Published in final edited form as:

Lang Learn Dev. 2017 ; 13(1): 100–126. doi:10.1080/15475441.2016.1180983.

Subtlety of Ambient-Language Effects in Babbling: A Study of English- and Chinese-Learning Infants at 8, 10, and 12 Months

Chia-Cheng Lee^a, Yuna Jhang^a, Li-mei Chen^b, George Relyea^d, and D. Kimbrough Oller^{a,c}

^aSchool of Communication Sciences and Disorders, The University of Memphis

^bDepartment of Foreign Languages and Literature, National Cheng Kung University

^cThe Konrad Lorenz Institute for Evolution and Cognition Research

^dSchool of Public Health, The University of Memphis

Abstract

Prior research on ambient-language effects in babbling has often suggested infants produce language-specific phonological features within the first year. These results have been questioned in research failing to find such effects and challenging the positive findings on methodological grounds. We studied English- and Chinese-learning infants at 8, 10, and 12 months and found listeners could not detect ambient-language effects in the vast majority of infant utterances, but only in items deemed to be words or to contain canonical syllables that may have made them sound like words with language-specific shapes. Thus, the present research suggests the earliest ambient-language effects may be found in emerging lexical items or in utterances influenced by language-specific features of lexical items. Even the ambient-language effects for infant canonical syllables and words were very small compared with ambient-language effects for meaningless but phonotactically well-formed syllable sequences spoken by adult native speakers of English and Chinese.

Introduction

Background

Ambient-language effects in babbling can be said to occur if the sounds of infant vocalizations shift in the direction of the infants' ambient language(s). Brown (1958) coined the term "babbling drift" to refer to such possible effects. The great bulk of research on possible ambient-language effects has focused on finding language-specific phonetic or prosodic elements in *pure babbling* during the first year or so. This focus implies that sounds of babbling are affected by infant hearing of language-specific *phonology* only rather than by language-specific *lexical items*. Ambient-language effects are said to occur in "babbling" or "prelinguistic utterances" in both textbooks reviewing these effects (e.g., Bauman-Waengler, 2012; Gleason & Ratner, 2013) and research articles reporting them (e.g.,

CONTACT: Chia-Cheng Lee, clee12@memphis.edu, The University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hlld.

Rvachew, Alhaidary, Mattock, & Polka, 2008; Whalen, Levitt, & Goldstein, 2007). These sources, with the exception of Engstrand, Williams, and Lacerda (2003), who studied babbling in 12 and 18 month olds, have not focused on the possibility that early ambient-language effects may be limited to infant words or emerging words rather than occurring in babbling per se. The present work provides evidence from the end of the first year of life suggesting ambient-language effects may occur discernibly only in emergent words and not in babbling.

Ambient-language effects even in the first year seem plausible partly because evidence suggests that auditory experience of language is essential in vocal development. Deaf infants show very late onset of canonical babbling (CB, which by definition includes well-formed consonant-vowel syllables such as [ba] or [ma]), a key feature of vocal development in the first year (Koopmans-van Beinum, Clement, & van den Dikkenberg-Pot, 1998; Oller & Eilers, 1988; Oller, Eilers, Bull, & Carney, 1985; Vinter, 1987). In addition, hearing-impaired infants show a less diverse phonetic repertoire than typically developing infants (Moeller et al., 2007; Stoel-Gammon & Otomo, 1986). Moreover, hearing experience through social interaction appears to help infants shape babbling and to facilitate phonological learning in their first year (Goldstein, King, & West, 2003; Goldstein & Schwade, 2008; Gros-Louis, West, Goldstein, & King, 2006).

Another source of evidence that suggests possible early effects of ambient language on infant vocal development comes from studies of speech perception in infancy. The head-turn and high-amplitude sucking techniques, as well as electrophysiological studies have shown that infants differentiate syllables from various languages at early ages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Morse, 1972; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). In addition, the experience of listening to the native language appears to influence speech perception tendencies of the infant by the second half-year of life, with increasing tendencies for infants to respond differentially to sound contrasts pertaining to their ambient language and to ignore contrasts not in the native language (Eilers, Gavin, & Wilson, 1979; Werker & Tees, 1984).

Furthermore, it is clear that infants in the first year learn to recognize words and word shapes. By 9 months they show listening preferences to items that are well-formed in their ambient language and that occur frequently as input (Jusczyk, 2002). They are able to learn to recognize new well-formed word shapes in a brief training session by 8 months (Aslin, Saffran, & Newport, 1998; Saffran, Johnson, Aslin, & Newport, 1999). And they show looking preference for familiar items that are named by 6 months, suggesting they have learned some word/entity associations (Bergelson & Swingle, 2012; Tincoff & Jusczyk, 1999). They can even learn *new* word associations in the laboratory by 8 months if the new words are presented with synchronous movement of the named object (Gogate, Walker-Andrews, & Bahrick, 2001).

Thus, there are several lines of evidence that infants in the first year learn about their ambient language by listening to it, and that therefore they *might* show influences of that ambient language in their own vocalizations. On the other hand, there exists a tradition of belief that would seem to deny the very possibility of babbling drift. A discontinuity was

proposed between babbling and speech by Jakobson (1941/1968), who thought babbling (in the first year of life) consisted of random sounds from all the world's languages, while early speech (second year of life) consisted of universal syllable types regardless of the infant's language background. This viewpoint seemed to rule out babbling drift. However, later studies have proven Jakobson to have been incorrect—babbling is far from random, and a tendency to produce very similar universal syllable types ([ba], [ma], [da], [na], [wa], and [ja] being universally or near universally reported) can be seen both in canonical babbling and in early speech (Oller, Wieman, Doyle, & Ross, 1976; Stoel-Gammon & Cooper, 1984; Vihman, Macken, Miller, Simmons, & Miller, 1985).

The question remains, however, whether there may be additional effects on some subset of the syllables of babbling such that certain language-specific (non-universal) features of the ambient language (such as trills, dark l's, palatal fricatives, etc., sounds that occur relatively rarely in languages) may come to be present in babbling. If so, there could be some syllabic/phonetic elements occurring in babbling of infants from one language background that would be absent in infants from another background. A similar question arises with regard to possible ambient-language effects based on prosodic (tonal, intonational, or rhythmic) features reflected in babbling.

The substantive question of babbling drift, whether syllabic or prosodic, proves difficult to investigate. Although everyone acknowledges that there must be ambient-language effects on vocalization at some point in the second or third year, there have been both empirical and methodological challenges to claims that babbling drift occurs very early in life (Engstrand et al., 2003; Navarro, Pearson, Cobo-Lewis, & Oller, 1998, 2005). The controversies revolve around the interpretation of existing data in the context of crucial methodological difficulties (Oller & Eilers, 1998).

Interpretive and methodological difficulties in studying ambient-language effects

Determining what babbling is—A key issue in the study of ambient-language effects concerns the seemingly simple question of isolating babbling from lexically influenced vocalizations (e.g., de Boysson-Bardies & Vihman, 1991). In other words, how can we distinguish between babbling and vocalizations that may constitute real words learned from the ambient lexicon? The question is relevant even if an infant sound is merely an attempt to approximate an item from the lexicon of the ambient language and even in the absence of any awareness by the infant of an association of the lexical item with a meaning. And the question is relevant to the interpretation of ambient-language effects even if adult listeners do not *recognize* the specific relation between an infant utterance and a lexical item in the language.

This issue is important because as indicated above, researchers have generally sought to find ambient-language effects in pure babbling, which is to say babbling utterances that have been influenced by the phonology of the ambient language only, and not by its lexicon. While the distinction between pure babbling and words is often fuzzy in infancy, there exists a well-formulated method to draw the fuzzy distinction. In extensive work on the early development of vocabulary in English, Vihman and her colleagues (Vihman, Ferguson, & Elbert, 1986; Vihman et al., 1985; Vihman & McCune, 1994) have laid out an operational

distinction: Roughly speaking, words, in accordance with this reasoning, consist of any infant utterance that has at some point been produced in a consistent context and that resembles a specific adult word. Babbles do not meet these criteria, because they consist of phonetic shapes that have not been observed to be produced in a consistent context and/or have not been observed to resemble an adult word of the infant's ambient-language environment. Of course this definition cannot provide unambiguous differentiation for all infant utterances. One could reasonably ask: How can we ever determine categorically and unambiguously that there is no consistent context of usage for any infant utterance, and how can we be sure there is no resemblance between any infant utterance and some lexical item? There is always the potential in babbling to find some very general context of usage as well as some resemblance of any babbled utterance to some word, and thus we can never conclusively determine that any utterance is pure babble. Furthermore, the form of a babbled utterance can be relatively distant from the form of any particular word in a language but at the same time be influenced by relatively high frequency of occurrence of some general feature for words in that language (such as the English "ing" ending or the [i] ending of many two-syllable words). In this way an utterance that may seem to be pure babble could bear some language-specific influence from the lexicon of that language, without being obviously based on any individual word. This ambiguity about what babbling constitutes presents a potential interpretive problem in all research on ambient-language effects. Having emphasized these ambiguities, we hasten to add that the Vihman procedure provides a workable approach for laboratory research to drawing a distinction between words and babble, a distinction that caregivers must be able to make in order to understand their infants' emerging talk.

Our approach for the present study is to simply require listeners to judge whether or not each infant utterance under study is or is not a word in one of the languages in question, an approach that assumes the listeners should have some of the same capabilities for drawing a distinction between words and babbles as caregivers, at least in cases where they speak the ambient language of the infant. In this article, we introduce an additional step, which is to categorize infant utterances in terms of whether they include any canonical syllables (well-formed syllables which in infant vocalizations usually have a consonant-vowel structure such as [bi], [do], or [na]), the kinds of syllables that are overwhelmingly used in real words in all spoken languages. By adding this layer of canonical vs. non-canonical analysis, we offer the possibility of specifically evaluating listener ambient-language detection for infant utterances that may have an especially high probability of reflecting ambient-language *lexical* properties such as word shape—empirical evidence that utterances with canonical syllables are in fact far more likely to be judged as words than utterances without them will be presented below. These precautions do not, of course, eliminate the problem of interpretation posed by the fact that we can never be sure that any infant utterance is pure babble. But they do provide a useful breakdown of data to evaluate the extent to which ambient-language effects could be more easily detected in cases where lexical influence is likely to be prominent.

Procedural difficulties in studying ambient-language effects—In addition to the problem of differentiating babbling from early speech, all the common methods of studying

ambient-language effects are subject to other procedural difficulties. Three primary methods have been used: (1) forced-choice language-background judgments by human listeners, who judge whether utterances are from infants of one language background or another (i.e., an identification task) or whether a presented pair of utterances or utterance sequences are from infants of the same or different language backgrounds (i.e., a same-different task) (Atkinson, MacWhinney, & Stoel, 1970; Engstrand et al., 2003; Navarro et al., 1998, 2005; Thevenin, Eilers, Oller, & Lavoie, 1985); (2) human coding by phonetic transcription or other auditory descriptions of features such as syllabic stress or intonation for infants of differing language backgrounds (de Boysson-Bardies & Vihman, 1991; Edwards & Beckman, 2008; Lee, Davis, & MacNeilage, 2010; Lléo, Prinz, Mogharbel, & Maldonado, 1996); and (3) instrumental acoustic analysis, addressing features such as pitch, formant frequencies, or voice onset time, where vocalizations from infants of different ambient languages are quantitatively compared (Chung et al., 2012; de Boysson-Bardies, Halle, Sagart, & Durand, 1989; Eady, 1980; Eilers, Oller, & Benito-Garcia, 1984; Enstrom, 1982; Halle, de Boysson-Bardies, & Vihman, 1991; Levitt & Wang, 1991; Preston, Yeni-Komshian, & Stark, 1967; Rvachew et al., 2008; Whalen et al., 2007; Whalen, Levitt, & Wang, 1991).

