Learning to Match Auditory and Visual Speech Cues: Social Influences on Acquisition of Phonological Categories

ARTICLE *in* CHILD DEVELOPMENT · NOVEMBER 2014

Impact Factor: 4.92 · DOI: 10.1111/cdev.12320

2 AUTHORS:

Nicole Altvater-Mackensen
Max Planck Institute for Human Cognitive an…
11 PUBLICATIONS **21** CITATIONS
SEE PROFILE

Tobias Grossmann
University of Virginia
52 PUBLICATIONS **1,204** CITATIONS
SEE PROFILE

CHILD DEVELOPMENT

Child Development, xxxx 2014, Volume 00, Number 0, Pages 1–17

Learning to Match Auditory and Visual Speech Cues: Social Influences on Acquisition of Phonological Categories

Nicole Altvater-Mackensen and Tobias Grossmann
Max Planck Institute for Human Cognitive and Brain Sciences

Infants' language exposure largely involves face-to-face interactions providing acoustic and visual speech cues but also social cues that might foster language learning. Yet, both audiovisual speech information and social information have so far received little attention in research on infants' early language development. Using a preferential looking paradigm, 44 German 6-month olds' ability to detect mismatches between concurrently presented auditory and visual native vowels was tested. Outcomes were related to mothers' speech style and interactive behavior assessed during free play with their infant, and to infant-specific factors assessed through a questionnaire. Results show that mothers' and infants' social behavior modulated infants' preference for matching audiovisual speech. Moreover, infants' audiovisual speech perception correlated with later vocabulary size, suggesting a lasting effect on language development.

Within the 1st year of life, infants become attuned to the phonological characteristics of their native language and gain profound knowledge about its sound system. This includes changes to the categorical perception of speech sounds: Infants lose their sensitivity to non-native sound contrasts while refining their sensitivity to those sound contrasts that are phonemic in their native language (e.g., Kuhl et al., 2006; Werker & Tees, 1999; see Saffran, Werker, & Werner, 2006, for a review). This process of perceptual narrowing occurs earlier for vowels than for consonants (e.g., Polka & Werker, 1994) and is modulated by various factors in the speech input that infants receive, such as the statistical distribution and frequency of sounds (Anderson, Morgan, & White, 2003; Maye, Werker, & Gerken, 2002) and their acoustic characteristics (Narayan, Werker, & Beddor, 2010; Polka & Bohn, 1996).

Yet, there is more to speech than just sounds. Naturally, infants learn language in interactions with their caregivers. This face-to-face interaction provides further sources of information: First, infants will not only receive acoustic information but also visual input on the mouth gestures that are associated with the production of a certain sound. Second, caregivers might convey additional social information that captures infants' attention and fosters learning (see Kuhl, 2007). Both audiovisual speech cues and social information have so far received relatively little attention in the research on infants' early language learning. The current study therefore addresses the question of whether mothers' behavior influences infants' perception of audiovisual speech. In the following, we will briefly summarize previous findings on infants' perception of audiovisual speech and on social factors that influence phoneme learning before we describe the specifics of the current study in more detail.

Audiovisual Speech Perception in Infants

In the first study that investigated bimodal speech perception in infancy, Kuhl and Meltzoff (1982) presented 4.5- to 5-month-olds with side-by-side videos of a female articulating two vowels while they heard one of the vowels spoken in synchrony with the facial movements. Results showed that infants prefer to look at the matching face, that is, the video of the female mouthing the heard sound, suggesting that the sensitivity to the correspondence between auditory and visual speech cues develops early in the 1st year of life. This finding has later been replicated with a different vowel contrast (Kuhl & Meltzoff, 1988), with speakers of different gender (Patterson & Werker, 1999), and with

This research was supported by funding from the Max Planck Society awarded to Tobias Grossmann. We thank Caterina Böttcher for her help in collecting and coding the data.

Correspondence concerning this article should be addressed to Nicole Altvater-Mackensen, Max Planck Research Group *Early Social Development*, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany. Electronic mail may be sent to altvater@cbs.mpg.de.

 0009-3920/2014/xxxx-xxxx
DOI: 10.1111/cdev.12320

infants as young as 2 months of age (Patterson & Werker, 2003). Furthermore, findings by Bristow et al. (2009) suggest that infants have cross-modal representations of speech sounds that link visual and auditory information. Using event-related potentials, this study found that 2.5-month-olds detect correspondences between an auditory sound and a preceding auditory or visually presented sound, and that brain responses were similar for sounds that matched within and across modalities. In addition to being able to detect links between auditory and visual speech cues, infants have been shown to integrate cross-modal information to create a fused percept. Specifically, work by Burnham and Dodd (2004) and Kushnerenko, Teinonen, Volein, and Csibra (2008) suggest that infants experience the McGurk effect (McGurk & MacDonald, 1976), that is, perceive the merged sound /da/ when hearing /ba/ and seeing /ga/.

Despite this early sensitivity for audiovisual speech cues, infants' perception appears substantially modulated by language experience. Pons, Lewkowicz, Soto-Faraco, and Sebastian-Galles (2009) showed that perceptual narrowing does not only apply to the auditory but also to the audiovisual domain. They tested English- and Spanish-learning infants on the /b/–/v/ contrast, which is phonemic in English but not in Spanish. In line with the perceptual narrowing account, this study showed that Spanish-learning infants succeeded at audiovisual matching in this task at the age of 6 months but not at the age of 11 months, while English-learning infants were able to correctly map sound and visual gesture at both ages. This suggests that infants' sensitivity to the correspondence of audiovisual cues that are not relevant to their native language declines, concurring with earlier findings by Weikum et al. (2007) showing that infants lose their sensitivity to the visual differences between languages in the course of native language attunement.

