Saffran, 2003). On this account, infants initially rely on language-universal cues – such as sensitivity to conditional statistical information – to segment words from fluent speech. Once they have identified and stored a set of word forms, they can identify the acoustic features that are consistent across them (e.g., Lew-Williams and Saffran, 2012). For example, if infants are exposed to a set of words in which stress consistently occurs on the first syllable, they will acquire a trochaic bias (Thiessen and Saffran, 2007). Once infants have discovered the acoustic features that are consistent in their proto-lexicon, they can use these features as cues to subsequent word segmentation (e.g., Johnson and Jusczyk, 2001).

This transition is from language-general to language-specific cues is thought to take place between 7 and 9 months. While 7-month-old infants rely on conditional statistical information to segment fluent speech, 9-month-old infants favor lexical stress, even if segmenting on the basis of stress contradicts conditional statistical information (Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003). Recent research by Höhle et al. (2009), however, indicates that infants as young as 6 months are familiar with the predominant prosodic structure of words in their native language. Höhle et al. suggest that 6 months is below the age at which infants are able to segment words from fluent speech via conditional statistical cues. If so, the statistical bootstrapping account of infants' prosodic learning is necessarily incorrect. Instead, this would suggest that language-specific prosodic cues may be the earliest cue infants use to segment words from fluent speech. Additionally, it would suggest that knowledge about the prosodic form of words arises from some source other than statistical learning, perhaps such as learning solely from words in isolation.

However, the claim that infants below 6 months are unable to segment speech on the basis of conditional statistical information may be incorrect. Evidence suggests that young infants and even neonates are sensitive to conditional statistical information (Kirkham et al., 2002; Teinonen et al., 2009; Kudo et al., 2011). Further, one prior experiment indicates that 5- to 6-month-old infants are able to segment fluent speech via conditional statistical information (Johnson and Tyler, 2010). In Experiment 1, we seek to provide additional evidence that infants are able to segment fluent speech below 6 months of age. Additionally, we will investigate whether infants at this young age prioritize conditional statistical information over lexical stress as a cue to word segmentation, consistent with the statistical bootstrapping account. In Experiment 2, we will investigate whether infants in this age range are capable of learning to use lexical stress as a cue to word segmentation.

EXPERIMENT 1A

Within the word segmentation literature, it is commonly held that infants develop the ability to segment fluent speech by 7.5 months, citing a seminal study by Jusczyk and Aslin (1995). Before this age, researchers have asserted that infants lack the ability to extract words from fluent speech on the basis of statistical structure (e.g., Höhle and Weissenborn, 2003). Others have proposed that the ability to segment words from fluent speech via transitional probabilities is intact earlier (e.g., Thiessen and Saffran, 2003; Johnson and Tyler, 2010). Evidence from neuroimaging is consistent with this claim (e.g., Teinonen et al., 2009; Kudo et al., 2011). The goal of Experiment 1A was to provide further behavioral evidence that infants are capable of segmenting fluent speech via conditional statistical information below 6 months. To do so, we exposed 5-month-old infants to an artificial language in which the only cue to segmentation is higher conditional relations between syllables within words relative to syllables spanning word boundaries (part-words). If the ability to segment speech does not emerge until later than 7 months, these 5-month-old infants should not discriminate between words and part-words following familiarization with this fluent speech stream. However, if the ability to parse speech on the basis of statistical cues is intact at an earlier age, infants should discriminate between words and part-words.

MATERIALS AND METHODS

Participants

Data were obtained from 10 participants between the ages of 5.0 and 5 months, 14 days ($M = 5.10$). To obtain data from 10 infants, it was necessary to run 13 infants. The additional three infants were excluded for crying during the testing session (1), average looking times of less than 3.0 s (1), or experimenter error (1). A sample size of 10 infants was used based on a power analysis using an effect size calculated from Thiessen and Saffran's (2003) Experiment 3, of which this experiment is a replication with a younger age group.

Stimuli

The stimuli used in this experiment were identical to those used in Thiessen and Saffran's (2003) Experiment 3. Infants were exposed to an artificial language containing four bisyllabic nonsense words: *diti*, *bugo*, *dapu*, and *dobi*. The language was synthesized using MacinTalk, and all syllables were produced with neutral stress. This language was constructed such that two of the words – *dapu* and *dobi* – occurred twice as often (90 times) as the other two words (*diti* and *bugo*, each of which occurred 45 times). This ensures that test item foils can be constructed that differ solely on

their conditional probabilities, rather than on the frequency with which infants hear them (for discussion, see Aslin et al., 1998). Words occurred in a pseudo-random order, with the constraint that no word could follow itself. Syllable-to-syllable transitional probabilities were 100% within a word, and 33% at word boundaries. Because there were no pauses or other acoustic cues to word boundaries in this artificial language, the conditional probabilities (high within a word, low at boundaries) provided the only cue to word segmentation.

Two kinds of test items were created to assess infants' ability to segment the language: words and part-words. The word test items were the infrequent words (*diti* and *bugo*) from the artificial language. Part-words were syllable conjunctions that occurred across the two more frequent words (*bida* and *pudo*). During the infants' exposure to the artificial language, both words and part-words occurred equally often. Therefore, any difference in infants' responses to these two kinds of test items is not due to the frequency with which they have heard the words or part-words.

Procedure

Infants were tested individually in a sound-attenuated testing room, seated on a caregiver's lap 150 cm away from a 32″ LCD monitor. An experiment outside the testing room observed the infant over closed-circuit video and recorded the duration of his or her gaze at the central monitor using the Habit X software (Cohen et al., 2004). To eliminate bias, parents were asked to wear headphones, and the experimenter was blind to the nature of the stimuli being presented. Two speakers situated next to the central LCD monitor were used to present the audio stimuli.

At the beginning of the experiment, the infants' attention was attracted to the central LCD monitor by the presentation of a colorful Winnie the Pooh video, accompanied by an attention-getting phrase. Once the infant looked at the central monitor, the video was replaced by a static image of a checkerboard, and the artificial language began to play. The checkerboard remained on screen, and the language continued to play, for 2 min. At the end of this time, the attention-getting movie reappeared on the screen.

Once infants focused their gaze on the central monitor, the test phase began. During this phase, 12 test trials were presented. Six of these trials were word trials, and six were part-word trials. Each test item occurred on three trials during the testing phase. Test trials were presented in random order. A test trial began with the attention-getting movie playing on the central monitor drawing the infants' gaze forward. When the observing experimenter pressed a key indicating that the infant had fixated, the monitor displayed a video of a looming green ball on a black background, while the speakers began to play the test item (either word or part-word) separated by 1.4 s pauses. For as long as the infant maintained their gaze on the central monitor, the test trial continued, up to a maximum of 20 s. When the infant looked away for more than two consecutive seconds, the test trial ended and the attention-getting video reappeared on the central monitor.

RESULTS

If infants were able to successfully segment the artificial language, they should respond differentially to word test trials than to part-word test trials (e.g., Saffran et al., 1996). While in principle, any group-level preference is indicative that infants are able to differentiate the items, the experiments most similar to this one have resulted in a novelty preference (e.g., Thiessen and Saffran, 2003, 2007). If infants in this experiment behave in the same way, they should look longer at test items that violate their expectations (i.e., part-words) than at test items that fit what they have learned (i.e., words).

The results were consistent with prior experiments using these stimuli. Infants in this experiment displayed a novelty preference, listening longer to part-words ($M = 8.10$ s, SE = 0.90) than words ($M = 6.78$ s, SE = 1.34; See **Figure 1**). A paired-samples t-test (all t-tests reported here and in subsequent experiments are two-tailed) revealed that the difference in listening times as a function of test item type was significant, $t(9) = 2.609$, $p < 0.05$. After familiarization, 5-month-old infants distinguished between words and part-words, indicating that they had succeeded in parsing the speech signal.

DISCUSSION

The fact that infants were able to segment the artificial language used in this experiment is inconsistent with the common assertion that speech segmentation does not begin until around 7 months of age (e.g., Jusczyk and Aslin, 1995; Höhle and Weissenborn, 2003). Instead, it is consistent with prior results indicating that infants are sensitive to conditional statistical information from a young age (Kirkham et al., 2002; Teinonen et al., 2009; Johnson and Tyler, 2010; Kudo et al., 2011). Indeed, to our knowledge the infants in this experiment are younger than any prior group of infants in a behavioral word segmentation experiment. The fact that they successfully segmented raises the possibility that word segmentation may begin at younger ages than previously thought in native language environments. Moreover, the 5-month-olds in this experiment are demonstrating sensitivity to conditional statistical information at a younger age than any prior experiment

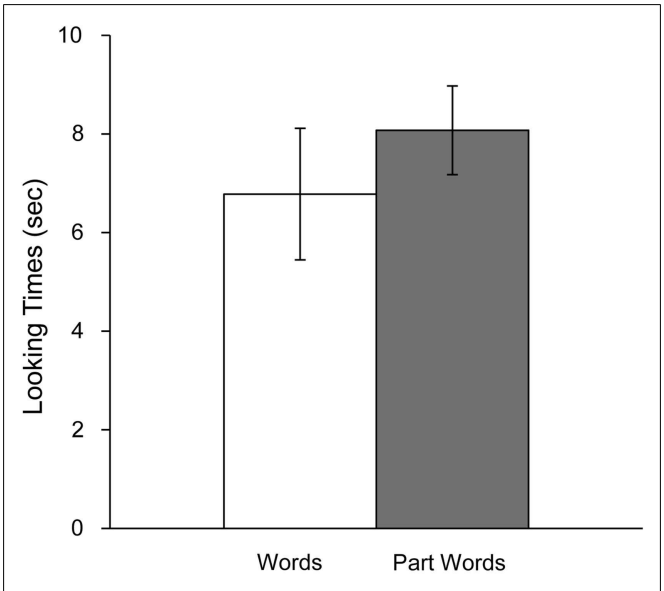


FIGURE 1 | Looking times to words and part-words in Experiment 1A. Error bars indicate standard error.

has found sensitivity to language-specific acoustic cues to segmentation, such as lexical stress patterns. As such, these results are consistent with the hypothesis that conditional statistical information is one of the first cues available to infants as they begin to discover word forms in speech.

EXPERIMENT 1B

Experiment 1A demonstrated that 5-month-old infants are able to segment word forms from speech solely on the basis of conditional probability information. In Experiment 1B, we were interested in how infants of this age behave when statistical cues to word identity are placed in direct conflict with lexical stress, an acoustic cue thought to be very salient to infants (e.g., Gleitman et al., 1988; Echols and Newport, 1992). Much research attests to infants' early sensitivity to prosodic information (e.g., Mehler et al., 1988) and preference that emerges at 9-months in English-exposed infants for trochaic words (consisting of a strong/weak pattern) over iambic words (weak/strong; Jusczyk et al., 1993). Additionally, 7.5-month-old infants in English-speaking environments are so reliant on lexical stress that they display a trochaic bias during segmentation, such that when exposed to passages containing the sequence "guiTAR#is," they segment the trochaic sequence "TARis" from fluent speech even when it when occurs less frequently than the iambic sequence "guiTAR" (Jusczyk et al., 1999).

In the present experiment, we were interested in whether infants would extract units from familiarization on the basis of conditional information (i.e., extract syllable pairings characterized by high transitional probabilities) or on the basis of lexical stress cues (i.e., trochees following the dominant pattern of English). Based on the prior finding that 7-month-olds ignore stress cues, segmenting items on the basis of conditional information (Thiessen and Saffran, 2003), we predicted that 5-month-old infants in this study would also extract units according to this language-universal strategy rather than on lexical stress, which requires language-specific knowledge about words. If infants of this age segment statistical words rather than trochaic disyllables, this would provide strong support for the idea that conditional information is one powerful language-universal cue that could be recruited to acquire language-specific knowledge such as the preferred position of stressed syllables within word forms. In contrast, if these infants extract trochees from the speech stream, even when they are characterized by low transitional probabilities, this would be consistent with the early rhythmic segmentation hypothesis, proposed by Nazzi and colleagues (e.g., Nazzi and Ramus, 2003; Nazzi et al., 2006; Höhle et al., 2009; Mersad and Nazzi, 2011). According to this hypothesis, early segmentation is based on the rhythmic unit of the native language, which derives from infants' early sensitivity to language rhythm.

MATERIALS AND METHODS

Participants

Data were obtained from 20 participants between the ages of 5.0 and 5 months, 15 days ($M = 5.9$). Half of these infants were exposed to a trochaic artificial language, and half to an iambic artificial language. To obtain data from 20 infants, it was necessary to run 23 infants. The additional three infants were excluded (two from the trochaic condition, one from the iambic condition) for crying during the testing session. A sample size of 10 infants for each language was used based on a power analysis of Thiessen and Saffran's Experiment 2, of which this experiment is a replication with a younger age group.

Stimuli

The artificial language used in this experiment had the same lexical items, word order, and statistical structure as the language used in Experiment 1A. Two versions of this language were used. In the trochaic language, lexical stress occurred in word-initial position, while in the iambic language lexical stress occurred in word-final position. For an illustration of the competing segmentations indicated by transitional probabilities and lexical stress in the iambic language, see **Figure 2**.

Lexical stress was created by altering three parameters of the stimuli: pitch contour, amplitude, and duration. The pitch contour in the stressed syllables was based on the pitch contours of an adult native English speaker producing the lexical items. The pitch peak of the vowels varied between 255 and 270 Hz, compared to a monotonic 200 Hz for the unstressed syllables. The pitch contour varied as a function of whether the syllable began with a voiced or a voiceless consonant. For voiced consonants, the pitch contour traced an inverted parabola, peaking near the midpoint of the vowel. For voiceless consonants, the pitch contour began near the peak, and traced a falling plateau. The amplitude of all stressed consonants was increased uniformly by 4 dB. The duration of the stressed syllables was altered by lengthening only the vowels. The average duration of the stressed syllables was 310 ms, compared to 185 ms for unstressed syllables. These languages were identical to those used in Thiessen and Saffran's (2003) Experiments 1 and 2. The duration of both languages was 140 s. The test items used were identical to those used in Experiment 1A.

Procedure

The procedure of this experiment was identical to that used in Experiment 1A.

RESULTS

If infants segment fluent speech via sensitivity to conditional statistical information, they should show the same pattern of preference in the test phase, regardless of whether they heard the trochaic or iambic language, because the conditional statistical information is identical across these two languages. However, if infants segment the artificial language via lexical stress, they should show the opposite pattern of preference across the two languages, because lexical

…diTIbuGOdaPUdoBIbuGOdaPUdiTI…

… di_TIbu_GOda_PUdo_BIbu_GOda_PUdi_TI…

…diTI_BUgo_daPU_doBI_buGO_daPU_diTI…

FIGURE 2 | Top: an excerpt of the iambic familiarization stream used in Experiment 1B; capitalized syllables represent stress. Middle: segmentation based on transitional probabilities. Bottom: segmentation based on trochaic bias.

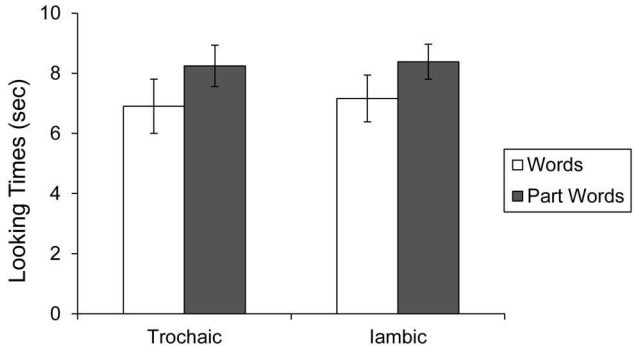


FIGURE 3 | Looking times to words and part-words in the trochaic and iambic conditions of Experiment 1B. Error bars indicate standard error.

stress occurs in word-initial position in the trochaic language and word-final position in the iambic language.

To determine whether preference for type of test items (words vs. part-words) differed as a function of condition (trochaic vs. iambic language exposure), a 2 × 2 ANOVA (Test Item × Condition) was performed (**Figure 3**). The main effect for test item (listening to words vs. part-words) was significant, $F(1, 18) = 11.98$, $p < 0.05$, indicating that infants exposed to both languages listened longer to part-words than words. Infants exposed to the trochaic language listened to part-words for 8.25 s (SE = 0.69) and to words for 6.90 s (SE = 0.90). Infants who were exposed to the iambic language listened to part-words for 8.39 s (SE = 0.58) and to words for 7.16 s (SE = 0.78). The main effect of condition (trochaic vs. iambic exposure) was not significant, $F(1, 18) = 0.041$, $p = 0.84$, indicating that infants listened similar lengths of time regardless of which language they heard. The interaction between test item and condition was also not significant, $F(1, 18) = 0.027$, $p = 0.87$, meaning that direction of preference for test items did not differ based on language exposure.

DISCUSSION

The results of Experiment 1B indicate that, regardless of whether they heard a language made up of trochaic words or iambic words, infants showed the same preference at test. This indicates that infants segmented the same items from both the trochaic and the iambic language. The fact that infants in both groups preferred part-words, as did infants in Experiment 1A, further supports this conclusion. This consistent preference across the trochaic and iambic language indicates that infants segmented the same items from both familiarization streams. The only cue to segmentation that is identical across the streams is the conditional statistical information, indicating that infants segmented on the basis of statistical cues. If infants had relied on lexical stress, they would segment the two languages differently (e.g., Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003).

These results support our prediction that 5-month-olds should rely on conditional statistical information over lexical stress, as do 7-month-olds infants (Thiessen and Saffran, 2003). This is consistent with proposal that use of statistical cues to segment speech develops earlier than use of acoustic cues such as lexical stress. More broadly, this developmental timetable is consistent with the hypothesis that sensitivity to conditional statistical information allows infants to discover a set of lexical forms, which in turn allow infants to identify language-specific acoustic cues such as lexical stress. Rather than statistical cues and acoustic cues being in conflict (as they are artificially placed in the iambic familiarization stream), conditional statistical information may actually allow infants to discover the dominant rhythmic patterns of their native language (Thiessen and Saffran, 2007).

EXPERIMENT 2

Experiments 1A and 1B established that (1) the ability to segment fluent speech on the basis of conditional information is present as early as 5 months of age and (2) that these infants segment on the basis of statistical cues rather than lexical stress cues when they are placed in conflict, replicating the findings of Thiessen and Saffran (2003) with 7-month-olds. By 9 months of age, the weight infants place on conditional statistical cues vis a vis lexical stress has changed, and they rely on stress cues to a greater extent than conditional statistical information. Thiessen and Saffran (2007) suggest that this developmental progression is due to statistical learning. Statistical learning plays two roles in this progression. The first, as demonstrated in Experiment 1, is that infants are able to use conditional statistical information to extract a set of lexical forms from fluent speech. The second is that statistical learning allows infants to identify the commonalities across these word forms, which relies upon distributional (as opposed to conditional) statistical information.

This hypothesis suggests that, once infants have discovered a set of word forms, they integrate information across them. Consider what would happen, for example, if an infant were familiar with the three words *baby*, *diaper*, and *shoe*, and integrated across these word forms. Integrating information across these word forms will emphasize information that is consistent across word forms, while de-emphasizing information that is inconsistent (e.g., Thiessen and Pavlik, 2012). In this case, there is no consistent phonemic information across the three known words, but all three begin with a stressed syllable. Integrating information across a lexicon like this should lead infants to discover that lexical forms can vary in their phonemic identity, but show a consistent word-initial stress pattern. The fact that this pattern is not tied to any particular set of phonemes suggests that it should be widely generalizable, even to new instances. As such, this information could serve to bias subsequent segmentation of novel words. For this hypothesis to be correct, two conditions must be met that would allow infants to learn a lexical stress pattern by 6 months of age (Höhle et al., 2009). First, infants must be able to segment words from fluent speech, via sensitivity to conditional statistical cues, before 6 months. Second, infants must be capable of learning from the distribution of lexical stress in word forms with which they are familiar before 6 months.

Given that Experiment 1 demonstrated that infants are sensitive to conditional statistical regularities in linguistic input, a natural subsequent question to ask is whether infants at this age are also sensitive to distributional statistical regularities in linguistic input. Thus, in Experiment 2, we ask whether 5-month-olds' learning

abilities satisfy the second condition, and they are able to identify a common acoustic feature across lexical forms to which they are exposed. If so, they should be able to discover a prosodic commonality across the word forms to which they are exposed in a laboratory setting. To test this possibility, we exposed 5-month-old infants to lists of trochaic words in isolation and then presented them with either a stream of trochaic or iambic speech. In prior research with 7- and 9-month-old infants, exposure to a list of this kind has been sufficient to allow infants to learn the relation between lexical stress and word position, and to being to use lexical stress as a cue to word segmentation (Thiessen and Saffran, 2007). Note that in prior experiments, English-learning infants have been able to learn both a trochaic and an iambic bias. In this experiment we only exposed infants to a trochaic bias. Previously, we have found that 7- and 9-month-old infants are able to learn an iambic bias that contradicts their native language. Therefore, it is likely that if 5-month-olds – who have less familiarity with the trochaic pattern of English than 7- or 9-month-olds – are able to learn a trochaic pattern, they would also be able to learn an iambic pattern. In this experiment, then, we assess whether 5-month-old infants are able to adapt to the distribution of lexical stress across familiar word forms and acquire a trochaic segmentation bias.

MATERIALS AND METHODS

Participants

Data were obtained from 20 participants between the ages of 5.0 and 5 months, 16 days ($M = 5.10$). Half of these infants were exposed to a trochaic artificial language, and half to an iambic artificial language. To obtain data from 20 infants, it was necessary to run 29 infants. The additional nine infants were excluded (five from the trochaic condition, four from the iambic condition) for crying or squirming during the testing session (4), looking times of less than 3 s to the test trials (3) and experimenter error (2).

Stimuli

The trochaic and iambic language, and the test items, used in this experiment were identical to those used in Experiment 1B. Before exposure to the to-be-segmented artificial language, infants heard a list of 30 CVCV bisyllabic nonsense words, repeated twice, for a total of 60 words. Each word in this list was stressed on its first syllable, and there was a pause of 1.4 s between each word; the total length of the 60 word set was 126 s. Lexical stress was created through the alteration of three parameters: pitch contour, amplitude, and duration. The list was identical to that used in Thiessen and Saffran (2007). All of the words in this list were different from the four words that occurred in the familiarization stream.

Procedure

The procedure used in this experiment was identical to that used in Experiment 1, with the exception that before the presentation of the to-be-segmented artificial language, infants were exposed to a list of 60 trochaic words (all infants heard the same 60 words), paired with the image of a static checkerboard on the central LCD monitor.

RESULTS

We compared listening times to words and part-words for infants exposed to both the trochaic language and the iambic language. If infants fail to learn from exposure to a list of trochaic items, they should segment fluent speech – like infants in Experiment 1 – via conditional statistical cues, and show the same pattern of preference after exposure to both the trochaic and iambic segmentation stream. However, if infants learn that lexical stress is a cue to word-initial position, they may begin to use lexical stress as a cue to word segmentation (e.g., Johnson and Jusczyk, 2001). If so, infants should segment different items from the trochaic segmentation stream than from the iambic segmentation stream, and show a different pattern of preference at test after exposure to these two languages.

To assess these possibilities, we performed a 2 (Test item) × 2 (Condition) ANOVA to determine whether infants showed the same, or a significantly different, preference for test items as a function of which segmentation stream they heard. The main effect for test item (listening to words vs. part-words) was not significant, $F(1, 18) = 0.24$, $p = 0.63$, indicating that infants exposed to both languages listened for similar times to words and part-words. The main effect of condition (trochaic vs. iambic exposure) was also not significant, $F(1, 18) = 0.075$, $p = 0.40$, indicating that infants listened for similar lengths of time regardless of which language they heard. However, the interaction between test item and condition was significant, $F(1, 18) = 17.69$, $p < 0.01$, meaning that the direction of preference for test items differed depending on the language to which infants were exposed.

To better understand this interaction, we performed planned t-tests comparing listening times to test items in the two conditions. Infants exposed to the trochaic language listened to part-words for 7.04 s (SE = 0.53) and to words for 5.89 s (SE = 0.51). A paired t-test revealed that this difference was significant, $t(9) = 2.93$, $p < 0.05$. Infants who were exposed to the iambic language listened to part-words for 6.65 s (SE = 0.54) and to words for 7.55 s (SE = 0.59; See **Figure 4**). A paired t-test revealed that this difference was significant, $t(9) = 3.12$, $p < 0.05$. These results indicate that infants show a different preference for test items after listening to the trochaic and iambic languages, as would be expected if they had learned to treat lexical stress as a cue to word segmentation. Because the placement of lexical stress differs across the two familiarization streams, relying on

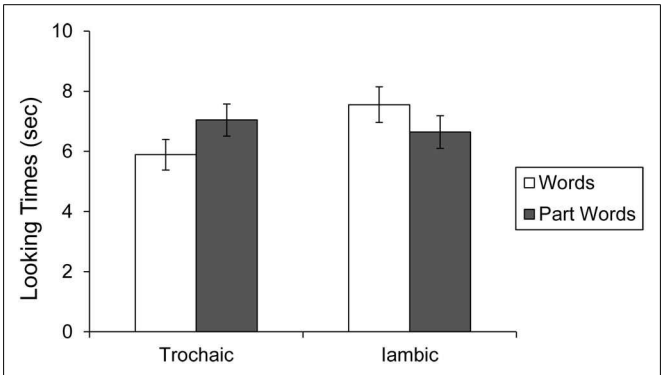


FIGURE 4 | Looking times to words and part-words in the trochaic and iambic conditions of Experiment 2. Error bars indicate standard error.

stress as a cue to segmentation should lead infants to segment different items from them, and therefore prefer different test items.

DISCUSSION

The fact that infants show a different pattern of preference after listening to the trochaic and iambic familiarization streams indicates that they segmented different items from the two streams. Because the only cue to word boundaries that differs across the two familiarization streams is lexical stress, the different pattern of preference across the two streams indicates that they learned a trochaic lexical stress pattern from exposure to the list of trochaic items, and used this pattern to subsequently segment the fluent speech. This result is consistent with prior experiments demonstrating that infants who rely on lexical stress as a cue to segmentation extract segment words from trochaic input, and actually mis-segment iambic input by treating stressed syllables as word onsets (e.g., Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003). Further, it replicates prior work with 7-month-olds demonstrating that infants – even 5-month-old infants – can learn to use lexical stress as a cue to word segmentation upon exposure to lexical forms that consistently exemplify a stress pattern (Thiessen and Saffran, 2007).

Prior work suggests that 7-month-old English-learning infants are able to learn an iambic stress pattern in addition to the trochaic stress pattern infants learned in this experiment (Thiessen and Saffran, 2007). We did not assess whether the 5-month-olds in this experiment would be able to learn an iambic pattern, which contradicts the predominant pattern of English words. The fact that 7-month-olds can learn such a pattern suggests that 5-month-olds may be able to do so as well, given that 5-month-olds have even less familiarity with the predominant pattern of English to overcome. This is not to suggest that infants at this age are completely unfamiliar with the preferred lexical stress pattern of their native language (e.g., Friederici et al., 2007). To the extent that 5-month-olds are familiar with any lexical items, they have likely already begun to identify some acoustic regularities across those forms. As these results demonstrate, exposure to lexical forms that show a consistent acoustic pattern allows infants to use that consistent information as a cue to subsequent word segmentation, a cue that infants did not rely upon in the absence of such exposure.

GENERAL DISCUSSION

Since the initial demonstration that 8-month-old infants are capable of extracting word forms in fluent speech solely by sensitivity to conditional statistical information, the question of how this sensitivity to statistical information might contribute to language acquisition has been a central one in the field of language development. The current experiments are relevant to that question in two ways. First, they reinforce the claim that sensitivity to statistical information is apparent for linguistic input at a younger age than the commonly cited 7–8 months (c.f. Johnson and Tyler, 2010, for a comparable demonstration with slightly older infants). This suggests that infants may have more opportunity to learn from statistical information than previously thought. Second, they suggest that sensitivity to statistical information can play an important role in helping infants adapt to the acoustic structure of their native language.

One argument against sensitivity to statistical information playing an important role in language acquisition is that real language is more complex than the kinds of artificial stimuli used in laboratory settings, and that statistical learning may not be sufficiently powerful or informative in the face of such complexity (e.g., Johnson and Tyler, 2010). Consistent with this, adults do not weight conditional statistical information very strongly as a cue to word boundaries, instead relying on language-specific segmentation cues (e.g., Mersad and Nazzi, 2011). Similarly, 8- and 9-month-old infants weight language-specific cues, such as lexical stress, more strongly than conditional cues (e.g., Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003). A related argument is that statistical learning develops later than other cues to word segmentation, and is thus not central to the process of language development. For example, some proposals have suggested that the earliest tools for word segmentation are prosodic cues (e.g., Johnson and Jusczyk, 2001; Nazzi et al., 2006). Indeed, German-learning infants have been found to use lexical stress as a cue to word segmentation by 6 months (Höhle et al., 2009), younger than any prior demonstrations that infants were able to use conditional statistical information to segment fluent speech.

The current results present an opportunity to reconsider the relative age at which infants are sensitive to prosodic vs. conditional statistical cues to word segmentation. The 5-month-olds in Experiments 1 and 2 are able to segment words from fluent speech via sensitivity to conditional statistical information, opening the possibility that sensitivity to conditional statistical cues plays a role in learning from a very young age (c.f. Kirkham et al., 2002). Moreover, despite their success at segmenting on the basis of statistical cues, 5-month-old infants do not appear to have developed a trochaic bias. This is consistent with the claim that sensitivity to conditional statistical information develops earlier than sensitivity to language-specific prosodic patterns (Thiessen and Saffran, 2003). Conditional statistical information is potentially available in every linguistic environment, and available without prior knowledge about the acoustic regularities that characterize the language. From our perspective, sensitivity to conditional statistical information is one of a small set of language-universal cues that help infants extract a set of lexical items from the input (for discussion, see Thiessen and Erickson, in press). Once infants have extracted a small set of lexical items, they can begin to learn the language-specific acoustic regularities that will subsequently inform segmentation (Thiessen and Saffran, 2007). If this account is correct, infants would necessarily be able to segment input via conditional statistical information before showing the ability to take advantage of language-specific cues.

This developmental account involves two different aspects of sensitivity to statistical information. Sensitivity to conditional statistical information is one of a small set of language-universal cues that can help infants to extract lexical items from fluent utterances (Thiessen and Erickson, in press). Because items with high conditional probabilities are more likely to be real words in the language than groupings with low conditional probabilities, sensitivity to conditional statistical information helps to guide infants toward discovering a set of lexical items. These lexical items, in turn, are

likely to follow the predominant phonological characteristics of the lexical forms in the native language (e.g., Swingley, 2005). This is especially true of the words in infant-directed speech, which appear to exaggerate the regularities present in adult-directed speech (e.g., Fernald and Simon, 1984; Kelly and Martin, 1994).

Once infants have extracted a small set of lexical items from the input, they can learn the phonological regularities that characterize words in the native language. Doing so entails taking advantage of distributional statistical information. Distributional statistical information relates to the frequency and variability of exemplars in the input (e.g., Zhao et al., 2011). It is especially useful in discovering the central tendency or prototypical configuration of some set of exemplars. One linguistically relevant application of this sensitivity to distributional information is category learning. Sensitivity to the frequency of exemplars along a perceptual continuum (e.g., voice onset time) is informative about category boundaries because categories often involve crowds of exemplars near the center of categories, and a sparser group of exemplars at the ambiguous region between categories (e.g., Maye et al., 2002). Sensitivity to variability is similarly informative for category learning; when exposed to distributions with high variability, learners accept a wider range of exemplars as members of the category (e.g., Clayards et al., 2008). A related example of sensitivity to distributional information is the discovery of the prototypical configuration of a set of exemplars. For example, exposure to a set of words allows infants to discover the phonological regularities that characterize those words (Chambers et al., 2003; Saffran and Thiessen, 2003; Thiessen and Saffran, 2007; Thiessen and Yee, 2010).