All three methods are difficult to implement: confounds and other design limitations in all these methods can yield reasons to doubt outcomes. One problem is possible bias in observers (whether forced-choice listeners, transcribers, or acoustic analysts) not effectively blinded to the ambient language of the individual infants or utterances they examine, because recordings presented to observers may provide indications (independent of the infant utterances) of the ambient language. A second problem concerns potential bias of the individual(s) who select utterances from the language groups for presentation to observers. A third problem concerns yoking of observers to groups, a flaw that applies only to methods 2 and 3 since in method 1, observers hear utterances from both groups. A fourth problem concerns recordings made in different settings (e.g., Engstrand et al., 2003; Rvachew et al., 2008), where there may be differences in the recordings (e.g., noise level differences) that are confounded with possible ambient-language differences.

Another problem that affects all these kinds of studies is sample size. It is clear that infants vary within ambient language on phonetic and prosodic characteristics of their babbling as well as their rate of babbling and vocabulary learning (Vihman et al., 1985; Vihman & McCune, 1994). Consequently, comparison of small sample sizes from each ambient-language always leaves open the possibility that group differences obtained may result from mere sampling error.

Working with large samples and controlling for the biases discussed above is costly. For example, an optimal approach requires selection of utterances from recordings either by individuals with no awareness of the possible hypotheses or by automated procedures with no human intervention. Utterances from recordings would optimally be extracted such that no background voices or differentiating noises can be heard. In addition, utterance samples from each language group should always be dissociated (unyoked) from individual observers or selectors of utterances. Such controls can be instituted with much greater ease now than in the past, given improvements in the availability of digital recording and processing technologies. But even now the costs of implementing optimal procedures in such research

are high. As time passes, automated methods for utterance selection will surely allow larger sample sizes and make research on ambient-language effects far less vulnerable to the methodological concerns we have raised.

The following offers more detailed perspectives on method 1, forced-choice language-background judgments, its associated difficulties, and reasons that reported results from key studies have been doubted by some in the field.

Early support for the hypothesis of babbling drift was provided by Weir (1966), who described her experience with recordings of Mandarin-, English-, and Russian-learning infants between 5 and 17 months. She concluded from this informal research that “we were able to identify the Chinese-learning infants by distinct pitch patterns (p.155)”. In retrospect, it seems likely that Weir’s casual experience was influenced by her awareness of where the recordings had been made in each case, perhaps in part because the full recordings included the voices of adults speaking in the background in Mandarin, Russian, or English. Her apparent awareness (i.e., her lack of blinding to group) may have engendered potential bias in judgment. The likely correctness of this suspicion is supported by further work with Weir’s own dataset by Atkinson et al. (1970) using forced-choice language-background judgments. In this work listeners were presented isolated vocal segments from infants only, with no background voices, to ensure that the listeners were blind and could only have identified language background if the infants truly produced phonetic patterns specific to the ambient language. The researchers found, after implementing these controls, that the listeners could not reliably identify language backgrounds of the infants even at 16–17 months.

More recent interest in ambient-language effects has been particularly spurred on by results of de Boysson-Bardies, Sagart, and Durand (1984), who reported perhaps the most widely cited study of ambient-language effects, a study based on forced-choice language-background judgments of 15-s utterances produced by infants learning French, Arabic, and Cantonese. The authors reported better than chance performance of listeners for utterances produced at 6 and 8 months but a lack of consistent performance of listeners at 10 months. The result is puzzling because it suggests a reversal of the presumed ambient-language effect at the oldest age tested, a finding inconsistent with the babbling drift hypothesis. The authors suggested that ambient-language effects were discerned by prosodic features, but the authors also said the utterances from 10-month-olds were “poorer in prosodic cues (p.10)” than utterances from other age groups. But there are other reasons for skepticism about the conclusion of babbling drift in this study. The authors gave neither an indication that the person or persons who selected the utterances was/were blind to the research hypothesis, nor that there was a strict protocol to minimize possible bias in selection of utterances across the three languages. These concerns are amplified by the fact that there were only 32 utterances altogether presented to the listeners for the three languages. Furthermore, the text of the article does not rule out the possibility that adult voices or background noises may have been included in these long utterances, and that these voices might have served as cues by which listeners could identify language background. Finally, the recordings were made in different locations, presumably with the same equipment, but there is no indication in the text that gain settings were matched across locations, that microphone distance from the infant was

controlled, nor that any method was used to ensure that listeners would not be able to detect differences in the signal-to-noise ratio or other background noises in the recordings, differences that may have been correlated with the language backgrounds.

It is notable that other studies that have exercised caution with regard to the issues of possible biased utterance selection and possible extraneous voices in the background of utterances presented to listeners for forced-choice judgment did not find reliable ambient-language effects (e.g., Atkinson et al., 1970; Thevenin et al., 1985).

Rationale

The three methods, as indicated above, are all difficult to implement, and all of them are subject to the problems of determining whether infant utterances are babbles or words (or something intermediate between them). Additional work is needed to help clarify the nature of early ambient-language effects. In this continuing effort, we think it may be possible to turn the problem of differentiating babbling from early words into an advantage. By studying ambient-language effects in such a way that influences of the lexicon are addressed explicitly, we should be able to elevate the importance of the research. Instead of looking for pure phonological effects, we can evaluate the integrated effects of the ambient language, both phonological and lexical, on the vocalizations of the infant. It makes sense that infants are increasingly driven to adapt for communication as they grow, and they need increasingly to adapt to the lexicon to do so. But we cannot be sure when that adaptation in vocal production begins. This fact imposes a methodological difficulty that we wish to take into explicit consideration in our study. The approach we have adopted, then, continues efforts of Engstrand et al. (2003) and Navarro et al. (1998, 2005), who studied infants older (past the first year) than the ones in this article and addressed possible ambient-language effects separately for items deemed to be babbles and items for which it was determined that they were or could have been words. In addition, we have provided another layer of analysis, a layer associated with the possibility that infant utterances with canonical syllables might more readily reflect ambient-language properties than utterances without canonical syllables. Thus, we offer two ways (splitting the data by words vs. non-words and by utterances with and without canonical syllables) to consider possible effects of the lexicon on emerging ambient-language effects.

The research for this article was based on method 1, a forced-choice judgment paradigm. We asked adult listeners to determine whether randomly presented infant utterances pertained to a particular ambient language *and* whether they consisted of language-specific lexical items. We chose method 1 in part for practical reasons—it offers the possibility of assessing a large number of utterances with multiple listeners. Methods 2 and 3 (acoustic analysis and phonetic transcription, respectively) require considerably more time per utterance and thus restrict the number of utterances that can be evaluated. Of course, automated acoustic analysis could in the future be very efficient, but no reliable automated acoustic system has yet been developed for study of ambient-language effects. In addition, while methods 2 and 3 clearly both have merits in terms of potentially determining sources of ambient-language effects, it may be that human audition implementing intuitive forced-choice judgments has an advantage over any currently available acoustic or transcriptional method in terms of

identifying such effects, because the ear can assess utterances at a gestalt level, whereas current acoustic analysis methods require parameterization of acoustic features, and similarly phonetic transcription forces the listener out of the intuitive mode in order to allow the characterization of phonetic features and sounds. Human hearing must have been evolved to recognize many subtleties of speech including factors involving foreign accent (a factor that is by definition related to detection of language-specific phonetic or prosodic elements), and consequently it could be true that no current acoustic analysis or transcription method could match the ear in detecting ambient-language effects in infancy. Still, we grant that, whenever practical, it is worthwhile to use a variety of methods to assess possible ambient-language effects, and our laboratory has been involved in work with all three approaches (Eilers et al., 1984; Oller & Eilers, 1982; Thevenin et al., 1985). In this article, we provide acoustic analysis of fundamental frequency (F0) on the subset of utterances that showed the most reliable ambient-language effects.

We extracted a large sample of infant utterances without other voice overlay from English- and Chinese-learning infants at 8, 10, and 12 months. Besides Mandarin, the primary language of most educated people in Taiwan, there is an additional Chinese language, Taiwan Southern Min, sometimes called Taiwanese. In our study the Chinese recordings came from homes where Mandarin was present in all cases, but where Southern Min was also often present. By requiring listeners to categorize randomly presented utterances as English or Chinese and at the same time requiring them to indicate whether they recognized each utterance as a word in either language, we addressed potential ambient-language effects in a way that may shed light on how such auditorily discernible effects emerge, whether through pure phonological adaptations or through a combination of phonological and lexical learning. The criterion for treating an infant utterance as a word for analysis was that at least two listener judges had to deem that utterance to be a word during the forced-choice listening task. The criterion for identifying infant utterances as having canonical syllables was selected by the first author (not a listener in the forced-choice listening task) *before* any of the forced-choice listening tasks were conducted. By identifying utterances with canonical syllables, we provided the basis for an additional layer of analysis on the possible effect of the emergent lexicon on ambient-language effects.

The decision to have listeners in our study indicate possible words among the infant utterances is in part inspired by thoughts of Vihman and colleagues (e.g., Vihman & de Boysson-Bardies, 1994) who have provided a schematic model of perceptual-motor interaction in the first year, a model that emphasizes a role for “salient adult words” associated with the “articulatory filter” (DePaolis, Vihman, & Keren-Portnoy, 2011, 2013; Majorano et al., 2014). Before producing canonical syllables, infants have heard canonical syllables in speech for months, and it is reasonable to assume they will have been aware of salient adult words containing such syllables. We hypothesize that as infants begin to produce canonical syllables, they notice similarities of their own syllables to adult words and may begin to be influenced in their own syllable productions by these adult words. Across time, the infant’s vocal activity yields certain well-practiced syllable and syllable sequences forming motor production routines that can be thought of as a repertoire of vocal motor schemes (McCune & Vihman, 1987, 2001). When hearing adult speech, an infant filters the speech through his or her “articulatory filter,” according to this thinking. If infants in the

latter months of the first year do indeed establish a working relation between their own babbling sounds and salient adult words (as the model suggests), then it seems possible that some of their productions may begin in that period to resemble some of those adult lexical items. Listeners knowing the ambient language might be able to recognize these special productions as pertaining to the ambient language even if the phonetics of the infant productions consist entirely of universal or near universal syllable shapes. For example, if an English-learning infant says [babi], it could be interpreted by an American listener as the English word “baby,” and the infant’s ambient language could thus be identified, even though both syllables [ba] and [bi] could also occur in babbling of infants exposed to Chinese and many other languages. The point is that English has a very commonly occurring word with the shape CVCi, while Chinese does not. Thus, we hypothesize, even without any difference in the syllable repertoire available to English and Chinese-learning infants, there could be lexically-based items in babbling repertoires that could supply information to listeners about the ambient language.

Methods

Stimuli

Infant recordings and utterance selection—The data for the study were acquired from the archives of two longitudinal investigations on typically developing infants from monolingual English homes in Memphis and from Chinese (Mandarin and Southern Min) middle-class households in Tainan, Taiwan. In each case audio recordings with high fidelity equipment had been made across the first year of life for infants in each group. In both cases the recordings were made during natural interactions of parents and sometimes experimenters with infants. The data selected here overlapped with but were not identical to the data used in any of several prior studies drawn from these archives (Buder, Chorna, Oller, & Robinson, 2008; Chen & Kent, 2009, 2010; Oller et al., 2013; Ramsdell, Oller, Buder, Ethington, & Chorna, 2012).