Moreover, infants' ability to match auditory and visual speech cues does not necessarily extend to all sound contrasts and integration of cross-modal speech information might initially still be limited. For example, Mugitani, Kobayashi, and Hiraki (2008) found that Japanese 8-month-olds have difficulties mapping noncanonical speech sounds like whistles to the corresponding mouth gesture, and findings by MacKain, Studdert-Kennedy, Spieker, and Stern (1983) indicate that 5- to 6-month-olds do not always consistently match longer sound sequences like a heard disyllable to the appropriate face articulating the disyllable. Furthermore, the integration of auditory and visual speech cues does not seem to be mandatory in young infants (see Desjardins & Werker, 2004) and even toddlers and children up to 9 years of age are still less likely to experience the McGurk effect than adults, suggesting that the processes leading to audiovisual integration of speech are not fully developed until late childhood (Desjardins, Rogers, & Werker, 1997; McGurk & MacDonald, 1976; Nath, Fava, & Beauchamp, 2011).

Given that language experience shapes the perception of audiovisual speech, the question arises which factors are influencing this development. Research on auditory speech perception suggests that statistical and distributional information plays a significant role (e.g., Maye et al., 2002). This information presumably also modulates audiovisual speech perception (see Pons et al., 2009; Teinonen, Aslin, Alku, & Csibra, 2008). However, the learning of audiovisual relations critically relies on the visibility of the mouth gesture accompanying a sound. Thus, face-to-face interactions are presumably essential to the development of audiovisual speech perception. This makes it especially important to investigate whether there are any specific characteristics of mother–infant interactions that foster learning of audiovisual speech categories.

Social Influences on Phoneme Learning

So far, no study has directly examined social influences on infants' audiovisual speech perception and only few studies have investigated social influences on auditory phoneme learning. A relatively well-investigated social cue is infant-directed speech (IDS). IDS is a special speech style that is used to converse with infants and that is characterized by higher pitch, exaggerated pitch contours, slower speaking rate, and shorter sentences (Fernald & Simon, 1984; Fernald et al., 1989). This type of speech seems to be highly engaging as infants prefer to listen to infant-directed over adult-directed speech (Cooper & Aslin, 1990) and even show a lasting preference for persons that use IDS over persons that use adult-directed speech (Schachner & Hannon, 2011). Indeed, IDS seems to promote the formation of native-language sound categories. For instance, 6- to 8- and 10- to 12-month-old Chinese infants with mothers who use more stretched vowel spaces in their IDS are better able to discriminate the native /tç^{h}i/-/çi/ contrast (Liu, Kuhl, & Tsao, 2003). This suggests that more pronounced IDS helps infants to focus on the relevant cues that distinguish native phonemes.

In addition to IDS, face-to-face interactions might enhance language learning. Kuhl, Tsao, and Liu (2003) exposed 9-month-old English-learning infants to Mandarin Chinese. One group of infants received the language input in face-to-face sessions from a native language tutor while reading books, and two other groups of infants were exposed to video- or audio-taped versions of the live interactions. After 12 sessions, infants from the live-interaction group had learned to discriminate a Mandarin Chinese sound contrast, while the infants from the video and audio groups were not sensitive to the non-native sound difference. Kuhl et al. therefore suggested that social information plays a crucial role in triggering the learning of a language's sound system.

Similarly, contingent responding seems to foster phoneme learning. Goldstein, King, and West (2003) show that 8-month-olds rapidly restructure their own babbling based on their mothers' responses. Those mothers who respond contingently to their infants' babbling elicit more native-like productions from their infants than mothers who respond in a random, noncontingent way (see also Goldstein & Schwade, 2008). Applying a similar line of reasoning to infant speech perception, Elsabbagh et al. (2013) investigated whether the contingency of mothers' responses influences perceptual reorganization. Their results show that 6-month-olds with mothers who show more contingent responses to their infants' actions are no longer sensitive to non-native speech contrasts, while 6-month-olds with less contingently behaving mothers still showed non-native sound discrimination. This suggests that contingent responding helps infants to ignore irrelevant sound contrasts and to focus on those sound differences that are important in the native language.

Taken together, these studies indicate that the characteristics of mothers' speech and behavior influence infants' learning of phonemic categories in speech perception and production. Although the precise mechanisms that facilitate learning are still unclear, it seems reasonable that enhanced auditory cues that are characteristic for IDS highlight relevant sound contrasts and thereby facilitate their learning (but see Benders, 2013; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). Contingent (vocal) responding and more specifically being imitated, however, might help infants to associate visual, acoustic, and articulatory information to create multisensory representations of sounds (see Ray & Heyes, 2011; we will come back to this point in more detail in the Discussion).

Both IDS and contingent responding might also aid learning more generally by increasing infants' arousal and attention. If heightened attention facilitates learning, then infants who are more attentive or easier to engage might also be better (language) learners. To our knowledge, thus far there is no work that directly tested this prediction. However, Kuhl, Coffey-Corina, Padden, and Dawson (2005) found that autistic children who are not engaged by IDS are worse in discriminating (native) speech sounds than autistic children who are engaged by IDS. Work by Conboy, Brooks, Meltzoff, and Kuhl (2008) further suggests that attention influences phoneme learning. They found that 10-month-old English-learning infants' overall attention to and their shared attention with a Spanish-speaking tutor predicted their success in discriminating phonemes in the non-native language. These findings indicate that difficulties in perceiving speech contrasts might be related to difficulties in eliciting or maintaining attention to (linguistic) stimuli more generally.