As these examples illustrates, distributional statistical learning differs from conditional statistical learning in its "output." Whereas conditional learning results in the segmentation of a discrete item from a larger continuous array of stimuli (such as words from a sentence), distributional learning results in a combination of information from multiple stimuli into a central tendency or prototypical configuration (e.g., Zhao et al., 2011; Thiessen et al., in press). There are several prior models of this kind of information integration, primarily models of long-term memory that combine information across prior instances to identify commonalities (e.g., Hintzman, 1984; McClelland and Rumelhart, 1985). Two of the processes invoked by these models are of particular importance: similarity-based activation, and summation of information across prior instances (for discussion, see Thiessen and Pavlik, 2012). The effect of similarity means that when information is presented, the most similar stored exemplars are most activated and have the greatest influence on the response to the current information. The information in activated memories is then summated, such that information that is consistent across prior activated memories is reinforced, while inconsistent information tends to be canceled out, and an average (weighted toward the most highly activated memories) or prototype can be identified (e.g., Hintzman, 1984). These processes can account for a wide variety of distributional learning phenomena, including category learning, acquired distinctiveness, and the role of variability in facilitating learning of non-adjacent relations (Thiessen and Pavlik, 2012).

Sensitivity to distributional statistical information, achieved by integrating information across many individual exemplars to yield a central tendency, can explain English-learning infants' acquisition of a trochaic bias. For example, if infants extract the words *BAby*, *DIAper*, and *SHOE*, there is no consistent phonemic information. However, each of the words has a word-initial stress pattern. Integrating across these lexical forms would yield a representation that is not specific to any particular set of phonemes (i.e., is widely generalizable), but strongly indicates that lexical stress is associated with word-initial position. Once infants detect this distributional regularity, it alters their segmentation of subsequent speech (e.g., Thiessen and Saffran, 2007). Experiment 2 demonstrates that even 5-month-olds are capable of this kind of distributional learning. Exposed to a set of lexical items in isolation, 5-month-olds were able to integrate information across these exemplars to identify the only feature consistent across all of them: their lexical stress pattern.

From this perspective, infants' and adults' use of phonological cues is not a sign that statistical learning is unimportant for language acquisition. Instead, sensitivity to phonological cues emerges from earlier sensitivity to conditional statistical information in a developmental progression. The cues to which infants are sensitive early in life, such as conditional statistical information or utterance boundaries (e.g., Christophe et al., 2001; Seidl and Johnson, 2006), require no prior experience with or knowledge about a specific language to use. These cues allow infants to discover a set of word forms even before they are familiar with language-specific acoustic cues to word boundaries (e.g., Thiessen and Saffran, 2003). Once infants have discovered a set of words, they can identify language-specific acoustic cues by taking advantage of distributional information about those word forms (Thiessen and Saffran, 2007; Lew-Williams and Saffran, 2012).

The fact that 5-month-old infants are sensitive to both the conditional and distributional regularities necessary to discover a phonological regularity such as lexical stress raises a developmental question: why have 5-month-olds not learned the trochaic pattern of English already? Most prior research indicates that infants do not discover this regularity until some time around 7 months (e.g., Jusczyk et al., 1999; Thiessen and Saffran, 2003). If we are right that discovering such phonological regularities requires infants to first identify a set of lexical items, the lack of a trochaic bias at 5 months likely indicates that infants have yet to become familiar with a sufficient number of words. Even though infants at this age are capable of segmenting fluent speech in a laboratory setting, they may not yet have extracted many words from natural linguistic input. There are several reasons why real languages present a greater challenge than the artificial systems used in experiments like these, including its greater degree of (both inter- and intra-speaker) variability, less robust conditional statistical cues, and a far greater number of lexical items repeated less closely together than in a laboratory setting. These factors may require that infants experience many more repetitions of a word in a natural language to segment it from fluent speech than is necessary in segmentation experiments.

As this discussion indicates, much remains unknown about the exact age at which infants begin to segment words from fluent speech, and the number of lexical forms they are able to extract from fluent native language input (for discussion, see Swingley,

2005). Nevertheless, the current experiments are informative with respect to the relative ordering of the acquisition of different cues to word segmentation. The present studies replicate and extend prior work by Thiessen and Saffran (2003) demonstrating that sensitivity to conditional statistical information in speech is early developing, and appears to emerge – at least in English-learning infants – before the development of the trochaic bias. Though 5-month-olds do not display a trochaic bias, they are able to segment speech via sensitivity to conditional statistical information. Further, they are able to learn a trochaic bias through exposure to a set of words that follow a consistent trochaic pattern. This is also consistent with the hypothesis that segmenting word forms via a domain-general process such as statistical learning is potential mechanism by which infants can develop language-specific acoustic biases (e.g., Thiessen and Saffran, 2007; Thiessen and Erickson, in press).

The ability to segment fluent speech on the basis of the probabilistic relation between sequences of speech sounds is an example of conditional statistical learning. The ability to learn the relation between lexical stress and word position on the basis of a set of exemplars following a particular prosodic pattern is an example of distributional statistical learning. These processes are typically studied and modeled in isolation (e.g., Perruchet and Vinter, 1998; Frank et al., 2010; Thiessen and Pavlik, 2012). But as these experiments indicate, distributional learning constrains subsequent statistical learning, as infants extract items that are consistent with the phonological pattern they have learned. Moreover, we propose that in the course of natural language acquisition, conditional statistical learning influences distributional learning. Infants are able to discover phonological patterns through the lexical forms that they learn via sensitivity to conditional statistical information. To fully understand the role of statistical learning in language acquisition, it will be necessary to develop models and theories that more thoroughly explore how sensitivity to conditional and distributional statistical learning interact to allow infants to adapt to the structure of their native language.

REFERENCES

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychol. Sci.* 16, 298–304.

Chambers, K. W., Onishi, K. H., and Fisher, C. L. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87, B69–B77.

Christophe, A., Sebastian-Galles, N., and Mehler, J. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy* 2, 385–394.

Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108, 804–809.

Cohen, L. B., Atkinson, D. J., and Chaput, H. H. (2004). *Habit X: A New Program for Obtaining and Organizing Data in Infant Perception and Cognition Studies (Version 1.0)*. Austin: University of Texas.

Cutler, A., and Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Comput. Speech Lang.* 2, 133–142.

Cutler, A., and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 113–121.

Echols, C. H., Crowhurst, M. J., and Childers, J. B. (1997). Perception of rhythmic units in speech by infants and adults. *J. Mem. Lang.* 36, 202–225.

Echols, C. H., and Newport, E. L. (1992). The role of stress and position in determining first words. *Lang. Acquis.* 2, 189–220.

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113.

Fiser, J., and Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134, 521–537.

Frank, M. C., Goldwater, S., Griffiths, T., and Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition* 117, 107–125.

Friederici, A. D., Friedrich, M., and Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Curr. Biol.* 17, 1208–1211.

Giroux, I., and Rey, A. (2009). Lexical and sublexical units in speech perception. *Cogn. Sci.* 33, 260–272.

Gleitman, L. R., Gleitman, H., Landau, B., and Wanner, E. (1988). "Where learning begins: initial representations for language learning," in *Linguistics: The Cambridge Survey, Vol. 3, Language: Psychological and Biological Aspects*, ed. F. J. Newmeyer (New York: Cambridge University Press), 150–193.

Graf Estes, K., Alibali, M. W., Evans, J. L., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Hayes, J. R., and Clark, H. H. (1970). *Experiments in the Segmentation of Artificial Speech Analog. Cognition and the Development of Language*. New York: Wiley.

Hintzman, D. L. (1984). "Schema Abstraction" in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Dev.* 3, 262–274.

Höhle, B., and Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Dev. Sci.* 6, 122–127.

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 548–567.

Johnson, E. K., and Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Dev. Sci.* 13, 339–345.

Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–23.

Jusczyk, P. W., Cutler, A., and Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687.

Jusczyk, P. W., Houston, D., and Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207.

Jusczyk, P. W., and Thompson, E. J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. *Percept. Psychophys.* 23, 105–109.

Kelly, M. H., and Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: sensitivity to probabilities in perception, cognition and language. *Lingua* 92, 105–140.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for domain general learning mechanism. *Cognition* 83, B35–B42.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., and Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* 14, 1100–1106.

Lew-Williams, C., and Saffran, J. R. (2012). All words are not created equal: expectations about word length guide infant statistical learning. *Cognition* 122, 241–246.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

McClelland, J. L., and Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *J. Exp. Psychol. Gen.* 114, 159–197.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor to language acquisition in young infants. *Cognition* 29, 143–178.

Mersad, K., and Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Mem. Cognit.* 39, 1085–1093.

Mirman, D., Magnuson, J., Estes, K., and Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition* 108, 271–280.

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagle, S., and Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Lang. Speech* 48, 279–298.

Nazzi, T., Iakimova, I., Bertoncini, J., Frédonie, S., and Alcantara, C. (2006). Early segmentation of fluent by infants acquiring French: emerging evidence for crosslinguistic differences. *J. Mem. Lang.* 54, 283–299.

Nazzi, T., and Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Commun.* 41, 233–243.

Norris, D., and McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol. Rev.* 115, 357–395.

Orban, G., Fiser, J., Aslin, R. N., and Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2745–2750.

Perruchet, P., and Vinter, A. (1998). PARSER: a model of word segmentation. *J. Mem. Lang.* 39, 246–263.

Saffran, J. R. (2001). Words in a sea of sounds: the output of statistical learning. *Cognition* 81, 149–169.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, J. R., and Thiessen, E. D. (2003). Pattern induction by infant language learners. *Dev. Psychol.* 39, 484–494.

Seidl, A., and Johnson, E. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Dev. Sci.* 9, 565–573.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cogn. Psychol.* 50, 86–132.

Teinonen, T., Fellman, V., Naatanen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10:21. doi:10.1186/1471-2202-10-21

Thiessen, E. D., and Erickson, L. C. (in press). "The statistical approach to word segmentation," in *Statistical Approaches to Language Acquisition*, ed. T. H. Mintz.

Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (in press). The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.*

Thiessen, E. D., and Pavlik, P. I. (2012). iMinerva: a mathematical model of (distributional) statistical learning. *Cogn. Sci.* 1–34.

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.

Thiessen, E. D., and Saffran, J. R. (2004). Spectral tilt as a cue to word segmentation in infancy and adulthood. *Percept. Psychophys.* 66, 779–791.

Thiessen, E. D., and Saffran, J. R. (2007). Learning to learn: infants' acquisition of stress-based strategies for word segmentation. *Lang. Learn. Dev.* 3, 73–100.

Thiessen, E. D., and Yee, M. N. (2010). Dogs, bogs, labs, and lads: what phonemic generalizations indicate about the nature of children's early word-form representations. *Child Dev.* 81, 1287–1303.

Weber, C., Hahne, A., Friedrich, M., and Friederici, A. D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: neurophysiological evidence. *Brain Res. Cogn. Brain Res.* 25, 180–187.

Zhao, J., Ngo, N., McKendrick, R., and Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychol. Sci.* 22, 1212–1219.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 May 2012; accepted: 13 December 2012; published online: 17 January 2013.

Citation: Thiessen ED and Erickson LC (2013) Discovering words in fluent speech: the contribution of two kinds of statistical information. Front. Psychology ***3****:590. doi: 10.3389/fpsyg.2012.00590*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 01 October 2012
doi: 10.3389/fpsyg.2012.00374

Statistical speech segmentation and word learning in parallel: scaffolding from child-directed speech

***Daniel Yurovsky*[1]*, *Chen Yu*[2] *and *Linda B. Smith*[2]**

[1] *Department of Psychology, Stanford University, Stanford, CA, USA*
[2] *Department of Psychological and Brain Sciences and Program in Cognitive Science, Indiana University, Bloomington, IN, USA*

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Toni Cunillera, University of Barcelona, Spain
Elika Bergelson, University of Pennsylvania, USA

***Correspondence:**
Daniel Yurovsky, Department of Psychology, Jordan Hall, Building 01-420, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA.
e-mail: yurovsky@stanford.edu

In order to acquire their native languages, children must learn richly structured systems with regularities at multiple levels. While structure at different levels could be learned serially, e.g., speech segmentation coming before word-object mapping, redundancies across levels make parallel learning more efficient. For instance, a series of syllables is likely to be a word not only because of high transitional probabilities, but also because of a consistently co-occurring object. But additional statistics require additional processing, and thus might not be useful to cognitively constrained learners. We show that the structure of child-directed speech makes simultaneous speech segmentation and word learning tractable for human learners. First, a corpus of child-directed speech was recorded from parents and children engaged in a naturalistic free-play task. Analyses revealed two consistent regularities in the sentence structure of naming events. These regularities were subsequently encoded in an artificial language to which adult participants were exposed in the context of simultaneous statistical speech segmentation and word learning. Either regularity was independently sufficient to support successful learning, but no learning occurred in the absence of both regularities. Thus, the structure of child-directed speech plays an important role in scaffolding speech segmentation and word learning in parallel.

Keywords: statistical learning, speech segmentation, word learning, child-directed speech, frequent frames

INTRODUCTION

Human language is richly structured, with important regularities to be learned at multiple levels (Kuhl, 2004). For instance, the human vocal apparatus can produce a staggering variety of sounds distinguishable from each other by prelinguistic infants (Eimas et al., 1971). However, only a tiny fraction of these become meaningful units – phonemes – within a particular language. Similarly, these phonemes can be strung together into an infinite number of sequences, but only a tiny fraction of these are words. Thus, infants must also solve the problem of parsing a continuous sequence of phonemes into word units. Further, some of these words refer to objects in the visual world, and so, for these segmented words, infants must solve the word-world mapping problem. In addition, speakers may refer to the same object with different words in different contexts, and different word orderings and stress patterns can radically alter an utterance's meanings, so children must organize sounds, segments, and meanings at the levels pragmatics, syntax, and prosody as well.

An emerging theoretical consensus is that many or even all of these problems may be solved through a process of statistical learning – tracking predictive relationships between elemental units (although, cf. Marcus, 2000; Waxman and Gelman, 2009). In order to determine their native language phonemes, infants may track the distribution of tonal and formant frequencies in their input (Maye et al., 2002; Pierrehumbert, 2003). Similarly, infants may learn word boundaries by tracking sequential syllable statistics (Saffran et al., 1996), learn word-world mappings by tracking word-object occurrence statistics (Smith and Yu, 2008; Vouloumanos and Werker, 2009), and learn grammar by tracking sequential and non-adjacent dependencies between word types (Gómez and Gerken, 2000; Saffran et al., 2008). Because statistical learning at each level assumes the availability of primitives at the level below and shows how to arrive at primitives for the level above, a complete statistical account of language learning must bridge these levels. Therefore, a critical question for statistical theories of language acquisition is how learners connect these primitives.

One possibility is that the infants learn each level sequentially, proceeding from the bottom up. Learning at each level would build the units over which the next level operates, and thus higher levels would have to wait until (at least some of) the primitives at the lower levels had been acquired. This hypothesis is intuitive, and makes several predictions consistent with the extant literature. First, it predicts a developmental trajectory in statistical learning abilities: phoneme learning should come first, followed by speech segmentation, followed by word-world mapping, followed by syntax. Indeed, this is the general trend observed in infant statistical learning experiments. At 6 months, infants show sensitivity to phoneme distributions (Maye et al., 2002), at 8 months they can segment continual speech into words (Saffran et al., 1996), at 12 months they can map words onto objects using co-occurrence information (Smith and Yu, 2008), and at 18 months they can learn non-adjacent syntactic dependencies (Gómez, 2002). Second, this account predicts that infants should be able to extract regularities at one level, and use them subsequently to learn at the next higher level. This has been confirmed by recent empirical findings from

Saffran and colleagues (Graf Estes et al., 2007; Hay et al., 2011) showing that statistically coherent word segments extracted from continuous speech subsequently act as superior labels in subsequent word learning. It is also supported by recent computational models showing that regularities at multiple levels can be learned serially from child-directed speech (Yu et al., 2005; Christiansen et al., 2009; Räsänen, 2011).

Alternatively, learners could acquire structure at each level in parallel. Because regularities at each level are statistically interrelated, partial acquisition of the structure at any level would reduce ambiguity at every other level (Feldman et al., 2009; Johnson et al., 2010). However, this aggregate ambiguity reduction comes at a cost: if units are uncertain at every level, demands on attention and memory are likely to skyrocket. Thus, an abundance of structure helpful for ideal learners might easily overload cognitively constrained statistical learners (Fu, 2008; Frank et al., 2010). This tradeoff is evident in recent experiments investigating simultaneous statistical speech segmentation and word learning. In these experiments, adult learners engaged in a standard statistical speech segmentation task with one addition: word-onsets occurred in a small window around the onset of visual objects. Under these conditions, adults succeeded at both segmenting the speech stream, and mapping the words onto their correct referents (Cunillera et al., 2010a,b; Thiessen, 2010). However, in identical experiments, 8-month-olds failed to acquire either regularity (Thiessen, 2010). Further, when the task is made slightly more difficult – presenting multiple objects at once (as in Yu and Smith, 2007) – adults fail to learn word-object mappings from continuous speech (Frank et al., 2007). Thus, while parallel statistical learning might provide a significant advantage, it could be outside the processing limits of human learners (cf. Fiser and Aslin, 2002, for an example of parallel learning in a purely visual task). However, these demands on cognitive processing could be alleviated in another way: human learners could be scaffolded by other properties of natural language (Vygotsky, 1978; Mintz, 2003). The studies in this paper provide evidence for just such a solution in the context of parallel speech segmentation and word learning.

In typical statistical learning experiments, regularities in the input are constructed in such a way as to isolate the problem of interest. For instance, in statistical speech segmentation tasks, each word typically occurs with equal frequency and is equally likely to follow each other word (e.g., Saffran et al., 1996; Graf Estes et al., 2007). In statistical word learning tasks, each word and object typically occur with equal frequency, and each incorrect mapping has equal statistical support (e.g., Yu and Smith, 2007; Smith and Yu, 2008; Vouloumanos and Werker, 2009). But this structure differs in a number of ways from the structure of natural language input, and these difference are likely to matter (Kurumada et al., 2011; Vogt, 2012). For instance, referential utterances in child-directed speech often come from a small set of stereotyped naming frames, e.g., "*look at the dog*" (Cameron-Faulkner et al., 2003). Children are remarkably sensitive to this structure: 18-month-old infants orient faster to the referent of a label embedded in such statistically frequent naming frames than they do to a label uttered in isolation (Fernald and Hurtado, 2006). Do these frequent frames help learners segment a stream of sounds into *and* to map these words onto referents?

We pursued this question in two steps. First, we sought to determine the statistical structure of the frames that characterize naming events to young children. To this end, we analyzed data from a corpus of child-directed speech recorded during naturalistic free-play interactions to discover the shared structure of common naming frames. Subsequently, we constructed an artificial language in which the strings were naming events that maintained the main regularities found in the natural speech corpus. We then embedded these naming events in a word-object mapping task in which each trial contained multiple naming events and multiple visual referents. Thus, to learn the language, participants would have to segment labels from continuous speech *and* map them to their statistically consistent referents. We then parametrically manipulated the artificial language to determine if and how the regularities in natural naming frames facilitate simultaneous speech segmentation and word learning. Our findings illustrate the importance of understanding the statistical properties of natural language contexts for drawing conclusions about statistical learning.

RESULTS

CORPUS ANALYSIS

To capture regularities in naming frame structure, we analyzed transcripts of child-directed speech from naturalistic free-play interactions between 17 parent-child dyads (Yu et al., 2008; Yu and Smith, 2012). This corpus contained 3165 parental speech utterances, 1624 of which contained the label of one of the toys in the room. Of these utterances, 672 (∼20%) were single-word utterances consisting of only the toy's label. Because the Experiments investigate the role of naming frames in parallel speech segmentation and word learning, these utterances were excluded from further analysis, but we return to them in the Discussion. The remaining 952 events were analyzed for consistent naming frame structure.

As shown in **Table 1**, 21 different naming frames cover more than 50% of all naming events. Together, these frames contain only 20 unique words and conform to two general regularities. First, in these frequent frames, the toy's label always occurs in the final position (see also Aslin et al., 1996). Second, only a small set of words – mostly articles – precede a toy's label (see also Shafer et al., 1998). Both regularities are also common in the remaining naming events, appearing in 50 and 63%, respectively. Because both final position (Endress et al., 2005) and onset cues (Bortfeld et al., 2005; Mersad and Nazzi, 2012) have previously been found to facilitate statistical sequence learning, each regularity could potentially scaffold statistical learners, buttressing them against the combinatorial explosion of parallel speech segmentation and word learning. Further, evidence from other studies suggests that redundant cues help children learn language (e.g., Gogate et al., 2000; Frank et al., 2009). Consequently the combination of both position and onset cues could play an additive role in speech segmentation and word learning.

EXPERIMENTS

To study joint speech segmentation and word-object mapping, we exposed adult participants to a series of individually ambiguous training trials based on the cross-situational learning paradigm (Yu

and Smith, 2007). On each trial, adults saw two objects and heard two phrases of continuous speech from an artificial language. In order to learn word-object mappings, they had to determine which phrase referred to which object, where the word boundaries were, and finally which words were Object Labels and which word were Frame Words. Crucially, the naming frames extracted from the natural child-directed speech corpus were encoded into the artificial language presented to participants (**Figure 1**).

Participants were assigned randomly to one of four language conditions. In the *Full* language condition, participants heard artificial language phrases containing both regularities found in natural naming frames. In the *Onset Only* language condition, Object Labels appeared in the middle of phrases instead of at the end, but they were always preceded by one of a small set of onset cue words. In the *Position Only* language condition Object Labels always appeared in utterance-final position, but were not preceded by a small set of onset cue words. Finally, in the *Control* language condition, neither regularity from the natural naming frames was provided. After training, participants were tested for their knowledge of both the words of the language (speech segmentation), and the word-object mappings. Additional details can be found in the section "Materials and Methods" below.

Table 1 | The 21 most frequent naming frames.

Phrase	Pct. of corpus
The OBJ	6.30
That is a OBJ	4.73
And the OBJ	4.31
A OBJ	4.10
It is a OBJ	3.78
This is a OBJ	3.57
And a OBJ	3.26
Can you say OBJ	2.94
Here is the OBJ	2.63
And OBJ	2.42
Where is the OBJ	1.89
That is the OBJ	1.79
Look at the OBJ	1.79
I have the OBJ	1.47
You want the OBJ	1.16
Color is the OBJ	1.16
Is that the OBJ	1.16
there is the OBJ	1.05
You put the OBJ	1.05
To put the OBJ	0.95
One is the OBJ	0.95
Total	52.42%

Two regularities are apparent in the most frequent naming frames. First, Object Labels occur reliably in final frame position. Second, labels are reliably preceded by a small set of onset cues (a, the, and, say).

Speech segmentation

On each segmentation test, participants were asked to indicate which of two sequences was more likely to be a word of the language. **Figure 2** shows how participants' segmentation of both Object Labels and Frame Words varied across language conditions. Overall, participants successfully segmented Object Labels only in the *Full* and *Position Only* language conditions. They segmented Frame Words successfully in the *Onset Only* language condition, and to a lesser extent in the *Position Only* and *Control* language conditions. Participants' segmentation accuracies were averaged across all words and submitted to a mixed 4 (Language) × 2 (Word Type) ANOVA. This analysis showed no main effect of language [$F(3,90) = 1.40$, $p = 0.25$] nor word type [$F(1,90) = 0.83$, $p = 0.37$], but did show a significant interaction [$F(3,90) = 5.39$, $p < 0.01$]. All segmentation accuracy were submitted to the Shapiro–Wilk test of normality (Shapiro and Wilk, 1965). Since none were found to be non-normal (all p's > 0.1), follow up analyses used t-tests. These follow up tests showed that Object Label segmentation was above chance in the *Full* [$M = 0.59$, $t(23) = 2.69$, $p < 0.05$] and *Position Only* language conditions [$M = 0.57$, $t(21) = 2.13$, $p < 0.05$], but not in the *Onset Only* [$M = 0.53$, $t(23) = 1.34$, $p = 0.19$] or *Control* language conditions [$M = 0.54$, $t(23) = 1.26$, $p = 0.22$]. Frame-word segmentation was above chance in the *Onset Only* language condition [$M = 0.68$, $t(23) = 5.39$, $p < 0.001$], trended toward significance in the *Position Only* and *Control* language conditions [$M_{PositionOnly} = 0.56$, $t(21) = 1.86$, $p = 0.08$; $M_{Control} = 0.55$, $t(21) = 1.93$, $p = 0.06$] and was indistinguishable from chance in the *Full* language condition [$M = 0.52$, $t(23) = 0.51$, $p = 0.62$]. Segmentation of Object Labels and Frame Words was correlated in *Position Only* language condition ($r = 0.48$, $p < 0.05$), but not in any of the other language conditions ($r_{Full} = -0.22$, $p = 0.29$; $r_{OnsetOnly} = 0.19$, $p = 0.39$; $r_{Control} = 0.23$, $p = 0.29$). Segmentation focus – and accuracy – thus varied across the conditions.

In the *Full* language condition, participants focused on and segmented only the Object Labels, learning little about the Frame

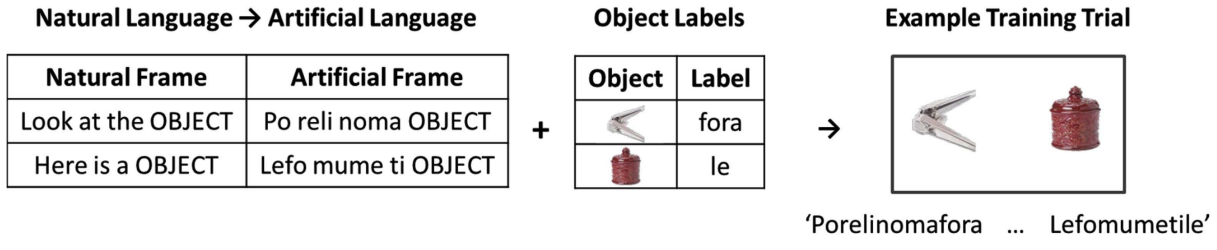


FIGURE 1 | An example training trial from the *Full* language condition. Trials were constructed by encoding naming event patterns from the child-directed speech corpus into the artificial language.

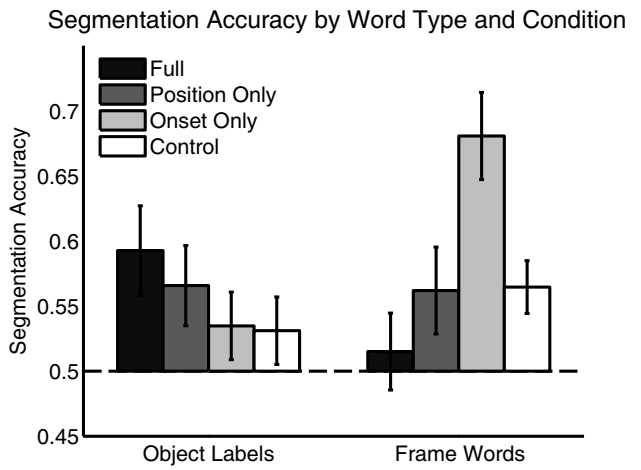


FIGURE 2 | Segmentation accuracy in each condition for both Object Labels and Frame words. Learners successfully segmented Object Labels in the *Full* and *Position Only* language conditions, and segmented Frame Words in the *Onset Only* language condition. Error Bars indicate ±1 SE.

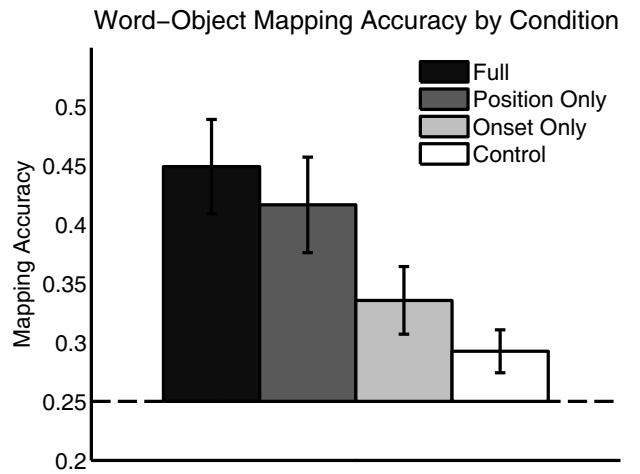


FIGURE 3 | Word-object mapping accuracy by condition. Participants mapped words onto object successfully in all but the *Control* language condition. Error Bars indicate ±1 SE.

Words. In the *Onset Only* language condition, participants segmented Frame Words very successfully, but failed to successfully segment the Object Labels. In the *Position Only* language condition, participants segmented Object Labels successfully and segmented Frame Words at near-significant levels. Further, segmentation accuracy for the two word types was correlated in this condition, suggesting that they supported each other. In the *Control* language condition, segmentation trended toward accuracy for the Frame Words and was at chance levels for Object Labels. Further, segmentation of the word types was uncorrelated, suggesting a less integrated segmentation strategy.

Word-object mapping

Participants were subsequently tested on their word-object mapping accuracy. On each test trial, they heard one word from training and were asked to select the most likely referent object from a set of four alternatives. As shown in **Figure 3**, participants learned a significant proportion of word-object mappings in all but the *Control* language condition, but were most successful in the *Full* and *Position Only* language conditions – the same languages in which they were most successful at Object Label segmentation. An ANOVA showed significant differences in mapping accuracy across conditions [$F(3,90) = 5.03$, $p < 0.01$]. Additional tests showed that accuracy was significantly above chance in all but the *Control* language condition [$M_{Full} = 0.45$, $t(23) = 4.98$, $p < 0.001$; $M_{PositionOnly} = 0.42$, $t(21) = 4.12$, $p < 0.001$; $M_{OnsetOnly} = 0.34$, $t(23) = 2.99$, $p < 0.01$; $M_{Control} = 0.29$, $t(23) = 1.78$, $p = 0.09$]. Further, accuracy was similar in the Full and *Position Only* language conditions [$t(44) = 0.57$, $p = 0.57$], and accuracy in both was significantly greater than in the *Control* language condition [$t_{Full}(46) = 3.69$, $p < 0.001$; $t_{PositionOnly}(44) = 2.92$, $p < 0.01$].Accuracy was significantly greater in the *Full* language condition than in the *Onset Only* language condition [$t(46) = 2.31$, $p < 0.05$], but accuracy did not differ between the *Position Only* and the *Onset Only* language conditions [$t(44) = 1.65$, $p = 0.11$]. Thus, participants were able to learn word-object mappings from continuous speech as long as either regularity from natural naming frames was present. However, the position regularity facilitated learning more than the onset cue regularity.

Correlations between speech segmentation and word-object mapping

Did segmentation and word-object mapping interact, bootstrapping each other? **Figure 4** shows correlations between each participant's average Object Label segmentation and average word-object mapping in each language condition. The two were positively correlated in the *Full* ($r = 0.51$; $p < 0.05$) and the *Position Only* language conditions ($r = 0.62$, $p < 0.01$), but were uncorrelated in the *Onset Only* ($r = -0.09$, $p = 0.67$) and *Control* language conditions ($r = -0.10$, $p = 0.56$). Thus, participants in the *Onset Only* language condition showed evidence of learning word-object mappings without fully segmenting the labels from the utterances.

DISCUSSION

Natural languages are richly structured, containing regularities at multiple hierarchal levels. Statistical learning approaches to language acquisition typically focus on one level at a time, showing how the primitives from the level below can be used to construct the primitives for the level above. Alternatively, statistical language learning at every level could proceed in parallel, exploiting statistical redundancies across levels (Feldman et al., 2009; Johnson et al., 2010). On this account, a child learning a word-referent mapping may not need to wait until she has fully learned the word. But uncertainty at multiple levels imposes significant attention and memory demands on learners, demands that may prevent learning altogether (Frank et al., 2007; Thiessen, 2010). In this paper, we suggest that these demands may be alleviated by other regularities

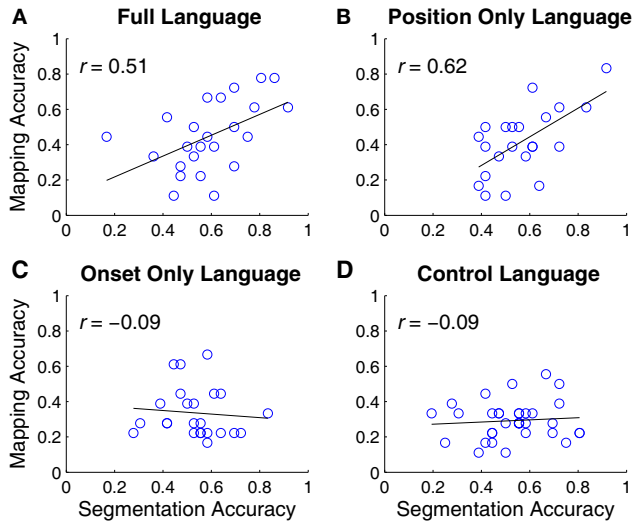


FIGURE 4 | Correlations between segmentation accuracy and word-mapping accuracy in each language condition. Learning the two regularities was positively correlated in the *Full* **(A)** and *Position Only* **(B)** language conditions, and uncorrelated in the *Onset Only* **(C)** and *Control* **(D)** language conditions.

in natural language input, for instance, frequent naming frames (Mintz, 2003).