Forty utterances from each recording of 8 English- and 8 Chinese-learning infants at three ages (8, 10, and 12 months, 3 males and 5 females in both cases) were extracted. There were 1920 utterances in total, half from English and half from Chinese recordings. Each utterance was extracted in accord with a breath-group criterion (Lynch, Oller, Steffens, & Levine, 1995), one utterance per breath group.

In order to limit the possibility of bias in utterance selection, we imposed a strict procedure. The first 40 canonical or precanonical vocalizations (excluding cry, laugh, or vegetative sounds) that met the following requirements were extracted from each recording: (1) no other voices could be heard within the extracted utterance; (2) no audio distortion (i.e., clipping of the amplitude spectrum) was present; and (3) audibility was sufficient so that the utterance would be deemed noticeable to a casual listener, and presumably to a caregiving interactor. The goal of the selection procedure was to ensure that each selected utterance would be free of any possible cue to the ambient language other than cues that might be found in the infant voice, and at the same time to ensure that each utterance was salient enough to present useful information about the infant vocalization capacity. Although we sought to limit possible bias in utterance selection by having strict criteria, it was not

possible to blind the selector (Lee) to the ambient language in the recordings. In future work we hope to employ automated selection procedures where potential selection bias can be reduced even further.

Noise-matching of the infant recordings—The two longitudinal studies from which the utterances were drawn were conducted independently. The equipment was different, the gain settings were different and variable, and the physical environments were different (home recordings in Taiwan and laboratory recordings in Memphis). Before presenting utterances to listeners for the present study, we took measures to ensure that these differences could not become the basis for an illusory differentiation of ambient-language background of the infants, when the differentiation might in fact be based on recognition of differences in the recordings. Thus, in a preparatory study, we evaluated the recording differences and found that the utterances from the audio recordings could indeed sometimes be differentiated from noise backgrounds alone. The recordings from Taiwan often had discernibly higher noise levels (lower S/N ratios) than those from Memphis (Lee, Jhang, Chen, & Oller, 2012).

We proceeded to an additional preparatory study to develop a procedure that would correct for the apparent noise-level differences (Lee, Jhang, Chen, & Oller, 2013a). First, 60 two-second background noise segments were extracted from the recordings (we used 30 Chinese and 30 English segments for this purpose), with all the recordings from all ages to be used in the present study being included in the selection of noise segments. Background noise segments were extracted such that no other sounds or voices could be heard in any segment. Second, 60 infant utterances were extracted, with 30 from each language group. Then a randomly selected noise segment from the Chinese segments was mixed with each of the English utterances. Then listeners were presented with three tasks in counterbalanced orders. In each task, they heard randomly ordered segments (half from the Chinese recordings and half from the English recordings) and were asked to judge each as having higher or lower noise content. In the first task only noise was involved, and the eight English and five Chinese listeners were significantly able to tell that one set of recordings (the Chinese recordings) had higher noise levels (65% correct, much higher than chance). In the second task the real infant utterances without noise addition (including the noise differences across the Chinese and English recordings) were presented, and listeners again could significantly discern (59% correct) the higher noise levels of one set of recordings (the Chinese recordings). But in the third task, where the noise levels had been equalized by mixing of background noise from the Chinese recordings into the English recordings, the listener performance fell to chance (<50% correct).

With these results in hand, we adjusted all the English utterances selected for the present study using the procedure of noise mixing. For each of the 960 English utterances, a noise segment was added, drawn at random from the noise background segments of the Chinese recordings, with all the Chinese recordings from all ages being included in the selection of noise segments. With these precautions we guarded in the present investigation against possible illusory differentiation of the utterances by ambient language when in fact the differentiation might have been based on noise-level differences.

Adult utterances composed of pseudo words—Previous published research has not to our knowledge ever provided a basis to reference reported effect sizes in ambient-language studies against effect sizes for detection of phonological/prosodic differences in the speech of mature native language users. In order to provide perspective on the magnitude of any possible effects of ambient-language identification based on infant utterances, we resolved to compare those effects with ambient-language identification of adult utterances lacking lexical identifying information. We thus constructed a set of pseudo words/sentences for both English and Mandarin, where each pseudo word was required to obey the phonotactic constraints of its language. The 40 English-like ($n = 20$) and Mandarin-like ($n = 20$) pseudo words/sentences were recorded by native speakers (Oller for English, Lee for Mandarin) first and then the recordings were used as models for each of the native speakers (five for English, five for Mandarin) to mimic as they made recordings of the pseudo words/sentences that would be used to present to listeners for ambient-language judgments in our study. The syllable sequences were pronounced with the same tones and intonations as occurred in the real conversations that had been used as models by Lee and Oller to construct the pseudo words/sentences (see “original transcript” Table 1). The stimuli were all phonotactically well-formed for the language, but completely meaningless. We took the precaution of ensuring that the listeners for the forced-choice judgments did not know the native speakers who recorded the sentences, so we could be sure that the listeners judged the ambient language of the pseudo sentences based on the content, not on the voices of speakers familiar to them. The conversations plus the pseudo words used to replace them are represented orthographically and in International Phonetic Alphabet (IPA) for our stimuli in English and phonetically using the Zhuyin fuhao (i.e., Mandarin phonetic alphabet, taught in all elementary schools in Taiwan) and IPA in Table 1.

Listener participants

Seven Chinese (Mandarin and Southern Min) and six American-English listeners participated in the forced-choice listening task (all females; ages 20–30). The English listeners were monolingual graduate students in the School of Communication Sciences and Disorders at the University of Memphis, while the seven Chinese listeners were native speakers of Mandarin and secondary speakers/listeners of Southern Min, three of whom were graduate students, the remainder undergraduates. The Chinese speakers were also speakers of English as a second language having studied in the Foreign Languages and Literature Department at the National Cheng Kung University in Tainan, Taiwan. According to self-report and report of the faculty supervisor in Taiwan (Chen), all the Chinese listeners had been studying English for at least 15 years, and they all had conversational English skills. The American-English speakers were aware that Mandarin is a tonal language, but they had never studied Mandarin or any other tone languages as foreign languages. All the listeners from both language backgrounds had received training in phonetic transcription. None of the listeners had a history of sensorineural hearing loss. The fact that the Chinese listeners knew both English and Chinese while the American listeners know only English imposes, of course, a requirement that we provide results for the listener groups separately for comparison.

Both Mandarin Chinese and Southern Min are tone languages, and both are Chinese languages, but they are not mutually intelligible. The two languages share many phonological and lexical characteristics, but the phonological system of Southern Min is more complex than that of Mandarin. Mandarin allows only /n/ and /ŋ/ in syllable-final position, while /n/, /ŋ/, /m/, /p/, /t/, /k/ are allowed at the end of syllables in Southern Min. Mandarin has four lexical tones: high-level, rising, falling-rising, and falling (Yip, 2002), while Southern Min has 7: high-level, high-falling, low-level, high-falling-short, low-rising, low-falling-short, and low-falling. Southern Min has a three-way stop voicing contrast, including voiced, aspirated, and unaspirated stops, while Mandarin only has aspirated and unaspirated stops (Chung, 1996). Southern Min is widely spoken in Taiwan, but Mandarin is the official language used in schools and is more frequently used than Southern Min (Ministry of Education, 2015).

Listening procedure

Listeners were presented with the 1920 utterances in random order in 16 sets of 120 utterances, each through earphones. The order of presentation of sets was counterbalanced. Listeners were told that in each set of utterances, half the utterances were from English-learning infants and half from Chinese-learning infants. For each utterance, listeners gave a forced-choice answer to the following two questions: “Is the utterance from a Chinese- or English-learning infant?” followed by, “Is the utterance a word or words?” If listeners could not recognize the utterance as a word(s), they answered “no”. Otherwise they entered a gloss of the word, for example, “doggie”. After the judgments for the infant utterances were completed, the listeners were also presented with the 40 adult pseudo words/sentences in random order. Here, the question was “Is this an English-like or Mandarin-like utterance?” For both infant utterances and adult pseudo words/sentences, listeners were allowed to listen to each utterance three times.

Data analysis

Exclusion of a small number of utterances and judgments—Three utterances from the English-learning infants were excluded because after applying the noise-mixing method, the background noise was so loud it effectively obscured the original utterances, yielding 1917 utterances for analysis. Twelve responses from Chinese listeners were missing due to technical problems, so a total of 24,909 responses were analyzed.

Subcategorization of infant utterances as words and as utterances containing canonical syllables (CB)—The data included a breakdown of infant utterances such that some were deemed to be words, and some not. We assigned an utterance to the category of words if at least 2 of the 13 listeners judged that particular utterance to be a word. The data were also broken down by whether each utterance was deemed to have a canonical syllable or not. For this breakdown, the lead author, who was not a listener in the forced-choice judgment task, listened to all 1920 utterances and judged them to include or not include a canonical syllable *before* any of the coding was conducted. To assess the reliability of canonical syllable subcategorization, we ran a separate study using the same set of 1920 infant utterances (Lee, Jhang, Chen, & Oller, 2013b). Agreement between the two coders (the first and second authors) on canonical babbling ratios (the number of canonical

syllables divided by the total number of syllables) on the forty-eight 40-utterance samples was $r = 0.91$ ($8 \text{ infants} \times 3 \text{ ages} \times 2 \text{ languages} = 48$).

To assess our assumption that utterances judged independently to have canonical syllables would be more likely to be categorized as words by the forced-choice listeners, we conducted odds tests. About 15% of utterances ($n = 286$) in the sample ($n = 1917$) were deemed to have canonical syllables. 6% of utterances ($n = 120$) were judged to consist of words. Yet 31% that were judged to have canonical syllables were also judged to be words ($n = 88$, $88/286 = 0.31$), and 73% that were judged to be words were also judged to have canonical syllables ($88/120 = 0.73$). The odds then that an utterance not judged to have canonical syllables would be categorized by the listeners as a word according to our criterion was 2% (32 words out of 1631 utterances without canonical syllables), while the odds that an utterance that was judged canonical would be categorized as a word according to our criterion was 31%, indicating that an utterance with canonical syllables was 15.7 times more likely to be judged a word than an utterance without canonical syllables. We also extracted F0 (fundamental frequency) for the final canonical syllable in each of these 88 utterances in Praat (Boersma & Weenink, 2014).

Statistical approaches—For the initial analysis (not distinguishing among the 1917 utterances between listener-identified words and non-words nor between utterances with or without canonical syllables), a Generalized Linear Mixed Models was used since outcomes were binary coded, 1 correct and 0 = incorrect judgments. Analysis was done using SAS Proc GLIMMIX. In this model, there were 2 within-subjects variables (infant ambient language [Chinese or English], infant age [8, 10, or 12 months]), 1 between-subjects variable (listener language), 2 random variables (individual infants and listeners), and 1 dependent variable (listeners' scores).

In these additional analyses, three mixed models analyses of variance were conducted. The first model consisted of two within-subjects variables (infant ambient language [Chinese or English] and word status [words or non-words]), one between-subjects variable (listener language), two random variables (individual infants and listeners), and one dependent variable (listeners' scores). The second model included two within-subjects variables (infant ambient language [Chinese or English] and canonical syllable status [CB or non-CB]), one between-subjects variable (listener language), two random variables (individual infants and listeners), and one dependent variable (listeners' scores). The third model included one within-subjects variable (canonical words [canonical words or non-canonical words]), one between-subjects variable (listener language), and one dependent variable (listeners' scores). The variables words and canonical words were highly correlated (words, Tolerance = 0.28, VIF = 3.5; canonical words, Tolerance = 0.27, VIF = 3.69), while the variable canonical syllable status in the main model caused over-fitting. Thus, in addition to the initial analysis for all the data, we conducted additional analyses to avoid the problems of collinearity and over-fitting.

