

Infants' brain responses to speech suggest Analysis by Synthesis

Patricia K. Kuhl¹, Rey R. Ramírez, Alexis Bosseler, Jo-Fu Lotus Lin, and Toshiaki Imada

Institute for Learning and Brain Sciences, University of Washington, Seattle, WA 98195

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2010.

Contributed by Patricia K. Kuhl, June 16, 2014 (sent for review January 30, 2014)

Historic theories of **speech perception** (Motor Theory and Analysis by Synthesis) invoked listeners' knowledge of speech production to explain speech perception. Neuroimaging data show that adult listeners activate motor brain areas during speech perception. In two experiments using **magnetoencephalography (MEG)**, we investigated motor brain activation, as well as **auditory brain activation**, during discrimination of native and nonnative syllables in infants at two ages that straddle the developmental transition from language-universal to **language-specific speech perception**. Adults are also tested in Exp. 1. MEG data revealed that 7-mo-old infants activate **auditory (superior temporal)** as well as motor brain areas (Broca's area, cerebellum) in response to speech, and equivalently for native and nonnative syllables. However, in 11- and 12-mo-old infants, native speech activates auditory brain areas to a greater degree than nonnative, whereas nonnative speech activates motor brain areas to a greater degree than native speech. This double dissociation in 11- to 12-mo-old infants matches the pattern of results obtained in adult listeners. Our infant data are consistent with Analysis by Synthesis: **auditory analysis** of speech is coupled with synthesis of the motor plans necessary to produce the speech signal. The findings have implications for: (i) perception-action theories of speech perception, (ii) the impact of "motherese" on early language learning, and (iii) the "social-gating" hypothesis and humans' development of social understanding.

brain imaging | phonetic perception

The development of human language presents a computational puzzle that is arguably among the most challenging in cognitive science. Children the world over acquire their native language in a relatively short time, mastering phonology and grammar via exposure to ambient speech, and producing speech patterns themselves by the age of 1 y that can be deciphered by their parents (1, 2). Machine-based systems have thus far been unable to learn a language based on acoustic input, although machines improve when "motherese" is used as training material (3, 4). Understanding the process by which the developing brain encodes language presents a problem of significant interest (5).

A central phenomenon in infant language development is the transition in phonetic perception that occurs in the second half of the first year of life. At birth and until about 6 mo of age, infants are capable of discriminating the consonants and vowels that make up words universally across languages. Infants discriminate phonetic differences measured behaviorally (6, 7) and neurally (8, 9), regardless of the language from which the sounds are drawn and the ambient language that infants have experienced. Infants are "citizens of the world." By 12 mo, a perceptual narrowing process has occurred; discrimination of native-language phonetic contrasts has significantly increased but perception of foreign-language speech contrasts shows a steep decline (6, 7, 10).

What is less clear is the mechanism underlying this initial phonetic learning. Behavioral studies uncovered two processes that affect infant learning, one computational and the other social (11). Infants' phonetic perception is altered by the distributional frequency of the speech sounds they hear (12, 13). A social process is also indicated. Infants exposed socially to a new

language at 9 mo in play sessions by a live tutor learn to discriminate foreign language sounds at levels equivalent to infants exposed to that language from birth; however, no learning occurs if the same material on the same schedule is presented via video (14, 15). The fact that social interaction is critical for phonetic learning led to the "social gating" hypothesis, the idea that a social setting provides essential motivational and informational enhancement for language learning in the first year of life (11, 16).

The Role of Action in Speech Perception

Auditory-evoked potential measures of this developmental change, obtained using event-related potentials (ERPs), reveal the expected change in neural discrimination; by 11 mo of age, the mismatch response (MMR)—a sensitive measure of auditory discrimination (17)—increases for native speech signals and decreases for nonnative speech signals (8). However, ERP methods cannot identify the cortical generators correlated with this developmental change, and in particular cannot investigate the potential role that motor brain systems play in the developmental transition. The action-perception link has a long history in the speech domain (18; see ref. 19 for discussion), with origins in the writings of von Humboldt (20), Helmholtz (21), and de Cordermo (22).

In the 1950s, Liberman et al. from Haskins Laboratories proposed a Motor Theory of speech perception, which argued—based on studies of "categorical perception"—that the perception of speech was accomplished via knowledge of the motor commands that produce speech (23). Variations on the Motor Theory (e.g., Direct Realism) argued that the perception of speech was accomplished through a direct link between auditory and gestural information (24).

Enthusiasm for motor theories was tempered in the 1970s by experimental findings on human infants and nonhuman animals (18, 25). Infants showed an enhanced ability to discriminate

Significance

Infants discriminate speech sounds universally until 8 mo of age, then native discrimination improves and nonnative discrimination declines. Using magnetoencephalography, we investigate the contribution of auditory and motor brain systems to this developmental transition. We show that 7-mo-old infants activate auditory and motor brain areas similarly for native and nonnative sounds; by 11–12 mo, greater activation in auditory brain areas occurs for native sounds, whereas greater activation in motor brain areas occurs for nonnative sounds, matching the adult pattern. We posit that hearing speech invokes an Analysis by Synthesis process: auditory analysis of speech is coupled with synthesis that predicts the motor plans necessary to produce it. Both brain systems contribute to the developmental transition in infant speech perception.

Author contributions: P.K.K. designed research; A.B., J.-F.L.L., and T.I. performed research; R.R.R. and T.I. analyzed data; and P.K.K. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. Email: pkkuhl@u.washington.edu.

sounds at the boundary between two speech categories (categorical perception) in the absence of speech production and regardless of language experience (26). In response to these findings, Motor Theory was revised to propose that representations of phonetic gestures are innate and universal and do not require speech production experience (27). However, nonhuman animals tested on the same speech stimuli demonstrated the same boundary phenomenon (28, 29).