The Current Study

The current study addresses the question of how sensitive infants are to audiovisual speech contrasts and how social factors influence phoneme learning. More specifically, we asked whether mothers' and infants' speech and behavior relates to infants' audiovisual speech perception. To assess audiovisual speech perception, we tested German 5.5- to 6-month-olds' ability to detect mismatches between auditory and visually presented native vowels. Infants were presented with different videos of a female mouthing a vowel that was either congruent or incongruent with the vowel they concurrently heard, while their looking times (LTs) to the videos were measured. If infants were sensitive to the congruency between auditory and visual speech cues, we expected them to look longer at matching than mismatching videos (see Mugitani et al., 2008). To ensure that infants were able to acoustically discriminate the vowels, we also presented them with an auditory discrimination task. Again, a preference paradigm was used. We measured infants' attention to trials in which two vowels were presented in alternation (alternating trials) and trials in which one vowel was repeated (nonalternating trials). If infants discriminated the vowels, we expected them to listen longer to alternating trials than nonalternating trials (see Best & Jones, 1998). We decided to test infants aged 5.5–6 months because they are old enough to have a sufficiently large attention span

to complete the experiment, while they are still young enough to be in the midst of native language attunement (see Polka & Werker, 1994, that perceptual narrowing for vowels occurs between 4 and 10 months of age).

After the speech perception experiment, we videotaped the mothers while they freely played with their infant to assess the infant directedness of their speech and their interactive behavior (see Henning, Striano, & Lieven, 2005), and we collected questionnaire data to assess infant-specific characteristics that might contribute to language learning like infants' ability to focus attention, their perceptual sensitivity, and their vocal productivity. Based on previous studies, we predicted that mothers' interactive style would be positively related to infants' performance in the language task; that is, those infants with more reactive mothers and with mothers who use more IDS would be better at detecting audiovisual mismatches. We further hypothesized that infants who are more sensitive to perceptual cues and more attentive to their environment might also perform better in the language task. To see whether any of these variables have a lasting effect on language development, we also assessed infants' vocabulary size 6 months after testing, that is, around their first birthday (see Tsao, Liu, & Kuhl, 2004). Table 1 provides an overview of the assessed variables and the results of the study.

Table 1
Overview of Assessed Variables and Their Influence on Audiovisual (AV) Speech Perception at 6 Months and Vocabulary Size at 12 Months of Age

	AV perception at 6 months	Vocabulary size at 12 months
Interactive behavior		
Mother	No	No
Infant	No	No
Imitation	Yes [R^2 = .102]	No
Mothers' speech style		
% of speech	No	No
% of vocal play	Yes [R^2 = .088]	Yes [R^2 = .114]
Pitch characteristics	No	No
Infant-specific factors		
Duration of orientation	Yes [R^2 = .099]	No
Perceptual sensitivity	No	No
Vocal productivity	Yes [R^2 = .109]	Yes [R^2 = .263]
Preference for matching AV speech at 6 months	n.a.	Yes [R^2 = .130]

Note. Effect sizes of correlations are given in square brackets.

Method

Participants

Forty-four German 6-month-olds (16 female) from a monolingual language environment participated in the experiment (age range = 5 months 11 days [5;11] to 6;04, M_{age} = 5;23). All infants were born full term with normal birth weight and had no reported hearing or vision impairment. Seven additional infants (four female) started to cry, and one additional infant was tested but excluded from analysis because his LTs were more than 2 *SD* away from the mean. Infants were recruited via the subject pool of the authors' institute. Parents gave informed consent to participate in the study and received 7.50 euro and a toy for their infant for participation.

Language Assessment

Stimuli

Visual stimuli for the discrimination task consisted of the video of a moving colored toy water wheel against a black background (Stager & Werker, 1997). Video frames were 1,200 pixels wide and 880 pixels high, resulting in a width of 32 cm and a height of 24 cm on screen. The video was accompanied by successive repetition of six different tokens of /a/ in nonalternating trials, and three tokens of each /a/ and /e/ or /a/ and /o/ in alternating trials (see Best & Jones, 1998, for the use of alternating and nonalternating trials to assess sound discrimination in infants). Tokens were separated by approximately 1.5 s of silence, leading to a trial length of 15 s. All vowels were spoken by a female native speaker of German using IDS (see below for further details on the acoustic characteristics of the stimuli).

Visual stimuli for the audiovisual matching task consisted of videos showing a woman articulating /a/, /e/, and /o/. Each video entailed six consecutive utterances of the respective vowel. Each utterance started and ended with the mouth completely shut in neutral position. Visual stimuli were hyperarticulated to mimic IDS. Each vowel articulation was separated by approximately 3 s in which the woman kept a friendly open face and smiled at the infant, leading to a trial length of 30 s. The eye gaze was always directed toward the infant. All videos were zoomed and cropped so that they only showed the woman's head against a light gray wall. Video frames were 1,024 pixels wide and 1,000 pixels high, resulting in a width of 27 cm and

a height of 26 cm on screen. Figure 1 shows an example frame of the mouth position for each of the fully articulated vowels.