We present results and figures using proportion correct. We also show sensitivity (d') and bias values (C) (Macmillan & Creelman, 2004) to indicate performance of individual listeners across the whole set of utterances and to indicate possible bias of listeners to judge

infant utterances predominantly as from English or from Chinese language backgrounds (accounting thus for possible “native language” or “own language” bias).

Results

The initial analysis mixed model results

The first analysis (Table 2, Column 1) showed no significant main effects of listener language or infant ambient language. There was, however, a significant main effect of infant age ($p < .001$). The listener language and infant ambient language main effects were not significant. Two two-way interactions were significant: infant age by infant ambient language ($p < .0001$) and infant ambient language by listener language ($p < .0001$). The listener language by infant age interaction was not significant. There was also a significant three-way interaction of infant age by infant ambient language by listener language ($p < .003$).

The additional analysis (first model, Table 2, Column 2) showed a significant main effect of word status ($p < .0001$), indicating that ambient-language detection was significantly higher for words than for non-words, but the main effects of infant ambient language and listener language were not significant. The two-way interactions were significant: infant ambient language by word status ($p < .0001$), listener language by word status ($p < .0001$), and infant ambient language by listener language ($p < .0001$). There was also a significant three-way interaction of word status by infant ambient language by listener language ($p < .0001$).

The additional analysis (second model, Table 2, Column 3) for effects of canonical syllables (CB) showed a significant main effect of canonical syllable status ($p < .0001$), indicating that ambient-language detection was significantly higher for utterances containing canonical syllables than for other utterances, but the main effects of infant ambient language and listener language were not significant. The interactions of infant ambient language by canonical syllable status ($p < .0001$) and infant ambient language by listener language ($p < .0001$) were significant, but listener language by canonical syllable status was not significant. There was no significant three-way interaction.

The additional analysis (third model, Table 2, Column 4) assessing effects of words containing canonical syllables (canonical words) in contrast with all other utterances showed a significant main effect of canonical words ($p < .0001$), indicating that ambient-language detection was significantly higher for canonical words than for other utterances, but the main effect of listener language was not significant. The interaction of listener language by canonical words was not significant.

Because the Chinese listeners knew both English and Chinese, they were in a position to judge word status in many more instances than the English listeners, and they could have had an advantage in determining ambient-language effects in non-words as well. Consequently, our interpretation of the above effects needs to consider this complexity. Below, we present the data in various graphic illustrations, and offer interpretations adjusted to account for the listener group differences.

The mixed model results also indicate whether ambient-language detection was significantly better than (or in some cases worse than) chance for the various subcategories of judgment. We will consider these outcomes along with the graphic illustrations.

Descriptive results in graphic form

The overall performance of the listeners for the entire set of utterances was 50% correct for English listeners and 52% correct for Chinese listeners. Figure 1A shows the performance of the two groups of listeners for the infant utterances at 8, 10, and 12 months, indicating a small, non-significant ambient-language effect overall. The blue symbols, representing English listeners, show a non-significant worse-than-chance performance at 8 months and a slightly better-than-chance performance at 10 and 12 months, whereas the red symbols, representing Chinese listeners, show a better-than-chance performance at all three ages. There was no significant infant age by listener language interaction and none of the differences from chance at any age or for either listener group was significant.

Figure 1B shows the performance of the two listener groups combined for utterances from English- and Chinese-learning infants at the three ages. The judgments on Chinese-learning infants (blue symbols) showed increased ambient-language detection over time, with significantly better- than-chance performance at 12 months ($\beta = -0.21$, $SE = 0.09$, $p < .03$). In contrast, utterances from English-learning infants (red symbols) showed no such consistent change and no significant differences from chance. The difference between infant groups on infant age was reflected in the significant infant age by infant ambient language interaction, mentioned above.

Figure 2 helps shed light on the complex results of Figure 1. Figure 2 shows the performance of the two listener groups for the utterances from the English- and Chinese-learning infants at each *age*. The English listeners performed *significantly worse* than chance with Chinese-learning infants at 8 months ($\beta = 0.33$, $SE = 0.10$, $p < .002$), but not significantly different from chance for any of the other possible comparisons. The Chinese listeners performed significantly better than chance at 10 months ($\beta = -0.24$, $SE = 0.09$, $p < .02$) and 12 months ($\beta = -0.29$, $SE = 0.09$, $p < .005$) for Chinese-learning infants.

The significant three-way interaction (infant age by infant ambient language by listener language) mentioned above is reflected in the differing patterns across the three panels of Figure 2. The Chinese listeners did better with the Chinese-learning infants at all ages. In contrast, the English listeners did better with the English-learning infants at 8 and 10 months but better with the Chinese-learning infants at 12 months. This pattern reveals a bias effect of listener language background (the native-language or “own language” bias effect). Chinese listeners tended to make more Chinese judgments and thus tended to be correct more often for Chinese-learning infants than for English-learning infants. Such misleading correctness on one language would happen even with no real ambient-language detection since there were an equal number of Chinese and English stimuli. Similarly, English listeners tended to make more English judgments (but only at 8 and 10 months), and thus were more often correct for English than for Chinese. This native-language bias effect may be thought to account for the significantly poorer than chance performance of English listeners on Chinese-learning infants at 8 months.

Table 3 shows the performance of individual listeners across the whole set of utterances for proportion correct, sensitivity (d'), and bias (C). d' reflects the difficulty of ambient-language detection; that is, the higher the d' , the better the listeners differentiated the two infant groups, with positive values reflecting correct identification and negative values incorrect identifications. C is used to measure listener bias toward one language over the other; $C = 0$ means that an individual does not show bias toward either of the languages, while values above or below 0 indicate bias. Nine of the 13 listeners showed d' above 0 but below 1, indicating that they were able to differentiate the two groups some of the time but not often. In contrast, four of the listeners showed d' below 0, indicating that they were consistently guessing the wrong language more often than guessing the right language. As for bias, only 7 of the 13 listeners actually showed a bias value corresponding to their own language—the others chose the non-native language more often than the native language. Although we noted an overall native-language bias effect in Figure 2, clearly only some of the listeners showed the effect.

Figure 3A shows the performance of individual *listeners* across the set of utterances. For the blue bars, representing English listeners, none performed better than chance. But for the red bars representing Chinese listeners, FHL performed significantly better than chance ($\beta = -0.13$, $SE = 0.05$, $p < .02$), while EHU performed significantly worse than chance ($\beta = 0.12$, $SE = 0.05$, $p < .03$). Thus, the Chinese listeners detected ambient language slightly better than English listeners on the entire set of utterances, but only one individual showed this effect significantly, while another individual showed the paradoxical significant tendency to identify the wrong language. Across infants, differences from chance in language judgment were larger than in the case of listeners (see Figure 3B). Seven of the infants were identified at significantly different from chance levels, but four were assigned to the wrong language. This paradoxical identification occurred for two English-learning infants (KA: $\beta = 0.19$, $SE = 0.09$, $p < .05$; KS: $\beta = 0.25$, $SE = 0.09$, $p < .02$) and two Chinese-learning infants (XJ: $\beta = 0.20$, $SE = 0.09$, $p < .04$; YL: $\beta = 0.34$, $SE = 0.09$, $p < .002$). On the other hand, the right language was identified for only two Chinese- and one English-learning infants at statistically significant levels (EA: $\beta = -0.24$, $SE = 0.09$, $p < .02$; QC: $\beta = -0.28$, $SE = 0.09$, $p < .006$; SH: $\beta = -0.35$, $SE = 0.09$, $p < .002$).

Figure 4A (and see Table 2, Column 2) shows performance broken down by *listener groups* for infant utterances where at least two of the thirteen listeners judged utterances as words as opposed to all other utterances (non-words). This figure illustrates the significant interaction of listener language by word status ($p < .0001$). 120 of the 1917 utterances, only 6% (120/1917), were judged to be words by this criterion, with 70 Chinese (CH) utterances and 50 English (EN) utterances being deemed words. Only English listeners did significantly better than chance for words ($\beta = -0.49$, $SE = 0.10$, $p < .0001$). Interestingly, the English listeners showed performance above 50% on both groups of infants for words (66% for English-learning infants and 57% for Chinese-learning infants, the left-hand blue bar representing the combination). The Chinese listeners showed 68% correct for words produced by Chinese-learning infants, while obtaining only 36% correct for English-learning infants' words (the combination being represented by the red bar on the left in the figure). In the next paragraph we offer a suggestion to explain the unexpected interaction of listener language by word status: the surprises were that English listeners (even though they

only knew English) did better than chance on words from infants of *both* language backgrounds, whereas Chinese listeners did not perform above chance (even though they knew both languages).

The mixed model results in Figure 4A may have been skewed by the 28 “mama” utterances (21 from English-learning infants and 7 from Chinese-learning infants). “mama” was by far the most frequently occurring word identified by the listeners. Because “mama” means “mother” in both English and Chinese, the listeners tended to guess their own native language, and consequently the English listeners tended to get many more answers correct on “mama”, while Chinese listeners tended to get many more “mama” utterances wrong. We also examined other utterances portrayed by the listeners as various canonical forms that could be words in one of the languages or the other (e.g., “aba” could mean “father” in Chinese, “dada” could mean “father” in English). In no case was the imbalance of distribution for production of these canonical forms by English-learning infants vs. Chinese-learning infants so strong as in the “mama” case. Thus, we presume that the unexpected imbalance in the data on “mama” produced mostly by one English-learning infant in a single session ($n = 16$ at 10 months) may have played a role in the unexpected interaction of listener language by word status.

Figure 4B (and see Table 2, Column 2) shows the same utterance breakdown as 4A, words ($n = 120$) vs. non-words ($n = 1797$). Consistent with the pattern in Figure 4A, language detection for words was significantly better than for non-words. In this case, the judgments for Chinese-learning infants (but not English-learning infants) on words showed significantly better than chance language detection ($\beta = -0.53$, $SE = 0.10$, $p < .0001$). Both groups of *listeners* showed higher than 50% correct on words of Chinese-learning infants (a combined effect of 57% for English listeners and 68% for Chinese listeners, represented by the left-hand red bar in Figure 4B). Neither words nor non-words from English-learning infants showed significantly above-chance language detection.

How might we explain the better performance on words from Chinese than from English infants? We speculate that: (1) English listeners made “correct” identifications through native-language bias on the large number of “mama” utterances from English infants, and these “correct” identifications were somewhat counterbalanced by the many “incorrect” judgments by Chinese listeners on these same utterances; (2) the remaining words (not “mama”) from English-learning infants may have provided no real information to either listener group regarding ambient language; and (3) the words from Chinese-learning infants appear to have included real ambient-language indicators since *both* groups of listeners, and especially the Chinese listeners, performed better than chance on them.

Figure 4C (and see Table 2, Column 3) shows the performance broken down by *listener groups* for the subset of utterances identified by the first author *before* any of the coding was conducted as having at least one canonical syllable (CB, $n = 286$, CH $n = 163$; EN $n = 123$) vs. all other utterances (non-CB, $n = 1631$). 88 of the 120 words were deemed CB, so they are also represented here among the 286 canonical utterances. Ambient-language detection was better for CB than for non-CB ($p < .0001$). Both the English listeners ($\beta = -0.18$, $SE = 0.08$, $p < .04$, 51% for English-learning infants and 58% for Chinese-learning infants,

combined effect represented by the left-hand blue bar, Figure 4C) and Chinese listeners ($\beta = -0.19$, $SE = 0.08$, $p < .02$, 46% for English-learning infants and 64% for Chinese-learning infants, left-hand red bar) did significantly better than chance. In contrast, non-CB were not identified reliably by either group of listeners. The data from the two listener groups combined showed significantly better-than-chance performance on CB ($\beta = -0.18$, $SE = 0.07$, $p < .01$), but not on non-CB ($\beta = -0.03$, $SE = 0.06$, $p = .59$).