CORPUS ANALYSIS

Analyzing the structure of natural naming events is an important step toward modeling children's word learning. Because consistency in naming event structure constrains the space of potential solutions, the same mechanism that fails in an unstructured environment may successfully extract words from fluent speech and map them to their referent objects when additional regularities are present. Our analysis showed, first, that a large proportion of naming events in naturalistic free-play are single-word utterances (see also Fernald and Morikawa, 1993; Brent and Siskind, 2001). These utterances could simplify later speech segmentation and give infants a leg up in later word learning (Brent and Siskind, 2001; Lew-Williams et al., 2011).

Second, our analysis revealed two regularities common to over 50% of naming events: labels occur in final phrasal position, and are preceded by an onset cue. We hypothesize that these regularities, like single-word utterances, could also scaffold statistical learning. Specifically, the information encoded in frequent naming frames may allow learners to identify the utterances most likely to be naming events and to spot the label within each frame, potentially without fully segmenting the other words. That is, word-referent mapping may begin before children know exact word boundaries (Yu et al., 2005).

EXPERIMENTS

Encoding these regularities into an artificial language, we tested this idea empirically. Exposing adult participants to artificial languages constructed from a corpus of child-directed speech, we were able to determine the independent and joint contributions of the two regularities apparent in the corpus. Keeping constant the words that make up naming phrases, we altered only their order across conditions. If parallel speech segmentation and word-object mapping rely on environmental cues to reduce cognitive load, this should be reflected in the learning rates across our four conditions.

In the *Full* language condition, which gave strong cues to the frame position of Object Labels as well as to their onset, participants successfully segmented labels from continuous speech and mapped them onto their referent objects. This success came in spite, or perhaps because, of chance-level performance on Frame Word segmentation. That is, participants were able to focus their attention on only the relevant portion of the speech steam (see also Cunillera et al., 2010a). These results, along with the strong correlation between word segmentation and word-object mapping, suggest that participants became attuned to the positional regularity and effectively ignored large portions of the speech input. This reduction in cognitive load may have supported learning.

The *Position Only* language condition, in contrast, removed the onset cue by moving words in the cue set to the beginning of each sentence. In this condition, participants also successfully segmented Object Labels from continuous speech, although at slightly a reduced level. In trade, they performed at a near-significant level on Frame Word segmentation. Also, unlike in the *Full* language condition, segmentation of Object Labels and Frame Words was highly correlated, suggesting an interaction between the processes. Nonetheless, despite these differences, participants in the *Position Only* language condition performed well on the test of word-object mapping. Thus, removing the onset cue forced participants to actively process more of the speech stream, but the presence of the position cue kept cognitive load low enough to enable learning. These results are consistent with previous work showing that utterance-final position facilitates language learning (Echols and Newport, 1993; Goodsitt et al., 1993; Endress et al., 2005; Frank et al., 2007).

Removing the position regularity from the *Full* language yielded the *Onset Only* language condition. In this condition, Object Labels were preceded by a small set of onset cues, but occurred always in medial phrasal position. Without labels in final position, participants performed at chance on tests of Object Label segmentation. However performance on Frame Word segmentation reached levels unseen in the other conditions. Surprisingly, although participants did not show knowledge of correct Object Label segmentation, they did succeed in mapping words to objects at above chance (albeit reduced) levels. Thus, an onset cue alone was sufficient to enable word learning. This is consonant with other work showing that familiar words can act as onset cues, giving infants a wedge into speech segmentation (Bortfeld et al., 2005; Mersad and Nazzi, 2012).

Finally, when naming phrases contained all of the same words but neither of the cues found in the child-directed speech corpus, participants showed poor learning of both kinds of statistics. Thus, in the *Control* language condition, participants were unable to cope with the cognitive load inherent in the simultaneous segmentation and word learning.

CONCLUSION

We began by considering the relationship between statistical speech segmentation and statistical word learning. While previous work has demonstrated a serial link (e.g., Graf Estes et al., 2007; Mirman et al., 2008), in which word candidates generated via statistical segmentation are privileged in statistical word learning, a robust parallel demonstration has remained elusive (Frank et al., 2007; Thiessen, 2010). Perhaps the computational resources required by the tasks are simply too costly to allow their simultaneous resolution. We proposed that construction of previous artificial languages may have averaged out the very regularities that support a parallel solution in naturalistic environments. To borrow from J. J. Gibson, "it's not [just] what is inside the head that is important, it's what the head is inside of."

Analysis of a corpus of child-directed speech from free-play found two potential sources of such scaffolding. First, Object Labels occurred consistently in the final position of naming phrases. Second, these labels were consistently preceded by one of a small set of onset cue words, predominantly articles. We constructed artificial languages following a 2×2 design to produce all possible presence/absence combinations of these regularities. Adult participants were exposed to an ambiguous word-object mapping task in the cross-situational word learning paradigm (Yu and Smith, 2007) in which labels were embedded within continuous speech phrases. These experiments allowed us to determine the independent and joint contributions of the two natural naming regularities. Although these studies use adult language learners as a proxy for child language learners (Gillette et al., 1999), future studies will need to ask this question more directly, using infant participants and measuring learning on-line over the course of training. This will allow finer-grained analysis of the relative time-course of acquisition of each regularity, making clearer whether learning is serial, parallel, or a mixture of both. Further, while the two major regularities found in the corpus have been observed in other corpora, further analyses will need to determine how naming frames change over development, and how these frames contribute to speech segmentation and word learning. Finally, it is important to know to what extent these kinds of frames characterize other languages. Although surely specific frames will differ from language to language, there are reasons to expect common regularities to generalize. For instance, Aslin et al. (1996) analyzed Turkish child-directed speech and found that mothers consistently placed target objects in final position even though this is ungrammatical.

These results highlight the importance of studying statistical language learning in the context of real language input. Although statistical learning is often studies under "unbiased" assumptions about input distributions (e.g., uniform word frequency), these assumptions can be a poor proxy for real-world input (e.g., Zipfian frequency). Sometimes, as in the *Full* language condition, natural input distributions facilitate statistical learning (see also, Johns and Jones, 2010; Kurumada et al., 2011). However, in other cases, natural input statistics make pure statistical learning difficult or impossible (e.g., Johnson and Tyler, 2010; Medina et al., 2011; Vogt, 2012). In such cases, we may be led to understand how other properties of the environment – or of children's and adults' perceptual systems – take up the slack. For instance, a number of previous studies highlight the importance of redundant information in language learning (e.g., Gogate et al., 2000, 2001; Frank et al., 2009; Goldstein et al., 2010; Grassmann and Tomasello, 2010; Smith et al., 2010; Riordan and Jones, 2011). In all of these cases, a difficult statistical language learning problem is made easier by the addition of redundant information, often information from a second sensory modality. For instance, the addition of a pointing (Grassmann and Tomasello, 2010) or synchronous motion (Gogate et al., 2000). This redundant information may make the regularity easier to notice. In other cases, this highlighting is accomplished with a single modality – e.g., presenting the label in a familiar voice (Bergelson and Swingley, 2012) or prosody (Thiessen et al., 2005; Shukla et al., 2011). Finally, in some cases this simplification may be accomplished by the child's own perception/action system, which may act as a filter on the visual (Yurovsky et al., 2012; Yu and Smith, 2012).

Language learning is a process of navigating uncertainty, of leveraging partially learned regularities to learn other regularities (Gleitman, 1990; Smith, 2000). Consequently, there many many routes for breaking into language, and the route that learners adopt is likely to depend on the statistics in their input. For instance, in the *Full* language condition, participants learned word-object mappings by segmenting Object Labels but ignoring Frame Words. In contrast, participants in the *Position Only* language condition segmented both kinds of words, and participants in the *Onset Only* language condition learned word-object mappings but segmented only the Frame Words. In concert with previous research indicating that learners can ignore irrelevant statistical information (Cunillera et al., 2010a; Weiss et al., 2010), and focus on reliable statistical information (Smith, 2000; Colunga and Smith, 2005), these results present a picture of language acquisition as an adaptive process in which learners focus on and exploit the regularities most useful for the task at hand. Thus, the timing with which different regularities are acquired is likely to vary as a function of each learner's input. There may thus be cases, as Peters (1977) suggested, in which children "learn the tune before the words."

MATERIALS AND METHODS

All experiments reported in this paper were approved by the Human Subjects Office at the Indiana University Office of Research Administration. Informed consent was obtained from all participants prior to their participation in these experiments.

CORPUS ANALYSIS

Data

Transcripts of child-direct speech for naming frame analysis were drawn from free-play interactions between 17 mothers and their 17–19-month-old children. These dyads were seated across from each other and asked to play with three novel toys for 3 min at a time. They were given three such sets of toys, resulting in nine total minutes of interaction. Parents were taught labels for each of these toys (e.g., "*dax*," "*toma*") and asked to use these if they wished to refer to them by name. No other instructions were given.

Audio recordings of each parent's speech were automatically partitioned into individual utterances using a threshold of 1 s of speech silence. This approach provides a consistent, objective cutoff and obviates the reliability issues involved in human coding. For the purpose of speech segmentation, the importance of

utterance boundaries is that they provide salient stops that disambiguate word boundaries. Because previous research shows that pauses on the order of 100 ms (Ettlinger et al., 2011) and 400 ms (Finn and Hudson Kam, 2008) affect adult speech segmentation, and pauses on the order of 500 ms (Mattys et al., 1999) affect infant statistical speech segmentation, 1 s is a conservative estimate of the length of pauses that would provide disambiguating information to children.

These utterances were then transcribed by human coders into English. Naming frame regularities were extracted using a six-word window made up of three words on either side of a toy's label. If fewer than three words preceded or followed a label in any given utterance, blanks were inserted to fill out the window (e.g., "_ _ the *toma* is blue _"). Next, individual toy labels were replaced with a common token (OBJ), and the frequency of each resulting multi-word frame was computed.

EXPERIMENTS

Participants

Ninety-two undergraduate students from Indiana University participated in exchange for course credit. All participants were self-reported native speakers of English. These participants were divided into four approximately equal groups, each exposed to one of the artificial languages.

Materials

Stimuli for the experiment consisted of 18 unique objects (from Yu and Smith, 2007), and 38 unique words. Eighteen of these words acted as labels for the novel objects, and the other 20 were mapped onto the words contained in the 21 most frequent frames found in the corpus analysis. Half of the words of each type were one syllable (CV) long, and the other half were two syllables (CVCV) long, necessitating the construction of 57 unique syllables. These syllables were created by sampling 57 of the 60 possible combinations of 12 constants and 5 vowels. Syllables were assigned to words randomly, so that nothing about a word's phonetic properties could be used to distinguish Object Labels from Frame Words.

Words were then concatenated together without intervening pauses to create artificial language equivalents of each of the 21 frequent frames in the corpus. Participants were exposed to synthesized versions of these phrases constructed with MBROLA (Dutoit et al., 1996). This produced utterances in which no prosodic or phonetic properties could be used to determine word boundaries, forcing participants to rely on statistical information. Speech was synthesized using the *us*1 diphone database – an American female speaking voice. Each consonant was 94 ms long with a pitch point of 200 Hz at 10 ms. Each vowel was 292 ms long with a 221 Hz pitch point at 108 ms and a 200 Hz pitch point at 292 ms. Each syllable was separated from the next by a 1 ms pause and each utterance ended with a 20 ms pause. These values were chosen to produce speech with a natural sound and cadence.

Design and procedure

Participants were told that they would be exposed to scenes consisting of two novel objects, and a phrase referring to each of them.

Table 2 | The 2 × 2 design of the artificial language experiment.

	Final position	**Middle position**
Preceding cue	*Full Language* "Look at *the* OBJ" Onset *H*: 1.45, Offset *H*: 0	*Onset Only Language* "At *the* OBJ look" Onset *H*: 1.45, Offset *H*: 3.50
No cue	*Position Only Language* "*The* look at OBJ" Onset *H*: 2.71, Offset *H*: 0	*Control Language* "*the* look OBJ at" Onset *H*: 2.71, Offset *H*: 3.50

Phrasal position of the Object Label varies along the rows; presence of the onset cue varies along the columns.

Each phrase would contain exactly one word labeling an on-screen object, along with several function words corresponding to the grammar of the artificial language. Participants had to determine which phrase referred to which object, how the phrases they heard should be segmented into words, and which of these words referred to which of the objects. Next, participants observed an example trial using English words and familiar objects to demonstrate the task. Importantly, the example contained both an object-final phrase ("observe the tractor") and an object-medial phrase ("and the dog over there") to prevent participants from expecting any particular positional regularity.

After the example, participants observed 108 training trials, each containing 2 objects and 2 spoken artificial language phrases (**Figure 1**). Trials began with 2 s of silence, each phrase was approximately 2 s in length, and 3 s of silence succeeded each phrase, resulting in trials approximately 12 s long. Each object appeared 12 times, and each naming frame occurred a number of times proportional to its appearance in the child-directed speech corpus. The entire training set ran just over 20 min.

After training, participants were tested first for speech segmentation and then word-object mapping. On each segmentation test trial, a participant heard 2 two-syllable words: a word from the experiment and a foil created by concatenating the first syllable of one word and the second syllable of another (following Fiser and Aslin, 2002). They were asked to indicate which of the words was more likely to be part of the artificial language (2AFC Test). Six correct Object Labels were tested against 6 Object foils, and 6 correct Frame Words were tested against 6 Frame foils, resulting in 72 total segmentation trials. Each possible word occurred an equal number of times in testing, preventing participants from using test frequency as a cue to correctness. Tests for Object Labels and Frame words were interspersed in a different random order for each participant.

Subsequently, participants were tested on their knowledge of word-object mappings. On each test trial, participants heard one of the Object Labels and were asked to select its correct referent from a set of four alternatives (4AFC Test). All of the labels were tested once in random order.

To assess the independent and joint contribution of both the final position and onset cue regularities, one group of participants was exposed to each of the four possible presence/absence combinations of these cues. Materials and procedure were identical for each of the groups except for the order of words within

each artificial language naming phrase (**Table 2**). To quantify the in-principle difficulty of segmenting each language, we compute the binary entropy of the Frame Words in the positions preceding and following an Object Label in each language condition. Entropy (H) quantifies the variability of a distribution, integrating both the number of unique alternatives and the relative frequency of each alternative (Shannon, 1948). When there is no variability, e.g., when the only possibility is an utterance boundary, entropy is zero. As the number of alternatives increases and their frequencies become more uniform, entropy increases. Onset and Offset entropies for each language are also found in **Table 2**.

ACKNOWLEDGMENTS

This research was supported by a National Science Foundation Graduate Research Fellowship to Daniel Yurovsky and National Institutes of Health R01HD056029 to Chen Yu. The authors wish to thank Damian Fricker for collecting much of the data, and Morten Christiansen, Jenny Saffran, Michael C. Frank, and Roberta Golinkoff for discussion.

REFERENCES

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., and Bever, T. G. (1996). "Models of word segmentation in fluent maternal speech to infants," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Hillsdale: Erlbaum), 117–134.

Bergelson, E., and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3253–3258.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., and Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychol. Sci.* 16, 298–304.

Brent, M. R., and Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition* 81, B33–B44.

Cameron-Faulkner, T., Lieven, E., and Tomasello, M. (2003). A construction based analysis of child directed speech. *Cogn. Sci.* 27, 843–873.

Christiansen, M. H., Onnis, L., and Hockema, S. A. (2009). The secret is in the sound: from unsegmented speech to lexical categories. *Dev. Sci.* 12, 388–395.

Colunga, E., and Smith, L. B. (2005). From the lexicon to expectations about kinds: a role for associative learning. *Psychol. Rev.* 112, 347–382.

Cunillera, T., Càmara, E., Laine, M., and Rodríguez-Fornells, A. (2010a). Speech segmentation is facilitated by visual cues. *Q. J. Exp. Psychol.* 63, 260–274.

Cunillera, T., Laine, M., Càmara, E., and Rodríguez-Fornells, A. (2010b). Bridging the gap between speech segmentation and word-to-world mappings: evidence from an audiovisual statistical learning task. *J. Mem. Lang*. 63, 295–305.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and van der Vrecken, O. (1996). The MBROLA project: towards a set of high-quality speech synthesizers dree of use for non-commercial purposes," in *Proceedings of the Fourth International Conference on Spoken Language Processing*, eds H. T. Bunell and W. Isardi (Wilmington, DE: Dupont Institute), 1393–1396.

Echols, C. H., and Newport, E. L. (1993). The role of stress and position in determining first words. *Lang. Acquis.* 2, 189–220.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306.

Endress, A. D., Scholl, B. J., and Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *J. Exp. Psychol. Gen.* 134, 406–419.

Ettlinger, M., Finn, A. S., and Hudson Kam, C. L. (2011). The effect of sonority on word segmentation: evidence for the use of a phonological universal. *Cogn. Sci.* 36, 655–673.

Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). "Learning phonetic categories by learning a lexicon," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds A. D. De Groot and G. Heymans (Austin, TX: Cognitive Science Society), 2208–2213.

Fernald, A., and Hurtado, N. (2006). Names in frames: infants interpret words in sentence frames faster than words in isolation. *Dev. Sci.* 9, F33–F40.

Fernald, A., and Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Dev.* 64, 637–656.

Finn, A. S., and Hudson Kam, C. L. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition* 108, 477–499.

Fiser, J., and Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467.

Frank, M. C., Goldwater, S., Griffiths, T. L., and Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition* 117, 107–125.

Frank, M. C., Mansinghka, V., Gibson, E., and Tenenbaum, J. B. (2007). "Word segmentation as word learning: integrating stress and meaning with distributional cues," in *Proceedings of the 31st Annual Boston University Conference on Language Development*, eds H. Caunt-Nulton, S. Kulatilake, and I. Woo (Boston, MA: Boston University), 218–229.

Frank, M. C., Slemmer, J. A., Marcus, G. F., and Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Dev. Sci.* 12, 504–509.

Fu, W. (2008). Is a single-bladed knife enough to dissect human cognition? Commentary on Griffiths et al. *Cogn. Sci.* 32, 155–161.

Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition* 73, 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Lang. Acquis.* 1, 3–55.

Gogate, L. J., Bahrick, L. E., and Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Dev.* 71, 878–894.

Gogate, L. J., Walker-Andrews, A. S., and Bahrick, L. E. (2001). The intersensory origins of word comprehension: an ecological-dynamic systems view. *Dev. Sci.* 4, 1–37.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., et al. (2010). General cognitive principles for learning structure in time and space. *Trends Cogn. Sci. (Regul. Ed.)* 14, 249–258.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436.

Gómez, R. L., and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci. (Regul. Ed.)* 4, 178–186.

Goodsitt, J. V., Morgan, J. L., and Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *J. Child Lang.* 20, 229–252.

Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Grassmann, S., and Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Dev. Sci.* 13, 252–263.

Hay, J. F., Pelucchi, B., Graf, K., and Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cogn. Psychol.* 63, 93–106.

Johns, B. T., and Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychon. Bull. Rev.* 17, 662–672.

Johnson, E. K., and Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Dev. Sci.* 13, 339–345.

Johnson, M., Frank, M. C., Demuth, K., and Jones, B. K. (2010). Synergies in learning words and their referents. *Adv. Neural Inf. Process Syst.* 23, 1018–1026.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.

Kurumada, C., Meylan, S., and Frank, M. C. (2011). "Zipfian frequencies support statistical word segmentation," in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, eds L. Carlson, C. Hölscher, and T. Shipley (Austin, TX: Cognitive Science Society), 2667–2672.

Lew-Williams, C., Pelucchi, B., and Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Dev. Sci.* 14, 1323–1329.

Marcus, G. F. (2000). Pabiku and ga ti ga: two mechanisms infants use to learn about the world. *Curr. Dir. Psychol. Sci.* 9, 145–147.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cogn. Psychol.* 38, 465–494.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

Medina, T. N., Snedeker, J., Trueswell, J. C., and Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9014–9019.

Mersad, K., and Nazzi, T. (2012). When mommy comes to the rescue of statistics: infants combine top-down and bottom-up cues to segment speech. *Lang. Learn. Dev.* 8, 303–315.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90, 91–117.

Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition* 108, 271–280.

Peters, A. M. (1977). Language learning strategies: does the whole equal the sum of the parts? *Language* 53, 560–573.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and the acquisition of phonology. *Lang. Speech* 46, 115–154.

Räsänen, O. (2011). A computational model of word segmentation from continuous speech using transitional probabilities of atomic acoustic events. *Cognition* 120, 149–176.

Riordan, B., and Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: comparing feature-based and distributional models of semantic representation. *Top. Cogn. Sci.* 3, 303–345.

Saffran, J. R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., and Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 621, 606–621.

Shafer, V., Shucard, D., Shucard, J., and Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *J. Speech Lang. Hear. Res.* 41, 874–886.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Shukla, M., White, K. S., and Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6038–6043.

Smith, L. B. (2000). "How to learn words: an associative crane," in *Breaking the Word Learning Barrier*, eds R. M. Golinkoff and K. Hirsh-Pasek (Oxford: Oxford University Press), 51–80.

Smith, L. B., Colunga, E., and Yoshida, H. (2010). Knowledge as process: contextually-cued attention and early word learning. *Cogn. Sci.* 34, 1287–1314.

Smith, L. B., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568.

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cogn. Sci.* 34, 1093–1106.

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 49–67.

Vogt, P. (2012). Exploring the robustness of cross-situational learning under Zipfian distributions. *Cogn. Sci.* 36, 726–739.

Vouloumanos, A., and Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Dev. Psychol.* 45, 1611–1167.

Vygotsky, L. (1978). *Mind and Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

Waxman, S. R., and Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends Cogn. Sci. (Regul. Ed.)* 13, 258–263.

Weiss, D., Gerfen, C., and Mitchel, A. (2010). Colliding cues in word segmentation: the role of cue strength and general cognitive processes. *Lang. Cogn. Process.* 25, 402–422.

Yu, C., Ballard, D. H., and Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cogn. Sci.* 29, 961–1005.

Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420.

Yu, C., and Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition* 125, 244–262.

Yu, C., Smith, L. B., and Pereira, A. F. (2008). "Grounding word learning in multimodal sensorimotor interaction," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, eds B. C. Love, K. McRae, and V. M. Sloutsky (Austin, TX: Cognitive Science Society), 1017–1022.

Yurovsky, D., Smith, L. B., and Yu, C. (2012). *Does Statistical Word Learning Scale? It's a Matter of Perspective.* Austin, TX: Cognitive Science Society, 1209–1213.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 July 2012; accepted: 11 September 2012; published online: 01 October 2012.

*Citation: Yurovsky D, Yu C and Smith LB (2012) Statistical speech segmentation and word learning in parallel: scaffolding from child-directed speech. Front. Psychology **3**:374. doi: 10.3389/fpsyg.2012.00374*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 06 February 2013
doi: 10.3389/fpsyg.2013.00024

The segmentation of sub-lexical morphemes in English-learning 15-month-olds

Toben H. Mintz[1,2]*

[1] Department of Psychology, University of Southern California, Los Angeles, CA, USA
[2] Department of Linguistics, University of Southern California, Los Angeles, CA, USA

Edited by:
Jutta L. Mueller, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
LouAnn Gerken, University of Arizona, USA
Barbara Hoehle, University of Potsdam, Germany
Marieke Van Heugten, CNRS/EHESS/ENS, France

***Correspondence:**
Toben H. Mintz, Department of Psychology, University of Southern California, SGM 501, MC-1061, Los Angeles, CA 90089-1061, USA.
e-mail: tmintz@usc.edu

In most human languages, important components of linguistic structure are carried by affixes, also called bound morphemes. The affixes in a language comprise a relatively small but frequently occurring set of forms that surface as parts of words, but never occur without a stem. They combine productively with word stems and other grammatical entities in systematic and predictable ways. For example, the English suffix *-ing* occurs on verb stems, and in combination with a form of the auxiliary verb *be*, marks the verb with progressive aspect (e.g., *was walking*). In acquiring a language, learners must acquire rules of combination for affixes. However, prior to learning these combinatorial rules, learners are faced with discovering what the sub-lexical forms are over which the rules operate. That is, they have to discover the bound morphemes themselves. It is not known when English-learners begin to analyze words into morphological units. Previous research with learners of English found evidence that 18-month-olds have started to learn the combinatorial rules involving bound morphemes, and that 15-month-olds have not. However, it is not known whether 15-month-olds nevertheless represent the morphemes as distinct entities. This present study demonstrates that when 15-month-olds process words that end in *-ing*, they segment the suffix from the word, but they do not do so with endings that are not morphemes. Eight-month olds do not show this capacity. Thus, 15-month-olds have already started to identify bound morphemes and actively use them in processing speech.

Keywords: language acquisition, morphology, infancy, speech perception, lexicon, psycholinguistics

INTRODUCTION

In most human languages important components of linguistic structure are carried by affixes, or bound morphemes. The affixes in a language comprise a relatively small but frequently occurring set of forms that surface as parts of words, but never occur without a stem. While bound morphemes always occur as part of a larger word, they are viewed as having an independent status by virtue of the fact that they combine productively with stems and other grammatical elements in systematic and predictable ways. For example, any English verb root that is inflected with the suffix *-ing* and is preceded by a form of the auxiliary verb, *be*, results in a verb form that is marked with particular tense and aspect: present progressive (e.g., *she is reading*). Mastering the morphological system of a language thus involves acquiring the generalizations about the relationships between formal elements (e.g., auxiliary-*be* and *-ing*), as well as the semantic and functional properties of the language that are represented in the morphological system (e.g., mood, aspect, number, etc.). However, before learners can acquire morphological facts about their language, they must first identify the sub-lexical combinatorial units: they must identify the bound morphemes.

Children's first productive use of bound morphemes (and functional categories more broadly, including function words) is delayed relative to their initial production of content words. For example, children typically produce their first words at approximately 12 months, but it is not until they combine words, between 18 and 24 months, that children learning English begin to produce morphemes when they are required (Brown, 1973; de Villiers and de Villiers, 1973), and even then, mastery may be limited to a small number of forms.

From perception and comprehension studies, there is also evidence that infants learning English have started to form representations of sub-lexical morphemes, and have learned something about the patterns in which the morphemes normally occur, by the time they start producing two-word combinations (Santelmann and Jusczyk, 1998; Golinkoff et al., 2001; Soderstrom et al., 2002). For example, Santelmann and Jusczyk (1998) showed that 18-month-old infants preferred to listen to grammatical sentences in which a word ending in the morpheme *-ing* followed the function word *is* (1a), over ungrammatical sentences in which the word followed the function word *can* (1b).

(1) a. At the bakery, everybody is baking bread.
b. *At the bakery, everybody can baking bread.

However, Santelmann and Jusczyk did not find such a differential preference in 15-month-olds. Similarly, for the inflection *-s* (plural and third person singular), Soderstrom (2003) and Soderstrom et al. (2002) showed that 19-month-olds noticed when normal dependencies between the affix and nearby function

words were violated, but 16-month-olds did not. However, Soderstrom et al. (2007) reported some conditions under which even 16-month-olds show a sensitivity to a misplaced *-s* affix. Taken together, these experiments demonstrate that by 18 months, English-learning infants have learned morphosyntactic patterns involving a range of sub-lexical morphemes, and suggest that infants' sensitivity to some of these patterns is developing at 16 months. As a consequence, these studies also provide evidence concerning when learners represent affixes as distinct forms – that is, separate from the stems to which they are attached – since infants must first segment the affixes as distinct units before learning patterns to which they contribute.

Similar experiments with infants learning German (Höhle et al., 2006), Dutch (van Heugten and Johnson, 2010), and French (van Heugten and Shi, 2010; Nazzi et al., 2011) have broadly replicated the finding that infants between 17 and 24 months are becoming sensitive to morphosyntactic patterns involving affixes, and to functional elements more broadly (van Heugten and Shi, 2009; Shi and Melançon, 2010). At the same time, these cross-linguistic studies provided further insights into the distributional and linguistic factors that influence how infants process morphosyntactic dependencies. However, these studies leave open the question of infants' representations of sub-lexical morphemes in the developmental period before they show sensitivity to dependencies between morphosyntactic units. That is, it is not clear precisely *why* 15-month-olds failed to respond differently to (1a) and (1b) in Santelmann and Jusczyk's (1998) study. There is evidence that between 11 and 14 months, infants acquire representations of function words (Shi et al., 2006a,b), so 15-month-olds' behavior is not likely to be due to an inability to distinguish *is* in (1a) from *can* in (1b). However, it could be that 15-month-olds simply do not represent *-ing* as a discrete unit, and therefore have no way of representing patterns and dependencies involving that morpheme. On the other hand, they might have a discrete representation of *-ing*, but have not yet learned the dependency patterns in which *-ing* participates. Resolving this question is important for understanding the time-course of infants' morphosyntactic development, as well as for providing a basis for further research into the mechanisms of infants' morphosyntactic acquisition.

A recent study of French-learning infants is relevant to this question. Marquis and Shi (2012) familiarized French-learning 11-month-olds to a pseudo-root (i.e., a nonsense syllable). They then recorded infants' listening times to passages containing the pseudo-root "inflected" with the actual French suffix /e/ and to sentences with an unfamiliarized pseudo-root, also ending in /e/. Infants listened longer to the sentences containing the inflected familiarized pseudo-root, suggesting that infants segmented the /e/ ending from the rest of the word and recognized the familiar stem. Different infants who were tested on familiarized and unfamliarized pseudo-roots inflected with /u/, which is not a French affix, did not listen preferentially to either stimulus type. Thus, the response of infants who preferred the familiarized vs. unfamiliarized stems with the /e/ suffix cannot be attributed to phonetic similarity of the familiarized and tested forms; rather, infants' behavior was apparently guided by factors relating to the status of /e/ as a morpheme. Marquis and Shi's study provides the earliest evidence for infants' segmentation of sub-lexical morphemes.

Marquis and Shi's (2012) results demonstrate that infants have begun to form representations of bound morphemes by the end of the first year of life, at least in the case of infants learning French. In considering the question of English-learners' representation of *-ing*, it is tempting to extend this finding to English, and conclude that English 15-month-olds must therefore represent *-ing* as a discrete form. However, there are important differences between French and English that might affect how Marquis and Shi's conclusions from French generalize to English. Foremost is that the inflectional system of French is overall richer than that of English. French marks both grammatical gender and number, and has gender and number agreement between nouns, pronouns, determiners, and adjectives. These properties might lead French-learning infants to attend to, detect, and process suffixes at an earlier age compared to infants learning English and other languages in which overt morphology is relatively impoverished. It is therefore important to verify the finding in other languages.

There are also methodological considerations that limit the generalizability of Marquis and Shi's (2012) findings. In their experiments, infants were familiarized to a pre-segmented stem, and only had to process and recognize that stem in combination with a suffix. If infants' early representations of sub-lexical forms are fragile, their ability to detect and process bound morphemes may be limited. The processing demands of tracking one pre-segmented stem over the course of an experiment may be simple enough for detection of the morpheme and subsequent segmentation of the stem, but sub-lexical processing could be hindered in more complex situations. Replicating the finding with different experimental designs, especially those that place more demands on processing and memory resources, is important for establishing the robustness of infants' early representations of morphology. In each experiment in the current study, infants were exposed to a multitude of stems inflected with *-ing*. In order to show evidence of morphological segmentation they had to segment the stems from these forms, remember them over the course of the familiarization period, and then recognize them during the test trials. While infants would not need to segment and retain every stem in order to show a reliable segmentation effect, they would have to track several, thus increasing complexity and resource demands. Furthermore, requiring infants to perform the segmentation during the familiarization phase rather than at test – reversing the method of Marquis and Shi – could increases task difficulty as well. When the bare stem is given first it can aid infants in detecting the relevant words in the test passages, making the task of detecting the stem in the inflected form somewhat easier. However, when the inflected forms are given first (particularly when they are in passages, as in Experiments 2–4), infants do not have this extra guide to morphological segmentation.