Audio stimuli for the audiovisual matching task consisted of six different tokens each of /a/, /e/, and /o/, spoken in IDS. Stimuli were recorded from the same woman who produced the visual stimuli and the auditory stimuli of the discrimination task. The length of the vowels was timed to match the length of the mouthing in the videos. The final stimuli were created by dubbing the audio recordings of the vowels onto the videos of the woman mouthing the vowels. This ensured that both matching and mismatching trials paired different auditory and visual tokens. For matching trials, visual and auditory vowels were the same; that is, seen /a/ was accompanied by heard /a/, seen /e/ by heard /e/, and seen /o/ by heard /o/. For mismatching trials, visual and auditory vowels did not fit, that is, seen /a/ was accompanied by heard /e/, seen /e/ by heard /a/, seen /a/ by heard /o/, and seen /o/ by heard /a/ (see Mugitani et al., 2008, for the use of matching and mismatching trials to assess sensitivity to audiovisual congruency in infants). Note that visual and auditory stimuli used in matching and mismatching trials were identical; only their pairing changed across trial types.

Three additional familiarization trials were created using different recordings of the same woman uttering each vowel twice in a block of three repetitions followed by an engaging smile and raise of her eyebrows. All stimuli were digitally recorded in a quiet room with a sampling rate of 24 frames per second and 44.100 Hz. Vowels were matched in volume (mean intensity: /a/ = 76.0 db, /e/ = 77.3 db, /o/ = 77.1 db), fundamental frequency (mean pitch: /a/ = 191.5 Hz, /e/ = 196.8 Hz, /o/ = 199.8 Hz), and length (mean duration: /a/ = 1.83 s, /e/ = 1.79 s, /o/ = 1.84 s). Furthermore, /e/ and /o/ matched in vowel height (mean F1: /a/ = 986.3 Hz, /e/ = 433.2 Hz, /o/ = 462.8 Hz) and differed similarly from /a/ in vowel backness (mean F2: /a/ = 1597.8 Hz, /e/ = 2754.3 Hz, /o/ = 956.2 Hz).

Procedure

Infants were seated on their parent's lap in a quiet experimental room, facing a 52-cm-wide and 32.5-cm-high TV screen at a distance of 40 cm from the screen. Parents wore headphones playing music intermixed with speech during the experiment and were instructed to interact as little as possible with their infant. A camera mounted below the screen recorded infants' eye movements during the experiment. Auditory stimuli were presented via loudspeakers that were located behind the screen. Stimuli were presented using the Presentation® software (http://www.neurobs.com). Based on the video image, the experimenter started a trial when the infant was looking at the screen and continued to indicate throughout the trial whether the infant was looking at the screen or away by pressing a button on a keyboard. Each trial lasted until the infant was looking away for more than 2 consecutive seconds or until completion. In between trials, a flashing light was displayed in silence to reorient infants toward the screen.

Each infant was first presented with the three familiarization trials showing the woman uttering /a/, /e/, and /o/, to familiarize infants with the speaker and her characteristics. This was immediately followed by the discrimination task. Infants were presented with a total of eight nonalternating /a/ trials, four alternating /a/–/e/ trials, and four alternating /a/–/o/ trials. Half of the alternating trials started with /a/; the other half started with /e/ or /o/, respectively. Trials were ordered so that each nonalternating trial was followed by an

Figure 1. Example of the mouth position for fully articulated /a/, /e/, and /o/.

alternating trial and so that /a/–/e/ and /a/–/o/ trials were evenly distributed across the 16 discrimination trials.

The discrimination task was immediately followed by the audiovisual matching task. Infants were presented with nine matching trials, three trials each for /a/, /e/, and /o/ and eight mismatching trials. Two mismatching trials each paired visual /a/ with auditory /e/, visual /e/ with auditory /a/, visual /a/ with auditory /o/, and visual /o/ with auditory /a/. Trials were ordered so that no more than two consecutive trials contained the same auditory or visual stimulus and so that the different vowels as well as matching and mismatching trials were evenly distributed across the 17 audiovisual trials. On average, the experiment took approximately 10 min.

Data Analysis

Based on the online codings of the experiments, we calculated the mean LT to alternating and nonalternating trials in the discrimination task and to matching and mismatching trials in the audiovisual matching task for each infant. We then calculated difference scores assessing infants' preference for alternating over nonalternating trials in the discrimination task (Pref.alt = LTalternating − LTnonalternating) and infants' preference for matching over mismatching trials in the audiovisual matching task (Pref.match = LTmatching − LTmismatching).

To assess reliability of the online codings, data from 15% of the infants were reassessed offline using a digital video scoring system. A trained coder indicated for each 40-ms frame of the video whether the infant was looking at the screen or away. The coder was blind to experiment phase and trial type. The coding output was aligned with information about the phase of the experiment and the auditory stimulus presented. Mean LTs to the different trial types and difference scores were calculated as described above and compared to the online codings. Reliability between online and offline codings was 99%, $r = .994$, $p < .001$.

Interaction Assessment

Procedure

After the infants had taken part in the experiment, parents and infants were given a short break before their interaction was recorded. Recordings were made in the playing area of the laboratory. Mother and infant were seated on a play mat and provided with a choice of small toys and play books. The mother was informed that the recording was meant to assess infants' natural reaction to their parents' voice and face, and was asked to play with the infant like she would usually do at home. No further instruction was given. Two cameras mounted in the corners of the playing area recorded separate videos of mother and infant while they were freely playing. After the experimenter had started the recordings, she left the room for the time of the recording. Each mother–infant pair was recorded for approximately 5 min.

Data Analysis

The recordings of mothers and infants were aligned to create a time-locked video of the interaction of each mother–infant pair. Two coders assessed each interaction based on a set of questions evaluating mothers and infants mutual attention, responsiveness, and engagement on a scale from 1 (*very low*) to 5 (*very high*). After any disagreement between coders (difference > 1 for any individual score) was resolved by discussion, the ratings of both coders were pooled to calculate a mean interaction score for mother and infant. In addition, mothers were classified as being imitators, that is, mothers imitated their infants at least once during the free-play session, or nonimitators, that is, mothers did not imitate their infants at all.