Figure 4D (and see Table 2, Column 3) shows the performance broken down by *infant groups* for CB and non-CB as in Figure 4C. CBs from Chinese-learning infants were judged significantly better than chance ($\beta = -0.43$, $SE = 0.09$, $p < .0001$) (57% for English listeners and 68% for Chinese listeners, combined in the left-hand red bar in Figure 4D), but neither CBs nor non-CBs were identified significantly differently from chance for English-learning infants (CB: 51% for English listeners, 46% for Chinese listeners, left-hand blue bar; non-CB: 53% for English listeners, 50% for Chinese listeners, right-hand blue bar). To explain the significant interaction ($p < .0001$) of infant ambient language by canonical syllable status seen in Figure 4D, we invoke the native-language bias effect. In this data breakdown, there were more CB utterances from Chinese-learning infants than English-learning infants (CH $n = 163$; EN $n = 123$) which, considering the native-language bias effect in listeners, may explain why Chinese-learning infants showed a higher proportion of “correct” answers than English-learning infants on CBs (left-hand red bar).

The results in Figure 4A–D suggest that infant utterances identified as including lexical items (words) and infant utterances including canonical syllables (CBs) can sometimes provide reliable cues about language background. Still the patterns of outcome were complex when broken down by listener- and infant-language background. The patterns may have been influenced by the native-language bias effect and by the fact that the distribution of both words and CBs was not even across the two language groups, presumably a result of sampling error.

Figure 5 shows three types of results: (1) overall correctness of each listener language group on the entire set of utterances (left pair of bars) based on the main mixed model analysis (Table 2, Column 1); (2) overall correctness on the subset of utterances that were both CBs and words (i.e., canonical words, Table 2, Column 4, represented in the middle pair of bars in Figure 5, $n = 88$, where EN $n = 36$, CH $n = 52$); and (3) on pseudo words/sentences pronounced by adult native speakers (right pair of bars, $n = 20$ for both languages). For these three types of results: (1) *Neither* group of listeners showed significantly above chance performance on the entire set of utterances (left pair of bars); (2) *both* groups of listeners showed significantly above chance performance on the canonical words (middle pair of bars) (EN: $\beta = -0.18$, $SE = 0.08$, $p < .03$; CH: $\beta = -0.20$, $SE = 0.07$, $p < .009$), although even here the overall proportion correct was only $\sim .60$; and (3) the total number of stimuli for adult pseudo words/sentences was too small to run a mixed-model analysis, so we used a binomial test for each listener’s results with means across the listeners represented by the right-hand pair of bars. In stark contrast to results where listeners judged infant utterances (left-hand and middle pairs of bars), listeners in both groups judged language background on the pseudo words/sentences pronounced by adults overwhelmingly correctly, with $\sim .95$ of

utterances assigned to the right language. Every listener showed significantly better than chance performance ($p < .0001$).

Follow-up explorations

To help clarify possible sources of the significantly better-than-chance performance by listeners on canonical words, we performed a number of explorations in the data, listening to utterances, observing the word “characterizations” by the listeners and the CB characterizations by the authors, and we attempted to find explanations for why some utterances showed higher or lower correct identification than others. These explorations yielded nothing to which we can confidently attribute the correct or incorrect ambient-language judgments.

One of these efforts provided hints worth mentioning. After listening to the canonical words, we hypothesized that perhaps subtle effects of tone in Chinese and of the trochaic stress tendency in English were reflected in the canonical words. If so, we reasoned, the occurrence in infant language input of high-level and rising tones in final syllables of Chinese words might have produced particularly high pitch on the final syllables of infant utterances deemed words by the listeners, while lower pitch on the final syllable may have corresponded to the trochaic pattern of stress in English words. Similarly, we reasoned that infant language input with rising tone in final syllables of Chinese words might have produced a greater tendency in final syllables identified as words from the Chinese-learning infants to have more rising fundamental frequency (F0).

To evaluate this possibility, we performed an acoustic analysis on the 88 canonical words, extracting F0 for the final canonical syllable in each one. Indeed, the utterances from Chinese-learning infants tended to have higher mean final-syllable F0 than utterances from English-learning infants (371 vs. 347 Hz). This effect was marginally significant by independent samples t -test ($p = .06$). We also compared the degree of F0 fall (start F0 minus end F0) across the final syllables of the utterances. The t -test here showed no notable tendency for the words from Chinese-learning infants to have more rising F0 ($p = .48$). The tendencies for the F0 to be higher on final syllables of utterances from the Chinese-learning infants may indeed suggest that in canonical words, F0 may have played a role, although a weak one in correct identification. We say a “weak role” because the listeners did tend to judge high final F0 (by median split) as Chinese (59% of cases), but they did not tend to judge low final F0 as English (46% of cases) in the infant utterances.

Thus, our acoustic analysis, similar to the outcome with our other explorations, did not yield a clear explanation for how the listeners may have correctly identified ambient language for canonical words. We should reemphasize that the sample of canonical words was small ($n = 88$), and only 4 of the 16 infants contributed more than 70% of these 88 utterances. Furthermore, as indicated earlier, a small number of utterance shapes (in particular “mama”) accounted for a disproportionate number of the canonical words, and a disproportionate number of these came from a single infant at 10 months. The results on these words, then, may be skewed by idiosyncrasies of particular utterance shapes from a few of the infants.

Discussion

General findings

The data depicted in Figure 5 provide a summarial perspective on the results. The listeners in this study did not strongly identify ambient language for the overall dataset ($n = 1917$) of infant utterances extracted from recordings of 8 English-learning infants vs. utterances extracted from recordings of 8 Chinese-learning infants. The lack of overall ambient-language detection applied both to 6 listeners of English-language background and to 7 listeners of primarily Chinese-language background, who also spoke English. At the same time, the subset of infant utterances that were judged as including a word in either language and as including at least one canonical syllable (canonical words) were identified at reliably better than chance levels by both listener groups, although correctness of identification was only about 60%. The inclusion of judgments by the same listeners on pseudo words/sentences produced by unfamiliar native-speaking adults of the two languages helps put the phrase “only about 60%” in perspective. The pseudo words were constructed to mimic the phonetics/phonotactics of the two languages, and they were produced by native adults with natural-sounding prosodic features, while including not a single real word from either language. Here the listeners performed at levels approaching the ceiling for the task ($> 95\%$ correct), even though the natural-sounding utterances were completely meaningless. We know of no prior study illustrating this perspective on effect size for ambient-language detection: the phonetics/phonotactics of English and Mandarin provide very clear evidence of the language source of utterances to adult listeners—and this clarity was just as great for the English-background listeners (who did not know Chinese) as for the Chinese-background listeners (who spoke both languages). Thus, even for infant utterances judged as canonical words, ambient-language identification was far below the level of language identification based on the phonetics/phonotactics of adult speakers. Post hoc explorations of language detection in the infant utterances did not yield clear indications about how listeners may have discerned ambient-language effects in canonical words.

We think the very limited nature of ambient-language detection here deserves attention, because the present study implemented a variety of precautions (large number of utterances, matching of background noise, elimination of utterances with overlaid voices, etc.) designed to prevent possible biases in the results, precautions that appear not to have been taken in widely cited studies reporting ambient-language effects (e.g., de Boysson-Bardies et al., 1984, 1989).

But the story is complex for the present results when we examine them in detail, and they leave open a more nuanced interpretation of the relation between results of the present report and prior reports of significant overall ambient-language effects. For example, there was a significant infant age effect for our overall data set, suggesting some improvement in listeners' scores for older infants, and in fact, utterances from the Chinese-learning infants were identified at reliably better than chance levels (55–56% correct) according the mixed-model analysis at 12 months (Figure 1B). To complicate the issue, however, listeners' judgments tended to favor their own primary language—although they were told before making any forced-choice judgments that half the utterances would be from English-

learning infants and half from Chinese-learning infants, both listener groups judged the majority of utterances to be of their own language background. This pattern of native-language bias appears to have yielded, for example, a statistically significant tendency of English-background listeners to reliably judge 8-month old Chinese-infant utterances falsely (Figure 2).

Variability of performance by the individual *listeners* (Figure 3A) and especially on the individual *infants* (Figure 3B) based on the overall data set highlight concerns about small numbers of infants and/or listeners in such research. Only two listeners showed overall reliable difference from chance, while judgments on seven *infants* differed reliably from chance. Yet these differences provide no obvious support for the occurrence of ambient-language effects because more than half these cases corresponded to paradoxical outcomes where the identifications were statistically reliably *incorrect*. The results suggest listeners may sometimes detect systematic differences between infant groups that correspond to a specific language in some way, but not always in the way predicted by the expectation of babbling drift. Why judgments sometimes show this paradoxical pattern is a mystery to us. The result supports the methodological conclusion that a very small number of infants in studies of ambient-language effects may well produce a misleading outcome. Regarding listeners, the same conclusion applies; we found one listener to show significantly correct language identification, and one to show significantly paradoxical below chance identification.

The most interesting positive finding concerns infant utterances identified as words and/or identified as including canonical syllables (CBs) (Figure 4). While the results in this realm were also complex, the overall patterns suggest statistically reliable success of listeners in identifying ambient language for words and for utterances with canonical syllables, but not for non-words or utterances without canonical syllables. The pattern is reminiscent of the outcome of Engstrand et al. (2003), who found ambient-language detection (with 12 and 18 month-olds) only in cases where utterances were judged either to be words or likely words. The result suggests that the study of ambient-language effects in “babbling” may offer the unexpected opportunity to evaluate subtle effects of the emergent lexicon in vocalization of infants.

The outcomes regarding words also hint at the possibility that prior reports of ambient-language effects in presumed babbling may have been influenced by judgments, transcriptions and/or acoustic measurements of infant utterances thought by laboratory staff to be babbling when in fact they were either lexical items or utterances that had been influenced by language-specific lexical patterns. We expound upon this speculation in the following section.

The fuzziness of the distinction between babbling and words

Let us illustrate the problem of differentiating words from babbling using the ambient languages considered in this article. Suppose an English-learning infant produces a two-syllable sequence interpreted transcriptionally as [bobi]. Suppose further that neither the parents nor laboratory personnel highly familiar with the infant could assign that form to any word present in the ambient language, and that the word has not been identified with any

particular context. This utterance should, as we understand Vihman et al.'s (1985) and Vihman and McCune's (1994) criteria, be deemed to be a babble, not a word. And yet there is a basis upon which the shape of this utterance might have been influenced by other very common word-shapes specific to the English lexicon (and especially common in child-directed speech), words such as doggy, birdie, baby, or kitty, all of which share with [bobi] the template CV₁Ci, where V₁ = i, and which are associated with an English-specific diminutive implication (Chinese does not have a diminutive word-final [i]).

Now consider a similar possible circumstance with Chinese as the ambient language. Suppose the child produces a two-syllable utterance interpreted phonetically as [atu], and again let us assume there is no basis for claiming the item to be a word, based on the Vihman criteria. A syllable with the form /α1/ (1 = high-level tone) functions as a prefix in Chinese in many two-syllable items. For names of people such as /hwa2/ or /mej3/ (2 = rising tone, 3 = dipping tone), one can prefix the /α1/ syllable to produce /α1hwa2/ or /α1mej3/ to address the person in the familiar (with an implication of familiarity or intimacy such as that associated with "tu" in Spanish). These forms are added to many names when speaking to children. Also, the /α1/ prefix is associated with common words such as /α1i2/ ("aunt"), as well as /α1ko3/ and /α1maw1/ (which loosely translated mean something like "John Doe", more literally translated as "some dog" or "some cat"). The /α1/ prefix is also used commonly as an onset for many animate nouns when speaking to a child (e.g., /α1te^hαŋ2/ for "giraffe" or /α1to1/ to designate a particular well-known cartoon character "Doraemon").