In summary, Marquis and Shi's (2012) findings provide important evidence that infants can represent sub-lexical morphemes well in advance of their ability to track the dependency patterns in which they occur. However, typological differences between English and French, as well as the single methodological context of the findings only provide indirect evidence with respect to morphological representations in English-learners. Thus, the question of whether English-learning 15-month-olds treat *-ing* as

a distinct form [and, thus, their apparent insensitivity to the violation in (1b)] remains open. The present study provides a more direct assessment of English-learning 15-month-olds' morphological representations. Experiments 1–3 use multiple designs and stimuli sets to provide converging evidence that English-learning 15-month-olds treat *-ing* as a distinct unit. Evidence for a discrete representation is inferred from infants' ability to segment *-ing*, in contrast to non-morpheme suffixes, from the ends of novel words. Motivated by the formal similarities of sub-lexical segmentation and word segmentation, Experiment 4 goes on to test for evidence of sub-lexical segmentation in 8-month-olds, who have been shown to segment words from continuous speech (Jusczyk and Aslin, 1995; Saffran et al., 1996; Jusczyk et al., 1999; Pelucchi et al., 2009).

EXPERIMENT 1

This experiment tested the hypothesis that English-learning 15-month-olds represent the suffix *-ing* as a distinct entity, and that the representation as a distinct form influences infants' parsing and representation of words.

Infants were familiarized to novel words, spoken in isolation. Some of the words ended in the English morpheme, *-ing* (e.g., *lerjoving*), and others ended with the phoneme sequence /ɑt/ (*-ot*, e.g., *jemontot*), while others did not systematically share an ending. The prediction was that if 15-month-olds represented the suffix *-ing* as a distinct entity, then they would be more likely to segment *-ing* from the ends of the novel words than they would the pseudo-suffix *-ot*. As a consequence of the segmentation process, infants would then store a representation of the resulting isolated novel "stems" (*-ing stems*, e.g., *lerjov*, in the example above). Since, by hypothesis, infants would not perform this kind of sub-lexical segmentation with words ending in *-ot* (or would be considerably less likely to), they should not form sub-lexical representations of the stems of words ending in *-ot* (*-ot stems*). As a result, infants should find *-ing* stems more familiar than *-ot* stems after familiarization. Differences in responses were tested using a version of the Head-turn Preference Procedure (HPP; Kemler Nelson et al., 1995).

MATERIALS AND METHODS

Subjects

All experiments reported in this paper were approved by the University of Southern California's Institutional Review Board. Subjects were recruited by telephone from a database of parents who had expressed interest in having their infant participate in research in our lab. At least one parent of each infant provided informed consent before the infant participated in the experiment. At the conclusion of each test session, we gave the parent a t-shirt for their child that read, "Graduate of the University of Southern California Language Development Lab," as a token of our appreciation.

Data for 24 English-learning 15-month-olds were analyzed (mean age 14:25, range 14:15–15:10). An additional 15 infants were tested but were excluded from the data analysis due to failure to complete the experiment (6), failure to attend for more than 1 s to at least three test trials per block (5), excessive fussiness (2), parental interference (1), infant moved out of view (1). Twelve subjects were randomly assigned to familiarization group A; the remaining 12 were assigned to familiarization group B.

Stimuli and design

Familiarization and test stimuli were recorded by a female, native American English speaker, who was blind to the purpose of the study. Recordings were made in a sound attenuating booth, using a Shure SM58 microphone. Stimuli were digitized directly to a computer, at a sampling rate of 44.1 kHz. Three instances of each of the familiarization and test items were recorded. All stimuli were recorded during the same recording session.

Familiarization stimuli. Familiarization stimuli consisted of two sets, A and B, each consisting of 16 nonce words. In each set, five words ended in the English suffix *-ing*, five ended in the non-morphological ending *-ot* (/ɑt/), and the remaining six words were "uninflected" – that is, ending in a phoneme sequence that was not shared by other familiarization words. The goal in including the uninflected fillers was to add some variety to the familiarization material to help maintain infants' engagement in the experiment. With respect to the design of the experiment, words ending in *-ing* and *-ot* were treated as a pseudo-stem plus an *-ing* or *-ot* suffix. Pseudo-stems in *-ing* words are called *-ing stems* and pseudo-stems in *-ot* words are called *-ot stems*. Sets A and B were designed to counterbalance stems and endings, such that *-ing* stems in one set were *-ot* stems in the other set. The "uninflected" words in both sets were the same. **Table 1** shows the complete set of familiarization stimuli for Experiment 1.

Four of the pseudo-stems were bisyllabic and the remainder were monosyllabic. Stem length was included as a variable in order to increase the variety of the familiarization material, and also to investigate the influence of word complexity on infants' ability to detect suffixes. For bisyllabic stems, stress was controlled such that trochaic and iambic stems occurred equally often with *-ing* and *-ot* endings (see **Table 1**).

Test stimuli. Test stimuli consisted of the 10 pseudo-stems that were "inflected" in the familiarization sets, but now without the suffixes (e.g., *gorp, rimp, gemónt,* etc.). There were four unique test stem types, characterized by their value on two dimensions: number of syllables, and stem status. Stems were either monosyllabic or bisyllabic (derived from bisyllabic and trisyllabic familiarization

Table 1 | Familiarization material for experiment 1.

Set A			Set B		
Gorping	Rimpot	Choon	Gorpot	Rimping	Choon
Feming	Genot	Wug	Femot	Gening	Wug
Fejing	Sibot	Zimp	Fejot	Sibbing	Zimp
Gemónting	Jivántot	Pux	Gemóntot	Jivánting	Pux
Lérjoving	Káfteeot	Grífdon	Lérjovot	Káfteeing	Grífdon
		Bincáde			Bincáde

Half the subjects heard Set A, half heard Set B. For bisyllabic words, the stressed syllable is indicated with the accent mark.

words, respectively), and were either *-ing* stems or *-ot* stems. While the test stimuli were identical for all infants, the status of the stem – that is, whether it was an *-ing* stem or *-ot* stem – depended on the infant's familiarization set. This design feature counterbalanced stem status for each test stem. **Table 2** shows the test stems, organized by number of syllables and stem status.

Acoustic properties. To ensure that any differences in infants' ability to segment *-ing* and *-ot* could not be due to acoustic differences between the endings, the mean amplitude and duration of *-ing* and *-ot* in tokens of the familiarization materials were measured using Praat (Boersma and Weenink, 2009). Since each word was realized in three tokens, acoustic measures were averaged across the three tokens for each word. **Table 3** presents the mean values for each suffix, organized by word stem. **Figure 1** depicts these means graphically, indicating the affix type. As the table and figure show, the endings are not systematically different as a function of either dimension nor simple combination of dimensions.

Procedure and apparatus

Each infant was tested separately while seated on a caretaker's lap in the center of a sound-attenuated room. The caretaker listened to masking music over close-fitting headphones, in order not to hear the experimental material. An experimenter observed the infant's looking behavior through a closed-circuit television monitor in an adjacent room. The experimenter registered the infant's head-turn responses into a computer that controlled all aspects of the experiment.

At the start of the familiarization phase, a red light positioned at eye level on the wall directly in front of the infant flashed repeatedly. When the infant oriented toward the light, the familiarization material was played on two loudspeakers mounted on the walls to the left and right of the infant. When the familiarization stream started, the center light was extinguished and a light mounted above one of the loudspeakers flashed. It continued to flash until the infant first looked toward it, then looked away for two consecutive seconds. The side light was then extinguished and the center light flashed again until the infant oriented to the neutral center position. This process was repeated for the duration of this phase, randomizing the side on which the light flashed. The interactions with the lights kept the infants engaged, and established the contingency between their looking behavior and the activation of the lights.

The familiarization material played continuously, during the entire familiarization phase, and was not dependent on the infants' orientation once the trial began. The 16 familiarization words were presented in five blocks, with the order of words randomized within each block, and with a different random order for each infant. There was a 300 ms silence between each word. Since there were three recorded versions of each word (see section Stimuli and Design), the computer randomly selected one of the three tokens on each presentation. Half the subjects heard word set A words, and the other half heard set B words. The total familiarization period lasted approximately 80 s.

Table 2 | Test stimuli for experiment 1.

MONOSYLLABIC TRIALS		
-ing stems for group A	*Gorp, fem, fej*	*-ot* stems for group B
-ot stems for group A	*Rimp, gen, sib*	*-ing* stems for group B
BISYLLABIC TRIALS		
-ing stems for group A	*Gemónt, lérjov*	*-ot* stems for group B
-ot stems for group A	*Jivánt, káftee*	*-ing* stems for group B

Each row specifies the stems used in one test trial.

Table 3 | Measurements of duration and intensity of the English affix and pseudo-affix used in Experiment 1.

	Duration (s)		Amplitude (dB sones)	
Stem	**-ing**	**-ot**	**-ing**	**-ot**
Fej	0.36	0.27	63.44	67.34
Fem	0.39	0.45	64.50	61.35
Gemont	0.31	0.32	63.28	68.26
Gen	0.31	0.38	66.24	55.69
Gorp	0.38	0.33	66.89	66.92
Jivant	0.38	0.43	64.39	60.32
Kaftee	0.38	0.43	62.81	60.52
Lerjov	0.30	0.40	62.55	66.77
Rimp	0.40	0.35	66.80	59.63
Sib	0.40	0.38	66.68	62.54
Mean	0.36	0.38	64.76	62.93

For each row, measurements were averaged from the three recorded tokens of the relevant word form (ending in -ing or -ot).

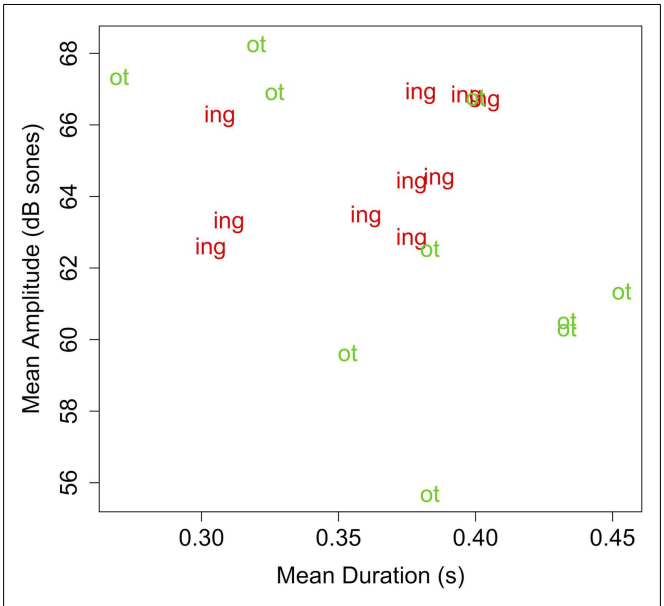


FIGURE 1 | Plot of duration (s) by amplitude (dB sones) for suffixes in Experiment 1. Each data point represents the mean of the duration and amplitude of the affix, averaged across the three tokens of a familiarization word.

A brief contingency training phase immediately followed the familiarization phase. Here, presentation of the auditory stimuli was also contingent on the infant orienting to the flashing side light. The auditory stimulus was always a 440 Hz pure tone lasting 1 s. Presentation started when the infant oriented toward the flashing side light, and the tone was repeated until the infant looked away for two contiguous seconds. This phase consisted of four such trials. Its purpose was to prepare the infant for the test phase that immediately followed, in which auditory stimulus presentation was similarly contingent on orienting to the flashing light.

The test phase was similar to the contingency training phase except that in each test trial, a sequence of stems was played. **Table 2** shows the four trial types that determined which particular sequences of stems was played. Trial types were defined by the length in syllables of the stems, and the ending that was associated with the stems during familiarization. Stems were played in the order shown, with and ISI of 300 ms. The sequence was repeated within a test trial until the infant looked away for two consecutive seconds, or after 15 repetitions of the sequence. Test trials were presented in two blocks, with trial order randomized within blocks, for a total of eight test trials per infant. The computer recorded the duration of each trial. The progression from one trial to the next was no different for trials within a block compared to the transitions from the first to the second block.

In all phases of the experiment the stimulus presentation side on a given trial was randomly selected. However, the selection was constrained such that stimuli would not be presented to the same side in more than three consecutive trials.

If infants segment the suffix *-ing* from familiarization words, then the *-ing* stems should be relatively familiar to them, as they are an outcome of the segmentation process. If infants do not segment the pseudo-suffix, *-ot*, then the *-ot* stems should be relatively less familiar. Differences in familiarity are predicted to result in differences in listening times to the two types of stimuli.

RESULTS

Listening times under 1 s were replaced with the listening time for the same stimuli in the alternate block. This criterion was used to identify trials in which infants looked away before they heard at least one entire stem in the trial, as such trials were not thought to be informative about the representations of interest. This resulted in one replacement for a bisyllabic *-ot* stem trial, and one replacement for a monosyllabic *-ing* stem trial. However, as described in the subject selection section, infants who maintained a head-turn for less than 1 s on more than one trial per block were not included in the data analysis.

The data were first submitted to an analysis of variance (ANOVA) with stem type (*-ing* or *-ot*) and length in syllables (1 or 2) as within-subjects factors, and familiarization group (A or B) as a between subjects factor. Since there were no significant main effect or interactions involving familiarization group, all further analyses combined group A and B, to increase power. In the resulting 2 × 2 ANOVA, there were no main effects, but there was a significant interaction between stem type and number of syllables in the stem [$F(1,23) = 4.47$, $p = 0.046$].

In order to understand this interaction, infants' mean listening times to *-ing* stems and *-ot* stems were compared separately for monosyllabic and bisyllabic stems. For the monosyllabic stems, infants' mean listening times to *-ing* and *-ot* stems were 12.70 s (SE = 1.18) and 11.60 s (SE = 1.27) respectively. A paired *t*-test showed that these listening times were not significantly different [$t(23) = 0.79$, $p = 0.44$]. However, for bisyllabic stems, infants listening significantly longer to *-ot* stems ($M = 14.1$s, SE = 1.3) compared to *-ing* stems [$M = 10.9$s, SE = 1.1; $t(23) = 2.42$, $p = 0.024$, $d = 0.56$]. Sixteen out of the 24 infants listened longer to bisyllabic *-ot* stems. **Figure 2** depicts listening times to each stem type, organized by length in syllables.

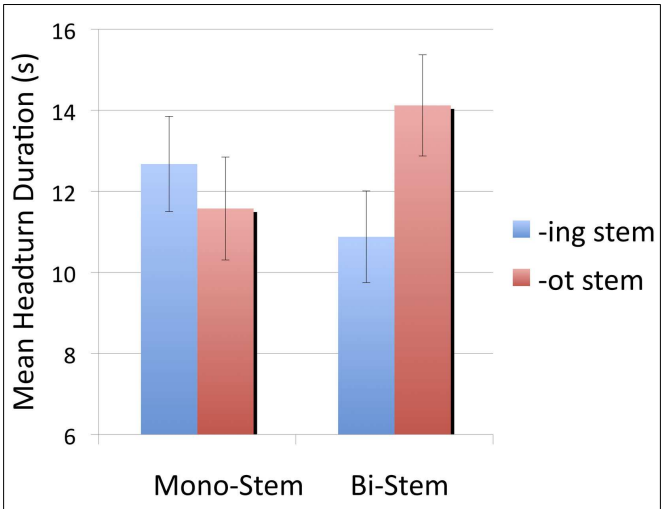


FIGURE 2 | Mean listening times for Experiment 1, organized by stem length in syllables and stem status. Error bars show standard errors.

DISCUSSION

Overall, this experiment provides evidence that by 15 months, English-learning infants treat *-ing* in a special way, such that when they hear a word that ends in that sequence, they segment it from the rest of the word. The evidence comes from comparing test trials in which subjects heard stems to which they were familiarized in words ending in *-ing* vs. words ending in *-ot*. When the stems were bisyllabic, subjects listened longer to the *-ot* stems. Under the assumption that infants had segmented the morphemic stems many times during familiarization, and thus experienced them as an entity distinct from the larger word, the listening differences are consistent with a novelty preference for the *-ot* stems, which, by hypothesis, the subjects had not previously segmented from the familiarization words.

It is not clear why such a difference was not observed for monosyllabic stems. One possibility is that the longer words were more salient in the familiarization phase, and were fore grounded against a background of shorter words. Infants may not have processed the words with monosyllabic stems to the same degree as the words with bisyllabic stems, and therefore may not have segmented either *-ing* or *-ot* from those words. In general, the variable length of the novel words might have disrupted infants ability to segment morphemes across all words (Johnson and Tyler, 2010), and the longer, trisyllabic words (i.e., with bisyllabic stems) may have

been more effective in capturing infants' attention[1]. Infants' ability to segment *-ing* from monosyllabic stems is explored further in Experiment 3.

As the measurements graphed and shown in **Figure 1** and **Table 3** indicate, there are no obvious differences in acoustic salience could have influenced sub-lexical segmentation in a way that would have given rise to the observed results in Experiment 1. Nevertheless, it is worth replicating the finding with different stimuli. With this in mind, Experiment 2 replicates the general finding from Experiment 1 with a different pseudo-affix and a slightly modified design.

EXPERIMENT 2

Experiment 1 provided evidence that is consistent with the interpretation that 15-month-olds preferentially segment *-ing* (as opposed to non-morphemic endings) from words, indicating that they represent *-ing* as a distinct entity. However, the experiment contrasted *-ing* with just one pseudo-affix, *-ot*. It is possible that *-ing* was intrinsically easier for infants to segment than *-ot*, although the acoustic measures do not support this possibility (see **Table 3**). Nevertheless, in order to be confident that the results were not due to some idiosyncratic property of *-ot*. Experiment 2 replicated the general design, but with the pseudo-affix *-dut*. The most obvious difference between the two pseudo-affixes is that *-dut* begins with a stop consonant, whereas *-ot* (like *-ing*) begins with a vowel. At a phonological level, the presence of an onset makes *-dut* more complete as a syllable, compared to *-ot* (and *-ing*), and therefore might increase the chances that the pseudo-suffix will be segmented from the rest of the word (Hayes, 2009). The acoustic properties of *-dut* and *-ing* in Experiment 2 are presented and discussed below.

To make the infants' experience more like one in a normal language context, the familiarization material presented the novel words in English sentences – e.g., *I see you lérjoving!* – rather than in isolation as in Experiment 1. Situating the novel words in simple sentences made the familiarization stimuli more natural than a list of isolated words. The natural contexts could lead to a greater engagement of language processing mechanisms, for example, those involving word segmentation, syntactic and semantic processing. Detecting and segmenting sub-lexical forms might then be enhanced by greater overall linguistic processing. On the other hand, the natural contexts are also more complex, with more material to process in a given utterance, and a greater demand on resources (assuming that subjects are carrying out processing at these other linguistic levels to some degree). We might, hence, observe an advantage for sub-lexical segmentation of forms that are more familiar to infants based on their experience with English, such as the suffix *-ing*.

MATERIALS AND METHODS

Subjects

Subject recruitment procedures were identical to those used in Experiment 1.

Thirty infants averaging 15 months of age participated in the experiment (mean age 15 months 3 days, range 14:15–15:18). Fifteen were randomly assigned to familiarization group A and the remaining subjects were assigned to familiarization group B. An additional 28 subjects were tested, but were excluded from the study due to failure to complete the experiment (15), failure to orient for at least 2 s in at least three trials per block (2), parental interference (3), excessive fussiness (6), equipment failure (1), and experimenter error (1).

Stimuli and design

The nonsense words were the trisyllabic words from Experiment 1. Each nonce word occurred in two different sentences, yielding a total of eight unique familiarization sentences. In all sentences, the nonce word was the final word in the sentence and was in the syntactic position of a verb. Two counterbalanced sets of familiarization sentences (set A and set B) were created. The sentences in set A are given in **Table 4**. Set B was created from set A by exchanging *-dut* and *-ing* endings on the nonce words in the sentences in **Table 4**. For example, the sentence *I see you lérjoving* in set A corresponded to *I see you lérjovdut* in set B.

The familiarization sentences were recorded by a female native English speaker, who was blind to the predictions of the experiment. The speaker was trained to produce the sentences with normal prosody that was appropriate for a simple declarative sentence or a question. The sentences were compiled into three lists, each listing the sentences in a different random order. The speaker was recorded reading each list, resulting in three separate instances of each familiarization sentence, from which the most natural sounding version was selected for use in the experiment.

Test items were the four bare nonce stems: *lérjov, gemónt, káftee, jivánt*. For a given subject, half the test stems were *-ing* stems, and half the stems were *-dut* stems. Due to the counterbalancing procedure, the *-ing* stems for subjects in group A were the *-dut* stems for subject in group B, and vice versa. Hence, any overall differences in infants' responses to *-ing* stems and-*dut* stems could not to be due to idiosyncrasies of the test items themselves, but rather must be related to differences in the test items' distribution in the familiarization strings.

Recall that the stress pattern was trochaic (strong-weak) for half of the nonce stems and iambic (weak-strong) for the other. Stress is known to be a factor in infant speech processing (Jusczyk et al., 1993; Echols et al., 1997; Thiessen and Saffran, 2003; Curtin et al., 2005; among others), and hence could influence sub-lexical segmentation. Consequently, stress pattern was incorporated as a controlled variable in the experimental design. The stress pattern

Table 4 | Familiarization sentences for subjects in group A, in Experiment 2.

Sentences with *-ing* words	Sentences with *-dut* words
I see you *lérjoving*!	Does Sam want to go *káfteedut*?
Johny likes *gemónting*!	I want to go *jivántdut*!
Do you want to go *lérjoving*?	Harold likes *káfteedut*!
Can you see me *gemónting*?	Can you see Sally *jivántdut*?

Familiarization sentences for subjects in group B were identical, except that -ing stems and -dut stems were switched.

[1]I am grateful to an anonymous reviewer for suggesting this interpretation.

for one nonce stem from each stem category (*-ing* and *-dut*) was trochaic and the other was iambic.

Test items were recorded by the same trained speaker who recorded the familiarization sentences. The stems were produced with list intonation, and each word was recorded three times and digitized onto the computer that controlled the experiment. When playing test items, the computer randomly selected one of the three instances of the item to play.

Acoustic properties. Although instances of *-ot* and *-ing* in Experiment 1 did not differ, overall, in the dimensions of intensity and duration (see **Table 3**), it is possible that some other factors made *-ot* particularly resistant to segmentation. The pseudo-affix used here, *-dut*, is more well-formed as a syllable than *-ot* due to the presence of an onset (Hayes, 2009), and should not be resistant to segmentation on phonological grounds. To compare acoustic intensity of *-dut* and *-ing*, the mean intensity for the two endings was measured in each familiarization sentence using the Praat software package (Boersma and Weenink, 2009). Each novel word occurred in two familiarization sentences, so measurements for each word were averaged across its two tokens. **Table 5** reports these means for each word, and **Figure 3** plots the endings on the two dimensions. (Items from Experiment 3 are also shown.) Clearly, on these acoustic measures, *-ing* and *-dut* are not systematically different. Thus, not only is the pseudo-suffix a CVC syllable, it is matched with *-ing* in duration and intensity. Thus, on acoustic-phonetic grounds, the pseudo-suffix should be just as easy to segment from the pseudo-stem as the actual English suffix.

Procedure and apparatus

The apparatus that was used in Experiment 1 was used in Experiment 2, however the procedure varied in several ways. First, the familiarization stimuli were presented in six blocks, rather than five. Subjects thus heard an additional repetition of each novel word in this experiment. The total duration of the familiarization phase was approximately 90 s. Familiarization utterances were presented with an ISI of 200 ms.

Table 5 | Duration and intensity measurements for *-ing* and pseudo-suffixes on target words in Experiments 2–4.

	Duration (s)		Amplitude (dB sones)	
Stem	**-ing**	**-dut**	**-ing**	**-dut**
Gemont	0.25	0.28	64.58	64.89
Jivant	0.26	0.33	64.78	64.22
Kaftee	0.33	0.34	64.72	61.61
Lerjov	0.27	0.29	65.27	47.98
Fem	0.20	0.25	62.96	72.72
Gorp	0.27	0.22	67.22	70.50
Riz	0.27	0.23	69.34	73.02
Mean	0.27	0.23	65.91	65.30

Bisyllabic stems were used in Experiment 2 and 4, and monosyllabic stems were used in experiment 3. Values are averaged across the two tokens of each word.

The test phase also differed from Experiment 1 in that here, each test trial repeated only one stem, rather than multiple stems of the same type. Thus, there were four unique test trials, together constituting every combination of stem type (*-ing* vs. *-dut*) and stress pattern (trochaic vs. iambic). Due to the counterbalanced design, *-ing* stems for group A subjects were *-ot* stems for group B subjects, and vice versa. As in Experiment 1, test trials were presented in two blocks, with order randomized within blocks.

All other aspects of the procedure were identical to Experiment 1.

RESULTS AND DISCUSSION

Test trials with a listening time under 1 s were replaced with the listening time for the same stimulus in the other block. Data for one *-ing* stem trial and one *-dut* stem trial were modified in this way.

The data were first submitted to a 2 × 2 × 2 ANOVA with stem type (*-ing* or *-dut*) and stem stress pattern (trochaic or iambic) as within-subjects factors, and counterbalance group (A or B) as a between subjects factor. Since the group variable did not interact with any other variable, data from the two groups were combined in subsequent analyses, to increase power. A 2 × 2 ANOVA was performed, with stem type (*-ing* or *-dut*) and stress pattern (trochaic or iambic) as within-subject variables. As predicted, there was a main effect of stem type, with infants listening on average for 9.02 s (SE = 0.34) to *-ing* stems compared to 8.04 s (SE = 0.33) to *-dut* stems [$F(1,29) = 5.30$, $p = 0.029$, $\eta_p^2 = 0.154$][2]. Twenty two out

[2]In this and subsequent experiments, the number of subjects that contributed to the data analyses was greater than in Experiment 1. This is because we anticipated relatively high dropout rates based on piloting, so the research assistants were given a high quota for the number of subjects to test in a given experiment. As a result,

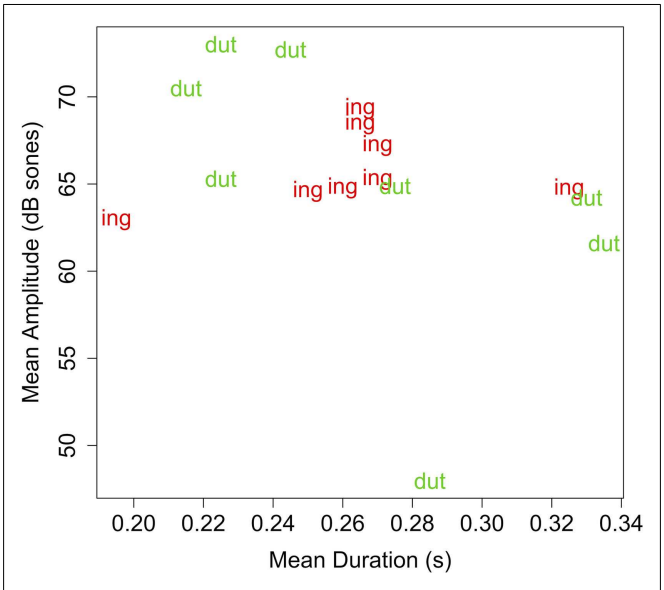


FIGURE 3 | Plot of duration (s) by amplitude (dB sones) for suffixes in Experiments 2–4. Each data point represents the two measurements of the affix (of the type designated by the label) of a token of a familiarization word.

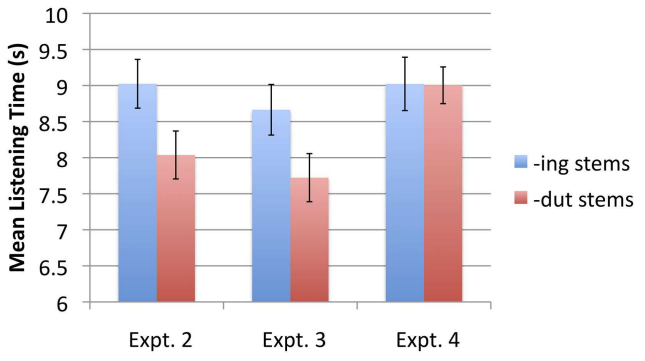


FIGURE 4 | Mean listening times to *-ing* stem and *-dut* stems in Experiments 2–4.

of the 30 infants showed this pattern. There was no other significant main effect or interaction. **Figure 4** graphs the mean listening times to *-ing* stems and *-dut* stems.

As in Experiment 1, infants responded differently to stems to which they were familiarized in words that ended in the English suffix, *-ing*, compared to stems to which they were familiarized in words that ended in a pseudo-suffix. However, here infants listening longer to *-ing* stems compared to the pseudo-suffix stems, whereas in Experiment 1 infants listened longer to the pseudo-suffix stems. The preference for familiarity here vs. novelty in Experiment 1 is not surprising when one considers the differences in design across the two experiments. In Experiment 1, infants were familiarized to the inflected words in isolation, whereas in this experiment, the words were embedded in English sentences. It is a reasonable assumption that 15-month-olds processed the additional rich structure in the familiarization input to some degree – segmenting words (Aslin et al., 1998), categorizing words (Höhle et al., 2004; Gerken et al., 2005; Mintz, 2006; Shi and Melançon, 2010), and accessing word meanings. Stimulus complexity has been proposed as an important influence on infants' preference for novelty or familiarity in experimental paradigms such as the HPP: Higher complexity during familiarization and learning phases is associated with a preference for more familiar test material, as long as that complexity is within the domain of what infants can process and represent (Hunter et al., 1983; Hunter and Ames, 1988; Kidd et al., 2012). Hence, the increase in complexity and variety in the familiarization material from Experiment 1 to Experiment 2 is consistent with a shift from a novelty preference in Experiment 1 to a familiarity preference in Experiment 2.

The results of Experiment 2 thus provide further support for the hypothesis that 15-month-olds treat the suffix *-ing* as a distinct element. Experiments 1 and 2 compared sub-lexical segmentation with *-ing* and two different pseudo-suffixes. In both cases, the results indicated that infants segmented stems and endings differently when the ending was the English suffix vs. the non-English pseudo-suffixes.

we ended up with more subjects than in the initial study. However, restricting the data analysis to the first 12 subjects per counterbalance condition in Experiments 2–4 (as in Experiment 1) yields an identical pattern of results to the reported ones that include additional infants.

Table 6 | Familiarization sentences for subjects in group A, in Experiment 3.

Sentences With *-ing* Words	Sentences with *-dut* Words
I see you *feming*!	Does Sam want to go *sibdut*?
Johny likes *gorping*!	I want to go *rizdut*!
Do you want to go *feming*?	Harold likes *sibdut*!
Can you see me *gorping*?	Can you see Sally *rizdut*?

Familiarization sentences for subjects in group B were identical, except that -ing stems and -dut stems were switched.

In Experiment 1, however, the segmentation differences were only found for bisyllabic stems. Infants did not show evidence of a different pattern of sub-lexical segmentation with monosyllabic stems. One explanation was that when listening to a list of isolated words, the trisyllabic words (with bisyllabic stems) may have stood out against a background of mono- and bisyllabic words, and captured infants attention more than the bisyllabic words. In contrast to the relatively unnatural familiarization scenario in Experiment 1 (a long list of isolated words), Experiment 2 exposed infants to the novel words in a much more natural context, which might more fully engage language processing mechanisms and in turn facilitate the detection of familiar suffixes in bisyllabic words. Experiment 3 tests this prediction by exposing 15-month-olds to bisyllabic nonsense words in an experimental design that is similar to Experiment 2.

EXPERIMENT 3

MATERIALS AND METHODS

Subjects

Subject recruitment procedures were identical to those used in the previous experiments.

Data for 34 infants averaging 15 months of age (mean age 15 months 1 day, range 14 months 13 days to 15 months 14 days) were analyzed. Data from 19 additional infants were excluded due to failure to complete the experiment (13), excessive fussiness (3), parental interference (2), and experimenter error (1).