To assess the infant directedness of mothers' speech, the audio track of each interaction was extracted from the video and annotated using the Praat software (Boersma & Weenink, 2005; http://www.praat.org). The start and end points of each individual utterance was marked and it was classified as speech or vocal play. The category vocal play included all utterances that were speech-like yet did not contain actual words but rather resembled babbling. Based on the annotations, the proportion of time that mothers talked and used vocal play during the free-play session was calculated. Because not all speech utterances had measurable pitch contours, we randomly selected a sample of 10 speech utterances per mother from which we calculated the mean pitch and the mean pitch range.

The final sample for the interaction analysis included 42 infants (15 female, age range = 5;11–6;04, M_{age} = 5;23). One infant had to be excluded because the father rather than the mother visited the laboratory with the infant, and during one recording a technical failure occurred.

Assessment of Infant-Specific Factors

At the end of each session, mothers were asked to fill out a German version of the Revised Infant Behaviour Questionnaire (Gartstein & Rothbart, 2003). The questionnaire assesses different dimensions of infant temperament. Each dimension is captured by 12 questions on infants' behavior during daily routines and activities. The primary caregiver rates the frequency of each behavior on a scale from 1 (*never*) to 7 (*always*). The following three dimensions were included in the analysis: duration of orientation (assessing the infant's attention to a single object or activity for extended periods of time) as a measure of infants' ability to focus attention, vocal reactivity (assessing the amount of vocalization exhibited by the infant during daily activities) as a measure of infants' vocal productivity, and perceptual sensitivity (assessing the detection of slight, low-intensity stimuli from the external environment) as a measure of infants' sensitivity to perceptual input.

The final sample for the analysis of infant-specific factors included 35 infants (15 female, age range = 5;11–6;04, M_{age} = 5;23). Nine infants (four female) had to be excluded because parents did not send back the questionnaire within 3 weeks after the experimental session.

Assessment of Later Vocabulary Size

Around the infants' first birthday, that is, approximately 6 months after the experimental session, we assessed infants' vocabulary size by means of a standardized German questionnaire on child development for 12-month-olds (Elternfragebogen [ELFRA] 1; Grimm & Doil, 2000). Mothers were asked to fill out a subpart of the questionnaire assessing infants' receptive and productive vocabulary. The list contains a total of 164 words from 13 semantic classes, such as animals, body parts, and activities. To estimate infants' vocabulary size, we calculated how many words mothers marked as being understood by the infant.

The final sample for the vocabulary analysis included 34 infants (12 female, age range = 5;11–6;04, M_{age} = 5;23). Ten infants (four female) had to be excluded because parents did not send back the questionnaire.

Results

Forty-four infants contributed data from the language tasks. For some infants, we did not obtain interaction data (2), questionnaire data (9), or vocabulary data (10; see the Method section for more detail). Because we did not want to disregard 25% of the data, we separately investigated the influence of mothers' interactive behavior and infant-specific factors on audiovisual speech perception and the factors influencing later vocabulary development, including all infants who contributed data for each specific data set. (Note that results are similar when including only those 32 infants who contributed data for all four parts, that is, language tasks, interaction, and questionnaire data at 6 months and vocabulary data at 12 months.)

Infants' Sensitivity to Audiovisual Mismatches

First, we examined whether infants were able to detect mismatches in audiovisual speech. A one-sample t test on the difference score between LTs to matching and mismatching trials showed that infants preferred matching trials, indicating that they were sensitive to the congruency between auditory and visually presented vowels, $t(43) = 3.097$, $p = .003$, $d = 0.467$. 32 of 44 infants looking longer at matching than mismatching trials, exact binomial $p = .004$. A one-sample t test on the difference score between LTs to alternating and nonalternating trials confirmed that infants also acoustically discriminated the vowels in the discrimination task, $t(43) = 6.368$, $p < .001$, $d = 0.960$. 37 of 44 infants looking longer at alternating than nonalternating trials, exact binomial $p < .001$. Figure 2 displays the mean difference in LT between alternating and nonalternating trials in the discrimination task and between matching and mismatching trials in the audiovisual matching task.

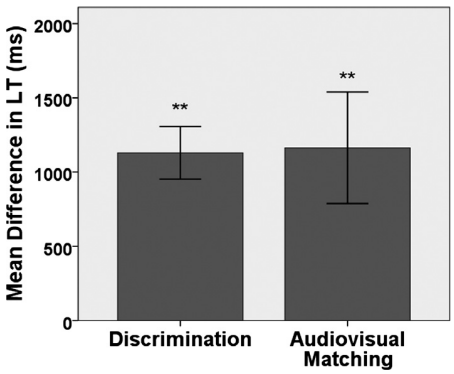

Figure 2. Mean difference in looking time (LT) to alternating versus nonalternating trials in the discrimination task and matching versus mismatching trials in the audiovisual matching task. Error bars indicate ± 1 *SE*.
**$p < .01$.