During this period, Massachusetts Institute of Technology scientists Stevens and Halle proposed an alternative theory, one influenced by machine-learning algorithms (30). This theory, Analysis by Synthesis (AxS), proposed that speech perception involved a dual “hypothesize and test” process. Bottom-up analysis and top-down synthesis jointly and actively constrained perceptual interpretation. On this account, listeners generate an internal model of the motor commands needed to produce the signal: in essence, a “guess” or prediction about the input. This hypothesis, based on listeners’ experience producing speech, is tested against incoming data.

A revival of these historical theories occurred in the 1990s because of the influence of neuroscience and machine-learning algorithms. The discovery of mirror neurons in monkey cortex played a role. Data showed cells in monkey F5 (argued to be homologous to Broca’s area in humans) that responded to both the execution of motor acts and the visual perception of those acts (31). This finding led to discussion about the brain systems that underpin humans’ exquisite social understanding (32) and language learning (33; but see ref. 34).

Bayesian frameworks describing “priors” that constrain the interpretation of incoming information began to reference AxS models (35). The advent of functional MRI (fMRI) enabled experimental investigations of AxS by examining brain activation patterns during speech perception in adult listeners. These studies set the stage for the present investigation, which poses the question of brain activation patterns during speech perception with infants using magnetoencephalography (MEG) technology.

Cortical Speech Motor Areas Activated During Speech Perception: Adult Evidence

In adults, phonetic tasks activate left hemisphere areas implicated in speech production, including Broca’s, the cerebellum, premotor cortex (PMC), and anterior insula, in addition to auditory brain regions, such as the superior temporal gyrus (STG) (36–40). These data were interpreted as consistent with the idea that speech production experience enables generation of internal motor models of speech, which are compared with incoming sensory data, as envisioned by AxS (41–43). Wilson et al. (44) compared fMRI activation under three conditions: passive listening to speech syllables, production of speech syllables, and passive listening to nonspeech. Activation in the ventral PMC occurred in both listening and speaking conditions, and was greatly reduced in response to nonspeech.

In several studies, activation of inferior frontal (IF) areas was associated with nonnative speech perception and learning (41, 45–50). Callan et al. (41), using fMRI, tested a single speech contrast (/r-l/) in adults for whom the contrast was native (English) versus nonnative (Japanese). Callan et al. reported greater activity in auditory areas (STG) when the signal was a native contrast, and greater activity in motor brain areas (Broca’s, the PMC, and anterior insula) when the same signal was a nonnative phonetic contrast. In addition, adult studies indicated that audiovisual tests of the McGurk Illusion activate a network of motor areas, including the cerebellum and cortical motor areas involved in planning and executing speech movements (51).

Taken together, the results of adult studies have been interpreted as support for the idea that, when listening to speech, adults generate internal motor models of speech based on their experience producing it. Several models of adult speech perception integrate this notion, arguing that internal motor models assist perception in a way that is consistent with the AxS conception (41, 51–53).

The Present Study: Posing the Question in Infants

Our goal in the present study was to use MEG technology to examine cortical activity in auditory and motor brain areas during native and nonnative speech discrimination in infants. Previous work shows that infants activate motor brain areas in distinct ways when listening to speech as opposed to nonspeech. Imada et al. (54) tested newborns, as well as 6- and 12-mo-old infants, using speech syllables, harmonic tones, and pure tones, examining activation in superior temporal (ST) and IF brain areas. Newborns showed brain activation in ST for all three categories of sounds, but no activation in the IF area. However, at 6- and 12-mo of age, infants’ brains showed synchronous activation in the ST and IF areas for speech signals but not for nonspeech signals. Similar activation in Broca’s area has been reported in 3-mo-old infants listening to sentences (55). These studies demonstrate that speech perception activates infants’ motor brain areas, but do not explicate their role in infant speech perception.

The key strategy in the present study was to compare brain activation in auditory and motor brain areas for the native vs. nonnative MMR, before and after the developmental change in speech perception. Based on previous data in adults (41) and our own data in infants with ERPs (8), we hypothesized a specific pattern of results: early in development, when perceptual discrimination is universal, infants’ MMR responses in auditory and motor brain areas should be similar for native and nonnative speech contrasts. By 12 mo, after the developmental transition, infants’ MMR responses in auditory and motor brain systems should show a double dissociation: activation should be greater in auditory brain areas for the native MMR compared with the nonnative MMR, whereas activation should be greater in motor brain areas for the nonnative MMR compared with the native MMR. The pattern in 12-mo-olds was expected to match that of adult listeners.

We report two MEG studies. Both studies tested infants at two ages that bracket the developmental transition in speech perception. The studies used different sets of native and nonnative syllables, different MEG facilities, and different methods of MEG data analysis to provide a rigorous test of the hypothesis. Both MEG studies used infant head models developed in our laboratory to improve localization of activity in the infant brain (56). Exp. 1 additionally examined the pattern of activation for native and nonnative MMR responses in adult listeners.

Experiment 1

Introduction. The goal of Exp. 1 was to examine neural activation in auditory (ST) and motor (IF) brain areas during native and nonnative phonetic discrimination. We examined brain activation at three ages (7 mo, 11 mo, and adulthood), focusing on differences in neural activation for the native vs. nonnative MMRs indicating phonetic discrimination.

Methods. Subjects. Twenty-five 7-mo-old and 24 11-mo-old English-learning infants, and 14 English native-speaking adults participated in the experiment. Infants had no reported hearing problems and no history of ear infections. Adults reported having no hearing or neurological problems. The mean age and SD for the three age groups, after rejection (see below), were as follows: 222.4 ± 3.8 d for 7 infants aged 7 mo, 328.4 ± 6.4 d for 10 infants aged 11 mo, and 26.6 ± 3.8 y for 10 adult subjects. Written informed consent in accordance with the Human Subjects Division at the University of Washington was obtained from the parents and from adult participants.