Although we assume for our example that [atu] and [bobi] would not be treated as words by parents, the disyllabic word shapes provide a rational basis for listeners to guess the language background with which they are associated. Even though the phonetics/phonotactics of both languages allow *all* the syllable types involved in both example words, the [a] prefix is presumably much more frequent in Chinese, and the [i] ending much more common in English. So listeners might judge, with better than chance accuracy, [atu] as based on Chinese and [bobi] as based on English, even though neither lexical form actually occurs (at least not frequently) in either language. Thus, the difference in frequency with which infants from the two languages might produce the two forms could be indirectly lexical in origin, because there are real words in each language that might predispose the child, under the influence of the ambient language, to selectively produce syllable sequences such as those in the examples. If so, the apparent ambient-language identification would not be purely segmental/phonetic, but would instead be at least partially lexical.

The examples can also be considered in light of language-specific prosody. English-learning infants' preference for trochaic patterns (e.g., strong/stressed syllables precede weak/unstressed syllables) in English words (e.g., "baby" or "happy") has been well proven in both speech perception and production research (Johnson & Seidl, 2009; Jusczyk, Cutler, & Redanz, 1993; Vihman, DePaolis, & Davis, 1998). If [bobi] is pronounced with a trochaic stress pattern and [atu] is pronounced with two high-level tones, the likelihood of judging the former as English and the latter as Chinese may increase, because the two prosodic patterns occur at relatively high frequency in English and Chinese respectively. On the other hand, laboratory staff following the Vihman criteria might not treat these utterances as words because infants might not use them sufficiently consistently in specific contexts to meet the

criteria. The words *could* be pure babbling, but how would we know for sure? These examples are intended to illustrate that it is hard to draw a clear distinction between babble and words because infants may be *targeting* words (or be influenced by the patterns of previously targeted words) even in apparent babbling. Engstrand et al. (2003) and Navarro et al. (1998, 2005) argue for more direct attempts, in research on ambient-language effects, to assess the role of the lexicon in apparently babbled utterances. These studies made such attempts by asking judges to indicate whether infant utterances were words or might be words (Engstrand et al., 2003) or by obtaining parent opinion about words occurring in their infants' recordings as well as laboratory listener opinion about words during their forced-choice task (Navarro et al., 1998, 2005).

These considerations illustrate that there will always be a hard problem of methodology in studies of ambient-language effects as long as we attempt to draw a hard distinction between babbling and words. Our inclination is to give up the attempt at drawing a hard distinction in favor of addressing lexicality directly and accepting the fuzziness of the distinction in vocal development and early word learning. Results from the present study, plus those of Engstrand et al. and Navarro et al., along with Vihman's work on establishing criteria for identifying words as they emerge, provide foundations for the next phase of such effort, hopefully opening the door to a more informative way of studying ambient-language effects as a general adaptation of infant vocalizations to the auditory-communicative environment. Further, our findings are consistent with Vihman and colleagues' (Vihman & de Boysson-Bardies, 1994) schematic model of perceptual-motor interaction in the first year, a model that emphasizes a role for "salient adult words" associated with the "articulatory filter". Our data imply that detectable ambient-language effects are commonly, if not necessarily, the product of influence from "salient adult words" rather than from an abstraction of the phonetics of the ambient language that is independent of the lexicon (and especially that is independent of the "salient adult words").

Salient limitations of the present work

Even though our study assessed listener judgments of nearly 2000 utterances, the sample size still leaves room for concerns. Infant vocalization patterns change markedly from session to session of recording even within the same day (see, e.g., Oller et al., 2007). Consequently, the number of utterances studied for each infant at each age ($n = 40$), especially given that the utterances were produced in sequence, may not have been large enough to be representative of the repertoires of each child at each age. This concern is particularly notable with regard to the numbers of utterances identified as including canonical syllables and those identified as words. A very few of the infants accounted for the great majority of utterances deemed to be canonical words (one English-learning infant EA: $n = 24$; two Chinese-learning infants: JL $n = 11$, SH $n = 14$, ZL $n = 13$, accounting for 62 of the 88 such utterances).

The most salient (though not the only) example of possible disproportionate influence in the study based on a single outlier sample concerns 17 [mama] utterances that occurred in a single recording from one 10-month-old English-learning infant (EA). As argued above, these utterances may have had disproportionate effects on the data, with English listeners

guessing right, and Chinese listeners guessing wrong, because [mama] sounds like a word in both languages. This pattern of judgment (given the relatively small set of utterances deemed to be words and/or deemed to have canonical syllables) may have skewed the data: (1) in Figure 4A, where English listeners appeared to outperform Chinese listeners on words (English listeners getting many right on [mama], Chinese listeners getting many wrong on [mama]); (2) in Figure 4B, where judgments on utterances of English-learning infants were less accurate than on utterances of Chinese-learning infants (the judgments on English-learning infants perhaps being deflected downward toward chance because of the counterbalancing effect of many [mama] utterances being judged correctly by English listeners but incorrectly by Chinese listeners), and (3) in Figure 4D, where since [mama] consists of canonical syllables, the same patterns on utterances of English-learning infants (less accurate) and Chinese-learning infants (more accurate), may have occurred at least in part due to disproportionate influence of the disproportionate number of [mama] utterances from a single English-learning infant. It clearly would be preferable to increase and/or diversify sampling in order to limit the effects of such (presumable) distributional anomalies, and thus to provide more representative perspective on the ambient-language effects.

Listener knowledge of the two languages may have played a role in this task as well. Chinese listeners (knowing both languages) performed a little better than English listeners at all ages, and although the difference was not statistically reliable, we hope to consider the issue more thoroughly in future studies where we can systematically employ listeners with variable language backgrounds (e.g., balanced bilinguals, English-dominant bilinguals, Chinese-dominant bilinguals, and monolinguals of both languages).

Another limitation in the present study is that the individual (Lee) who selected the utterances from the two groups for presentation to observers was aware of the purpose of the study and of course had to listen to the tapes with their background voices in order to extract utterances. For future ambient-language studies we suggest an automated procedure to label and extract utterances. With the development of automated recognition of utterances in large-scale all-day recordings (Warren et al., 2010; Xu, Richards, & Gilkerson, 2014), we may not be far away from being able to rely on a totally objective method of utterance selection for ambient-language studies.

We did not have information available from the parents on whether each utterance was a word or not, the method of study recommended by Vihman et al. (1985). In an optimal procedure, primary caregivers would review selected utterances with the researchers after each recording session.

The two sets of recordings were made in homes in Taiwan and in a laboratory in Memphis. There is reason to be concerned about differences in recording outcomes across these differing environments especially for overall volubility (see e.g., Lewedag, Oller, & Lynch, 1994). In addition we cannot rule out the possibility that interaction patterns may be different at home and in a laboratory (Jaffe et al., 2001). Furthermore, recording protocols would optimally be matched as closely as possible, for example, by instructing parents of both languages to interact in similar ways. In the present study, the protocols were similar, but had been constructed independently.

And, of course, it would have been preferable to have used the same recording equipment in both locations, with microphones at the same distances from infant mouths, and controlled calibration and gain settings. After-the-fact noise matching is not impossible, but it is tedious.

Conclusion

In summary, with a methodology designed to limit possible bias effects, we did not find significant differences from chance performance for language identification in our overall dataset of 1917 utterances from English- and Chinese-learning infants at 8, 10, and 12 months. The results support the conclusion that infant babbling is based primarily on vocalization patterns *common* to both of these very different ambient languages. However, reliable though small ambient-language identification effects were discernible for the small subset of utterances identified as words and/or as containing canonical syllables. The complexity of the results (with unexpected interactions of listener language by word status, infant ambient language by word status, and infant ambient language by canonical syllable status), and the fact that the results suggest the emergent lexicon may play a key role in ambient-language effects, inspires us to pursue further research at larger sample size, where the role of the lexicon is more directly addressed, and where automated procedures for utterance selection may be able to preclude any possibility of utterance-selection bias. We plan in the future to use such improved methods, especially by engaging parents in judgments about the likelihood that their infants' utterances may be influenced by words of the ambient language. With this approach we hope to evaluate the role of the lexicon in early ambient-language effects more precisely.

Acknowledgments

Thanks to Xuzi Zhou of the Computer Science Department at the University of Kentucky for technical assistance.

Funding

This research was supported by NIH R01 DC011027 to D. Kimbrough Oller, PI and by the Chiang Ching-Kuo Foundation for International Scholarly Exchange in Taiwan RG005-D-10 to Li-mei Chen.

References

- Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*. 1998; 9(4):321–324. DOI: 10.1111/1467-9280.00063
- Atkinson K, MacWhinney B, Stoel C. An experiment in the recognition of babbling. *Papers and Reports on Child Language Development*. 1970; 5:1–8.
- Bauman-Waengler, J. *Articulatory and phonological impairments: A clinical focus*. Upper Saddle River, NJ: Pearson Higher Ed; 2012.
- Bergelson E, Swingle D. At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*. 2012; 109(9):3253–3258. DOI: 10.1073/pnas.1113380109
- Boersma, P., Weenink, D. Praat: Doing phonetics by computer [Computer Program]. 2014. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Brown, R. *Words and things*. New York, NY: Free Press; 1958.
- Buder EH, Chorna L, Oller DK, Robinson R. Vibratory regime classification of infant phonation. *Journal of Voice*. 2008; 22(5):553–564. DOI: 10.1016/j.jvoice.2006.12.009 [PubMed: 17509829]

- Chen LM, Kent RD. Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language*. 2009; 36(1):73–84. DOI: 10.1017/S0305000908008878 [PubMed: 18752703]
- Chen LM, Kent RD. Segmental production in Mandarin-learning infants. *Journal of Child Language*. 2010; 37(2):341–371. DOI: 10.1017/S0305000909009581 [PubMed: 19490748]
- Chung H, Kong EJ, Edwards J, Weismer G, Fourakis M, Hwang Y. Cross-linguistic studies of children's and adults' vowel spaces. *The Journal of the Acoustical Society of America*. 2012; 131(1):442–454. DOI: 10.1121/1.3651823 [PubMed: 22280606]
- Chung, RF. The segmental phonology of Southern Min in Taiwan. Taipei, Taiwan: Crane Publishing; 1996.
- de Boysson-Bardies B, Halle P, Sagart L, Durand C. A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*. 1989; 16(1):1–17. DOI: 10.1017/S0305000900013404 [PubMed: 2925806]
- de Boysson-Bardies B, Sagart L, Durand C. Discernible differences in the babbling of infants according to target language. *Journal of Child Language*. 1984; 11(1):1–15. DOI: 10.1017/S0305000900005559 [PubMed: 6699104]
- de Boysson-Bardies B, Vihman MM. Adaptation to language: Evidence from babbling and first words in four languages. *Language*. 1991; 67(2):297–319. DOI: 10.1353/lan.1991.0045
- DePaolis RA, Vihman MM, Keren-Portnoy T. Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*. 2011; 34(4):590–601. DOI: 10.1016/j.infbeh.2011.06.005 [PubMed: 21774986]
- DePaolis RA, Vihman MM, Nakai S. The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*. 2013; 36(4):642–649. DOI: 10.1016/j.infbeh.2013.06.007 [PubMed: 23911593]
- Eady, SJ. The onset of language specific patterning in infant vocalization (Unpublished doctoral dissertation). Ottawa, Canada: University of Ottawa; 1980.
- Edwards J, Beckman ME. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development*. 2008; 4(2):122–156. DOI: 10.1080/15475440801922115 [PubMed: 19890438]
- Eilers RE, Gavin WJ, Wilson WR. Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*. 1979; 50(1):14–18. DOI: 10.2307/1129035 [PubMed: 446199]
- Eilers RE, Oller DK, Benito-Garcia CR. The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study. *Journal of Child Language*. 1984; 11(2):313–336. DOI: 10.1017/S0305000900005791 [PubMed: 6746779]
- Eimas PD, Siqueland E, Jusczyk P, Vigorito J. Speech perception in infants. *Science*. 1971; 171(3968):303–306. DOI: 10.1126/science.171.3968.303 [PubMed: 5538846]
- Engstrand O, Williams K, Lacerda F. Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12-and 18-month-olds. *Phonetica*. 2003; 60(1):17–44. DOI: 10.1159/000070452 [PubMed: 12802087]
- Enstrom DH. Infant labial, apical and velar stop productions: A voice onset time analysis. *Phonetica*. 1982; 39(1):47–60. DOI: 10.1159/000261650 [PubMed: 7089069]
- Gleason, JB., Ratner, NB. The development of language. Upper Saddle River, NJ: Pearson; 2013.
- Gogate L, Walker-Andrews AS, Bahrick LE. The intersensory origins of word-comprehension: An ecological-dynamic systems view. *Developmental Science*. 2001; 4(1):1–18. DOI: 10.1111/desc.2001.4.issue-1
- Goldstein MH, King AP, West MJ. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*. 2003; 100(13):8030–8035. DOI: 10.1073/pnas.1332441100
- Goldstein MH, Schwade JA. Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*. 2008; 19(5):515–523. DOI: 10.1111/j.1467-9280.2008.02117.x [PubMed: 18466414]
- Gros-Louis J, West MJ, Goldstein MH, King AP. Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*. 2006; 30(6):509–516. DOI: 10.1177/0165025406071914