Influence of Mothers' Interactive Behavior and Speech Style

To investigate the influence of interactive behavior and speech style on infants' sensitivity to audiovisual mismatches, we correlated the difference score from the audiovisual matching task with the assessed interaction and speech variables. Pearson correlations show that mothers' imitation behavior, $r(42) = .319$, $p_{(\text{one-tailed})} = .020$, $R^2 = .102$, and the amount of time that mothers used vocal play, $r(42) = .297$, $p_{(\text{one-tailed})} = .028$, $R^2 = .088$, were positively related to infants' preference for matching trials in the audiovisual speech perception task. The amount of time that mothers talked, the mothers' mean pitch and pitch range, and the mothers' and infants' interaction scores were not related to infants' preference in the audiovisual matching task ($ps > .30$). Note that we report one-tailed p values for the correlations. Based on previous research, we expected that imitation will have positive influences on language development and specifically on the learning of phonological categories (e.g., Ray & Heyes, 2011, and references therein). The same holds for the use of exaggerated facial and vocal cues in mothers' vocal play (e.g., Green, Nip, Wilson, Mefferd, & Yunusova, 2010). Using the more conventional two-tailed p value, the positive correlation between mothers' imitation behavior and infants' preference for matching trials would remain significant ($p = .04$), while the correlation between mothers' use of vocal play and infants' preference for matching trials would only approach significance ($p = .056$). Yet, given the consistent difference between high- and low-vocal-play groups reported in the following paragraph, we feel confident that mothers' use of vocal play is positively related to infants' sensitivity to the matching between auditory and visual speech cues. Figure 3 plots infants' preference for matching trials as a function of mothers' vocal play (see Figure 4 for differences in preference based on mothers' imitation behavior).

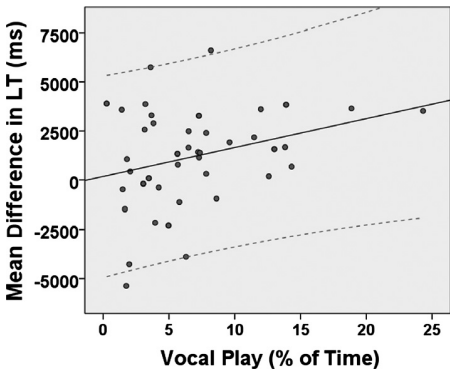

Figure 3. Mean difference in looking time (LT) to matching versus mismatching trials in the audiovisual matching task plotted against the percentage of time that mothers used vocal play during the free-play session. The continuous line depicts the linear regression line ($R^2 = .088$); the dotted lines indicate the 95% confidence interval.

To further investigate the influence of mothers' use of vocal play and imitation on infants' performance in the audiovisual matching task, we split infants into subgroups based on mothers' imitation (yes/no) and vocal play (high/low) behavior. One-sample t tests showed that infants whose mothers imitated them had a preference for matching trials, $t(19) = 5.924$, $p < .001$, $d = 1.325$, 18 of 20 infants looking longer at matching than mismatching trials, exact binomial $p < .001$. Those infants with mothers

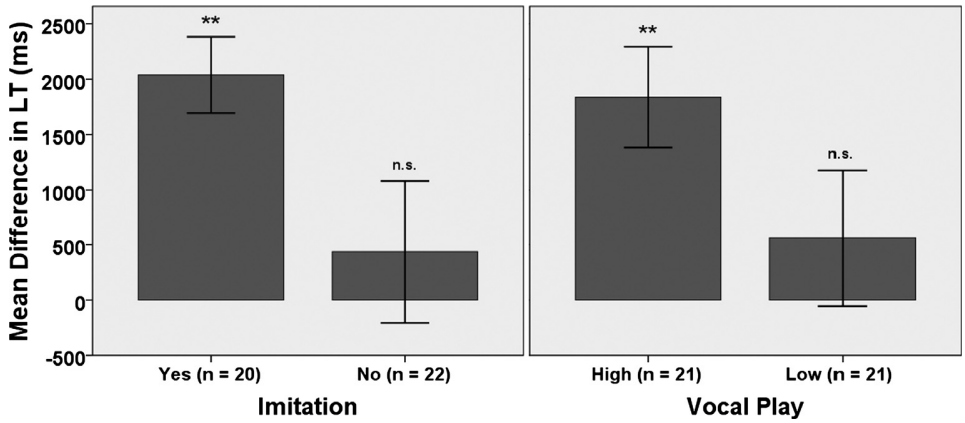

Figure 4. Mean difference in looking time (LT) to matching versus mismatching trials in the audiovisual matching task dependent on mothers' imitation and vocal play behavior (vocal play groups formed by median split). Error bars indicate ± 1 *SE*.
**$p < .01$.

who did not imitate them showed no preference in the audiovisual matching task, $t(21) = 0.678$, $p = .505$, $d = 0.145$, 13 of 22 infants looking longer at matching than mismatching trials, exact binomial $p = .523$. A chi-square test further supported the association between infants' preference for matching trials and mothers' imitation, $\chi^2(1) = 5.177$, exact $p = .035$, odds ratio = 6.231. Similarly, only infants whose mothers often used vocal play had a preference for matching trials, $t(20) = 4.042$, $p = .001$, $d = 0.882$, 19 of 21 infants looking longer at matching than mismatching trials, exact binomial $p < .001$. Those infants with mothers who only occasionally used vocal play showed no preference in the audiovisual matching task, $t(20) = 0.909$, $p = .374$, $d = 0.198$, 12 of 21 infants looking longer at matching than mismatching trials, exact binomial $p = .664$. A chi-square test further supported the association between infants' preference for matching trials and mothers' use of vocal play, $\chi^2(1) = 6.035$, exact $p = .032$, odds ratio = 7.125. Thus, only infants with mothers who imitated them and often used vocal play showed sensitivity to the congruency between auditory and visually presented vowels. Figure 4 plots infants' preference for matching trials depending on mothers' imitation and vocal play behavior.