Stimuli and design

The familiarization and test stimuli were prepared in the same manner as in Experiment 2. The structure of the familiarization material conformed to the structure in Experiment 2, except the nonce words were bisyllabic rather than trisyllabic, and the stress pattern for all nonce words was trochaic. As in Experiment 2, there were two counterbalanced familiarization sets, A and B, such that the *-ing* stems inset A were the *-dut* stems inset B, and vice versa. The familiarization items for set A are given in **Table 6**. The test items were the four nonce stems alone: *fem, gorp, sib,* and *riz*. *Fem,* and *gorp* were *-ing* stems for group A subjects, and *-dut* stems for group B subjects. Likewise, *sib* and *riz*, were *-ing* stems for group B subjects, but *-dut* stems for group A subjects.

Procedure

The procedure was identical to the procedure in Experiment 2, except that there were seven, rather than six familiarization blocks.

This is because the familiarization sentences were slightly shorter in duration, and the total duration of the familiarization period was kept to approximately 90 s.

RESULTS AND DISCUSSION

As in the prior studies, test trials with orientation times under 1 s were replaced with the subject's orientation time for the same stimulus in the other block. Data for three *-ing* stem trials were modified in this way.

For each subject, a mean orientation time for *-ing* stems was calculated by averaging orientation times to all *-ing* stem trials, across test blocks. An average orientation times to *-dut* stems was calculated in the analogous way, resulting in two data points per subject.

Subjects in the A and B familiarization groups did not differ in their overall response patterns to *-ing* vs. *-dut* stems [$t(32) = 1.33$, $p = 0.19$], so scores for the two groups were pooled. As in Experiment 2, infants listened significantly longer to *-ing* stems compared to *-dut* stems. Mean listening times were 8.7 s (SE = 0.35) and 7.7 s (SE = 0.334) for *-ing* and *-dut* stems, respectively [$t(33) = 2.34$, $p = 0.026$ two-tailed, $d = 0.47$]. Twenty two out of the 34 infants showed this pattern. **Figure 4** graphs the mean listening times to the two stem types.

Infants thus behaved similarly here when tested on monosyllabic stems as they did in Experiment 2 when tested on bisyllabic stems: They listened reliably longer to the *-ing* stems compared to the *-dut* stems. Thus, as in Experiments 1 and 2, infants segmented stems out of familiarized words that had the English suffix, but not stems that carried the pseudo-suffix. Here, however, infants showed this segmentation difference for bisyllabic words, whereas in Experiment 1 they did not. As discussed earlier, the structure of the familiarization material could have focused infants' attention on the more distinctive trisyllabic words, so that they were less likely to detect and segment *-ing* from monosyllabic stems. In addition, familiarizing infants to the nonce words in otherwise normal English sentences may have resulted in a greater engagement and activation of normal language processing mechanisms and representations, including the processing of familiar affixes such as *-ing*.

This experiment thus lends further support to the hypothesis that 15-month-old English-learners treat the English suffix *-ing* in a privileged way when processing speech. These findings are in accord with those of Marquis and Shi (2012), who showed that infants learning French represent elements of bound morphology by as early as 11 months. Marquis and Shi suggested that infants form distinct representations of bound morphemes, at least initially, simply because the forms are very frequent in their input. This explanation may be sufficient to account for the difference in segmentation between *-ing* and the pseudo-suffixes used here. However, a mechanism that considers the internal predictability of forms, perhaps in addition to their frequency, is also consistent with the present findings. For example, the word segmentation mechanism proposed by Saffran et al. (1996) segments sequences at junctures of low transitional probability between syllables. Sequences with high transitional probability may also be relatively high in frequency, but two sequences could be equal in frequency yet differ in internal transitional probabilities. Infants as young as 8-months appear to be sensitive to transitional probabilities, not just frequency (Aslin et al., 1998).

The functional similarity between word segmentation and the sub-lexical segmentation of bound morphemes – that is, extracting predictable sequences from larger sequences – could be mirrored by similarities in processing mechanisms. Since 8-month-old infants show evidence of statistically based word segmentation, it is thus possible that they also can detect highly regular patterns *within* words. Experiment 4 investigates this question by replicating the procedures and design of Experiment 2, but testing 8-month-old infants.

EXPERIMENT 4

MATERIALS AND METHODS

Subjects

Subject recruitment procedures were identical to those used in the previous experiments.

Thirty-six infants averaging 8 months of age (mean age 8 months 3 days, range 7 months 18 days to 8 months 20 days) were tested. Infants were randomly assigned to one of two familiarization groups, A or B, consisting of 18 infants each. Data from all 36 infants were analyzed.

Stimuli and design

The stimuli and design of the experiment was identical to Experiment 2.

Procedure and apparatus

The apparatus and testing procedure was identical to Experiment 2.

RESULTS AND DISCUSSION

As in the prior experiments, any test trial with an orientation times under 1 s was replaced with the subject's orientation time for the same stimulus in the other block. Data for one *-ing* stem test trial was modified in this way.

The data were first submitted to a 2 × 2 × 2 ANOVA with stem type (*-ing* or *-dut*) and stem stress pattern (trochaic or iambic) as within-subjects factors, and counterbalance group (A or B) as a between subjects factor. Since the group variable did not interact with any other variable, data from the two groups were combined to increase power. A 2 × 2 ANOVA was performed, with stem type (*-ing* or *-dut*) and stress pattern (trochaic or iambic) as within-subject variables. Neither main effect was significant, nor was the interaction (all Fs < 1). As shown in **Figure 4**, infants' listening times to *-ing* and *-dut* stems was 9.0 s (SE = 0.37) and 9.0 s (SE = 0.26), respectively.

Unlike in the previous experiments with 15-month-olds, there was no evidence that 8-month-olds treated *-ing* in a special way when processing the familiarization material. In principle, the mechanisms that are engaged in laboratory demonstrations of word segmentation in 7.5–8-month-olds could segment predictable sub-lexical patterns such as bound morphemes. However, this experiment provides no evidence that 8-month-olds are carrying out these kind of analyses. Of course, the design of the experiment assesses segmentation of suffixes indirectly, by measuring infants' responses to stems. It could be that infants segmented

-ing (but not *-dut*) during familiarization, but did not have sufficient exposure to the resulting stems to be able to recognize them during the test phase. Compared to word segmentation experiments, infants' exposure to individual test items is much less in the experiments reported here. For example, in Saffran et al.'s (1996) study, infants were tested on words they had heard 45 times. The number of exposures in the present study may have been sufficient for 15-month-olds, but not for 8-month-olds. On the other hand, it also is possible that 8-month-olds have not yet begun to form long-term representations of sub-lexical forms.

The design of this experiment could be modified to increase exposure to nonce words. However, this runs the risk of providing infants with distributional evidence that the pseudo-affixes are also affixes, and infants may then start segmenting pseudo-affixes as well. Indeed, in one experiment, Marquis and Shi (2012) demonstrated that with sufficient exposure to a pseudo-suffix, /u/, French-learning 11-month-olds started treating the ending similarly to the actual French suffix, /e/, in their experimental task.

GENERAL DISCUSSION

Taken together, the experiments in this study demonstrate that English-learning 15-month-olds represent the suffix *-ing* as a discrete unit. Thus, although previous experiments failed to find evidence that 15-month-olds have acquired morphosyntactic dependencies involving *-ing* (Santelmann and Jusczyk, 1998), infants may nevertheless be in the process of learning these dependencies at this age. Specifically, having a discrete representation of an affix allows infants to notice dependencies between that affix and other forms.

It is important to note that while this study supports the hypothesis that infants treat *-ing* as a discrete entity at 15 months, it would be premature to conclude that they have acquired the form *qua* suffix of English. That is, there is no evidence that these forms are fully morphological, in the sense that infants represent them as elements that participate in dependencies and that are associated with certain semantic properties. (Indeed, Santelmann and Juscyk's results suggest that infants have not yet learned basic patterns and dependencies involving *-ing*.) Initially, infants might represent bound morphemes as distinct entities simply by virtue of the fact that they occur frequently within words, as suggested by Marquis and Shi (2012). The results from the present study are entirely consistent with that proposal. In an examination of the input to the child Peter, in the Bloom corpus (Bloom et al., 1974, 1975) of the CHILDES database (MacWhinney, 2000), 2.2% of word tokens and 6.9% of word types spoken by adults to Peter ended in /Iŋ/ (regardless of whether the ending was a morpheme or not, as in *sing*). In contrast, only 0.6% of tokens and 0.5% of word types ended in /ɑt/, and there were no words that ended in the sequence /dʌt/[3].

Although Marquis and Shi (2012) discuss infants' early representations of bound morphemes in terms of the frequency of sub-lexical patterns, it is reasonable to conjecture that the detection of sub-lexical forms may also depend on transitional probabilities. That is, when a frequent form occurs in many different contexts, it might be more likely to be identified as a distinct form than a form of equal frequency that occurs in a more restricted set of contexts. The mechanisms for segmenting sub-lexical forms would then be computationally similar to mechanisms that have been proposed for detecting words in fluent speech (Saffran et al., 1996; Aslin et al., 1998; Pelucchi et al., 2009). While this may be so, Experiment 4 did not find evidence that 8-month-olds detected and segmented *-ing* from nonsense words, although infants had relatively few exposures to the novel forms compared to other experiments in word segmentation. Future research, using different methods, can further probe how early infants start to segment and represent bound morphemes as distinct forms.

Beyond distributional properties such as frequency and transitional probabilities, phonological factors could also influence infants' early representation of affixes. To the degree that affixes in a given language have phonotactic tendencies that infants can detect, once infants have segmented enough affixes to detect the patterns, they could use the tendencies as cues to guide further segmentation and the discovery of new affixes. This possibility raises a potential concern in this study: Although, as just reported, the frequency of /ɑt/ and /dʌt/ at the ends of words in children's input is very low or virtually absent, the two pseudo-affixes are not parallel in comparison to real English affixes when analyzed at a more general level. Specifically, no English inflectional suffix has a CVC structure, like /dʌt/ (although some derivational affixes do, e.g., *-tion*), but there are frequent affixes with a VC structure, like /ɑt/ (e.g., /Iŋ/, /əz/, /əd/). In principle, if infants are sensitive to these broader phonotactic properties of English inflectional affixes, the atypical structure of *-dut* could have caused infants to reject *-dut* as a possible suffix in Experiments 2 and 3.[4] This possibility offers another explanation for the differing results with respect to monosyllabic stems in Experiment 1 compared to Experiment 3: Infants may be relatively more likely to treat *-ot* as a possible suffix because of its phonological structure, and given the simpler overall structure of bisyllabic words, segmented both *-ing* and *-ot* from the shorter words in Experiment 1. Of course, this study was not designed to test these broader generalizations of phonological form. Nevertheless, to address this possibility, a followup study with adults was carried out; the experiment was designed to assess whether experienced English users show an advantage in segmenting *-ot* – which conforms to English inflection structure – from nonce word forms, compared to *-dut*, which does not. Fifteen native English speakers listened to the same nonce words that ended in *-dut* and *-ot* that were used in these studies, but the words were presented in a rapid sequence, with 1.1 s between word onsets. From time to time, two words in a row both ended in *-dut* or both in *-ot*. Participants had to press a key whenever they heard a word that rhymed with the word before it. The question of interest was whether participants differed in their accuracy in detecting rhymes with *-ot* compared to rhymes with *-dut*. A logistic regression with the ending (*-dut* vs. *-ot*) as a within-subjects variable did not reveal any difference in accuracy in detecting rhymes with *-dut* (on average 78%

[3]The CMU pronouncing dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) was used to identify orthographic forms corresponding to words that ended in the relevant phoneme sequences. The combined frequency of those forms was then tallied in the corpus of child-directed speech.

[4]This possibility was suggested by an anonymous reviewer.

detected) compared to rhymes with *-ot* (on average 68% detected; $p = 0.336$). So for adults, there is apparently no advantage for one form or the other with respect to ease of detection. Interestingly, there was a slight reaction time advantage for *-dut* rhymes (607 ms, measured from suffix onset) compared to *-ot* rhymes [653 ms; $t(14) = 2.20$, $p < 0.05$]. Although these findings from adults are hardly conclusive concerning infants' knowledge of inflections, they at least suggest that infants would not be biased against segmenting *-dut* compared to *-ot* from pseudo-stems, despite the fact that the former is atypical with respect to inflectional suffixes in English.

The modest but reliable speed advantage for detecting *-dut* over *-ot* in adults could be related to the fact that *-dut* is a complete syllable, whereas *ot* lacks an onset and is subject to resyllabification with segments at the end of the stem. Indeed, the motivating factor for using *-dut* in Experiments 2–4 was to use a pseudo-affix that was relatively easy to segment on structural grounds, thus providing a stronger test of infants' treatment of *-ing* as a privileged form. However, going beyond the methodological considerations of this study, perceptual factors relating to affix syllable structure is another way in which phonological variables could play a role in infants' acquisition of affixes: All else being equal, affixes that are subject to resyllabification might be harder to detect and take longer to acquire than affixes that are not. Cross-linguistically, there is some support for this notion. For example, Turkish morphemes are generally syllabic and contain unreduced vowels, and many have onsets. Children learning Turkish show productive use of morphemes somewhat in advance of children learning English (Aksu Koç and Ketrez, 2003). In the present study, although *-ing* lacks an onset, it stands out from most other inflectional morphemes in English in that it has a full vowel. It is also typically the first inflectional morpheme to be reliably produced when required by children learning English. It is possible, then, that while 15-month-olds have identified this "robust" morpheme as a distinct form, they have not yet formed independent representations of other English morphemes. Exploring this question by testing different morphemes will clarify the role of the perceptual properties of suffixes that may influence how bound morphemes are first represented.

Finally, in addition to the potential role of frequency in infants' acquisition of affixes (Marquis and Shi, 2012), more general distributional properties of a language's inflectional system may influence infants' detection of bound morphemes. As alluded to earlier, one might expect the developmental timing of the first representations of morphemes to depend on the richness of a language's overt morphological marking. Learners of languages with rich morphological marking (such as French) may begin to detect and represent sub-lexical forms in advance of their peers learning languages that are morphologically more "impoverished" (such as English). The acquisition of Turkish, again, provides some evidence for this view. Turkish makes extensive use of morphological marking, and children show productive use of morphemes as early as 17 months (Aksu Koç and Ketrez, 2003). However, such comparisons are complicated by the phonological and perceptual factors discussed earlier.

CONCLUSION

A significant component of language, both in structure and in content, resides in the sub-lexical combinatorial units – the bound morphemes. In acquiring a language, learners must acquire the semantic and structural properties of bound morphemes, but before doing so, they must identify what the relevant sub-lexical units are in their language. The experiments reported here demonstrate that English-learning 15-month-olds represent *-ing* as a distinct form. When processing novel words that end in *-ing*, they segment the suffix from the stem. This allows them to notice morphosyntactic and morphosemantic patterns that involve that form, and that will form a part of their acquired grammatical knowledge. In addition, by representing word stems as distinct forms, infants can then detect morphosyntactic patterns involving the stem, such as other inflectional paradigms. Thus, at an age where many learners are not yet combining words in their own speech, and before they use bound morphemes productively, infants have started to develop representations of the morphology of their language.

ACKNOWLEDGMENTS

I would like to thank Laura Steenberge and Christy Hardy for their assistance in collecting and coding the infant data, Felix Wang and Kalyn Reddy for running the adult experiment, and the many parents and families who volunteered to make this research possible. Preliminary data from some of these experiments was reported at the International Conference on Infant Studies, and at the Boston University Conference on Language Development. This research was supported in part by a grant from the National Institutes of Health (NICHD-R01HD040368).

REFERENCES

Aksu Koç, A., and Ketrez, F. N. (2003). "Early verbal morphology in Turkish: emergence of inflections," in *Mini-Paradigms and the Emergence of Verb Morphology*, eds D. Bittner, W. U. Dressler, and Kilani-Schoch (Berlin: Walter de Gruyter), 27–52.

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324.

Bloom, L., Hood, L., and Lightbown, P. (1974). Imitation in language development: if, when, and why. *Cogn. Psychol.* 6, 380–420.

Bloom, L., Lightbown, P., and Hood, L. (1975). Structure and variation in child language. *Monogr. Soc. Res. Child Dev.* 40, 1–97.

Boersma, P., and Weenink, D. (2009). *Praat: doing phonetics by computer (Version 5.1.43)[Computer program]*. Available at: http://www.praat.org/.

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Curtin, S., Mintz, T. H., and Christiansen, M. H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition* 96, 233–262.

de Villiers, J. G., and de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *J. Psycholinguist. Res.* 2, 267–278.

Echols, C. H., Crowhurst, M. J., and Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *J. Mem. Lang.* 36, 202–225.

Gerken, L., Wilson, R., and Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *J. Child Lang.* 32, 249–268.

Golinkoff, R. M., Hirsh-Pasek, K., and Schweisguth, M. A. (2001). "A reappraisal of young children's knowledge of grammatical morphemes," in *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition*, Vol. 1, eds J. Weissenborn and B. Höhle (Amsterdam: John Benjamins), 167–188.

Hayes, B. (2009). *Introductory Phonology*. Malden, MA: Wiley-Blackwell.

Höhle, B., Schmitz, M., Santelmann, L. M., and Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: evidence for lexical and structural influences on children's early processing capacities. *Lang. Learn. Dev.* 2, 277–300.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., and Schmitz, M. (2004). Functional elements in infants' speech processing: the role of determiners in the syntactic categorization of lexical elements. *Infancy* 5, 341–353.

Hunter, M. A., and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Adv. Infancy Res.* 5, 69–95.

Hunter, M. A., Ames, E. W., and Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Dev. Psychol.* 19, 338.

Johnson, E. K., and Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Dev. Sci.* 13, 339–345.

Jusczyk, P. W., and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–23.

Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev.* 64, 675–687.

Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cogn. Psychol.* 39, 159–207.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behav. Dev.* 18, 111–116.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* 7:e36399. doi:10.1371/journal.pone.0036399

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*: *The Database*. 3rd Edn, Vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates.

Marquis, A., and Shi, R. (2012). Initial morphological learning in preverbal infants. *Cognition* 122, 61–66.

Mintz, T. H. (2006). "Finding the verbs: distributional cues to categories available to young learners," in *Action Meets Word: How Children Learn Verbs*, eds R. M. Golinkoff and K. Hirsh-Pasek (New York: Oxford University Press), 31–63.

Nazzi, T., Barrière, I., Goyet, L., Kresh, S., and Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: evidence from the acquisition of subject-verb agreement in French. *Cognition* 120, 119–135.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev.* 80, 674–685.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Santelmann, L. M., and Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition* 69, 105–134.

Shi, R., Cutler, A., Werker, J., and Cruickshank, M. (2006a). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *J. Acoust. Soc. Am.* 119, EL61–EL67.

Shi, R., Werker, J. F., and Cutler, A. (2006b). Recognition and representation of function words in English-learning infants. *Infancy* 10, 187–198.

Shi, R., and Melançon, A. (2010). Syntactic categorization in french-learning infants. *Infancy* 15, 517–533.

Soderstrom, M. (2003). *The Acquisition of Inflection Morphology in Early Perceptual Knowledge of Syntax*. Doctoral Dissertation, The Johns Hopkins University. [Retrieved from ProQuest Dissertations and Theses. (UMI No. 3068215)].

Soderstrom, M., Wexler, K., and Jusczyk, P. W. (2002). "English-learning toddlers' sensitivity to agreement morphology in receptive grammar," in *Proceedings of the 26th Annual Boston University Conference on Language Development*, eds B. Skarabela, S. Fish, and A. H.-J. Do (Somerville: Cascadilla Press), 643–652.

Soderstrom, M., White, K. S., Conwell, E., and Morgan, J. L. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy* 12, 1–29.

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.

van Heugten, M., and Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *J. Mem. Lang.* 63, 197–209.

van Heugten, M., and Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Dev. Sci.* 12, 419–425.

van Heugten, M., and Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *J. Acoust. Soc. Am.* 128, EL223–EL228.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 July 2012; accepted: 10 January 2013; published online: 06 February 2013.

Citation: Mintz TH (2013) The segmentation of sub-lexical morphemes in English-learning 15-month-olds. Front. Psychology ***4****:24. doi: 10.3389/fpsyg.2013.00024*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 29 October 2012
doi: 10.3389/fpsyg.2012.00447

Infants generalize representations of statistically segmented words

Katharine Graf Estes*

Department of Psychology, University of California Davis, Davis, CA, USA

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Sarah Creel, University of California at San Diego, USA
Heather Bortfeld, University of Connecticut, USA

***Correspondence:**
Katharine Graf Estes, Department of Psychology, University of California Davis, One Shields Avenue, Davis, CA 95618, USA.
e-mail: kgrafestes@ucdavis.edu

The acoustic variation in language presents learners with a substantial challenge. To learn by tracking statistical regularities in speech, infants must recognize words across tokens that differ based on characteristics such as the speaker's voice, affect, or the sentence context. Previous statistical learning studies have not investigated how these types of non-phonemic surface form variation affect learning. The present experiments used tasks tailored to two distinct developmental levels to investigate the robustness of statistical learning to variation. Experiment 1 examined statistical word segmentation in 11-month-olds and found that infants can recognize statistically segmented words across a change in the speaker's voice from segmentation to testing. The direction of infants' preferences suggests that recognizing words across a voice change is more difficult than recognizing them in a consistent voice. Experiment 2 tested whether 17-month-olds can generalize the output of statistical learning across variation to support word learning. The infants were successful in their generalization; they associated referents with statistically defined words despite a change in voice from segmentation to label learning. Infants' learning patterns also indicate that they formed representations of across word syllable sequences during segmentation. Thus, low probability sequences can act as object labels in some conditions. The findings of these experiments suggest that the units that emerge during statistical learning are not perceptually constrained, but rather are robust to naturalistic acoustic variation.

Keywords: statistical learning, word segmentation, language acquisition, word learning, speech perception, generalization

INTRODUCTION

Very early in development, infants perform impressive feats of learning. Investigations of statistical learning have revealed that infants rapidly detect distributional patterns that are present in novel visual and auditory input (e.g., Saffran et al., 1996, 1999; Kirkham et al., 2002, 2007). Within the domain of language, statistical learning is hypothesized to support the acquisition of many levels of linguistic structure, from sounds (e.g., Maye et al., 2002), to words (e.g., Saffran et al., 1996; Graf Estes et al., 2007), to syntax (e.g., Gomez, 2002; Mintz, 2003; see recent reviews by Romberg and Saffran, 2010; Thiessen et al., in press). The experimental evidence leaves little doubt that infants can detect statistical regularities in linguistic input. However, there is much less evidence regarding the degree to which the mechanisms at work in statistical learning experiments can contribute to development. A crucial question remains: is statistical learning *useful* for language acquisition? It is not yet clear whether the representations that emerge from statistical learning possess the characteristics that are necessary to support language acquisition and processing.

Effective language processing requires that representations of words be appropriately abstract. They must not be limited to the specific perceptual details of a given word token. Rather, phonological representations must be flexible and generalizable across variation in how words sound because each token of a word varies based on characteristics such as the speaker's vocal tract, articulatory patterns, accent, speaking rate, and speaking register, as well as the surrounding words and prosodic patterns of the utterance (e.g., Peterson and Barney, 1952; see also reviews in K. Johnson, 2008; Luce and McLennan, 2008; Nygaard, 2008). This presents a significant challenge to young language learners who do not yet know which acoustic variations signify meaningful differences between words and which do not. The ubiquitous variation in speech also presents a challenge to statistical learning accounts of language acquisition. Recognizing sound sequences across acoustically distinct tokens is necessary in statistical learning. In order to track distributional information, infants must detect when the same phonemes, syllables, and/or words occur in different utterances. In addition, to take advantage of prior statistical learning, infants must identify previously discovered patterns when they occur in different contexts or voices. Generalizing from statistical learning experience is crucial for infants to build future learning from prior learning.

The present experiments investigate infants' ability to generalize statistical learning experience by examining statistical word segmentation, the process of using statistical cues to detect words in fluent speech. Infants were given the opportunity to segment words from a continuous speech stream based on patterns of syllable co-occurrences (i.e., transitional probabilities). Testing then

probed whether representations of statistically segmented words are robust to the challenges presented by acoustic variation.

For adults, word recognition is quite resilient to variations in the surface form characteristics of words, which are acoustic variations that do not signal differences in word meaning, such as voice, affect, and accent. These characteristics are encoded during speech processing, but adults adapt quickly (reviewed in Johnson, 2008; Luce and McLennan, 2008; Nygaard, 2008). However, recognizing words across surface form changes is difficult for infants. Houston and Jusczyk (2000) found that 7.5-month-olds failed to recognize words embedded in native language (English) passages of continuous speech when the voice during familiarization differed in gender from the voice used in testing. Infants successfully detected the words when the gender of the voice was consistent. Singh et al. (2004) reported that 7.5-month-olds failed to recognize words across variation in the speaker's affect. For example, infants familiarized with words in a happy voice recognized them when they were embedded in passages produced with happy affect, but not with neutral affect. Singh et al. (2008b) found that changes in voice pitch (but not amplitude) had a similar effect. Bortfeld and Morgan (2010) also reported that 7.5-month-olds have difficulty detecting familiarized words in passages when stress characteristics of the words change (i.e., from emphatic to non-emphatic stress, or vice versa) between familiarization and testing. These studies indicate that early native language word recognition is inhibited by many acoustic variations that are irrelevant to lexical identity, variations that would have little effect on mature speech processing.

Several factors influence infants' ability to generalize lexical representations across surface form variation. One important factor is the type of experience that infants have had with the words. When infants hear variable word tokens during familiarization, even 7.5-month-olds can detect those words in sentences across surface form changes (Houston, 2000; Singh, 2008). Infants' prior word knowledge also matters. Singh et al. (2008a) showed that young infants recognize words across changes in voice pitch if the words are highly familiar items like *Mommy* and *Daddy*, but not when words are unfamiliar. There are also developmental changes in the resilience of infant word recognition, so that by 10.5 months of age, infants can recognize words across changes in voice, pitch, and affective styles (Houston and Jusczyk, 2000; Singh et al., 2004, 2008b; see also Schmale and Seidl, 2009; Schmale et al., 2010 for effects of accent on infant word recognition). This increased sophistication is likely tied to infants' accumulation of varied experiences and increased word knowledge. By the end of the first year of life, infants' ability to recognize native language words expands; they are no longer misled by many surface form variations. This expansion occurs at around the same age that infants' speech perception narrows to focus on sound categories that are meaningfully distinct in their native language (e.g., Werker and Tees, 1984; Werker and Lalonde, 1988; reviewed in Saffran et al., 2006).

The studies investigating how infants cope with surface form variation during word recognition highlight a crucial process in language acquisition. To recognize words, infants must develop lexical representations that are abstract and flexible. They must attend to differences that make meaningful distinctions between words and generalize across irrelevant surface form variants. However, studies of the mechanisms that underlie word segmentation, such as statistical learning, have not explored the effects of acoustic variation.

Many statistical learning experiments present listeners with highly controlled speech streams, produced in a consistent voice throughout learning and testing (e.g., Aslin et al., 1998; Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003). Learners are not required to perform the acoustic generalizations that are necessary in natural language processing, so it remains unclear whether infants can generalize statistical learning experience. During statistical word segmentation, infants may form rigid representations that are constrained by the perceptual details of the input. This would suggest that statistical learning tasks measure lab-based mechanisms with little potential for the flexibility that language acquisition requires. Alternatively, infants may form representations of statistically defined words that are robust to acoustic variation. By the age that infants readily recognize native language words across changes in surface form (Houston and Jusczyk, 2000; Singh et al., 2004, 2008b), they may also readily recognize newly segmented words across variation. This finding would support the hypothesis that statistical learning can meet naturalistic language processing challenges. If statistical learning is a viable contributor to language acquisition, learners must form generalizable representations of the units they extract.

The present experiments investigate whether infants can generalize the representations that emerge during statistical learning. Across two experiments, infants heard the same statistical word segmentation experience. However, two different age groups were tested, 11- and 17-month-olds, with distinct methods designed to tap key learning processes occurring at each age.

During the first year of life, infants' ability to detect words in fluent speech develops substantially (e.g., Jusczyk, 1997). Therefore, Experiment 1 examined generalization in a traditional statistical word segmentation task with 11-month-olds. During the segmentation phase, infants listened to an artificial language in which the only reliable word boundary cue was transitional probability information. Transitional probability is a conditional probability statistic that indicates the predictive association between two elements. It is calculated based on the frequency of occurrence of a sequence *XY* divided by the frequency of *X* alone. When the sequence *XY* occurs reliably (as occurs within words), transitional probably is high, but when the sequence is inconsistent (as occurs across word boundaries), transitional probability is low. The artificial language exaggerated the pattern that occurs in natural languages (Harris, 1955): within words, syllable co-occurred consistently (i.e., perfect transitional probability); across word boundaries, transitional probability was substantially lower. Similar to prior statistical learning experiments (e.g., Saffran et al., 1996; Aslin et al., 1998), to demonstrate successful learning, infants must discriminate between the high probability words from the language and the low probability sequences that crossed word boundaries, termed part-words. In the present experiment, infants were required to generalize beyond the perceptual details of the segmentation speech stream. Specifically, the infants must segment the words from a language produced by a female voice, then recognize the words in a male voice during testing. If infants form generalizable representations, they should recognize the statistically defined words when they are presented in a new, acoustically distinct voice.

During the second year, a major developmental task is for infants to associate the sounds of words with their meanings. Therefore Experiment 2 tested 17-month-olds in a statistical word segmentation task integrated with a word learning task. Infants listened to an artificial language segmentation phase followed by a label-object association task. Integrating word segmentation and word learning presents an opportunity to investigate the nature of the representations that infants form during statistical learning. It is possible to examine how infants use the units that they discover. In a previous study employing this method, Graf Estes et al. (2007) found that infants took advantage of prior statistical learning to associate novel objects with their labels. They readily learned high probability words from the artificial language as object labels, but failed to learn low probability part-words as labels. Graf Estes et al. proposed that during statistical learning infants form candidate words that are ready to be associated with meanings.

In Graf Estes et al.'s (2007) study, the same female voice presented the segmentation phase and the object labels. Thus, it is not clear whether infants' representations of candidate words possess the flexibility necessary to facilitate word learning when surface form characteristics change. To investigate this process, the segmentation phase in Experiment 2 was presented in a female voice, but the labels were presented in a male voice. For one group of infants, the object labels were words from the language that the infants had prior opportunity to segment. Alternatively, the labels were part-word sequences that spanned word boundaries in the language (Experiment 2A). If statistical segmentation yields generalizable word like representations, these units should subsequently be available to support lexical functions, such as labeling objects. A follow-up experiment also tested infants' learning of the labels with no segmentation phase and therefore no prior exposure to the sequences (Experiment 2B).

The variation inherent to speech presents a substantial challenge to learning that learning theories must explain. The present experiments explore whether infants' statistical learning can meet this challenge. They present two approaches to investigating the abstractness of statistical learning. Experiment 1 tested whether during word segmentation, infants form generalizable acoustic representations of the units they detect. Experiment 2 addressed the underlying representations of statistically defined words, examining whether infants extract and store flexible word like representations that support learning of new object labels.

EXPERIMENT 1

Experiment 1 examined whether infants form generalizable representations during statistical word segmentation. In the *inconsistent voice condition*, infants listened to an artificial language produced in a female voice during the segmentation phase of the task. During the test phase, a male voice produced the test items. In the *consistent voice condition*, the segmentation phase was identical to the inconsistent voice condition. However, the test items were produced by the same female voice as infants heard during segmentation. The purpose of the consistent voice condition was to establish 11-month-olds' learning pattern for these stimuli when the voice is consistent from segmentation to testing. If infants learn the structure of the artificial language, they should show a difference in listening time between the low transitional probability part-words versus the high transitional probability words. In statistical learning experiments, infants typically display a novelty preference for the part-words (e.g., Saffran et al., 1996; Aslin et al., 1998).