Stimuli. Rivera-Gaxiola et al. (8) created three sounds for use in a double-oddball paradigm. This paradigm uses a phonetic unit common to English and Spanish as the standard sound, and two deviant sounds, one exclusive to English and the other to Spanish. A voiceless alveolar unaspirated stop, common to English and Spanish (perceived as /da/ in English and /ta/ in Spanish) served as the standard sound; a voiceless aspirated alveolar stop exclusive to English (/t^ha/) served as the English deviant sound; a prevoiced alveolar stop exclusive to Spanish

(/da/) served as the Spanish deviant sound. The stimulus duration was 230 ms for all stimuli. Mean intensity was adjusted to 65 dBA at the infant's ear and to a comfortable level for adults.

Stimulus presentation. The auditory stimuli were delivered to infants via a flat panel speaker (Right-EAR), positioned 2 m from the infant's face and to adults' right ear by a plastic tube connected to an earphone (Nicolet Biomedical). The inter-stimulus interval was 700 ms onset-to-onset for adults and 1,200 ms onset-to-onset for infants, consistent with the interstimulus intervals used in previous infant experiments (8, 57).

Experimental paradigm. The standard speech sound was presented on 80% of the trials, and the two deviant speech sounds were each presented on 10% of the trials. Presentation format was pseudo-random; no more than three standards in a row could be presented.

Experimental procedure. Adult participants watched a silent video during the MEG measurement. Infants were prepared for testing outside the magnetically shielded room (MSR) while an assistant who waved silent toys entertained them. During the preparation phase, we placed a lightweight soft nylon cap containing four coils on the infant's head; the coils were used to track the infant's head during recordings. Once prepared, infants were placed in a custom-made chair that was adjustable in height as well as forward and backward. The adjustable chair made it easy to place infants ranging in height and weight in the dewar in an optimal position for MEG recording. All infants were highly alert during the recording because the assistant continued to wave silent toys in front of the infant and use a silent video to entertain them throughout the 20-min session. Infants' head movements were variable throughout the session, but the head-tracking software allowed us to either calibrate for movement or reject epochs in the preprocessing stage, as described below.

MEG measurement. A whole-head MEG system (Elekta-Neuromag) with 306 superconducting quantum interference device sensors, situated in a MSR at the University of Washington Institute for Learning and Brain Sciences (I-LABS), was used. MEG recorded the brain magnetic responses to each of the three speech sounds every 1.0 ms, with antialiasing low-pass (330 Hz cut-off) and high-pass (0.1 Hz cut-off) filters for adult subjects, and every 0.5 ms, with antialiasing low-pass (660 Hz cut-off) and a high-pass (0.03 Hz cut-off) filters for the infant subjects. Real-time head-position tracking was performed for infants by recording the magnetic fields produced by four head coils installed in a soft close-fitting cap. To offline localize the head-coil positions, weak sinusoidal signals with high frequencies (293, 307, 314, or 321 Hz) were applied every 200 ms to the head coils during measurement. Localization of the head coils enabled us to estimate the infant's head position and orientation with respect to the sensor array coordinate system, which made it possible to recalculate and obtain the magnetic fields virtually, as though they originated from a stable head position at the origin of the MEG sensor array coordinate system.

MEG data analysis.

Preprocessing. Because we used the active shielding equipment in the MSR, we first applied MaxFilter software (Elekta-Neuromag) or the signal space separation (SSS) algorithm (58) to all of the raw MEG data. Second, we applied the temporal SSS (tSSS) procedure to all of the results from the basic SSS. Third, we applied the movement compensation algorithm, implemented in MaxFilter, to the results tSSS obtained (infants only). Adult subjects remained still in the MEG during measurement and no movement compensation was necessary.

Rejection and averaging. Epochs were rejected when MaxFilter could not locate the infant's head position or orientation, or the peak-to-peak MEG amplitude was over 6.0 pT/cm (gradiometer) or 6.0 pT (magnetometer) for the adults and 8.0 pT/cm (gradiometer) for the infants. After rejecting epochs by these criteria, we further rejected participants if they had fewer than 40 (infants) or 80 (adults) acceptable epochs for averaging. Epochs in response to the standard immediately before the deviants were averaged separately for each subject after the peak-to-peak amplitude rejection was performed.

Filters and baseline. After averaging, the low-pass filter with the cut-off frequency of 20 Hz was applied to all of the data and then the DC-offset during the baseline range from 100 ms before the stimulus onset to 0 ms was removed.

Minimum norm estimate. We used the minimum norm estimation (MNE) method (59), with a spherical brain model, to obtain the neural currents that produced the measured magnetic responses. For the adults, the sphere was approximated to each subject's brain images obtained by the 3.0T MRI system (Philips). For the infants, the sphere was fit to a standardized 6-mo-old infant brain template, created by I-LABS using 60 infant brains (56), which has a Montreal Neurological Institute-like head coordinate system (McGill University, Montreal, QC, Canada). MNE currents were calculated at 102 predetermined points on the approximated sphere from the MEG data at 102 gradiometer pairs. The MMR was calculated by subtracting the response (MNE currents) to the standard stimulus immediately before each of the corresponding two deviant stimuli, separately: these were defined as the native (English) MMR and the nonnative (Spanish) MMR. Following conventional MMR analyses, we compared the maximum peak magnitudes of the MNE current waveforms, between the language conditions, within the predetermined latency ranges and within the four predetermined regions of interest. The latency ranges were between 140 and 250 ms for the adults (60) and between 140 and 500 ms for the infants, reflecting data from our previous ERP studies using these same stimuli (8, 15). The MNE neural current magnitude was determined at each MNE current location at each sampling point. The four predetermined regions of interest were the left and right ST regions, and the left and right IF regions.

Results. The mean MEG waveforms for each age in response to the native (English) and nonnative (Spanish) deviants, averaged across all subjects, indicate high-quality MEG recordings for all ages, including infants. Fig. 1 provides representative data in the form of magnified waveforms over the left and right IF regions (Fig. 1, red circle) and the left and right ST regions (Fig. 1, green circle) in 7-mo-old infants (Fig. 1A), 11-mo-old (Fig. 1B) infants, and adults (Fig. 1C); the waveforms are averaged across all participants showing responses to the native (English) (Fig. 1, *Middle*) and nonnative (Spanish) (Fig. 1, *Bottom*) deviants.