Independent sample t tests confirmed that any difference between groups was not caused by differences in infants' overall attention to the audiovisual matching task (ps > .59). Separate one-sample t tests on the difference scores from the discrimination task confirmed that all groups acoustically discriminated the vowels (all ps < .01), ensuring that any difficulties to detect audiovisual mismatches cannot be attributed to difficulties in the acoustic discrimination of the vowels. Note that none of the assessed interaction and speech variables was related to infants' auditory speech perception (ps > .06). This suggests that mothers' interactive behavior and speech style did not influence infants' auditory speech perception. Furthermore, mothers' use of vocal play and their imitation behavior was not correlated (ps > .12), suggesting that imitating mothers do not necessarily also use more vocal play, or vice versa.

Influence of Infant-Specific Factors

To investigate the influence of infant-specific factors on infants' sensitivity to audiovisual mismatches, we correlated infants' duration of orientation, perceptual sensitivity, and vocal productivity scores from the questionnaire with their preference for matching trials in the audiovisual matching task. Pearson correlations show that infants' ability to focus their attention measured by their duration of orientation, $r(35) = .315$, $p_{(\text{one-tailed})} = .033$, $R^2 = .099$, and their vocal productivity, $r(35) = .331$, $p_{(\text{one-tailed})} = .026$, $R^2 = .109$, were positively related to infants' preference for matching trials in audiovisual speech perception. Perceptual sensitivity scores, however, were not related to infants' preference in the audiovisual matching task ($p > .39$). Again, we report one-tailed p values here because we expected a positive influence of infants' ability to focus attention on phoneme learning (Conboy et al., 2005) and infants' babbling on language development (McCune & Vihman, 2001). Note, however, that using the more conventional two-tailed p value, the positive correlation between infants' ability to focus attention and their preference for matching trials would only be marginally significant ($p = .066$), and the correlation between infants' vocal productivity and their preference for matching trials would also only be marginally significant ($p = .052$). Yet, the consistent difference between high- and low-vocal-productivity groups as well as between high and low duration of orientation groups reported in the following paragraph supports the assumption that infants' vocal productivity and infants' ability to focus attention are positively related to their sensitivity in matching auditory and visual speech cues. Figure 5 plots infants' preference for matching trials as a function of their duration of orientation and vocal productivity scores, respectively.

To further investigate the influence of infants' ability to focus their attention and their vocal productivity on performance in the audiovisual matching task, we split infants into subgroups based on their duration of orientation (high/low) and vocal productivity (high/low) scores. One-sample t tests showed that only infants with high duration of orientation scores had a preference for matching trials, $t(16) = 3.053$, $p = .008$, $d = 0.740$, 14 of 17 infants looking longer at matching than mismatching trials, exact binomial $p = .013$. Those infants with low duration of orientation scores showed no preference in the audiovisual matching task, $t(17) = 0.834$, $p = .416$, $d = 0.197$, 12 of 18 infants looking longer at matching than mismatching trials, exact binomial $p = .238$. Yet, a chi-square test did not support the association between infants' preference for matching trials and their duration of orientation, $\chi^2(1) = 1.126$, exact $p = .443$, odds ratio = 2.333. Similarly, only infants with high vocal productivity scores had a preference for matching trials,

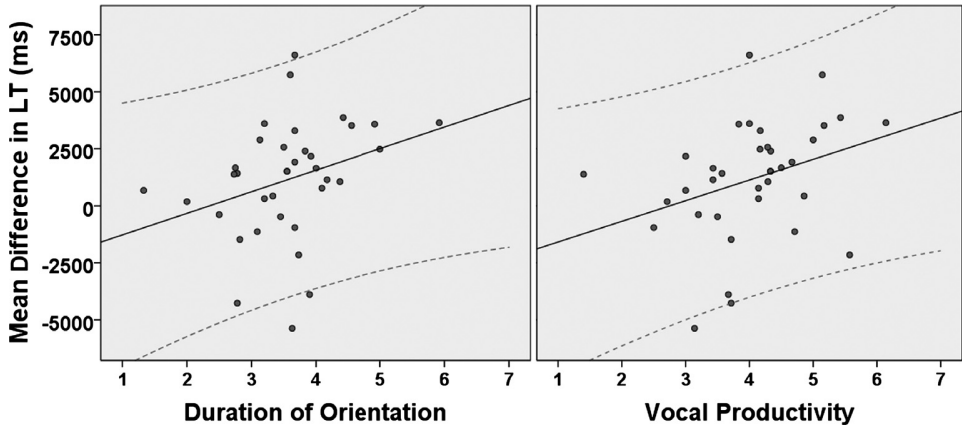

Figure 5. Mean difference in looking time (LT) to matching versus mismatching trials in the audiovisual matching task plotted against infants' duration of orientation and vocal productivity scores. The continuous line depicts the linear regression line (duration of orientation: R^2 = .099; vocal productivity: R^2 = .109); the dotted lines indicate the 95% confidence interval.