MATERIALS AND METHODS

Participants

Fifty-six infants were randomly assigned to the consistent and inconsistent voice conditions (28 infants per condition; 35 males and 21 females). The average age was 11.1 months (SD = 0.23; range 10.2–11.5 months). The infants were born full term and were free of vision and hearing problems, according to parental report. The infants all came from homes in which English was the predominant language spoken. Based on parental interviews, 15 of the infants had some exposure to a second language, 20 h per week or less ($n = 5$ in the consistent voice condition, $n = 10$ in the inconsistent voice condition). The results of the experiment are unchanged if the infants with second language exposure are excluded from the analyses. In the consistent voice condition, two additional infants were identified as outliers based on listening time differences to words versus part-words that were over 2.5 SD from the mean. These infants were excluded from analyses. An additional 17 infants were excluded because of fussiness ($n = 8$ in the consistent voice and $n = 9$ in the inconsistent voice conditions). The University of California, Davis Institutional Review Board approved the research protocol for Experiments 1 and 2. The parents of our participants gave informed consent.

Stimuli

The artificial language used in the segmentation phase was originally developed by Graf Estes et al. (2007). To control for infants' arbitrary listening preferences, there were two counterbalanced versions of the artificial language. The words in Language 1 were *timay, dobu, gapi,* and *moku;* the words in Language 2 were *pimo, kuga, buti,* and *maydo.* As shown in **Table 1**, the counterbalancing resulted in syllable sequences that acted as word test items in Language 1 and part-word test items in Language 2, and vice versa. The artificial language was recorded using a method that approximates the actions of a speech synthesizer. A female speaker recorded 3-syllable sequences, of which the medial syllables were excised and spliced to form the final speech stream (i.e., the recorded sequences *timaydo*, *maydobu*, *dobuga* were spliced to form the sequence *maydobu*). Recording 3-syllable sequences allowed for natural coarticulation of each syllable. Splicing the medial syllables to form a fluent sequence reduced the chance for the speaker to inadvertently introduce additional word boundary indicators. The speech stream contained no pauses or other reliable acoustic cues to word boundaries. The only reliable word boundary cues were the transitional probabilities of syllable sequences. The within-word

Table 1 | Word and part-word test items for Experiments 1 and 2.

	Words	Part-words
Language 1	*timay, dobu*	*pimo, kuga*
Language 2	*pimo, kuga*	*timay, dobu*

transitional probabilities were 1.0 (i.e., the syllables within each word always occurred together) and the across word probabilities ranged from 0 to 0.5. The duration of each speech stream was 5.5 min.

The artificial language was designed to equate the frequency of the word and part-word test items, but maintain the difference in their transitional probabilities. Using this design, it is possible to determine whether infants discriminate words from sequences that occur with equal frequency in the artificial language, but differ in their internal statistical structure (Aslin et al., 1998). To balance the frequency of the test items, the language contained two high frequency words that occurred 180 times in the speech stream (Language 1: *gapi* and *moku*; Language 2: *buti* and *maydo*) and two low frequency words that occurred 90 times (Language 1: *timay* and *dobu*, Language 2: *pimo* and *kuga*). This design yielded two part-words that occurred 90 times in the speech stream, occurring at the conjunction of the two high frequency words. For example, in Language 1, *gapi* preceded *moku* 90 times. Therefore, the part-word sequence *pimo* occurred the same number of times as the low frequency words (e.g., *timay*). The test items were the low frequency words and the part-words formed from the high frequency words (see **Table 1**). All occurred 90 times during the segmentation phase. However, the words had perfect transitional probability (transitional probability = 1.0) and the part-word sequences contained a dip in transitional probability between syllables (transitional probability = 0.5).

In the consistent voice condition, the same female speaker recorded the artificial language and the test items. In the inconsistent voice condition, a male speaker recorded the test items. The average fundamental frequency (*F*0, a measure of pitch) of the male voice test items was 121 Hz, which was substantially lower than the fundamental frequency of the artificial language (224 Hz) and the female voice test items (234 Hz). The test items were recorded in citation form, with a monotone speaking style in order to maintain similarity with the speech from the segmentation phase. Repetitions of the test items were separated by 750 ms of silence. All sounds were played at a level approximating conversational speech, around 65 dB.

Procedure

During the segmentation phase, each infant and his or her parent were allowed to move around a sound attenuated booth while playing quietly. The parent was instructed not to refer to the artificial language and to remain as quiet as possible. Following the segmentation phase, the parent and child were moved to a second sound attenuated booth. In the test booth, a television at the front of the room displayed visual animations and attention-getting stimuli and broadcast the sound sequences. The infant sat on the parent's lap approximately 1 m from the screen. A camera mounted below the television screen enabled the observer, located outside the booth, to monitor looking behavior. When the parent and child entered the test booth, the parent heard a brief reminder about the instructions for the test phase of the experiment. Because of this delay, the infant received a 30 s refamiliarization with the artificial language before testing. The refamiliarization was paired with a silent cartoon clip to maintain the infant's interest.

The program Habit X (Cohen et al., 2004) was used to present infants with the test items in an auditory preference procedure. As a protection against bias, the experimenter was blind to the identity of the materials being presented, and the parent listened to masking music over headphones. Test trials immediately followed the refamiliarization. Each trial consisted of repetitions of a word test item or a part-word test item. There were 16 test trials. The four test items (two words, two part-words) were presented in four randomized blocks.

To measure infants' listening time to the auditory test items, all items were paired with a visual animation of an orange oval turning in a circle on the screen. The presentation of the test trials was contingent on the infant's looking at the visual animation. Using a button press, the experimenter indicated how long the infant's attention remained fixated on the audio-visual item. The test item repeated until the infant looked away for 1 s or after a maximum listening time of 20 s. To regain the infant's interest, a cartoon played between trials.

The program Habit X tallied listening time to each test item. The dependent measure was based on listening time (indicated by attention to the audio-visual stimuli) to the word and part-word test items. The measure of listening time used here is similar to the central fixation procedure used by Shi and Werker (2001; Shi et al., 2006) and the visual fixation-based auditory preference procedure used by Cooper and Aslin (1990, 1994). It is also similar to the head turn preference procedure frequently used in statistical learning experiments (Saffran et al., 1996; Aslin et al., 1998; Johnson and Jusczyk, 2001).

RESULTS AND DISCUSSION

Preliminary analyses revealed that there were no differences in performance based on sex or artificial language version (Language 1 versus Language 2). Therefore, subsequent analyses collapsed across these variables.

Infants' learning was analyzed in a 2 (Condition: consistent voice vs. inconsistent voice; between subjects) × 2 (Trial type: word vs. part-word; within subjects) mixed ANOVA. There was no main effect of condition and no main effect of trial type, F's < 1. There was a significant interaction of condition by trial type, $F(1, 54) = 12.4$, $p = 0.001$, $\eta_p^2 = 0.19$ To explore the interaction, each condition was analyzed separately with a paired samples t-test comparing looking time to word versus part-word test trials. In the consistent voice condition, infants listened significantly longer to the part-words, $t(27) = 2.56$, $p = 0.016$, $d = 0.31$. In the inconsistent voice condition, infants listened significantly longer to the words, $t(27) = -2.41$, $p = 0.023$, $d = 0.25$. Listening time performance is illustrated in **Figure 1**. In the consistent voice condition, 20 of 28 infants showed the novelty preference for part-words. In the inconsistent voice condition, 17 of 28 infants showed the familiarity preference for words.

In both conditions, infants discriminated the word versus part-word test items, indicating that they learned the structure of the artificial language, and recognized the words from the speech stream. However, the infants showed different directions of preference. The part-word preference in the consistent voice condition follows the pattern of many statistical learning experiments (Saffran et al., 1996; Aslin et al., 1998; Johnson and Jusczyk, 2001;

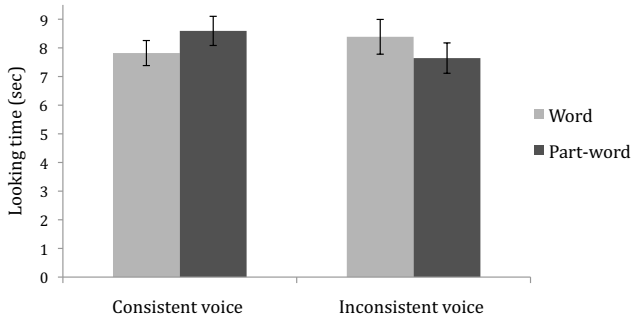


FIGURE 1 | Mean looking time (in seconds) to word versus part-word test trials. Error bars represent standard errors.

Thiessen et al., 2005; Experiment 2) and is typically interpreted as a novelty preference for the items that were not previously detected in the segmentation phase. The preference for word test items has been demonstrated in some experiments (Saffran, 2001; Thiessen et al., 2005, Experiment 1; Thiessen and Saffran, 2003, Experiment 1). According to Hunter and Ames's (1988) model of infants' attentional preferences, infants display novelty preferences when information has been thoroughly processed (see also Houston-Price and Nakai, 2004). Infants are likely to display familiarity preferences when a task is difficult. One characteristic that affects task difficulty is the match between the familiarization stimuli and test items (Hunter and Ames, 1988; Thiessen and Saffran, 2003). When test items are similar to the familiarization stimuli, the task is easier than when the test items differ from familiarization. The novelty preference displayed in the consistent voice condition and the familiarity preference displayed in the inconsistent voice condition suggest that recognizing the words in the familiar voice was easier for infants than recognizing the words in the novel voice.

In Experiment 1, 11-month-olds performed a linguistically relevant generalization across acoustic variation in a statistical learning task. The infants' representations of the statistically segmented word forms were sufficiently abstract to recognize the words when they were produced in a novel, acoustically distinct voice during testing. This is very close to the age at which infants readily recognize native language words across changes in affect (Singh et al., 2004) and speaker's voice (Houston and Jusczyk, 2000), 10.5 months. The similar age across experiments highlights the notion that infants are processing speech in a similar way when it is produced in their native language or an artificial language. In addition, our findings are consistent with a recent experiment by Vouloumanos et al. (2012), who found that adults readily identify statistically defined words across a change in voice. For highly experienced adult language processors, performance was not different when recognizing the words in the same voice or a different voice. For infants, the change in direction of preference suggests that generalizing across voices is more difficult than recognizing words when the voice is consistent. Yet the infants' representations of statistically segmented units are not limited by the perceptual details of their learning experience.

EXPERIMENT 2

Experiment 1 demonstrated that statistically segmented units are robust to surface form variation. Such generalization is necessary for recognizing words and accumulating information about the meanings and uses of words. However, the listening time measure used in Experiment 1, and in many other statistical learning experiments, is limited in what it can reveal about the representations that infants form during statistical learning. Listening preference measures are highly valuable tools. Infants' discrimination of high and low transitional probability sequences demonstrates that they are powerful learners, able to rapidly detect structure in linguistic input based on limited information. But infants' discrimination performance alone cannot tell us whether the representations formed during statistical learning are mere sounds, or whether they have any linguistic status (Saffran, 2001). To directly explore the nature of the representations that infants form during statistical learning, it is necessary to design tasks that test how infants apply the output of statistical learning to other linguistic processes. If the output of statistical word segmentation is word like units, infants should be able to use those units to perform the kinds of tasks that real words perform.

To address this issue, Graf Estes et al. (2007) designed a task that integrates statistical word segmentation with word learning. Infants first participated in a segmentation phase during which they heard an artificial language. The segmentation phase was immediately followed by a label-object association task, rather than a listening preference measure. The same (female) voice presented the segmentation phase and labeling task. The label-object association task presented a simplified word learning event (Werker et al., 1998). Infants habituated to two label-object pairs. After habituation, infants' learning was measured by the duration of their looking time on test trials in which they viewed the original label-object pairs or trials in which the original associations were violated. If infants have learned the labels, they should look longer on the trials in which the learned pairings were violated.

Using this method, Graf Estes et al. found that 17-month-olds readily learned statistically defined words as object labels. However, infants failed to learn labels that were part-words or non-words (novel sequences of syllables from the language). As in Experiment 1, the word and part-word test items occurred with equal frequency during the segmentation phase. Therefore, before they occurred as labels infants heard the words and part-words equally often, but the items differed in their internal transitional probabilities. Graf Estes et al. (2007) concluded that transitional probability information was weighted more heavily than frequency information in determining whether a sound sequence was a good potential object label. The findings also indicate that infants can use statistical learning to extract candidate words that are then available to be associated with meanings (for related findings with adults see Mirman et al., 2008 and Endress and Mehler, 2009 for a counterargument).

Experiment 2 used the method designed by Graf Estes et al. (2007) to examine infants' ability to use the output of statistical learning in a word learning task when infants must generalize across acoustic variability in order to do so. The participants were 17-month-olds because at this age, the process of associating

sounds with meanings is a major focus of language acquisition. This age group also allows for a direct comparison with previous experiments examining the connection between statistical word segmentation and word learning (Graf Estes et al., 2007; Hay et al., 2011).

The stimuli in Experiment 2 came from the inconsistent voice condition of Experiment 1. The segmentation phase was presented in a female voice and the label-object associations were presented in a male voice. For half of the infants, the labels were words from the artificial language. For the other half of the infants, the labels were part-words. If infants form generalizable representations of candidate words, Experiment 2 should replicate the findings from Graf Estes et al. (2007) when the voice changes from segmentation to label learning. Statistical word segmentation should support infants' learning of novel object labels when the labels are newly segmented words, but not when the labels are part-word sequences.

EXPERIMENT 2A

MATERIALS AND METHODS

Participants

Forty-four infants were randomly assigned to the word and part-word label conditions (22 infants per condition; 22 males, 22 females). The average age of the participants was 17.3 months (SD = 0.34; range 16.6–17.8 months). All infants were born full term and had no history of hearing or vision impairments. Based on parental interviews, eight infants had some exposure to a second language, 20 h per week or less ($n = 5$ in the word condition and $n = 3$ in the part-word condition). The results of the experiment are unchanged if infants with second language exposure are excluded from the analyses. Twenty-three additional infants were excluded because of fussiness ($n = 19$), moving out of the video frame ($n = 3$), and experimenter error ($n = 1$). In the part-word condition, one additional infant was identified as an outlier based on a looking time difference to same versus switch test trials that was greater than 2.5 SD from the mean. The infant was excluded from the analyses.

Stimuli

Word Segmentation Task. The artificial language was the same as the language used in Experiment 1. It was presented in a female voice. The test items were identical to the word and part-word sequences presented in the inconsistent voice condition (male voice) of Experiment 1.

Object Labeling Task. The novel objects, shown in **Figure 2**, were two computerized 3-D images designed to be visually complex and discriminable in shape and color. Each object was paired with an object label. For all infants, the labels were presented in a male voice. For half of the infants, the object labels were words from the artificial language (e.g., *timay* in Language 1). For the other half of the infants the object labels were part-words (e.g., *kuga* for Language 1). Because of the artificial language design (see Experiment 1), the word and part-word labels occurred equally frequently during the segmentation phase, but differed in their internal transitional probabilities.

Each infant participated in one of four testing conditions: half of the infants exposed to Language 1 received two word test items, and half received two part-word test items. Half the infants exposed to Language 2 received two word test items, and half received two part-word test items. The test items are shown in **Table 1**.

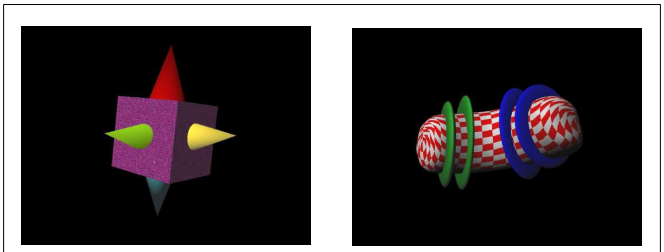


FIGURE 2 | Novel objects that received labels.

Procedure

The method for presenting the artificial language in the word segmentation phase was identical to the method described in Experiment 1. The infant listened to the language in a sound attenuated booth and heard a 30 s refamiliarization after being moved to the testing booth. Instead of measuring infants' discrimination of word and part-word test items, the infants immediately participated in a label-object association task. A version of the Switch task was used to test infants' learning of label-object pairings (Werker et al., 1998). It is a popular measure of early word learning with low task demands. Although the Switch task lacks the social referential context that is present in interactive word learning tasks, it retains a fundamental component of the word learning process – linking a sound sequence representation with a meaning representation (here, object identity). The measure has been used recently in several studies to investigate factors affecting early word learning (Fennell et al., 2007; Curtin, 2009; Rost and McMurray, 2009).

The program Habit X was used to present the label-object combinations in the Switch task. As a protection against bias, the experimenter was blind to the identity of the materials being presented, and the parent listened to masking music over headphones. The infant started the task with a familiarization trial that allowed the infant to become accustomed to the audio-visual stimuli presentation before the first habituation trial. The infant viewed a rotating gray screen presented on a black background accompanied by repetitions of the syllable "neem."

During the habituation phase, the infants viewed two label-object combinations. Each label-object combination was presented one at a time, with the order randomized by blocks. The object moved from side to side while its associated label played. Each label repetition was separated by 750 ms of silence. Presentation of the stimulus continued as long as the infant remained fixated on it. Trials terminated when the infant looked away for 1 s, or for a maximum of 20 s. A cartoon played between trials to guide the infant's attention back to the screen. The habituation criterion was satisfied when the infant's average looking time on three consecutive trials decreased to 50% of the average looking time on the first three habituation trials.

Test trials began immediately after the infant reached the habituation criterion or viewed a maximum of 25 habituation trials.

There were two types of test trials: on *same* test trials, the original label-object associations from habituation were maintained. On *switch* test trials, the label-object pairings were violated (e.g., object 1 was presented with label 2). There were four same and four switch test trials, organized in two counterbalanced testing orders. In both orders, the switch test trials occurred first, which provides infants with the best opportunity to display learning in case infants' attention wanes throughout testing. These test orders replicate the orders that Graf Estes et al. (2007) used. When the label voice matched the segmentation voice, they found that infants learned the word object labels, but not the part-word labels. Thus, although the test orders give infants the strongest chance to display learning, it is possible for infants to fail to display learning of the labels using test orders in which switch trials are presented first (see also Experiment 2B).

RESULTS AND DISCUSSION

Preliminary analyses revealed no significant differences in performance based on sex or language version (Language 1 versus 2). Therefore, subsequent analyses collapsed across these variables.

In the word label condition, infants reached the habituation criterion in a mean of 11.5 trials (SD = 5.8). In the part-word condition, infants reached the habituation criterion in a mean of 11.2 trials (SD = 5.3). There was no significant difference in the number of trials to reach habituation, $t(42) = 0.163$, $p = 0.872$, $d = 0.05$. One infant in the word label group and one in the part-word label group failed to habituate. The results of the analyses are unchanged if these infants are excluded.

Infants' learning was analyzed in a 2 (Label condition: word versus part-word; between subjects) × 2 (Trial type: same versus switch; within subjects) mixed ANOVA. There was no main effect of label condition, $F(1, 42) = 1.64$, $p = 0.207$, $\eta_p^2 = 0.04$ and no interaction of label condition by trial type, $F < 1$. There was a main effect of trial type, $F(1, 42) = 13.46$, $p = 0.001$, $\eta_p^2 = 0.24$. Follow-up paired samples t-tests confirmed that infants in the word label condition showed significantly longer looking on the switch test trials, $t(21) = 2.47$, $p = 0.022$, $d = 0.64$. Infants in the part-word label condition showed the same pattern, $t(21) = 2.94$, $p = 0.008$, $d = 0.69$. Fifteen of 22 infants in the word label condition and 17 of 22 infants in the part-word label condition looked longer on the switch test trials than the same trials. Looking time is illustrated in **Figure 3**.

The analyses of the same and switch test trials indicate that infants learned both the word and part-word labels; they detected when the label-object pairings were switched. In a previous experiment using a consistent voice, but the same task and test orders, infants who heard word labels showed significantly longer looking on switch trials, but infants who heard part-word labels did not (Graf Estes et al., 2007). However, it is theoretically possible that the difference in looking time to the same and switch trials occurred here because the test phase began with switch trials and a general decline in attention produced the effect. If the present findings occurred because of declining attention, looking time should also decline from habituation to the first block of test trials. In contrast, if infants learned the label-object pairings during habituation, they should dishabituate to the first switch trials even though the trials occurred later in the experiment. Similar to the analyses above, a 2 (Label condition: word vs. part-word) × 2 (Trial type: habituation versus first two switch trials) ANOVA was performed (Two infants who did not habituate were excluded, but the pattern is the same with these infants included.). **Figure 4** shows that across the word and part-word label conditions, infants dishabituated to the switch trials (main effect of trial type: $F(1, 40) = 8.3$, $p = 0.006$; no effect of

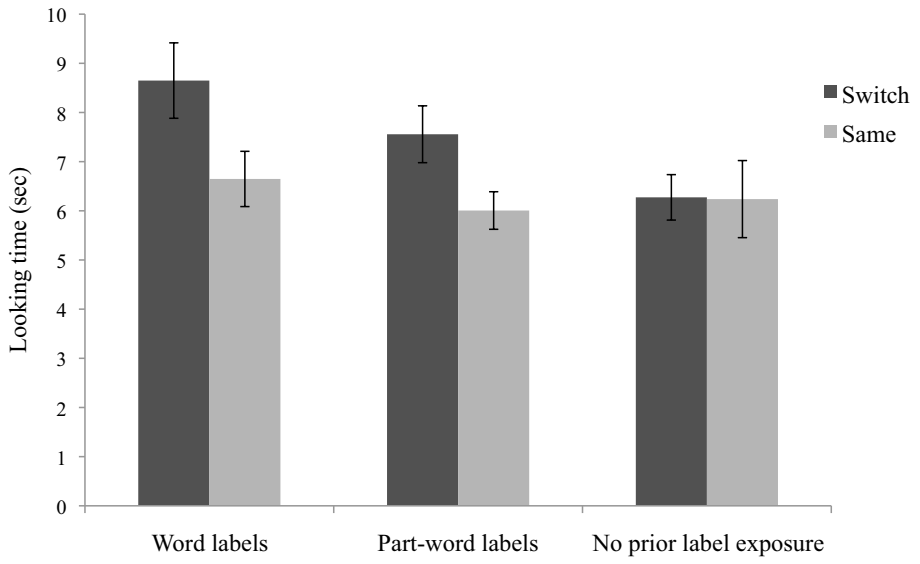


FIGURE 3 | Mean looking time (in seconds) to word and part-word labels (Experiment 2A), and labels with no prior segmentation phase exposure (Experiment 2B) to the same and switch test trials. Error bars represent standard errors.

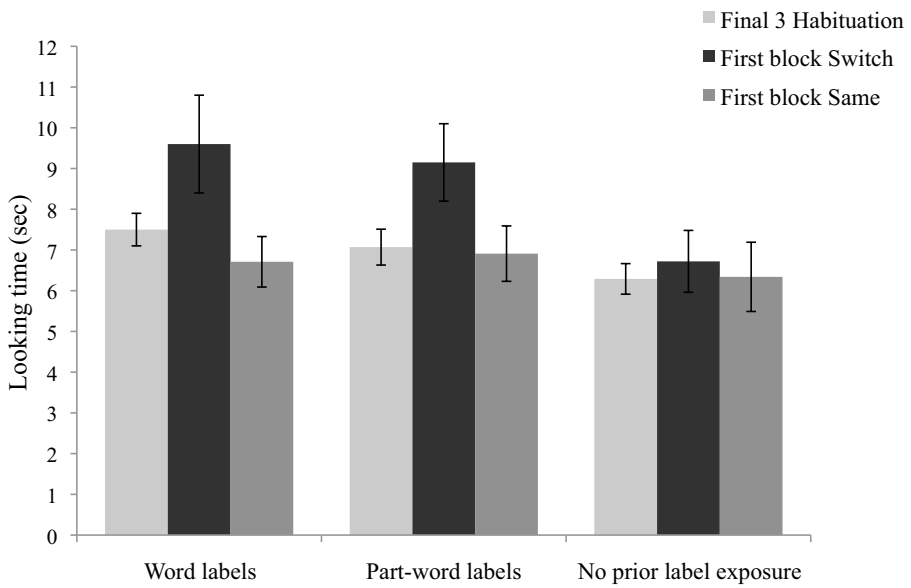


FIGURE 4 | Mean looking time (in seconds) to word and part-word labels (Experiment 2A), and labels with no prior segmentation phase exposure (Experiment 2B) during the final three habituation trials and the first block of switch and same test trials. Error bars represent standard errors.

label condition and no interaction, p's > 0.63). The analysis was repeated for the first block of same test trials versus habituation trials. There were no main effects of trial type or label condition, and no interaction (all p's > 0.35), indicating that looking time between habituation and the first same trials did not differ. This analysis suggests that infants looked longer on switch trials than same trials during testing because they detected that switch test trials differed from the label-object pairings shown during habituation.

EXPERIMENT 2B

Infants displayed evidence of learning both the word and part-word labels in Experiment 2A. This conflicts with previous evidence that infants learn word, but not part-word labels when the same voice presents the segmentation phase and the object labels (Graf Estes et al., 2007). Given the design of Experiment 2A, it is possible that infants learned more effectively from the male test voice than the female test voice that Graf Estes et al. (2007) used. The difference in performance could be unrelated to infants' statistical segmentation experience. Graf Estes et al. (2007) and Graf Estes and Hurley (in press) also reported that infants failed to learn these labels when they were presented in a monotone or adult-directed female voice with no segmentation phase. Experiment 2B tested whether infants readily learned the labels when they were presented in a male voice, without any prior segmentation experience. Infants only participated in the label-object association task, which should minimize any effects of fatigue during testing, thereby giving infants the best opportunity to display learning. If the male voice labels are simply easy to learn on their own, infants should look longer on the switch test trials than the same trials. However, if exposure to the speech stream before label exposure is important, infants should have difficulty learning the labels.

MATERIALS AND METHODS

Participants

Twenty-two infants participated in this task (10 females, 12 males). The average age was 17.1 months (SD = 0.31; range 16.7–17.9 months). Seven infants had some exposure to a second language. The results of this experiment are unchanged with these participants excluded. An additional seven infants were excluded because of fussiness ($n = 5$), moving out of the camera view ($n = 1$), or equipment or experimenter error ($n = 1$).

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 2A, except that infants did not participate in the segmentation phase. They went directly to the test booth and participated in the label-object association task. Infants were randomly assigned to hear the labels *timay* and *dobu* or *gapi* and *moku*.

RESULTS AND DISCUSSION

Preliminary analyses revealed no significant differences in performance based on sex or labels versions (*timay* and *dobu* vs. *gapi* and *moku*). Therefore, subsequent analyses collapsed across these variables.

Infants reached the habituation criterion in a mean of 10.2 trials (SD = 4.5). All infants met the habituation criterion. A paired samples t-test revealed that there was no difference in looking time on same versus switch test trials, $t(21) = 0.051$, $p = 0.960$, $d = 0.01$. Thirteen of 22 infants showed a switch test trial preference. There is no evidence that infants learned the labels in the absence of the opportunity to segment them from fluent speech

before they occurred as labels. They failed to display learning even though conditions were designed to minimize fatigue effects.

In contrast to Experiment 2A, **Figure 4** shows that infants in Experiment 2B did not look longer during the first block of switch trials compared to the final habituation trials. There was also no difference in looking time between the first block of same trials and the final habituation trials (p's > 0.61). This result further supports the argument that infants' differential attention to same and switch test trials for the word and part-word labels was not merely due to the test orders combined with a general decline in attention throughout testing.

Across Experiments 2A and 2B, infants learned the statistically defined words and part-words as labels, but failed to learn the same labels in the absence of prior exposure. Infants transferred statistical segmentation experience to support object label learning when it required generalizing beyond the acoustic characteristics of their input. The output of statistical learning is not bound by the perceptual details of the original familiarization stimuli. Rather, infants can perform this naturalistic generalization in service of a real language acquisition task, associating the sounds of words with their referents.

The results also show that infants form and store representations of sequences that are not word units, but rather occur across word boundaries. Although the part-word test items had low transitional probability relative to the words, several characteristics may have facilitated their use in the label learning task. Infants had ample opportunity to hear the part-words before they appeared as object labels; they occurred in segmentation phase 90 times across 5.5 min. In addition, the transitional probability of the part-word labels was 0.5, whereas other word boundary probabilities ranged from 0 to 0.26. These lower transitional probability sequences may have produced clearer word boundaries than the across word sequences that occurred as labels. In natural languages, word-internal transitional probabilities are rarely perfect. Some real words may contain probabilities closer to the 0.5 value of the part-words than the 1.0 value of the words examined here. Based on their frequency and transitional probability patterns, the part-words may have formed relatively coherent sequences, available to support label learning. However, this rationale and the present results conflict with Graf Estes et al.'s (2007) findings when the voice presenting the object labels matched the voice during segmentation. The General Discussion proposes an explanation for the divergent results. Nonetheless, the present findings demonstrate that infants form and retain representation of the sequences that cross word boundaries in addition to representations of the coherent, high transitional probability word units.

GENERAL DISCUSSION

In this series of experiments, infants participated in tasks tailored to investigate two different stages of language acquisition. Experiment 1 examined statistical word segmentation in 11-month-olds and found that infants can recognize statistically segmented words across variation in a speaker's voice. Experiment 2 examined whether 17-month-olds can generalize the output of statistical word segmentation across variation to support object label learning. The infants were successful; they associated referents with statistically defined words, as well as with frequently occurring sequences that spanned word boundaries in the speech stream. Across Experiments 1 and 2, infants heard the same stimuli, but testing tapped different language acquisition processes. Each experiment has an independent contribution to understanding the representations infants form during statistical learning. In addition, combining the methods of testing word segmentation and label learning following segmentation has revealed characteristics of learning that would not have been apparent from either experiment alone (see also Pelucchi et al., 2009; Hay et al., 2011).

In Experiment 1, infants' discrimination performance showed that they could recognize the statistically segmented words across a change in voice. Similar to many previous statistical word segmentation experiments, infants presented with a consistent voice attended longer to novel part-words than to words (Saffran et al., 1996; Aslin et al., 1998; Johnson and Jusczyk, 2001; Thiessen et al., 2005; Experiment 2). In contrast, infants who heard an inconsistent voice across segmentation and testing showed a familiarity preference for the words. Models of infants' attentional preferences explain that when a task is relatively easy, or infants have become highly familiar with the training stimuli, novel stimuli elicit greater attention than familiar stimuli. When a task is difficult, there is a greater likelihood that infants will demonstrate a familiarity preference for patterns that are consistent with their training stimuli (Hunter and Ames, 1988). A mismatch between familiarization and test (such as the change in voice in the inconsistent voice condition) is one characteristic that can make a task difficult. Thus, a conclusion from the segmentation task in Experiment 1 is that infants can generalize across statistical segmentation experience, but it is more difficult than recognizing words when the voice is consistent. The label learning measure in Experiment 2 did not reveal this difference in the ease of processing.

Around 11 months of age, infants can recognize native language words across variation in characteristics such as affect, pitch, and voice (Houston and Jusczyk, 2000, 2003; Singh et al., 2004, 2008b). Thus, in Experiment 1, infants showed flexibility in word recognition in statistical learning at around the same age as in their native language. It is not yet clear whether the full developmental trajectory of word recognition across variation is similar in native language word segmentation and statistical word segmentation of artificial languages. It remains to be tested whether younger infants (e.g., 7.5-month-olds) have difficulty recognizing statistically segmented words across variation, as they do for native language words (Houston and Jusczyk, 2000; Singh et al., 2004, 2008a,b; Bortfeld and Morgan, 2010). In addition, future experiments will be necessary to explore the range of flexibility of infants' representations of statistically segmented words. Vouloumanos et al. (2012) found that adults' representations are abstract, but within limits. In a statistical learning task, adults recognized words across a change in the speaker's voice and across some types of distortion. While adult native language word recognition withstands many forms of unnatural variation, such as distortion (Remez et al., 1981; Pisoni, 1996; Saberi and Perrott, 1999), it greatly disrupts infant word recognition (Zangl and Mills, 2007). The effects of unnatural variation on recognizing segmented words may be stronger than the effects of natural variation because infants lack experience with experimentally manipulated unnatural variations (e.g., time reversals or low-pass filtering). By 11 months of age,

infants may succeed in recognizing statistically segmented words across the change in voice because native language experience leads them to expect that the same word can sound different depending on who says it.