MNE current magnitudes were normalized by the mean baseline noise for further analyses. We compared the maximum peak magnitudes of the MNE neural current waveforms across four conditions: Age (7 mo, 11 mo, adult), Language (native English, nonnative Spanish), Hemisphere (left, right), and Region (ST, IF). A four-way ($3 \times 2 \times 2 \times 2$) repeated-measures ANOVA examined the effects of Age, Language, Hemisphere, and Region. The results revealed two significant main effects: Age, $F(2,24) = 6.005$, $P = 0.008$, and Region, $F(1,24) = 5.479$, $P = 0.028$ (Fig. 2A), as well as a significant interaction, Language \times Region, $F(1,24) = 6.194$, $P = 0.020$ (Fig. 2B).

There was no three-way interaction between Language, Region, and Age, perhaps because of amplitude differences across Age unrelated to our hypothesis. We conducted post hoc comparisons to test the hypothesis of developmental change in brain activation for the native vs. nonnative MMRs. The probability values obtained from the post hoc tests were Bonferroni-corrected.

We directly compared 11-mo-old infants and adults to assess the hypothesis that both groups exhibit the interaction between language and brain region (Fig. 2B). As predicted, analysis of 11-mo-old infants and adults revealed a significant Age main effect, $F(1,18) = 8.996$, $P = 0.023$ (corrected), as well as a significant interaction, Language \times Region, $F(1,18) = 8.569$, $P = 0.027$ (corrected). We also directly compared the 7- and 11-mo-old infants, and directly compared the 7-mo-old infants to adults. The results were as expected. Comparison of 7- and 11-mo-old infants revealed no significant main effects or significant interactions. Tests comparing the 7-mo-old infants to adults showed significant

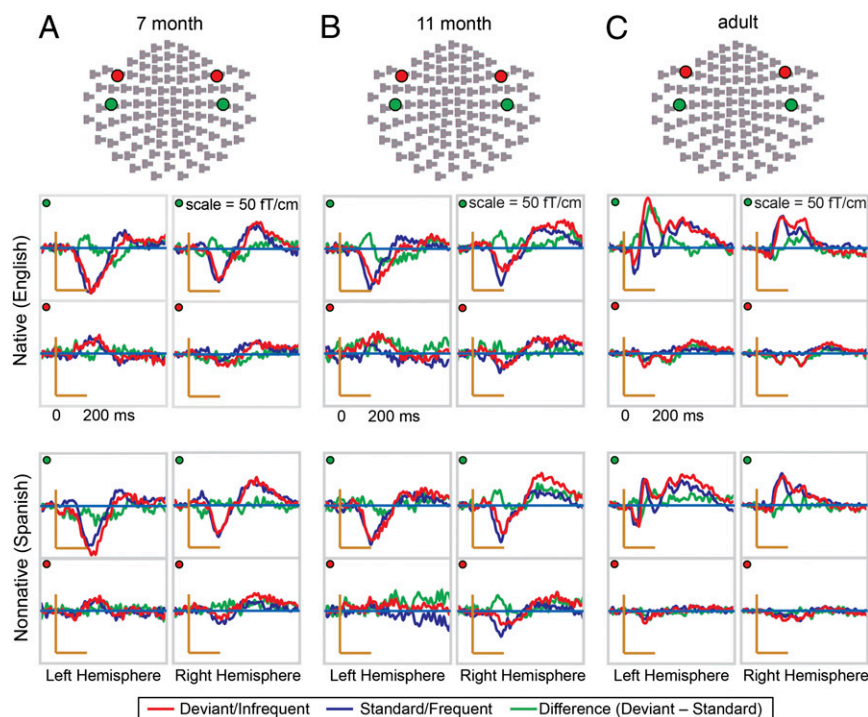


Fig. 1. Mean MEG magnetic waveforms from the left and right ST (green dot) and IF (red dot) regions for 7-mo-old infants (A), 11-mo-old infants (B), and adults (C), for native English (Middle) and nonnative Spanish (Bottom) deviants.

Age main effects, $F(1,15) = 8.134$, $P = 0.036$ (corrected), but no significant interaction. Our results indicate that in 12-mo-old infants and in adults, discrimination of native syllables evokes greater activation in the ST compared with nonnative syllables, whereas discrimination of nonnative syllables evokes greater activation in the IF compared with native syllables (Fig. 2B). This pattern is not shown at the age of 7 mo.

Discussion of Exp. 1

In Exp. 1, two groups of infants straddling the transition from a universal to a language-specific form of phonetic perception,

along with a group of adults, were tested on native and nonnative phonetic discrimination. The goal of the experiment was twofold: first, we tested the hypothesis that early in development speech activates not only cortical areas related to auditory perception, but also motor cortical areas. Second, we tested the hypothesis of developmental change: at 7 mo of age we expected equivalent activation for native and nonnative contrasts, but by the end of the first year we expected native and nonnative speech to activate auditory and motor brain areas differentially, with native activation greater than nonnative in auditory brain areas, and nonnative activation greater than native in motor brain areas, a pattern that would match that shown in adults.

Our results provided support for both hypotheses. At both ages, the IF area is activated in response to speech, supporting our first hypothesis. Moreover, at 7 mo infants' MMRs for native and nonnative contrasts were equivalent in auditory areas and also equivalent in motor brain areas, whereas at 11 mo of age, infants demonstrated the double dissociation we predicted: greater activation in auditory areas for native speech and greater activation in motor brain areas for nonnative speech, matching the pattern we obtained in adults.

Exp. 1's results suggest that brain activation patterns in response to speech change with language experience in both auditory and motor brain areas. We sought to replicate and extend this pattern of results in Exp. 2.