- Hallé PA, de Boysson-Bardies B, Vihman MM. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French Infants. *Language and Speech*. 1991; 34(4):299–318. [PubMed: 1843528]
- Jaffe J, Beebe B, Feldstein S, Crown CL, Jasnow MD, Rochat P, Stern DN. Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*. 2001; 66(2):i–viii. 1–149.
- Jakobson, R. *Child language, aphasia and phonological universals*. The Hague, The Netherlands: Mouton; 1941/1968.
- Johnson EK, Seidl AH. At 11 months, prosody still outranks statistics. *Developmental Science*. 2009; 12(1):131–141. DOI: 10.1111/desc.2009.12.issue-1 [PubMed: 19120421]
- Jusczyk PW. How infants adapt speech-processing capacities to native-language structure. *Current Directions in Psychological Science*. 2002; 11(1):15–18. DOI: 10.1111/cdir.2002.11.issue-1
- Jusczyk PW, Cutler A, Redanz NJ. Infants' preference for the predominant stress patterns of English words. *Child Development*. 1993; 64(3):675–687. DOI: 10.2307/1131210 [PubMed: 8339688]
- Koopmans-van Beinum, FJ., Clement, CJ., van den Dikkenberg-Pot, I. Influence of lack of auditory speech perception on sound productions of deaf infants. Paper presented at the International Society for the Study of Behavioral Development; Berne, Switzerland. 1998 Jul.
- Lee, CC., Jhang, Y., Chen, LM., Oller, DK. Syllabification and segmentation in English and Mandarin: A methodological study of possible ambient language effects. Paper presented at the 2012 American Speech-Language-Hearing Association (ASHA) convention; Atlanta, GA. 2012 Nov.
- Lee, CC., Jhang, Y., Chen, LM., Oller, DK. Effective blinding of observers for study of ambient language effects. Paper presented at the 2013 International Child Phonology Conference (ICPC); Nijmegen, The Netherlands. 2013a Jun.
- Lee, CC., Jhang, Y., Chen, LM., Oller, DK. Ambient language effects in Chinese- and English-learning infants at 8, 10, and 12 months. Paper presented at the 2013 American Speech-Language-Hearing Association (ASHA) convention; Chicago, IL. 2013b Nov.
- Lee SAS, Davis B, Macneilage P. Universal production patterns and ambient language influences in babbling: A cross-linguistic study of Korean- and English-learning infants. *Journal of Child Language*. 2010; 37(2):293–318. DOI: 10.1017/S0305000909009532 [PubMed: 19570317]
- Levitt AG, Wang Q. Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning Infants. *Language and Speech*. 1991; 34(3):235–249. [PubMed: 1843525]
- Lewedag VL, Oller DK, Lynch MP. Infants' vocalization patterns across home and laboratory environments. *First Language*. 1994; 14(40):049–65. DOI: 10.1177/014272379401404004
- Lléo C, Prinz M, Mogharbel C, Maldonado A. Early phonological acquisition of German and Spanish: A reinterpretation of the continuity issue within the principles and parameters model. *Children's Language*. 1996; 9:11–31.
- Lynch MP, Oller DK, Steffens ML, Levine SL. Onset of speech-like vocalizations in infants with Down syndrome. *American Journal on Mental Retardation*. 1995; 100(1):68–86. [PubMed: 7546639]
- Macmillan, NA., Creelman, CD. *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum; 2004.
- Majorano M, Vihman MM, DePaolis RA. The relationship between infants' production experience and their processing of speech. *Language Learning and Development*. 2014; 10(2):179–204. DOI: 10.1080/15475441.2013.829740
- McCune L, Vihman MM. Vocal Motor Schemes. *Papers and Reports on Child Language Development*. 1987; 26:72–79.
- McCune L, Vihman MM. Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*. 2001; 44(3):670–684. DOI: 10.1044/1092-4388(2001/054)
- Ministry of Education. The history of ministry of education in Taiwan. 2015. Retrieved December 1, 2015, from <http://history.moe.gov.tw/milestone.asp?YearStart=81&YearEnd=90&page=15>
- Moeller MP, Hoover B, Putman C, Arbataitis K, Bohnenkamp G, Peterson B, Stelmachowicz P. Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I –

- Phonetic development. *Ear and Hearing*. 2007; 28(5):605–627. DOI: 10.1097/AUD.0b013e31812564ab [PubMed: 17804976]
- Morse PA. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*. 1972; 14(3):477–492. DOI: 10.1016/0022-0965(72)90066-5 [PubMed: 4658485]
- Navarro, AM., Pearson, BZ., Cobo-Lewis, A., Oller, DK. Identifying the language spoken by 26-month-old monolingual and bilingual-learning babies in a no-context situation. In: Greenhill, M., Hughes-Littlefield, H., Walsh, H., editors. *Proceedings of the 22nd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadia Press; 1998 Mar. p. 557-568.
- Navarro, AM., Pearson, BZ., Cobo-Lewis, A., Oller, DK. Differentiation, carry-over and the distributed characteristic in bilinguals: Differentiation in early phonological adaptation. In: Cohen, J., McAlister, K., Rolstad, K., MacSwan, J., editors. *Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA: Cascadia Press; 2005 May. p. 1690-1702.
- Oller DK, Buder EH, Ramsdell HL, Warlaumont AS, Chorna L, Bakeman R. Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*. 2013; 110(16):6318–6323. DOI: 10.1073/pnas.1300337110
- Oller DK, Eilers RE. Similarity of babbling in Spanish- and English-learning babies. *Journal of Child Language*. 1982; 9(3):565–578. DOI: 10.1017/S0305000900004918 [PubMed: 7174757]
- Oller DK, Eilers RE. The role of audition in infant babbling. *Child Development*. 1988; 59(2):441–449. DOI: 10.2307/1130323 [PubMed: 3359864]
- Oller DK, Eilers RE. Interpretive and methodological difficulties in evaluating babbling drift. *Revue Parole*. 1998:7–8.
- Oller DK, Eilers RE, Bull DH, Carney AE. Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech, Language, and Hearing Research*. 1985; 28(1):47–63. DOI: 10.1044/jshr.2801.47
- Oller, DK., Iyer, SN., Buder, EH., Kwon, K., Chorna, L., Conway, K. Diversity and contrastivity in prosodic and syllabic development. In: Trouvain, J., Barry, WJ., editors. *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken; Germany: 2007 Aug. p. 303-308.
- Oller DK, Wieman LA, Doyle WJ, Ross C. Infant babbling and speech. *Journal of Child Language*. 1976; 3(1):1–11. DOI: 10.1017/S0305000900001276
- Preston M, Yeni-Komshian G, Stark R. Voicing in initial stop consonants produced by children in the pre-linguistic period from different language communities. *Annual Report, Neurocommunications Laboratory, Johns Hopkins University School of Medicine*. 1967; 1:307–323.
- Ramsdell HL, Oller DK, Buder EH, Ethington CA, Chorna L. Identification of prelinguistic phonological categories. *Journal of Speech, Language, and Hearing Research*. 2012; 55(6):1626–1639. DOI: 10.1044/1092-4388(2012/11-0250)
- Rivera-Gaxiola M, Silva-Pereyra J, Kuhl PK. Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*. 2005; 8(2):162–172. DOI: 10.1111/desc.2005.8.issue-2 [PubMed: 15720374]
- Rvachew S, Alhaidary A, Mattock K, Polka L. Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French. *Journal of Phonetics*. 2008; 36(4):564–577. DOI: 10.1016/j.wocn.2008.02.001
- Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. *Cognition*. 1999; 70(1):27–52. DOI: 10.1016/S0010-0277(98)00075-4 [PubMed: 10193055]
- Stoel-Gammon C, Cooper JA. Patterns of early lexical and phonological development. *Journal of Child Language*. 1984; 11(2):247–271. DOI: 10.1017/S0305000900005766 [PubMed: 6746776]
- Stoel-Gammon C, Otomo K. Babbling development of hearing-impaired and normally hearing subjects. *Journal of Speech and Hearing Disorders*. 1986; 51(1):33–41. DOI: 10.1044/jshd.5101.33 [PubMed: 3945058]
- Thevenin DM, Eilers RE, Oller DK, Lavoie L. Where's the drift in babbling drift? A cross-linguistic study. *Applied Psycholinguistics*. 1985; 6(1):3–15. DOI: 10.1017/S0142716400005981

- Tincoff R, Jusczyk PW. Some beginnings of word comprehension in 6-month-olds. *Psychological Science*. 1999; 10(2):172–175. DOI: 10.1111/1467-9280.00127
- Vihman MM, de Boysson-Bardies B. The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica*. 1994; 51(1–3):159–169. DOI: 10.1159/000261967 [PubMed: 8052670]
- Vihman MM, DePaolis RA, Davis BL. Is there a “trochaic bias” in early word learning? Evidence from infant production in English and French. *Child Development*. 1998; 69(4):935–949. DOI: 10.2307/1132354 [PubMed: 9768479]
- Vihman MM, Ferguson CA, Elbert M. Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*. 1986; 7(1):3–40. DOI: 10.1017/S0142716400007165
- Vihman MM, Macken MA, Miller R, Simmons H, Miller J. From babbling to speech: A reassessment of the continuity issue. *Language*. 1985; 61(2):397–445. DOI: 10.2307/414151
- Vihman MM, McCune L. When is a word a word? *Journal of Child Language*. 1994; 21(3):517–542. DOI: 10.1017/S0305000900009442 [PubMed: 7852471]
- Vinter S. Contrôle de premières productions vocales du bébé sourd. *Bulletin D’audiophonologie*. 1987; 3(6):659–670.
- Warren SF, Gilkerson J, Richards JA, Oller DK, Xu D, Yapanel U, Gray S. What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*. 2010; 40(5):555–569. DOI: 10.1007/s10803-009-0902-5 [PubMed: 19936907]
- Weir, RH. Some questions on the child’s learning of phonology. In: Smith, F, Miller, GA., editors. *The genesis of language: A psycholinguistic approach*. Cambridge, MA: MIT Press; 1966. p. 153–172.
- Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. 1984; 7(1):49–63. DOI: 10.1016/S0163-6383(84)80022-3
- Whalen DH, Levitt AG, Goldstein LM. VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*. 2007; 35(3):341–352. DOI: 10.1016/j.wocn.2006.10.001 [PubMed: 19641636]
- Whalen DH, Levitt AG, Wang Q. Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*. 1991; 18(3):501–516. DOI: 10.1017/S0305000900011223 [PubMed: 1761611]
- Xu D, Richards JA, Gilkerson J. Automated analysis of child phonetic production using naturalistic recordings. *Journal of Speech, Language, & Hearing Research*. 2014; 57:1638–1650. DOI: 10.1044/2014_JSLHR-S-13-0037
- Yip, M. *Tone*. Cambridge, UK: Cambridge University Press; 2002.