Experiment 2 combined statistical word segmentation with a label learning task in order to capture a more nuanced picture of statistical learning than the segmentation task alone can provide. This integration yields an understanding of the linguistic status of the representations that infants form by showing how the output of statistical learning can be used to support word learning. In this case, it revealed an unexpected pattern. In contrast to previous findings (Graf Estes et al., 2007), 17-month-olds learned low transitional probability part-word sequences as object labels in addition to statistically coherent, high transitional probability words. Infants' learning of the part-word labels suggests that they develop and store representations of syllable sequences that cross word boundaries in fluent speech in addition to the sequences that form words.

It is not yet clear why part-word sequences support label learning when infants must generalize their statistical segmentation experience across voices, but not when the voice is consistent throughout segmentation and label learning. One possible explanation is motivated by models of word segmentation and memory (see Thiessen et al., in press, for a more thorough discussion of the integration of memory and statistical learning models). In Perruchet and Vinter's (1998) Parser model of word segmentation, one process that contributes to learners' extraction of word units is interference. In Parser, sequences, or chunks, that occur together frequently build up activation. The reliability of a chunk also contributes to its strength of activation. Chunks that occur frequently, but unreliably (like part-words) will not emerge as units because of interference from learning the reliably occurring, high probability units (words). Part-words consist of syllables that belong to the words, so knowledge of the words inhibits learners from segmenting out the part-word sequences (see also Giroux and Rey, 2009). This helps to frame the prior finding that infants learn word labels, but not part-word labels (Graf Estes et al., 2007).

To consider why the change in voice affects label learning, one must also consider memory models that posit that each experience with a word affects its stored lexical representation. In episodic memory models, each exemplar (e.g., each token of a word) is stored as a memory trace and exemplars accumulate over time. When a retrieval cue is presented and the stored exemplars overlap greatly with it (e.g., a word is repeatedly produced in a consistent voice), there is a stronger activation than when the retrieval cue is dissimilar from previous experience (e.g., a word produced in a new voice; Hintzman, 1986; Goldinger, 1996, 1998).

Integrating the episodic memory and word segmentation models suggests the following hypothesis. When the voice is consistent from segmentation to labeling, infants activate detailed representations of the segmentation speech stream because of the high overlap between the retrieval cue (i.e., the label) and prior experience. Infants' representations of the highly reliable and frequent words are strong; these units can act as object labels. The part-words, although frequent, conflict with the word representations and are therefore not stored as units available for further processing. However, when the voice changes from segmentation to labeling, the mismatch means that activation of prior learning is weaker. Building from the role of interference in Parser, the reduced activation caused by the change in voice could free infants from the inhibition caused by the conflicting representations of the words and part-words. This could then allow infants to use their experience hearing other frequently occurring syllable sequences, like part-words, to promote label learning. This hypothesis leads to the prediction that other conditions that produce weak activation of statistical learning, such as introducing a delay between segmentation and labeling, should reveal stored representations of part-words.

Further consideration of word segmentation models provides additional context for the findings from Experiment 2 and offers new predictions. Clustering and bracketing models present two broad categories of word segmentation strategies that have been explored (Goodsitt et al., 1993; Brent, 1999). Clustering (or chunking) models share the concept that tracking probabilistic information leads learners to extract sequences that occur reliably, yielding statistically coherent word like units (see various instantiations by Perruchet and Vinter, 1998; Swingley, 2005; Giroux and Rey, 2009; Frank et al., 2010). In contrast, bracketing (or boundary-finding) models propose that learners track the relations between elements and infer boundaries between them at points of low probability (e.g., Elman, 1990; Cairns et al., 1997; Christiansen et al., 1998). Learners do not extract cohesive units, but detect areas of low predictability. Evidence that infants readily associate meanings with statistically defined words supports clustering accounts. It suggests that infants extract and store candidate words that are available to feed other linguistic processes.

Clustering models also shed light on why the part-words acted as good object labels. Giroux and Rey (2009) explained that according to clustering models, increased experience with a speech stream should lead to stronger differentiation of items that are and are not words because learning about words should interfere with representations of other frequently occurring sequences. With sufficient experience, words will become the units that are available in memory, not part-word sequences, or sublexical sequences (i.e., syllable pairs within trisyllabic words). Accordingly, they found that after a brief exposure to an artificial language, adults did not differ in their ability to distinguish words and sublexical sequences from part-words. However, after a long exposure, participants identified words more accurately than sublexical units. In contrast, bracketing models predict that increased duration of exposure should not produce stronger differentiation of words and sublexical units because the exposure to and representation strength of words and sublexical units are tightly linked.

Giroux and Rey's (2009) account raises the possibility that infants in Experiment 2 were still learning about the frequency and reliability of the words in the language. The learning was not sufficiently complete to produce full inhibition of the part-word sequences, at least not when the sequences changed in voice from segmentation to labeling, thereby reducing interference from the word sequences. With greater exposure, the clustering account suggests that infants should show stronger differentiation between word and part-word labels, as well as word and sublexical sequences. Bracketing models would not predict this change.

There is an apparent contrast between infants' performance in the segmentation task alone (Experiment 1) and in the segmentation task followed by the label learning task (Experiment 2). In the segmentation task, infants differentiated the word and part-word test items, but in the label learning task they did not. It seems unlikely that the age difference across experiments, 11 months versus 17 months, produced the contrasting patterns of performance. Previous studies suggest that children do not lose the ability to perform statistical word segmentation (Saffran et al., 1997; Graf Estes et al., 2007). Rather, the different patterns of learning across experiments reveal that while infants can generalize representations of statistically segmented words, generalization depends on context. Differences in the demands and goals of each task may have encouraged infants to interpret the same stimuli in different ways. The auditory preference task from Experiment 1 is well-suited to measuring infants' ability to discriminate sound sequences. It presents a within subjects comparison of attention to each test trial type. Hearing the test items in close succession may promote infants' attention to the differences between them. The auditory preference task revealed a rapid learning and generalization capability, evidenced by infants' differentiation of items with high and low transitional probability. However, the preference task was not equipped to explore whether the items that infants perceive to be different also differ in their linguistic status (but see Saffran, 2001). Integrating the segmentation and label learning tasks can show whether infants form representations during statistical learning that feed forward to support label learning. However, the design of the label learning task is not well-suited to a direct comparison of the ease or strength of learning because infants hear only one label type (words or part-words). Infants cannot compare the high and low probability test items as they can in the auditory preference task. In addition, the Switch task does not typically indicate precise differences in the strength learning. Infants either show a significant difference in attention to same versus switch test trials or they do not. Thus, it is possible that words and part-words do not serve as equally good object labels, but more sensitive methods (e.g., Yoshida et al., 2009) will be necessary to reveal the difference. This possibility is currently being tested.

The present experiments highlight the importance of using multiple methodologies to investigate a construct. Experiments 1 and 2 examined two interrelated aspects of statistical learning: statistical word segmentation and the representational status of statistically segmented sequences. The combination of findings from these experiments show that generalization in statistical learning is affected by the demands of the problem that infants must solve. The experiments also illustrate limitations of the methods used in each experiment. The auditory preference measure yielded two different statistically significant directions of preference. Although there are precedents for both novelty and familiarity preferences in statistical learning tasks, making the same conclusions (i.e., successful learning) from opposite results can present interpretational challenges. In addition, as discussed above, auditory preference tasks can reveal that infants successfully discriminate sound sequences, but cannot specify the nature of those representations. Integrating statistical word segmentation and word learning, as in Experiment 2, takes a significant step toward understanding the output of statistical learning. It revealed that infants detect and store generalizable representations of words and cross word sequences that can serve as object labels. However, the Switch task is limited in its ability to detect fine-grained differences in learners' representations of novel word forms.

Advances in infant testing methodologies may help to address some of the limitations of these behavioral methods and present additional means of exploring questions about statistical word segmentation. Neurophysiological measures have potential to reveal characteristics of learning that may be masked by behavioral methodologies. Recent studies indicate that measures of brain activity, event-related potentials (ERPs), can provide more sensitive measures of infant word segmentation than listening time measures (Kooijman et al., 2005, 2009; Goyet et al., 2010). ERPs have also provided some evidence that newborns can track transitional probabilities in speech streams (Teinonen et al., 2009). Furthermore, Cunillera et al. (2006) recorded ERPs during statistical word segmentation in adults. They concluded that the timing of adults' neural activity was consistent with the hypothesis that adults extract possible lexical units. Similar ERP evidence with infants would help to strengthen the claim that infants discover candidate words during statistical learning.

In conclusion, the results of the present experiments indicate that during statistical learning, infants form representations that are sufficiently abstract and flexible to recognize them across acoustic variation. Infants can perform this generalization to recognize words and to support other linguistic processes, in this case, associating the sounds of words with meanings. These findings suggest that statistical learning can withstand acoustic challenges present in infants' language environments, which support the case for statistical learning as a viable contributor to language acquisition.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Science Foundation (BCS0847379). I would like to thank Carolina Bastos, Stephanie Chen-Wu Gluck, and the members of the Language Learning Lab at the University of California, Davis for their assistance with this research. I would also like to thank Erik Thiessen for helpful comments on an earlier draft of this manuscript. Thanks also the parents and infants who generously contributed their time.

REFERENCES

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324.

Bortfeld, H., and Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cogn. Psychol.* 60, 241–266.

Brent, M. R. (1999). Speech segmentation and word discovery: a computational perspective. *Trends Cogn. Sci. (Regul. Ed.)* 3, 294–301.

Cairns, P., Shillcock, R., Chater, N., and Levy, J. (1997). Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation. *Cogn. Psychol.* 33, 111–153.

Christiansen, M. H., Allen, J., and Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: a connectionist model. *Lang. Cogn. Process.* 13, 221–268.

Cohen, L. B., Atkinson, D. J., and Chaput, H. J. (2004). *Habit X: A New Program for Obtaining and Organizing Data in Infant Perception and Cognition Studies* (Version 1.0). Austin: University of Texas.

Cooper, R. P., and Aslin, R. N. (1990). Preference for infant-directed

speech in the first month after birth. *Child Dev.* 61, 1584–1595.

Cooper, R. P., and Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Dev.* 65, 1663–1677.

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., and Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res.* 1123, 168–178.

Curtin, S. (2009). Twelve-month-olds learn novel word-object pairings differing only in stress pattern. *J. Child Lang.* 36, 1157–1165.

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211.

Endress, A. D., and Mehler, J. (2009). The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. *J. Mem. Lang.* 60, 351–367.

Fennell, C. T., Byers-Heinlein, K., and Werker, J. F. (2007). Using speech sounds to guide word learning: the case of bilingual infants. *Child Dev.* 78, 1510–1525.

Frank, M. C., Goldwater, S., Griffiths, T. L., and Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition* 117, 107–125.

Giroux, I., and Rey, A. (2009). Lexical and sublexical units in speech perception. *Cogn. Sci.* 33, 260–272.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436.

Goodsitt, J. V., Morgan, J. L., and Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *J. Child Lang.* 20, 229–252.

Goyet, L., de Schonen, S., and Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: an ERP study. *Brain Res.* 1332, 75–89.

Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Graf Estes, K., and Hurley, K. (in press). Infant-directed prosody helps infants map sounds to meanings. *Infancy*.

Harris, Z. S. (1955). From phoneme to morpheme. *Language* 31, 190–222.

Hay, J., Pelucchi, B., Graf Estes, K., and Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cogn. Psychol.* 63, 93–106.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428.

Houston, D. M., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1570–1582.

Houston, D. M., and Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 1143–1154.

Houston-Price, C., and Nakai, S. (2004). Distinguishing novelty and familiarity effects in infants preference procedures. *Infant Child Dev.* 13, 341–348.

Hunter, M. A., and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Adv. Infancy Res.* 5, 69–95.

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 548–567.

Johnson, K. (2008). *Speaker Normalization in Speech Perception.* Malden: Blackwell Publishing.

Jusczyk, J. (1997). *The Discovery of Spoken Language.* Cambridge, MA: MIT Press.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition* 83, B35–B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., and Johnson, S. P. (2007). Location, location, location: development of spatiotemporal sequence learning in infancy. *Child Dev.* 78, 1559–1571.

Kooijman, V., Hagoort, P., and Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cogn. Brain Res.* 24, 109–116.

Kooijman, V., Hagoort, P., and Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy* 14, 591–612.

Luce, P. A., and McLennan, C. T. (2008). *Spoken Word Recognition: The Challenge of Variation.* Malden: Blackwell Publishing.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90, 91–117.

Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition* 108, 271–280.

Nygaard, L. C. (2008). *Perceptual Integration of Linguistic and Nonlinguistic Properties of Speech.* Malden: Blackwell Publishing.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev.* 80, 674–685.

Perruchet, P., and Vinter, A. (1998). PARSER: a model of word segmentation. *J. Mem. Lang.* 39, 246–263.

Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184.

Pisoni, D. B. (1996). Word identification in noise. *Lang. Cogn. Process.* 11, 681–687.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–950.

Romberg, A. R., and Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cogn. Sci.* 1, 906–914.

Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Dev. Sci.* 12, 339–349.

Saberi, K., and Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature* 398, 760–760.

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition* 81, 149–169.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., and Tunick, R. A. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* 8, 101–105.

Saffran, J. R., Werker, J. F., and Werner, L. A. (2006). "The infant's auditory world: hearing, speech, and the beginnings of language," in *Handbook of Child Psychology, Vol 2, Cognition, Perception, and Language*, 6th Edn, eds D. Kuhn, R. S. Siegler, W. Damon, and R. M. Lerner (Hoboken, NJ: John Wiley & Sons Inc.), xxvii, 1042, 2058–2108.

Schmale, R., Cristia, A., Seidl, A., and Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy* 15, 650–662.

Schmale, R., and Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Dev. Sci.* 12, 583–601.

Shi, R., and Werker, J. F. (2001). Six-month old infants' preference for lexical words. *Psychol. Sci.* 12, 70–75.

Shi, R., Werker, J. F., and Cutler, A. (2006). Recognition and representation of function words in english-learning infants. *Infancy* 10, 187–198.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition* 106, 833–870.

Singh, L., Morgan, J. L., and White, K. S. (2004). Preference and processing: the role of speech affect in early spoken word recognition. *J. Mem. Lang.* 51, 173–189.

Singh, L., Nestor, S. S., and Bortfeld, H. (2008a). Overcoming the effects of variation in infant speech segmentation: influences of word familiarity. *Infancy* 13, 57–74.

Singh, L., White, K. S., and Morgan, J. L. (2008b). Building a word-form lexicon in the face of variable input: influences of pitch and amplitude on early spoken word recognition. *Lang. Learn. Dev.* 4, 157–178.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cogn. Psychol.* 50, 86–132.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10, 21. doi:10.1186/1471-2202-10-21

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71.

Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (in press). The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.*

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.

Vouloumanos, A., Brosseau-Liard, P. E., Balaban, E., and Hager, A. D. (2012). Are the products of statistical learning abstract or stimulus-specific? *Front. Lang. Sci.* 3:70. doi:10.3389/fpsyg.2012.00070

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., and Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Dev. Psychol.* 34, 1289–1309.

Werker, J. F., and Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Dev. Psychol.* 24, 672–683.

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63.

Yoshida, K. A., Fennell, C. T., Swingley, D., and Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Dev. Sci.* 12, 412–418.

Zangl, R., and Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6-and 13-month-old infants. *Infancy* 11, 31–62.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 August 2012; accepted: 05 October 2012; published online: 29 October 2012.

*Citation: Graf Estes K (2012) Infants generalize representations of statistically segmented words. Front. Psychology **3**:447. doi: 10.3389/fpsyg.2012.00447*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 28 February 2013
doi: 10.3389/fpsyg.2013.00057

Acoustic analyses of speech sounds and rhythms in Japanese- and English-learning infants

Yuko Yamashita[1]*, Yoshitaka Nakajima[2]*, Kazuo Ueda[2], Yohko Shimada[3], David Hirsh[4], Takeharu Seno[5] and Benjamin Alexander Smith[6]

[1] *Graduate School of Design, Kyushu University, Fukuoka, Japan*
[2] *Department of Human Science, Center for Applied Perceptual Research, Kyushu University, Fukuoka, Japan*
[3] *Graduate School of Asian and African Studies, Kyoto University, Kyoto, Japan*
[4] *Faculty of Education and Social Work, University of Sydney, Sydney, NSW, Australia*
[5] *Faculty of Design, Institute for Advanced Study, Kyushu University, Fukuoka, Japan*
[6] *Department of Design, Architecture and Planning, University of Sydney, Sydney, NSW, Australia*

Edited by:
Claudia Männel, Max-Planck-Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Yang Zhang, University of Minnesota, USA
Josiane Bertoncini, CNRS – Université Paris Descartes, France
Ryoko Mugitani, Nippon Telegraph and Telephone Corporation, Japan

****Correspondence:***
Yuko Yamashita, Graduate School of Design, Kyushu University, 4-9-1 Shiobaru Minami-ku, Fukuoka 815-0032, Japan.
e-mail: yukoy6633@gmail.com;
Yoshitaka Nakajima, Department of Human Science, Kyushu University, 4-9-1 Shiobaru Minami-ku, Fukuoka 815-0032, Japan.
e-mail: nakajima@design.kyushu-u.ac.jp

The purpose of this study was to explore developmental changes, in terms of spectral fluctuations and temporal periodicity with Japanese- and English-learning infants. Three age groups (15, 20, and 24 months) were selected, because infants diversify phonetic inventories with age. Natural speech of the infants was recorded. We utilized a critical-band-filter bank, which simulated the frequency resolution in adults' auditory periphery. First, the correlations between the power fluctuations of the critical-band outputs represented by factor analysis were observed in order to see how the critical bands should be connected to each other, if a listener is to differentiate sounds in infants' speech. In the following analysis, we analyzed the temporal fluctuations of factor scores by calculating autocorrelations. The present analysis identified three factors as had been observed in adult speech at 24 months of age in both linguistic environments. These three factors were shifted to a higher frequency range corresponding to the smaller vocal tract size of the infants. The results suggest that the vocal tract structures of the infants had developed to become adult-like configuration by 24 months of age in both language environments. The amount of utterances with periodic nature of shorter time increased with age in both environments. This trend was clearer in the Japanese environment.

Keywords: infant vocalization, speech development, spectral fluctuations, factor analysis, speech rhythm

INTRODUCTION

During the first 2 years of age, the human speech production mechanism develops rapidly. Various anatomic structures of the vocal tract grow to 55–80% of adult size by 18 months of age (Vorperian et al., 2005). Corresponding to the growth of the vocal tract as well as the control of places and manners of articulation, infant vocalization changes from cooing (vowel-like sounds) to babbling (e.g., da-da or ma-ma), and then to words similar to adult speech during the first 2 years of life.

A number of studies have explored the acoustic characteristics in infant speech spectra, such as formants or spectral peaks of vowels (e.g., Buhr, 1980; Lieberman, 1980; Bond et al., 1982; Kent and Murray, 1982; Gilbert et al., 1997; Rvachew et al., 2006; Ishizuka et al., 2007); these acoustic characteristics reflect the development of the vocal tract and the acquisition of places and manners of articulation. For example, Gilbert et al. (1997) explored developmental characteristics of formant 1 (F1) and formant 2 (F2) produced by four young English-learning children between 15 and 36 months of age. The results revealed that F1 and F2 were relatively stable during the period of 15–21 months and their frequencies decreased significantly between 24 and 36 months. Gilbert et al. (1997) suggested that the vocal tract length and pharyngeal space increased whereas nasal cavity influence decreased, which would probably result in relatively stable F1 and F2 during the period of 15–21 months. Bond et al. (1982) analyzed F1 and F2 of English front and back vowels between 17 and 29 months, and showed that vowel formants shifted in accordance with vowel space expansion with age. Ishizuka et al. (2007) also explored longitudinal developmental changes (4–60 months of age) in spectral peaks of vowels with two Japanese-learning infants. The results showed that a categorically separated vowel space is formed by around 20 months of age, and that the speed of vowel space expansions is rapid by around 24 months of age. These studies supported the view that there are rapid developmental changes in acoustic characteristics during the first 2 years of age corresponding to anatomical development of the vocal tract and manners and places of articulation.

In addition to the acoustic characteristics in infant speech, increasing attention has been devoted to temporal periodicity (e.g., Oller, 1986; Davis and MacNeilage, 1995; Davis et al., 2000; Kouno, 2001; Nathani et al., 2003; Petitto et al., 2004; Dolata et al., 2008). This interest has been caused by the statement that consonant-vowel (CV) sequences in babblings are simply determined by open-close mandibular oscillation, which gives listeners

the perceptual impression of temporal regularity (e.g., Davis and MacNeilage, 1995; Oller, 2000). Dolata et al. (2008) explored the repetition of CV forms in reduplicative vocal babblings obtained from English-learning infants (7–16 months of age) and reduplicated syllables from adult speakers. The results showed that the mean syllable duration in vocal babblings was 329.5 ms and 95% of total durations were between 250 and 425 ms. For adult speakers, the mean syllable duration was 189 ms, which was shorter than that of infant utterances. Nathani et al. (2003) investigated normally hearing and deaf infants at prelinguistic vocal development. For normally hearing infants, the mean nonfinal syllable durations decreased from 378 to 316 ms, and final syllable durations decreased from 527 to 355 ms. Final syllable length ratios for normally hearing infants decreased across age whereas it was relatively stable for deaf infants. The results suggested that the rhythmic organization was influenced by the auditory status and the level of vocal development. Kouno (2001) reported that syllable duration of two- or three-syllable words gradually decreased to be less than 420 ms in babbling forms and less than 330 ms in word forms in Japanese-learning infants by around 20 months of age. Both studies (Kouno, 2001; Nathani et al., 2003) showed gradual development in that the syllable duration in infant vocalizations became shorter across age.

Some studies attempted to find language-related aspects of temporal periodicity in early word production period. A representative series of Vihman (1991), Vihman et al., 1998, 2006, Vihman and de Boysson-Bardies (1994) explored speech rhythm in infant production from different language backgrounds. For example, Hallé et al. (1991) investigated duration patterns in disyllabic vocalization in either word or babbling forms with Japanese- and French-learning infants by around 18 months of age. Final syllable lengthening, which reflected duration characteristics in French, was found in French-learning infants, whereas it was absent for Japanese-learning infants: Language-related aspects of prosodic patterns were already found in infant utterances in these linguistic environments. Vihman et al. (1998) examined disyllables obtained from English- and French-learning infants in the late single-word period (13–20 months of age). The tendency that the second vowel duration was longer than the first vowel duration was adult-like in French-learning infants, whereas each syllable was at considerably higher level of variability, which less closely matched to prosodic patterns in adult speech, in English-learning infants. There was also individual variability for English-learning infants. Vihman et al. (1998) considered children's differing learning strategies, and argued that each child filtered the input of language, and attempted to reproduce words based on their favored word production templates. Language-related aspects were found while there was variability of syllable duration in the early word production period.

Although these studies shed light on the developmental changes in acoustic characteristics and temporal periodicity, they had the following problems: (1) Formant frequency analysis (e.g., Buhr, 1980; Lieberman, 1980; Bond et al., 1982; Kent and Murray, 1982; Gilbert et al., 1997; Ishizuka et al., 2007), which was most frequently used, is employed basically to detect only vowel sounds in order to obtain knowledge for linguistic development. There has been a lack of acoustic analysis which measures the whole pattern of spectral fluctuations. (2) Speech samples to observe temporal periodicity were limited to disyllabic vocalizations (e.g., Hallé et al., 1991; Vihman et al., 1998; Davis et al., 2000). There was no automatic measurement to identify temporal periodicity, and thus phoneticians judged duration by looking at speech waveforms, which might have be subjective.

In the present study, a critical-band-filter bank was used to analyze the spectral fluctuations and temporal periodicity in infants' utterances. A practical way to analyze speech signals is to separate them into a certain number of narrow frequency bands as in a historical (traditional) vocoder system, and to observe the temporal power fluctuation in each frequency band. The notion of critical bands, which reflects basic characteristics of the auditory system (see, e.g., Fletcher, 1940; Zwicker and Terhardt, 1980; Patterson and Moore, 1986; Unoki et al., 2006; Fastl and Zwicker, 2007; Moore, 2012), seemed convenient for our present purpose, because the power fluctuations in 15–22 critical bands contain enough information to make speech almost fully intelligible. Ueda and Nakajima (2008) performed factor analyses of the spectral fluctuations in speech sounds of different languages, utilizing a critical-band-filter bank. The same three factors appeared in Japanese and English, which were replicated for a far smaller number of speech samples (see **Figure A1** in Appendix). The critical-band-filter bank analysis seemed applicable to Japanese- and English-learning infant speech in order to detect the whole pattern of spectral fluctuations. We were particularly interested in what age of life the factors as in adults' speech would appear in infant speech.

As a next step, we explored the temporal periodicity in infant speech obtained from Japanese- and English-learning infants. The speech samples in the current study were not limited to disyllabic vocalization. We used all the speech samples ($\geq$1.5 s) in order to explore the whole pattern of developmental changes. We utilized the temporal periodicity of the factor scores that summarizes power fluctuations of speech sounds in the outputs of critical-band filters, instead of measuring temporal intervals in speech waveforms by the eye. Japanese and English adult speech samples in a database were first analyzed, and the validity of this method was proved (see **Figure A2** in Appendix). Thus, we applied this method to identify the temporal periodicity in infant speech.

Three ages, 15, 20, and 24 months, were selected for the following reasons. The various vocal tract structures, predominantly pharyngeal/posterior structure, achieve 55–80% of the adult size by 18 months of age (Vorperian et al., 2005). In addition to the development of vocal tract, lexical development is in rapid progress from 12 to 18 months of age. Many infants over this period become capable of producing at least 50 meaningful words, which is so called "50-word stage" (MacNeilage et al., 2000). After "50-word stage," there is an explosion of phonetic diversification due to the better control of manners and places of articulations to produce a variety of consonant sounds, and expansion of the vowel spaces to include diverse vowel types (Kern et al., 2010). Thus, around the age of 15 months, the vocal tract is in the process of rapid development and this corresponds to a period of rapid lexical development (12–18 months), while infants from 20 to 24 months of age become capable of diversifying phonetic inventories and form some sentences to convey more complex messages. Thus,

the period of 15–24 months of age seemed appropriate to explore significant changes in infant speech development.

The questions of infant speech development were addressed as follows:

(1) How do spectral fluctuation and temporal periodicity in infant speech change between 15 and 24 months of age?
(2) Are the developmental changes of speech in the acoustic domain similar in Japanese- and English-learning infants?

MATERIALS AND METHODS

INFANT PARTICIPANTS

Participants included five typically developing infants at 15 months of age (three girls and two boys), five infants at 20 months of age (three girls and two boys), and five infants at 24 months of age (three girls and two boys) from Japanese-speaking families. Five typically developing infants at 15 months of age (three girls and two boys), five infants at 20 months of age (two girls and three boys), and four infants at 24 months of age (three girls and one boy) were from English-speaking families. The Japanese-learning infants were being raised by monolingual Japanese adult speakers. The English-learning infants were being raised by monolingual English adult speakers or adult speakers whose first language is English. For all Japanese-learning infants, their weight was over 8, 10, and 9 kg and height was over 76, 83, and 82 cm at 15, 20, and 24 months of age, respectively. For all English-learning infants, their weight was over 10, 11, and 10 kg and their height was over 78, 84, and 84 cm at 15, 20, and 24 months of age, respectively. This showed that all infants exhibited normal physical development. Parental consent forms and information sheets were provided to a parent of each infant. The procedures required for the project and the time involved were explained. Parental consent forms from each parent were received.

RECORDINGS

Utterances were recorded in a quiet room in each infant's home for about 2 h a month. Special care was taken to keep each infant in a normal environment at home. A digital sound recorder (Roland, R-09HR or TEAC, DR-07) was set to 44.1-kHz sampling and 16-bit linear quantization. The recorder was placed on a pillow in order to prevent vibration and reverberation. It was kept at least 1 m away from the infant in order to stabilize the recording level. The parent or parents were instructed to behave in a usual manner and to do daily activities during the recording process. No specific procedures to elicit infant vocalization were utilized.

SPEECH SAMPLES

One of the authors and two students in the Department of Acoustic Design and Human Science course at Kyushu University extracted utterances from each 2-h recording, using audio software (Syntrillium, Cool Edit 2000, or Adobe, Audition) based on the following criteria:

1. Silent parts of 75 ms before and after each utterance were included.
2. If a silent part between two potential utterances was shorter than 1200 ms, the whole pattern was considered a single utterance. Since we were particularly interested in rhythmic patterns in speech, we calculated autocorrelations of factor scores up to 1 s. This prohibited us from discarding silent intervals shorter than 1 s. For assurance, we included all silent intervals shorter than 1200 ms as part of the utterances to be analyzed.
3. If a single utterance was separated by adult speech or background noise, the separated parts were analyzed as different utterances.
4. If an utterance was overlapped by adult speech or background noise from toys or other objects, it was excluded from analysis.
5. Anomalous vocal signals, such as laughter, crying, squeals, growls, and shrieking were excluded.

We constructed a database consisting of utterances of Japanese- and English-learning infants. Speech samples longer than 1.5 s in this database represented 25, 30, and 54% of all utterances for Japanese-learning infants at 15, 20, and 24 months of age, respectively, and 23, 27, and 59% of all utterances for English-learning infants at 15, 20, and 24 months, respectively.

Table 1 presents information regarding the number of utterances and the average duration of utterances obtained for each infant. In total, 484, 474, and 586 utterances were collected from Japanese-learning infants at 15, 20, and 24 months, respectively; 529, 465, and 426 utterances were collected from English-learning infants at 15, 20, and 24 months, respectively.

SPEECH ANALYSIS

All the speech signals were analyzed using the same approach as in Ueda and Nakajima (2008). A bank of critical-band filters was constructed. The total passband of the filter bank ranged from 100 to 12,000 Hz, and the center frequencies of the filters ranged from 150 to 10,500 Hz. The cutoff frequencies of the critical-band filters were based on Zwicker and Terhardt (1980). Each filter was constructed as concatenate convolutions of an upward frequency glide and its temporal reversal. Both sides of the filters had slopes steeper than 90 dB/oct. Each filter output was squared, smoothed with a Gaussian window of $\sigma = 20$ ms, and sampled at every 1 ms. Factor analyses were performed based on the correlation matrices between the power fluctuations of the 22 critical-band filters. In each age/language group, the average levels of all the speech samples were adjusted to be equal to each other, and the adjusted samples were connected in time for factor analysis. The total duration of the connected signals was 667, 626, and 897 s for the Japanese-learning infants and 630, 512, and 763 s for the English-learning infants, at 15, 20, and 24 months of age, respectively. Correlation-based (normalized) analysis was performed; varimax rotation followed principal component analysis. The number of factors was set at two or three in order to compare the present results with Ueda and Nakajima's (2008) results.

In the following analysis, the autocorrelation functions were obtained in order to observe temporal periodicity in the factor scores. The correlation between the nth and the $(n+k)$th sample in a time series of N samples was calculated as follows:

$$r(k) = \frac{\sum_{n=1}^{N-k} (x_n - \bar{x}_1)\left(x_{n+k} - \bar{x}_{k+1}\right)}{\sqrt{\sum_{n=1}^{N-k} (x_n - \bar{x}_1)^2} \cdot \sqrt{\sum_{n=\kappa+1}^{N} \left(x_n - \bar{x}_{k+1}\right)^2}},$$

Table 1 | Number and average duration of utterances.