Experiment 2

Introduction. To validate Exp. 1's findings, we used a new set of native and nonnative speech stimuli and different MEG analysis methods. We examined the posterior ST and an expanded set of motor brain areas: Broca's, cerebellum, precentral gyrus, and left precentral sulcus. Adult data implicate these areas in speech processing (36–50, 61–63). We were especially interested in the cerebellum because our recent whole-brain voxel-based morphometry study demonstrated that concentrations of white- and gray-matter in cerebellar areas at 7 mo predict infants' language development at 1 y of age (64).

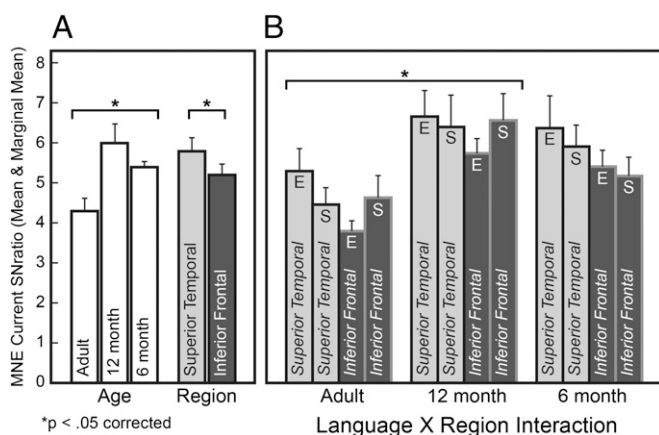


Fig. 2. The maximum peak magnitude of the MMR MNE-current normalized by the baseline noise magnitude. (A) Main effects of age and brain region: mean maximum peak magnitude recorded from 7-mo-old infants, 11-mo-old infants, and adults; mean maximum peak magnitude recorded from ST and IF regions. (B) Interaction effects: mean maximum peak magnitude for native English and nonnative Spanish recorded from ST (light gray) and IF (dark gray) regions from 7-mo-old infants, 11-mo-old infants, and adults. E, native English; S, nonnative Spanish.

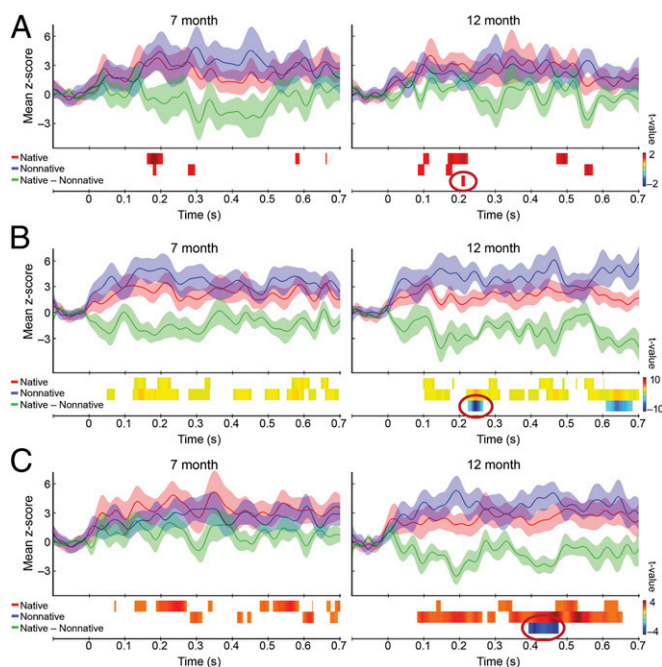


Fig. 3. Mean MMRs for native (red) and nonnative (blue) contrasts relative to prestimulus baseline, and for the native minus nonnative (green) for ST (A), Broca's area (B), and the cerebellum (C). Waveforms for 7-mo-old and 12-mo-old infants are shown on the left and right, respectively. Significant temporal clusters are shown below waveforms, FWER-corrected, $P < 0.05$. Positive t values (red) indicate native > nonnative; negative t values (blue) the reverse. Red circles highlight the timing of significant differential activation for native and nonnative contrasts in each brain area.

finally, in a significant temporal cluster in cerebellar cortex at 393–476 ms ($P = 0.02$, corrected).

Discussion of Exp. 2

Exp. 2 was designed to confirm and extend Exp. 1's findings on infants' brain responses to speech in auditory and motor brain areas. A new set of native and nonnative syllables, a new population of infants, and different MEG analysis methods were used. We tested a broader range of cortical areas involved in motor control than were tested in Exp. 1.

Exp. 2 replicated Exp. 1's findings of a developmental change in response to speech in both auditory and motor brain areas. By 12 mo of age, the native MMR is larger than the nonnative MMR in auditory cortex (ST), whereas the nonnative MMR is larger than the native MMR in motor brain areas (Broca's and the cerebellum). These response patterns in auditory and motor brain areas emerge between 7 and 12 mo. At 7 mo, native vs. nonnative MMRs do not differ in either auditory or motor brain areas.

Exp. 2 thus provides further evidence that in infants, by the end of the first year of life, a double dissociation occurs: activation in auditory brain areas is greater for native speech, whereas activation in motor brain areas is greater for nonnative speech.

General Discussion

The present studies focused on the earliest phases of language learning: infants' early transition in phonetic perception. Speech perception begins as a universal phenomenon wherein infants discriminate all phonetic contrasts in the world's languages. By the end of the first year, perception becomes more specialized. Foreign language phonetic contrasts that were discriminated earlier are no longer discriminated and native speech perception improves significantly over the same period.

The present studies focused on infants' brain responses to speech to examine cortical activation patterns that underlie the transition in infant speech perception. We used MEG technology with infants who straddled the transition in perception, testing 7- and 11-mo-old infants and adults in Exp. 1, and 7- and 12-mo-old infants in Exp. 2. Auditory (ST) as well as motor brain areas (inferior frontal in Exp. 1; Broca's, cerebellar, motor, and premotor areas in Exp. 2) were studied. We compared the discriminatory mismatch responses for native and nonnative phonetic contrasts, hypothesizing that at the earliest age tested, auditory as well as motor brain areas would be activated by hearing speech, and at 7 mo of age, activated equivalently for native and nonnative contrasts. A developmental change was predicted in 11- to 12-mo-old infants; we hypothesized that activation in the ST brain area would be greater for native as opposed to nonnative speech, whereas activation in Broca's area and the cerebellum would be greater for nonnative as opposed to native speech. We expected the pattern in infants at 1 y of age to match that obtained in adult listeners.