Figure 1.

(A) Ambient-language detection by infant age and *listener* language. (B) Ambient-language detection by infant age and *infant* ambient language. * $p < .05$.

Figure 2. Interaction among infant age, infant ambient language, and listener language in ambient-language detection. * $p < .05$.

Figure 3.

(A) Ambient-language detection by individual listeners. (B) Ambient-language detection by listeners on individual infants. * $p < .05$.

Figure 4.

(A) Ambient-language detection by *listener language background* on utterances identified as words vs. non-words. (B) Ambient-language detection by *infant language background* for infant utterances identified as words vs. non-words. (C) Ambient-language detection by *listener language background* on utterances identified as including canonical syllables and all other utterances. (D) Ambient-language detection by *infant language background* for infant utterances identified as including canonical syllables and all other utterances. * $p < .05$.

Figure 5.

Ambient-language detection broken down by listener language for utterances with different properties: Left bars, for all infant utterances, center bars for infant utterances judged to be words and to contain canonical syllables (canonical words), and right bars for adult utterances composed of phonotactically well-formed pseudo words of Mandarin and English. * $p < .05$.

Table 1

English and Mandarin pseudo words.

English pseudo words	Original transcript
1 ningra, mormy ['nɪŋ.ɹə, 'mo.ɹmi]	Morning, Myra.
2 youking lartam ethra? ['ju:kɪŋ 'la.ɹrəm 'eə.ɹə?]	You're taking Ecology Three Eleven, too?
3 Kolay itsik ['kʰowlej 'ɪtsɪk]	Looks like it.
4 manotta theyly istive. [mə'nɔrə 'ðejli 'ɪstɪv]	It's my only elective this term.
5 pendly? ['pʰɛnli?]	Really?
6 wuhpannend? [wɔ'pʰɛnənd?]	What happened?
7 nurustah recalsus pankit. [nu'ɹʌstə ɹɪ'kʰælsəs 'pʰɛŋkɪt]	Now I've got to pick up all the required courses I've been ignoring.
8 toonaferstees moo chun, teenam.[tʰu'næfəstɪz mu tʃʰʌn, tʰɪnæm]	University's just too much fun, I guess.
9 poo timit shardows? [pʰu 'tʰɪmɪt 'ʃa.ɹrowz]	Too many parties?
10 choo nemow...[tʃʰu nə'mow]	But you know
11 gar doo gawzee lerfuh, Elding? ['ga.ɹ ru 'gawzi 'lɜfə, 'eldɪŋ?]	Are you going to the job fair, Zelda?
12 boochet! [bu'tʃʰɛt]	You bet!
13 ruh jummy latev nasper[ɹə 'dʒʌmi 'læɹəv 'næspɜ]	I'm going to have no money left at all after this term.
14 fing ooter? [fɪŋ 'uɹə?]	You going?
15 wuh beed kust gow burgust [wu bɪd kʰʌst gaw 'bɜgəst]	I've gotta find work by July or August at the latest.
16 im yo jahssiet neeb linta [ɪm jow 'dʒəsɪt nɪb 'lɪntə]	I don't really need to find a job right away- -
17 mun thartahmun padmag [mʌn θɑ.tʰɪmən 'pʰædmæg]	I can always part-time in my dad's shop
18 ruhtay waff uh rouk, kahrry?['ɹʌɹej wæf ʌ ɹawk, 'kʰɑ.ɹɹi?]	But I do have to get going on my career, eh?
19 ruh ven aht ferkship mefehver [ɹʌ 'vɛn at 'fɜkʃɪp mə'fevɜ]	Or I'll end up in his workshop maybe forever.
20 wahn ee tanuhweed, klowa [wʌn i 'tʰɛnəwɪd 'kʰlowə]	What kind of business is it, anyway?

Mandarin pseudo words	Original transcript (English translation)
1 尸𐄂𐄂 尸𐄂𐄂 𐄂 [ʃaj2 ʃwə3]	您好。(Hello.)
2 𐄂𐄂𐄂 𐄂𐄂𐄂 𐄂𐄂𐄂 𐄂𐄂𐄂 [su3 tse4 se4 swən2]	我叫大明。(My name is Da Ming.)
3 尸𐄂𐄂𐄂 𐄂𐄂𐄂 𐄂𐄂𐄂 𐄂𐄂𐄂 𐄂𐄂𐄂 [ʃwan3 hən1 tɕʰjo4 kʰwan4 tsa4 sə2]	很高興見到您。(Nice to meet you.)
4 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [nye3 mə4 tsa4 sɿ2 nɟaw1 mɔ1 tɕʰwən4 mu1?]	你是做什麼工作的？(What do you do for living?)
5 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [tɕʰyən3 tsʰaŋ4 nɟəŋ1 tʰəŋ4 swən2 mjen1]	我是個建築師。(I am an architect.)
6 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [mje3 səŋ1 kʰəŋ4 ha4 tjo4 aŋ3 maŋ4 mjo3]	我們在建造體育場。(We are now building a gym.)
7 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [fo1 mjen1 mə1?]	真的嗎？(Really?)
8 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [kaŋ2 tɕʰwən4 pʰən3 hən1 jaj4?]	您在哪工作？(Where is the company you work with?)
9 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [ɤwəŋ4 tsʰaŋ4 nə4 tsaŋ2]	現代建築。(Modern architect.)
10 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [tɕʰa3 ɤan4 tʰəŋ1 tsʰwə3]	我是經理。(I am a manager.)
11 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [waj2 əŋ4 tɕjo2 lo1 lu1?]	您住倫敦嗎？(Do you live in London?)
12 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 𐄂𐄂𐄂𐄂 [nwan4 hə3 kʰən4 ɕjo2 lyɛ1 lə2]	不，我是臺灣人。(No, I am Taiwanese.)

	English pseudo words	Original transcript
13	ㄋㄨˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [no3 paŋ2 ʒwaj2 roə3 tɕyan2 e1]	我從臺北來的。(I am from Taipei.)
14	ㄋㄨˇ ㄋㄨˊ ㄋㄨˊ ㄋㄨˊ ㄋㄨˊ ㄋㄨˊ ㄋㄨˊ [nən2 nye2 ʒwəŋ4 tsʰu3 ən3?]	您來自哪裏？(Where are you from?)
15	ㄇㄜˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [sə3 ʒwan4 ny4 tsʰə2 pən2]	我住在德國。(I live in Germany.)
16	ㄇㄜˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [ʒə2 pən2 pʰan3 ja1 maŋ4 se1]	德國什麼地方？(Which part of Germany?)
17	ㄇㄜˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [ja4 lə3 ʒwəŋ4 tʰəŋ4]	漢堡附近。(Around Hamburg.)
18	ㄇㄜˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [ʒe4 tɕʰyən3 pʰo4 tɕʰyən4 he2 kʰən4 ni1?]	漢堡是大城市嗎？(Is Hamburg a big city?)
19	ㄇㄜˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [sən3 ke4 tjo3 tsʰwa3 tsʰwa4 ʒu1 ʒə2]	有近兩百萬居民。(Around two million people.)
20	ㄇㄜˇ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ ㄅㄨˊ [tsʰaŋ4 ʒən1 ləŋ1 kʰwaj2 mjoŋ4 pʰjo4 lwan1 kʰan2 kʰəŋ4]	是個非常漂亮的城市。(It's a very beautiful city.)

Table 2

Overall mixed model results.

	Column 1 Main model (including analysis of all 1917 utterances)	Column 2 Additional model 1 on word status (120 words vs. 1797 non-words)	Column 3 Additional model 2 on canonical syllable (CB) status (286 utterances with CB vs. 1631 utterances without CB)	Column 4 Additional model 3 on canonical words (88 canonical words vs. 1829 non- canonical words)
Main effects	infant age ** infant ambient language listener language (Displayed in Figure 5, left pair of bars)	word status ** (words [^] vs. non-words) infant ambient language listener language	canonical syllable status ** (CB [^] vs. non-CB) infant ambient language listener language	canonical words ** listener language (Figure 5 middle pair, ENL [^] vs. CHL [^])
Interactions	listener language × infant age (Figure 1A) ENL for 08 mo CHL for 08 mo ENL for 10 mo CHL for 10 mo ENL for 12 mo CHL for 12 mo infant ambient language × infant age ** (Figure 1B) ENI at 08 mo CHI at 08 mo ENI at 10 mo CHI at 10 mo ENI at 12 mo CHI at 12 mo [^] infant ambient language × listener language ** infant age × Infant ambient language × listener language * (Figure 2) ENI at 08 mo for ENL ENI at 08 mo for CHL [^] CHI at 08 mo for ENL CHI at 08 mo for CHL ENI at 10 mo for ENL ENI at 10 mo for CHL CHI at 10 mo for ENL CHI at 10 mo for CHL [^] ENI at 12 mo for ENL ENI at 12 mo for CHL CHI at 12 mo for ENL CHI at 12 mo for CHL [^]	listener language × word status ** (Figure 4A) ENL for words [^] ENL for non-words CHL for words CHL for non-words listener language × word status ** (Figure 4B) ENL for words ENL for non-words CHL for words [^] CHL for non-words infant ambient language × listener language ** infant ambient language × listener language × word status **	listener language × canonical syllable status (Figure 4C) ENL for CB [^] ENL for non-CB CHL for CB [^] CHL for non-CB infant ambient language × canonical syllable status ** (Figure 4D) ENI for CB ENI for non-CB CHI for CB [^] CHI for non-CB infant ambient language × listener language ** infant ambient language × listener language × canonical syllable status	listener language × canonical words
Random effects	individual listeners (Figure 3A) individual infants (Figure 3B)	individual listeners individual infants	individual listeners individual infants	

CHL = Chinese Listeners, ENL = English Listeners, CHI = Chinese-learning infants, ENI = English-learning infants, mo = months, CB = canonical syllables;

*
p < .05,

**
p < .001,

[^]
ambient language detection statistically significantly above chance,

[^]
statistically significantly below chance

Table 3

Ambient-language detection by individual listeners for the 1917 utterances, corresponding to the main model results in column 1, Table 2.

	ID	Proportion correct	Sensitivity, d' (Higher values indicate better language detection)	Bias, C (Positive values indicate bias toward Chinese)
English listeners	EBS	0.50	-0.01	0.02
	EKG	0.49	-0.05	0.01
	JER	0.49	-0.03	0.05
	LMM	0.51	0.08	-0.20
	MAC	0.51	0.10	-0.04
	SDM	0.50	0.02	-0.22
Chinese listeners	YSJ	0.50	0.10	0.00
	ALE	0.50	0.12	0.38
	DEM	0.52	0.24	-0.07
	EHU	0.47	-0.08	0.21
	FLC	0.49	0.09	0.10
	FHL	0.53	0.35	-0.08
	JWK	0.49	0.07	0.04