	Months of age	Number of utterances	Average duration of utterances (s)	Standard deviation (SD)
JAPANESE-LEARNING INFANTS				
JF2	15	98	1.98	1.67
JF6	15	132	1.08	1.03
JM3	15	111	0.99	0.64
JM4	15	69	1.18	0.78
JM7	15	74	1.67	1.36
Overall		484	1.38	1.07
JM1	20	90	1.15	0.96
JF2	20	102	1.22	0.86
JF3	20	95	1.16	1.14
JM3	20	85	1.79	1.26
JF1	20	102	1.30	1.10
Overall		474	1.32	1.06
JF2	24	101	1.83	0.9
JF3	24	130	1.49	0.74
JF6	24	124	1.39	0.67
JM1	24	123	1.52	0.8
JM3	24	108	1.45	0.65
Overall		586	1.53	0.75
ENGLISH-LEARNING INFANTS				
EF1	15	99	1.71	1.98
EF3	15	105	1.16	1.03
EM3	15	114	1.09	0.93
EF2	15	120	0.98	0.77
EM1	15	91	0.75	0.52
Overall		529	1.19	1.05
EF1	20	107	1.17	0.62
EF2	20	78	1.26	0.93
EM2	20	73	1.18	0.84
EM4	20	121	0.9	0.69
EM1	20	86	0.99	0.67
Overall		465	1.10	0.73
EF1	24	96	1.71	0.79
EF2	24	85	2.30	0.91
EF07	24	107	1.41	0.74
EM06	24	138	1.73	0.82
Overall		426	1.79	0.81

where $\bar{x}_1 = \frac{\sum_{n=1}^{N-k} x_n}{N-k}$, and

$$\bar{x}_{k+1} = \frac{\sum_{n=k+1}^{N} x_n}{N-k}.$$

The autocorrelation function of the temporal distance τ was defined as

$$R(\tau) = r\left(\tau \cdot f_s\right),$$

where f_s represents the sampling frequency; $R(\tau)$ was defined only when $\tau \cdot f_s$ was an integer.

In the factor analysis, factor scores were sampled at every 1 ms. We used speech samples ≥1.5 s and observed temporal periodicity in factor scores by calculating autocorrelations up to 1 s. There was always a factor including a frequency range of 1000–1600 Hz, and this factor seemed to be related to vowel-like sounds (Nakajima et al., 2012); the autocorrelation of this factor (factor scores as a function of time) was calculated for each utterance in order to observe a global pattern of temporal periodicity, if any. The amplitude of the first peak above zero was taken as the representative of an autocorrelation score. If there was no peak above zero, the autocorrelation function was considered to be without a peak.

RESULTS

FACTOR ANALYSES

Figures 1A–F show the results obtained from Japanese- and English-learning infants at 15, 20, and 24 months of age. Factor 1 related to a frequency range around 1600 Hz, factor 2 was bimodal surrounding factor 1, and was related to frequency ranges around 350 Hz and around 4000 Hz, and factor 3 was related to high frequency ranges.

The factor loadings of factors 1–7 or 1–8 whose original principal components always exhibited eigenvalues greater than 1, were observed. The cumulative contributions obtained from the data for each language/age group were 50–57% for the seven or eight components. For comparison with adult speech, two or three factors were chosen. The Cumulative contributions were 30–32% for the first three components. The first three components showed clear correspondence with the adults' results for Japanese-learning infants at 20 and 24 months, and English-learning infants at 24 months. The second or third factor did not show clear correspondence with any particular frequency ranges for Japanese-learning infants at 15 months or with English-learning infants at 15 and 20 months. For older infants, factor 1 was surrounded by factor 2, which was bimodal, and factor 3 was specifically related to the highest frequency range. If factor loadings are indicated against frequency represented logarithmically, the configurations of the three factors in the infant speech at 24 months of age are well in correspondence with those in the adult speech in both linguistic environments (see **Figure A3** in Appendix).

Peaks of the curves represented relatively high factor loadings, and we considered the crossover frequency of two adjacent curves as an indication of the boundary between the corresponding factors. **Table 2** shows the obtained boundaries as represented by the closest center frequencies. The first and second crossover points between factors 1 and 2 are indicated as the first and second boundary frequencies; the crossover points between factors 2 and 3 are indicated as the third boundary frequencies. If the boundary frequencies are difficult to observe, they are indicated as unclear.

It appears that the same factors as in the infant speech shifted downward (leftward) in logarithmic frequency in the adult speech (**Figure A3** in Appendix): The boundary frequencies (represented logarithmically) in the infant speech at 24 months were higher than those in the adult speech by a factor around 1.7 times. This indicates that the 24-month-old infants and the adult speakers used the articulation organs basically in the same way, and that the differences between the factor configurations were caused simply by the size differences – if the articulation organs are doubled in size, the frequencies indicating the factor locations are halved.

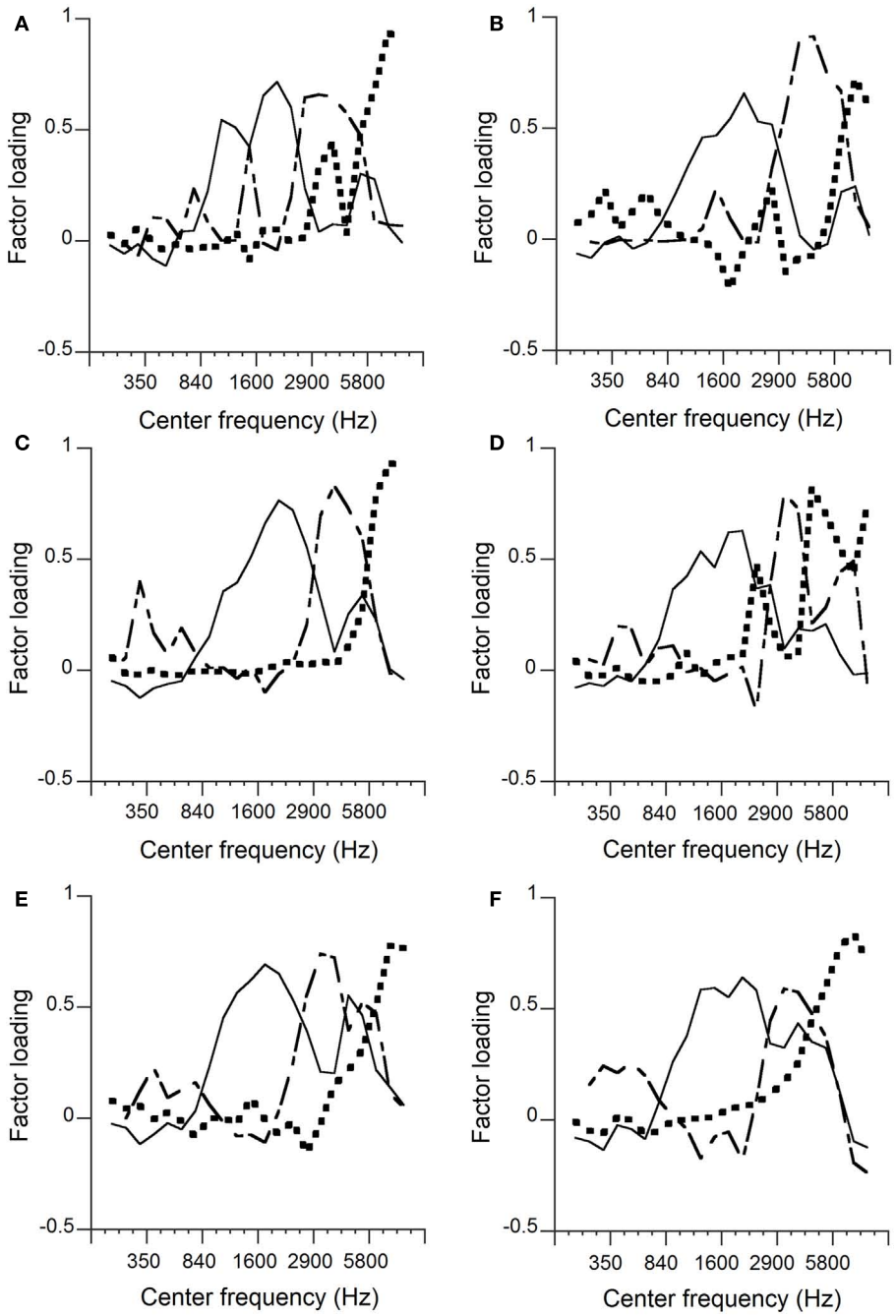


FIGURE 1 | Factor analyses. Japanese-learning infants at 15 months of age **(A)**, English-learning infants at 15 months **(B)**, Japanese-learning infants at 20 months **(C)**, English-learning infants at 20 months **(D)**, Japanese-learning infants at 24 months **(E)**, and English-learning infants at 24 months **(F)**. The solid lines, dashed lines, and dotted lines represent factors 1, 2, and 3, respectively.

AUTOCORRELATION ANALYSES

We adopted the two-factor analysis, which produced visually clear results in most cases. The cumulative contributions were 23–27% for the two principal components. There was always a factor including a frequency range around 1600 Hz, which was similar to one of the factors in the three-factor analysis. Infants' utterances $\geq$1.5 s were selected from speech samples so that at least 1500 factor scores, sampled at every 1 ms (as exactly as possible), were used for each autocorrelation analysis. **Figure 2** shows an example of an autocorrelation function from a Japanese-learning female infant at 24 months of age. The amplitude of the first peak above zero (0.36 s in **Figure 2**) was taken as the representative autocorrelation score. If there was no peak above zero, the autocorrelation score was considered as without a peak. For Japanese-learning

Table 2 | Boundary frequencies of the factor-related frequency bands observed in infants and adults.

Language	Months of age	Boundaries (Hz)		
		First	Second	Third
Japanese	15	Unclear	Unclear	Unclear
	20	840	2900	5800
	24	840	2500	5800
English	15	Unclear	Unclear	Unclear
	20	Unclear	Unclear	Unclear
	24	840	2500	4800
Japanese	Adult	450	1850	3400
English	Adult	450	1600	2500

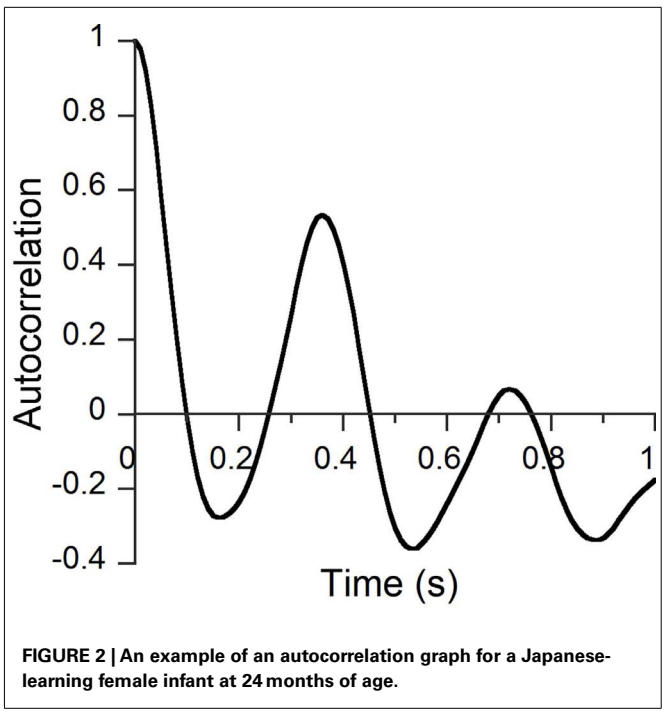


FIGURE 2 | An example of an autocorrelation graph for a Japanese-learning female infant at 24 months of age.

infants, the total numbers of utterances ≥1.5 s were 157, 141, and 311 at 15, 20, and 24 months of age, respectively. For English-learning infants, the total numbers of utterances ≥1.5 s were 118, 126, and 251 at 15, 20, and 24 months of age, respectively.

Figures 3A–C show the relative frequency distributions (%) of the first peaks for the Japanese- and English-learning infants at 15, 20, and 24 months of age. We focused on the first peaks located above 0.10 up to 0.40 s to explore the temporal periodicity, which was observed in previous studies (see, e.g., Kouno, 2001; Nathani et al., 2003; Dolata et al., 2008). As shown in **Figure 4**, 15.9, 22.7, and 29.9% of the first peaks were located in this range at 15, 20, and 24 months of age for the Japanese-learning infants, compared with 13.6, 15.9, and 23.1% for the English-learning infants. A chi-square test was carried out. For the Japanese-learning infants, the results showed that the relative frequency of the first peaks located above 0.10 up to 0.40 s increased across age, and the change was

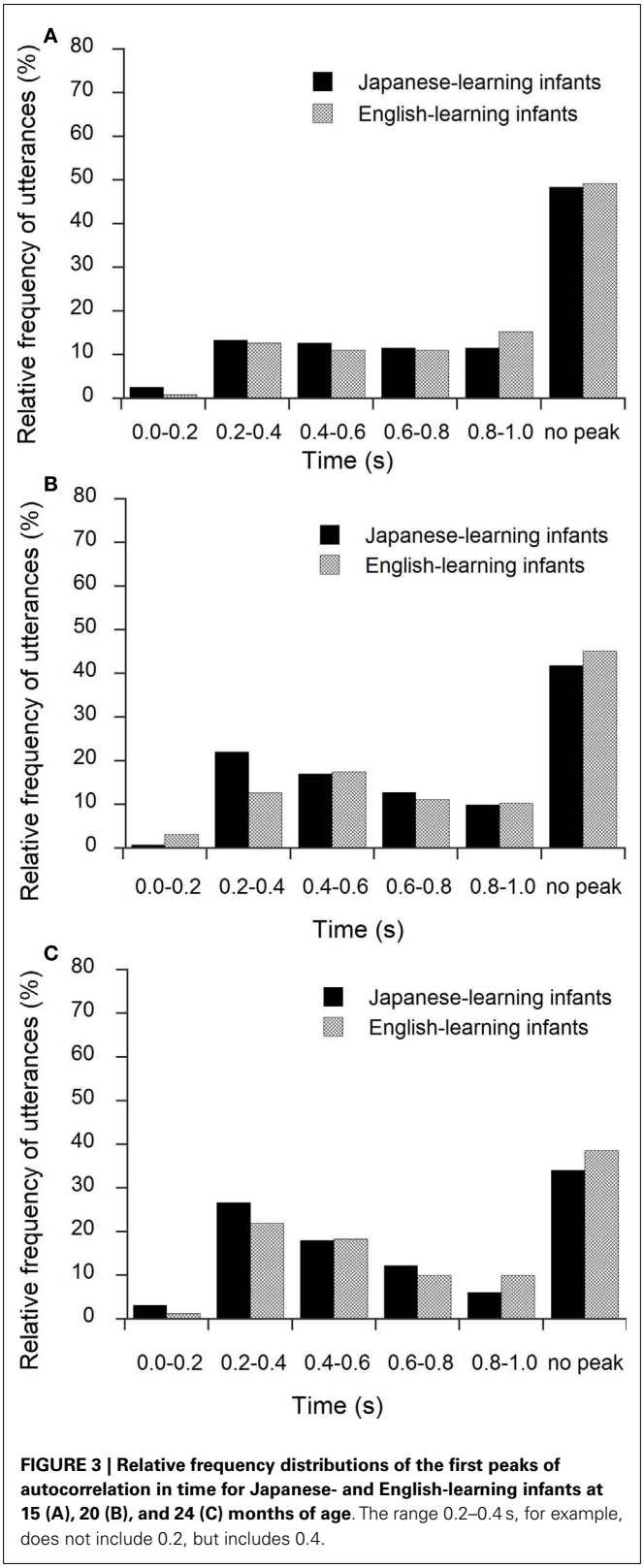


FIGURE 3 | Relative frequency distributions of the first peaks of autocorrelation in time for Japanese- and English-learning infants at 15 (A), 20 (B), and 24 (C) months of age. The range 0.2–0.4 s, for example, does not include 0.2, but includes 0.4.

statistically significant (15, 20, and 24 months of age; $\chi^2 = 11.35$, $df = 2$, $p < 0.01$). There was a similar trend in the English-learning infants, but it was not statistically significant.

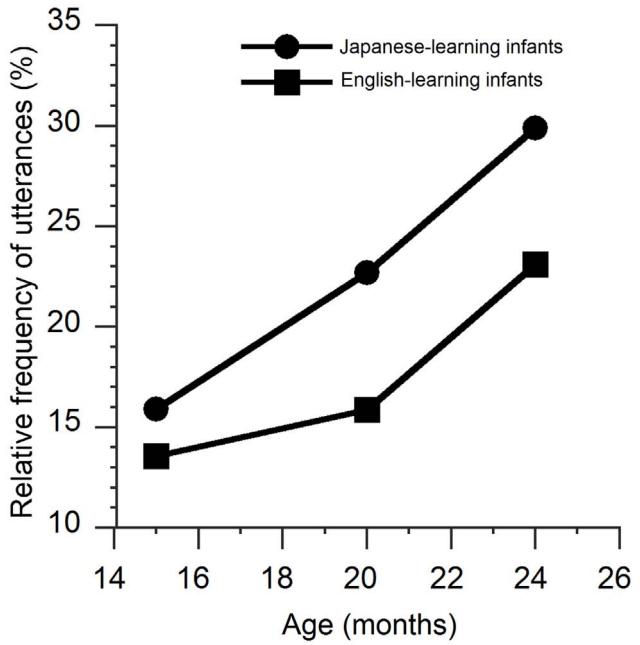


FIGURE 4 | Relative frequencies of the first peaks of autocorrelation located above 0.10 up to 0.40 s for Japanese- and English-learning infants across age (15, 20, and 24 months of age).

DISCUSSION

The purpose of the present investigation was to explore how the spectral fluctuations and the temporal periodicity of infant speech changed in Japanese- and English-learning infants between 15 and 24 months of age. The factor analyses of spectral fluctuations showed that three factors observed in adult speech appeared by 24 months of age in both linguistic environments. Those three factors were shifted to a higher range corresponding to the smaller vocal tract size of the infants (e.g., Goldstein, 1980; Vorperian et al., 2005). It is probable that the vocal tract structures of the infants had developed to adult-like configuration, but the whole vocal tract was still shorter than that of an adult. This corresponds to the vocal development study by Vorperian et al. (2005), which showed that the sizes of the various vocal tract structures grew rapidly to achieve 55–80% of that of the adult's by 18 months of age. The results also agree with previous studies (e.g., Bond et al., 1982; Ishizuka et al., 2007), which showed there were rapid vowel space expansions during the first 2 years of age.

Autocorrelations were calculated from temporal fluctuations of the factor scores. It should be pointed out that the present study included a variety of utterances; it differs from previous studies, in which speech samples were limited to disyllabic vocalizations (e.g., Hallé et al., 1991; Vihman et al., 1998; Davis et al., 2000). One of the reasons that the previous analyses were limited to disyllabic vocalizations was the difficulty of measuring temporal periodicity. Conventional methods for adult speech, which are based on phonological properties, such as syllable structure and vowel reductions (e.g., Ramus et al., 1999; Low et al., 2000; Deterding, 2001; Grabe and Low, 2002; White and Mattys, 2007), were not applicable to infants. Since phonological properties in infant utterances are obscure. Thus, measuring duration was a common way to explore temporal periodicity in infant utterances. As Roach (1982) pointed out, there was no automatic measurement to identify stressed syllables: Phoneticians needed to judge stressed syllables by looking at speech waveforms, which might be influenced by incidental characteristics such as vowel length or pitch. The present authors employed an automatic method to identify temporal periodicity; it is based on temporal fluctuations of factor scores (by calculating autocorrelations). This method made it possible to explore the whole patterns of temporal periodicity in infant utterances. The amount of utterances with periodic nature of shorter time (up to 0.4 s) increased with age. The result corresponds to syllable durations observed in previous studies (e.g., Kouno, 2001; Nathani et al., 2003; Dolata et al., 2008). It needs to be examined whether this trend reflects ambient language rhythm.

In conclusion, the present analysis of spectral fluctuation showed that three factors observed in adult speech appeared by 24 months of age in both linguistic environments. Those three factors were shifted to a higher frequency range corresponding to the smaller vocal tract size. The amount of utterances with periodic nature of shorter time increased with age in both linguistic environments. This trend seemed clearer in the Japanese environment, which should be examined further in the future.

ACKNOWLEDGMENTS

This work was supported by the Japan Society for the Promotion of Science [Grant-in-Aid for Scientific Research (S) (No. 19103003)], and the Kawai Foundation for Sound Technology and Music. The present research was a part of Kyushu University Interdisciplinary Programs in Education and Projects in Research Development (The Kyushu University Project for Interdisciplinary Research of Perception and Cognition). We would like to express our sincere gratitude to the parents and infants who were willing to participate in this study. Takuya Kishida, Bao Zhimin and Hirotoshi Motomura gave us technical assistance.

REFERENCES

Bond, Z. S., Petrosino, L., and Dean, C. R. (1982). The emergence of vowels: 17 to 26 months. *J. Phon.* 10, 417–422.

Buhr, R. D. (1980). The emergence of vowels in an infant. *J. Speech Hear. Res.* 23, 73–94.

Davis, B., and MacNeilage, P. (1995). The articulatory basis of babbling. *J. Speech Hear. Res.* 38, 1199–1211.

Davis, B. L., MacNeilage, P. F., Mayyear, C. L., and Powell, J. K. (2000). Prosodic correlates of stress in babbling: an acoustic study. *Child Dev.* 71, 1258–1270.

Deterding, D. (2001). The measurement of rhythm: a comparison of Singapore and British English. *J. Phon.* 29, 217–230.

Dolata, J. K., Davis, B. L., and MacNeilage, P. F. (2008). Characteristics of the rhythmic organization of vocal babbling: implications for an amodal linguistic rhythm. *Infant Behav. Dev.* 31, 422–431.

Fastl, H., and Zwicker, E. (2007). *Psychoacoustics: Facts and Models.* New York: Springer.

Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.* 12, 47–65.

Gilbert, H. R., Robb, M. P., and Chen, Y. (1997). Formant frequency development: 15 to 36 months. *J. Voice* 11, 260–266.

Goldstein, U. G. (1980). *An Articulatory Model for the Vocal Tract of Growing Children.* Ph.D. Dissertation, Cambridge: MIT.

Grabe, E., and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis," in *Papers in Laboratory Phonology*, eds C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 515–546.

Hallé, P., de Boysson-Bardies, B., and Vihman, M. (1991). Beginnings of prosodic organization: intonation and duration patterns of disyllables produced by Japanese and French infants. *Lang. Speech* 34, 299–318.

Ishizuka, K., Mugitani, R., Kato, H., and Amano, S. (2007). Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infants. *J. Acoust. Soc. Am.* 121, 2272–2282.

Kent, R. D., and Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *J. Acoust. Soc. Am.* 72, 353–365.

Kern, S., Davis, B., and Zink, I. (2010). "From babbling to first words in four languages: common trends, cross language and individual differences," in *Becoming Eloquent*, eds J. M. Hombert and F. d'Errico (Cambridge: John Benjamins Publishers), 205–232.

Kouno, M. (2001). *Onseigengo no ninshiki to seisei no mekanizumu: kotoba no jikanseigyokikou to sono yakuwari.* Tokyo: Kinseido.

Lieberman, P. (1980). "On the development of vowel production in young children," in *Child Phonology*, ed. G. H. Yeni-Komashian, J. F. Kavanagh, and C. A. Ferguson (London: Academic Press), 113–142.

Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. *Lang. Speech* 43, 377–401.

MacNeilage, P. F., Davis, B. L., Kinney, A., and Matyear, C. L. (2000). The motor core of speech: a comparison of serial organization patterns in infants and languages. *Child Dev.* 71, 153–163.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing.* Bingley: Emerald.

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, S., and Ohsaka, Y. (2012). Acoustical correlate of phonological sonority in British English. *Paper presented at the 28th Annual Meeting of the International Society for Psychophysics*, Ottawa, ON.

Nathani, S., Oller, D., and Cobo-Lewis, A. (2003). Final syllable lengthening (FSL) in infant vocalizations. *J. Child Lang.* 30, 3–25.

Oller, D. K. (1986). Metaphonology and infant vocalizations," in *Precursors of Early Speech*, eds B. Lindblom and R. Zetterstrom (New York: Stockton Press), 21–35.

Oller, D. K. (2000). *The Emergence of the Speech Capacity.* Mahwah: Lawrence Erlbaum Associates.

Patterson, R. D., and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency selectivity in Hearing*, ed. B. C. J. Moore (London: Academic Press), 123–177.

Petitto, L. A., Holowka, S., Lauren, E. S., Bronna, L., and Davis, J. O. (2004). Baby hands that move to the rhythm of language: hearing babies acquiring sign languages babble silently on the hands. *Cognition* 93, 43–73.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292.

Roach, P. (1982). "On the distinction between 'stress-timed' and 'syllable-timed' languages," in *Linguistic Controversies*, ed. D. Crystal (London: Edward Arnold), 73–79.

Rvachew, S., Mattock, K., Polka, L., and Menard, L. (2006). Developmental and cross-linguistic variation in the infant vowel space: the case of Canadian English and Canadian French. *J. Acoust. Soc. Am.* 120, 1–10.

Ueda, K., and Nakajima, Y. (2008). A consistent clustering of power fluctuations in British English, French, German, and Japanese. *Trans. Tech. Comm. Psychol. Physiol. Acoust.* 38, 771–776.

Unoki, M., Irino, T., Glasberg, B., Moore, B. C. J., and Patterson, R. D. (2006). Comparison of the roex and gammachirp filters as representations of the auditory filter. *J. Acoust. Soc. Am.* 120, 1474–1492.

Vihman, M. M. (1991). "Ontogeny of phonetic gestures: Speech production," in *Modularity and the Motor Theory of Speech Perception*, eds I. G. Mattingly and M. Studdert-Kennedy (New York: Lawrence Erlbaum Associates).

Vihman, M. M., and de Boysson-Bardies, B. (1994). The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica* 51, 159–169.

Vihman, M. M., Nakai, S., and De Paolis, R. A. (2006). "Getting the rhythm right: a cross-linguistic study of segmental duration in babbling and first words," in *Laboratory Phonology 8: Phonology and Phonetics*, eds L. Goldstein, D. Whalen, and C. Best (New York: Mouton de Gruyter), 341–366.

Vihman, M. M., Rory, D., and Barbara, L. D. (1998). Is there a "trochaic basis" in early word learning? *Child Dev.* 69, 933–947.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (2005). Development of vocal tract length during childhood: a magnetic resonance imaging study. *J. Acoust. Soc. Am.* 117, 338–350.

White, L., and Mattys, S. L. (2007). Calibrating rhythm: first language and second language studies. *J. Phon.* 35, 501–522.

Zwicker, E., and Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *J. Acoust. Soc. Am.* 68, 1523–1525.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2012; accepted: 25 January 2013; published online: 28 February 2013.

*Citation: Yamashita Y, Nakajima Y, Ueda K, Shimada Y, Hirsh D, Seno T and Smith BA (2013) Acoustic analyses of speech sounds and rhythms in Japanese- and English-learning infants. Front. Psychol. **4**:57. doi:10.3389/fpsyg.2013.00057*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

APPENDIX

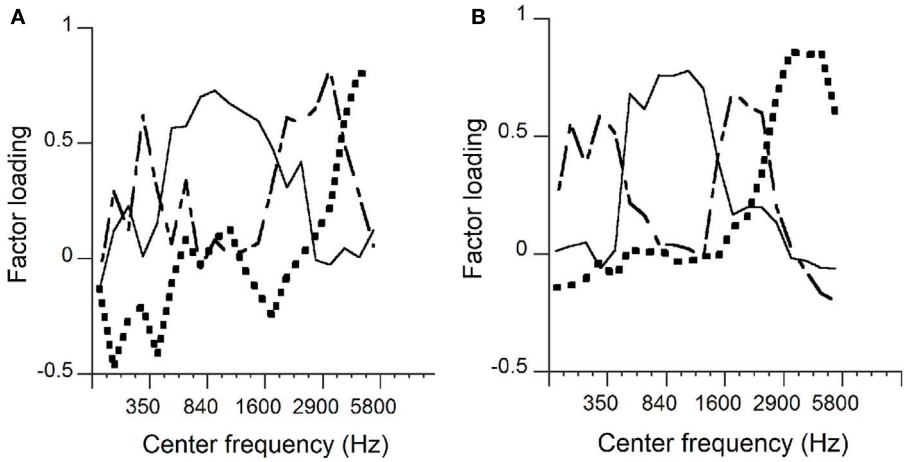


FIGURE A1 | The graph in (A) shows the results of factor analysis from adult Japanese speakers ($N = 10$), and the graph in (B) shows the results from adult English speakers ($N = 10$). The total number of sentences was 20 for each language group. The solid lines, dashed lines, and dotted lines represent factors 1, 2, and 3, respectively.

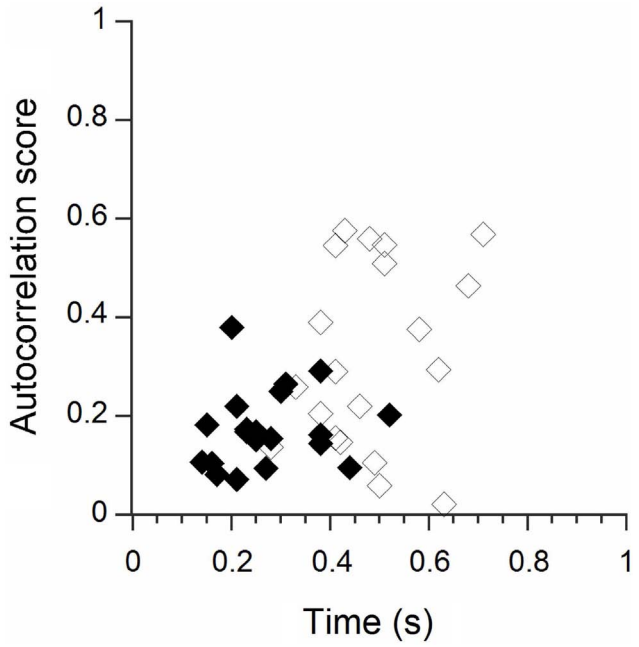


FIGURE A2 | The graph shows the autocorrelation score and time of the first peak for each sample of adult Japanese speakers (black square) and adult English speakers (white square). The acoustic method clarified difference in temporal periodicity between Japanese and English.

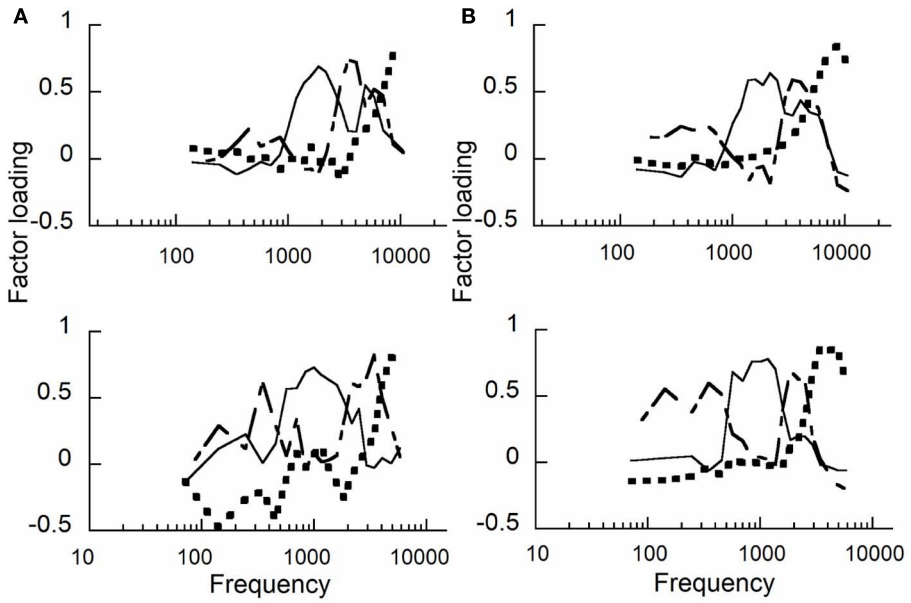


FIGURE A3 | The graphs in (A) show the results from Japanese-learning infants at 24 months (upper) and adult Japanese speakers (lower), and the graphs in (B) show the results from English-learning infants at 24 months (upper) and adult English speakers (lower). The solid lines, dashed lines, and dotted lines represent factors 1, 2, and 3, respectively. Adult speakers' data are from **Figure A1**. The use of logarithmic frequency scales is helpful to compare the configurations of the factors in infant and adult speech. The horizontal axis in the graph of adult speech was shifted by 1.7 times. If a point in an upper graph and another point in a lower graph agreed with each other on the horizontal location, the frequency in the upper graph is 1.7 times as high as that in the lower graph. The graphs showed that the configurations of the three factors in infant speech were in correspondence with those in adult speech. Roughly speaking, the frequency boundaries for the infant data were higher by a factor around 1.7 times. This tolerably corresponds to the fact that infants' articulation organs at this age are 55–80% in size compared with the adults' articulation organs (Vorperian et al., 2005).