The results of Exps. 1 and 2 support these hypotheses. In both experiments, 7-mo-old infants respond equivalently to native and nonnative contrasts in both auditory and motor brain areas. Moreover, by the end of the first year, infants' brain responses show a double dissociation: activation for native stimuli exceeds that for nonnative in auditory (ST) brain areas and activation for nonnative stimuli exceeds that for native in motor brain areas (IF in Exp. 1; Broca's and the cerebellum in Exp. 2).

Two key points related to these findings advance our understanding of infant speech processing, and raise additional questions. First, at the earliest age tested (7 mo of age), both native and nonnative speech activate motor brain areas, and equivalently. Thus, the activation of motor brain areas in response to speech at 7 mo of age is not limited to sounds that infants hear in ambient language. At this early age, infants' motor brain areas appear to be reacting to all sounds with speech-like qualities.

This result raises a question: How are infants capable of creating motor models for speech, and when in development might they begin to do so? Our working hypothesis is that infants are capable of creating internal motor models of speech from the earliest point in development at which their utterances resemble the fully resonant nuclei that characterize speech, which occurs at 12 wk of age. By 20 wk, experimental evidence indicates that infants imitate pitch patterns (73) and vowel sounds (74). We argue that speech production experience in the early months of life yields a nascent auditory-articulatory map, one with properties of an emergent "schema" that goes beyond specific action-sound pairings to specify generative rules relating articulatory movements to sound. This emerging auditory-articulatory map is likely abstract, but allows infants to generate internal motor models when perceiving sounds with speech-like qualities, regardless of whether they have experienced the specific sounds or not. Thus, we posit that infants' nascent speech motor experience is the catalyst—a "prior"—for the effects observed in the present experiments. Future MEG studies can manipulate the speech stimuli to test whether infants' speech production experience is an essential component of the motor brain activation we observed.

Second, our data demonstrate developmental change. By the end of the first year, infants' auditory and motor brain areas pattern differently for native and nonnative sounds. Auditory areas show greater activation for native sounds, whereas motor areas show greater activation for nonnative sounds, suggesting that these two brain systems are coding different kinds of information in response to speech. We offer the following tentative explanation: Hearing native speech throughout the first year increases infants' auditory sensitivity to native speech-sound differences, and strengthens sensory-motor pairings, as incoming auditory representations of native speech are linked to internally generated speech motor models based on infants' prior speech production experience. Language experience would thus serve to

strengthen knowledge of native-language speech, both perceptual and motor. By the end of the first year, the emergent schema relating speech motor movements to sound would describe native speech more accurately than nonnative speech, making it more difficult and less efficient to generate internal models for nonnative speech. Hence, our finding that nonnative speech elicits greater cortical activation (indicating greater effort) in Broca's area and the cerebellum than native speech.

Our results suggest that motor brain areas play a role in speech learning, one we believe is consistent with an AxS view. In what follows, we develop this position and its implications by discussing three points: (i) perception-action theories of developmental speech perception, (ii) the impact of motherese on language learning, and (iii) the social-gating hypothesis and the development of human social understanding.

Perception-Action Theories of Developmental Speech Perception.

Early theories of speech perception incorporated knowledge of speech production. The Motor Theory held that innate knowledge of speech gestures mediated the auditory perception of speech, even in infants (27). Analysis by Synthesis incorporated listeners' motor knowledge in speech perception as a "prior," and argued that speech production experience enabled listeners to generate internal motor models of speech that served as hypotheses to be tested against incoming sensory data (30).

Early models of developmental speech perception, such as our Native Language Neural Commitment concept (1), described a process of "neural commitment" to the auditory patterns of native speech. Revisions in the model, named Native Language Magnet-Expanded (NLM-e), described emergent links between speech perception and production (9). The present data will allow further refinement of the model by suggesting how speech perception and speech production become linked early in development: infants' brains respond to hearing speech by activating motor brain areas, coregistering perceptual experience and motor brain patterns. Given 20-wk-old infants' abilities to imitate vowels in laboratory tests by generating vocalizations that acoustically and perceptually correspond to the ones they hear (74), we posit that infants use prior speech production experience to generate internal motor models of speech as they listen to us talk.

To our knowledge, the present data suggest, for the first time, a developmental theory that refines previous theory by incorporating infants' nascent speech production skills in speech perception learning. On this view, both auditory and motor components contribute to the developmental transition in speech perception that occurs at the end of the first year of life.

The Impact of "Motherese" on Language Acquisition. Work in this laboratory has emphasized the role of language input to infants, especially the role of motherese, the acoustically exaggerated, clear speech produced by adults when they address young infants (75). Our most recent work in this arena recorded the entirety of infants' auditory experience at home over several days and demonstrated that two factors in language addressed to 11- and 14-mo-old infants predicted infants' concurrent babbling and their future language performance: (i) the prevalence of motherese speech style in speech directed to infants, and (ii) the prevalence of one-on-one (vs. group) linguistic interactions between adult and child (76). Based on the present study's

results, we speculate that motherese speech, with its exaggerated acoustic and articulatory features, particularly in one-on-one settings, enhances the activation of motor brain areas and the generation of internal motor models of speech. The present experiment used synthesized (standard) speech and infants were not in face-to-face interaction; infants instead heard speech via a loudspeaker. Future MEG experiments can manipulate these factors to test the idea that motherese, especially in face-to-face mode, alters motor brain activation in infants.

The "Social-Gating" Hypothesis and Human Social Understanding. In the last decade we described a social-gating hypothesis regarding early speech learning to account for the potent effects of social interaction on language learning (16). We demonstrated dramatic differences in foreign-language learning under social and nonsocial conditions: when 9-mo-old infants are exposed to a novel language during human social interaction versus exposure to the same material via television, phonetic learning is robust for infants exposed to live tutors and nonexistent via television (14). We argue that social interaction improves language learning by motivating infants via the reward systems of the brain (77) and by providing information not available in the absence of social interaction (11).

Theories of social understanding in adults (32) and infants (78) suggest that humans evolved brain mechanisms to detect and interpret humans' actions, behaviors, movements, and sounds. The present data contribute to these views by demonstrating that auditory speech activates motor areas in the infant brain. Motor brain activation in response to others' communicative signals could assist broader development of social understanding in humans.

Our findings offer an opportunity to test children with developmental disabilities, such as autism spectrum disorder (79, 80), whose social and language deficits are potentially associated with a decreased ability to activate motor brain systems in response to human signals.

Future imaging studies using infant MEG will allow us to test younger infants to determine the role of speech production experience, and target key brain areas argued to be involved in the interface between sensory and motor representations for speech (e.g., inferior parietal lobe) (81). MEG makes it possible to chart the developmental time course of activation in these areas very precisely. The use of causality analyses will help us build causal models of these processes in young infants.

Early theorists envisioned a potential role for motor systems in perceptual speech processing (23, 30). Developmental neuroscience studies using MEG now allow us to test formative hypotheses regarding human speech and its development. The present study illustrates a growing interdisciplinary influence among scientists in psychology, machine learning, and neuroscience (82), as foreshadowed by the concept of analysis by synthesis (30).

ACKNOWLEDGMENTS. The authors thank Denise Padden for valuable assistance throughout data collection and manuscript preparation; Andrew Meltzoff for helpful comments on earlier drafts; and Ken Stevens for long discussions about Analysis by Synthesis 20 y ago. The work described here was supported by a grant to the University of Washington LIFE (Learning in Informal and Formal Environments) Center from the National Science Foundation Science of Learning Center Program (SMA-0835854) (to Principal Investigator P.K.K.), and through private philanthropic support.

- Kuhl PK (2004) Early language acquisition: cracking the speech code. *Nat Rev Neurosci* 5(11):831–843.
- Saffran JR, Werker JF, Werner LA (2006) in *Handbook of Child Psychology*, eds Lerner RM, Damon W (Wiley, Hoboken, NJ), pp 58–108.
- de Boer B, Kuhl PK (2003) Investigating the role of infant-directed speech with a computer model. *Acoust Res Lett Online* 4(4):129–134.
- Vallabha GK, McClelland JL, Pons F, Werker JF, Amano S (2007) Unsupervised learning of vowel categories from infant-directed speech. *Proc Natl Acad Sci USA* 104(33):13273–13278.
- Kuhl PK (2010) Brain mechanisms in early language acquisition. *Neuron* 67(5):713–727.
- Kuhl PK, et al. (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev Sci* 9(2):F13–F21.
- Werker JF, Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav Dev* 7(1):49–63.
- Rivera-Gaxiola M, Silva-Pereyra J, Kuhl PK (2005) Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Dev Sci* 8(2):162–172.
- Kuhl PK, et al. (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci* 363(1493):979–1000.

10. Best CT, McRoberts GW, Sithole NM (1988) Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *J Exp Psychol Hum Percept Perform* 14(3):345–360.
11. Kuhl PK (2011) In *The Oxford Handbook of Social Neuroscience*, eds Decety J, Cacioppo J (Oxford Univ Press, New York), pp 649–667.
12. Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044):606–608.
13. Maye J, Werker JF, Gerken L (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3):B101–B111.
14. Kuhl PK, Tsao FM, Liu HM (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci USA* 100(15):9096–9101.
15. Conboy BT, Kuhl PK (2011) Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Dev Sci* 14(2):242–248.
16. Kuhl PK (2007) Is speech learning ‘gated’ by the social brain? *Dev Sci* 10(1):110–120.
17. Näätänen R, et al. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385(6615):432–434.
18. Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychon Bull Rev* 13(3):361–377.
19. Poeppel D, Monahan PJ (2011) Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Lang Cogn Process* 26(7):935–951.
20. von Humboldt W (1836) *On Language: The Diversity of Human Language Structure and its Influence on the Mental Development of Mankind*, trans Heath P. (1988) (Cambridge Univ Press, Cambridge, UK).
21. Helmholtz H (1867) *Handbuch der Physiologischen Optik* (Voss, Leipzig).
22. De Coudemoy G (1668) *Discours Physique de la Parole* (Florentin Lambert, Paris).
23. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74(6):431–461.
24. Fowler CA (2006) Compensation for coarticulation reflects gesture perception, not spectral contrast. *Percept Psychophys* 68(2):161–177.
25. Repp BH (1984) In *Speech and Language: Advances in Basic Research and Practice*, ed Lass NJ (Academic, New York), pp 243–335.
26. Eimas PD, Siqueland ER, Jusczyk P, Vigorito J (1971) Speech perception in infants. *Science* 171(3968):303–306.
27. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21(1):1–36.
28. Kuhl PK, Miller JD (1975) Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190(4209):69–72.
29. Kuhl PK, Padden DM (1983) Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J Acoust Soc Am* 73(3):1003–1010.
30. Stevens KN, Halle M (1967) In *Models for the Perception of Speech and Visual Form: Proceedings of a Symposium*, ed Waltham-Dunn W (MIT Press, Cambridge, MA), pp 88–102.
31. Rizzolatti G, Arbib MA (1998) Language within our grasp. *Trends Neurosci* 21(5):188–194.
32. Hari R, Kujala MV (2009) Brain basis of human social interaction: From concepts to brain imaging. *Physiol Rev* 89(2):453–479.
33. Pulvermüller F, Fadiga L (2010) Active perception: Sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci* 11(5):351–360.
34. Hickok G (2009) Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci* 21(7):1229–1243.
35. Yuille A, Kersten D (2006) Vision as Bayesian inference: Analysis by synthesis? *Trends Cogn Sci* 10(7):301–308.
36. Benson RR, et al. (2001) Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang* 78(3):364–396.
37. Démonet JF, et al. (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115(Pt 6):1753–1768.
38. Price CJ, et al. (1996) Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain* 119(Pt 3):919–931.
39. Zatorre RJ, Binder JR (2000) Functional and structural imaging of the human auditory system. *Brain Mapping: The Systems*, eds Toga AW, Mazziotta JC (Academic, San Diego), pp 365–402.
40. Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256(5058):846–849.
41. Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22(3):1182–1194.
42. Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage* 33(1):316–325.
43. Iacoboni M (2008) The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *J Physiol Paris* 102(1–3):31–34.
44. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7(7):701–702.
45. Callan DE, et al. (2003) Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage* 19(1):113–124.
46. Dehaene S, et al. (1997) Anatomical variability in the cortical representation of first and second language. *Neuroreport* 8(17):3809–3815.
47. Golestani N, Zatorre RJ (2004) Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage* 21(2):494–506.
48. Menning H, Imaizumi S, Zwitserlood P, Pantev C (2002) Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learn Mem* 9(5):253–267.
49. Nakai T, et al. (1999) A functional magnetic resonance imaging study of listening comprehension of languages in human at 3 tesla-comprehension level and activation of the language areas. *Neurosci Lett* 263(1):33–36.
50. Wang Y, Sereno JA, Jongman A, Hirsch J (2003) fMRI evidence for cortical modification during learning of Mandarin lexical tone. *J Cogn Neurosci* 15(7):1019–1027.
51. Skipper JL, van Wassenhove V, Nusbaum HC, Small SL (2007) Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17(10):2387–2399.
52. Levelt WJM (2001) Spoken word production: a theory of lexical access. *Proc Natl Acad Sci USA* 98(23):13464–13471.
53. Poeppel D, Idsardi WJ, van Wassenhove V (2008) Speech perception at the interface of neurobiology and linguistics. *Philos Trans R Soc Lond B Biol Sci* 363(1493):1071–1086.
54. Imada T, et al. (2006) Infant speech perception activates Broca’s area: A developmental magnetoencephalography study. *Neuroreport* 17(10):957–962.
55. Dehaene-Lambertz G, et al. (2006) Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc Natl Acad Sci USA* 103(38):14240–14245.
56. Akiyama LF, et al. (2013) Age-specific average head template for typically developing 6-month-old infants. *PLoS ONE* 8(9):e73821.
57. Bosseler AN, et al. (2013) Theta brain rhythms index perceptual narrowing in infant speech perception. *Front Psychol* 4(690):690.
58. Taulu S, Kajola M, Simola J (2004) Suppression of interference and artifacts by the signal space separation method. *Brain Topogr* 16(4):269–275.
59. Hämäläinen MS, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65(2):413–497.
60. Hari R, et al. (1984) Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neurosci Lett* 50(1–3):127–132.
61. Callan D, Callan A, Gamez M, Sato MA, Kawato M (2010) Premotor cortex mediates perceptual performance. *Neuroimage* 51(2):844–858.
62. Fedorenko E, Behr MK, Kanwisher N (2011) Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci USA* 108(39):16428–16433.
63. Booth JR, Wood L, Lu D, Houk JC, Bitan T (2007) The role of the basal ganglia and cerebellum in language processing. *Brain Res* 1133(1):136–144.
64. Deniz Can D, Richards T, Kuhl PK (2013) Early gray-matter and white-matter concentration in infancy predict later language skills: A whole brain voxel-based morphometry study. *Brain Lang* 124(1):34–44.
65. Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys Med Biol* 51(7):1759–1768.
66. Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53(1):1–15.
67. Gramfort A, et al. (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86(1):446–460.
68. Pascual-Marqui RD (2002) Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods Find Exp Clin Pharmacol* 24(Suppl D):5–12.
69. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12(5):535–540.
70. Vul E, Kanwisher N (2010) In *Foundational Issues in Human Brain Mapping*, eds Hanson S, Bunzl M (MIT Press, Cambridge, MA), pp 71–91.
71. Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164(1):177–190.
72. Ridgway GR, Litvak V, Flandin G, Friston KJ, Penny WD (2012) The problem of low variance voxels in statistical parametric mapping: A new hat avoids a ‘haircut’. *Neuroimage* 59(3):2131–2141.
73. Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. *Science* 218(4577):1138–1141.
74. Kuhl PK, Meltzoff AN (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change. *J Acoust Soc Am* 100(4 Pt 1):2425–2438.
75. Kuhl PK, et al. (1997) Cross-language analysis of phonetic units in language addressed to infants. *Science* 277(5326):684–686.
76. Ramirez-Esparza N, Garcia-Sierra A, Kuhl PK (2014) Look who’s talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Dev Sci*, 10.1111/desc.12172.
77. Stavropoulos KKM, Carver LJ (2013) Research review: Social motivation and oxytocin in autism—Implications for joint attention development and intervention. *J Child Psychol Psychiatry* 54(6):603–618.
78. Marshall PJ, Meltzoff AN (2014) Neural mirroring mechanisms and imitation in human infants. *Philos Trans R Soc Lond B Biol Sci* 369(1644):20130620.
79. Kuhl PK, et al. (2013) Brain responses to words in 2-year-olds with autism predict developmental outcomes at age 6. *PLoS ONE* 8(5):e64967.
80. Stavropoulos KK, Carver LJ (2014) Reward anticipation and processing of social versus nonsocial stimuli in children with and without autism spectrum disorders. *J Child Psychol Psychiatry*, 10.1111/jcpp.12270.
81. Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat Neurosci* 12(6):718–724.
82. Meltzoff AN, Kuhl PK, Movellan J, Sejnowski TJ (2009) Foundations for a new science of learning. *Science* 325(5938):284–288.