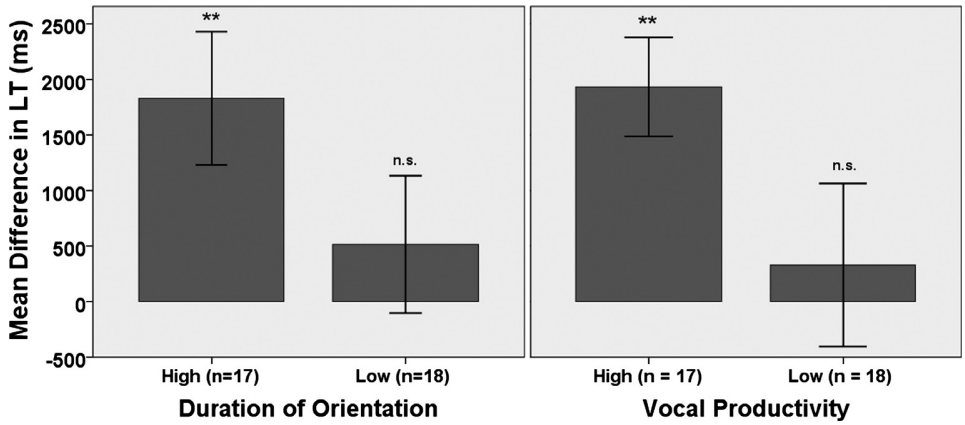

Figure 6. Mean difference in looking time (LT) to matching versus mismatching trials in the audiovisual matching task dependent on infants' duration of orientation and vocal productivity scores (groups formed by median split). Error bars indicate ± 1 *SE*.
**$p < .01$.

$t(17) = 4.339$, $p < .001$, $d = 1.023$, 16 of 18 infants looking longer at matching than mismatching trials, exact binomial $p = .001$. Those infants with low vocal productivity scores showed no preference in the audiovisual matching task, $t(16) = 0.449$, $p = .660$, $d = 0.109$, 10 of 17 infants looking longer at matching than mismatching trials, exact binomial $p = .629$. A chi-square test on the association between infants' preference for matching trials and their vocal productivity was marginally significant, $\chi^2(1) = 4.137$, exact $p = .060$, odds ratio = 5.6. Thus, infants who are well able to focus their attention and often vocalize tend to be more sensitive to the congruency between auditory and visually presented vowels. Figure 6 plots infants' preference for matching over mismatching trials depending on their duration of orientation and vocal productivity scores.

Independent sample *t* tests ensured that any difference between groups was not caused by differences in infants' overall attention to the audiovisual matching task (*p*s > .10). Separate one-sample *t* tests on the difference scores from the discrimination task confirmed that all groups acoustically discriminated the vowels (all *p*s < .002), excluding any difficulties in the acoustic discrimination of the vowels as a source for difficulties to detect audiovisual mismatches. Note that none of the assessed infant-specific variables were related to infants' preference for alternating trials in the auditory discrimination task (*p*s > .11), indicating that infants' ability to focus attention and their vocal productivity specifically influenced audiovisual but not auditory speech perception. Furthermore, infants' vocal productivity did not correlate with mothers' use of vocal play or imitation (*p*s > .28), suggesting that infants' higher

vocal productivity does not necessarily lead to more maternal vocalization and imitation, or vice versa.

Relation to Later Language Development

To explore whether any of the above-described variables have a lasting effect on language development, we investigated their influence on infants' vocabulary size 6 months after testing, that is, around their first birthday. Pearson correlations show that infants' preference for matching over mismatching trials in the audiovisual matching task, $r(33) = .360$, $p_{(\text{one-tailed})} = .020$, $R^2 = .130$, and their vocal productivity, $r(32) = .513$, $p_{(\text{one-tailed})} = .001$, $R^2 = .263$, as well as the amount of vocal play that mothers used during the free-play sessions, $r(32) = .338$, $p_{(\text{one-tailed})} = .029$, $R^2 = .114$, were positively related to later vocabulary size. No further variable—including their preference for alternating trials in the auditory discrimination task—was related to infants' later vocabulary scores, $ps_{(\text{one-tailed})} > .09$. Given that advanced development in auditory speech discrimination in the 1st year of life (Tsao et al., 2004), babbling characteristics (McCune & Vihman, 2001), and the quality of mothers' language input (Cartmill et al., 2013) positively relate to vocabulary development, we expected that infants' preference for matching trials as well as their vocal productivity and mothers' vocal play might also positively relate to infants' later vocabulary size. We therefore report one-tailed p values. Note that using the more conventional two-tailed p values, the positive correlation between infants' preference for matching trials ($p = .04$) as well as infants' vocal productivity ($p = .002$) and their later vocabulary size would still be significant, while the correlation between mothers' use of vocal play and infants' later vocabulary size would only be marginally significant ($p = .058$). Figure 7 plots infants' vocabulary size as a function of the mean difference in LTs to matching and mismatching trials in the audiovisual matching task. Figure 8 plots infants' vocabulary size as a function of infants' vocal productivity scores and the amount of vocal play that mothers used during the free-play sessions.

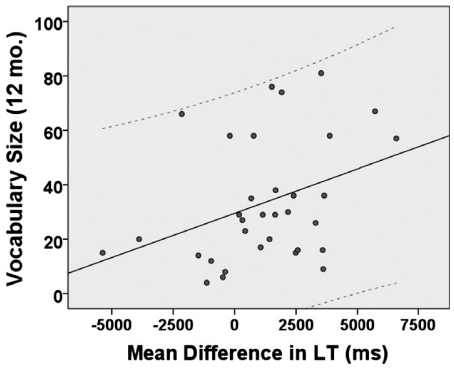

Figure 7. Mean vocabulary size at 12 months of age (in words) plotted against infants' mean difference in looking time (LT) to matching versus mismatching trials in the audiovisual matching task. The continuous line depicts the linear regression line ($R^2 = .130$); the dotted lines indicate the 95% confidence interval.

Discussion

The current study investigated potential social and infant-specific factors that might contribute to early audiovisual phoneme learning. Testing 5.5- to

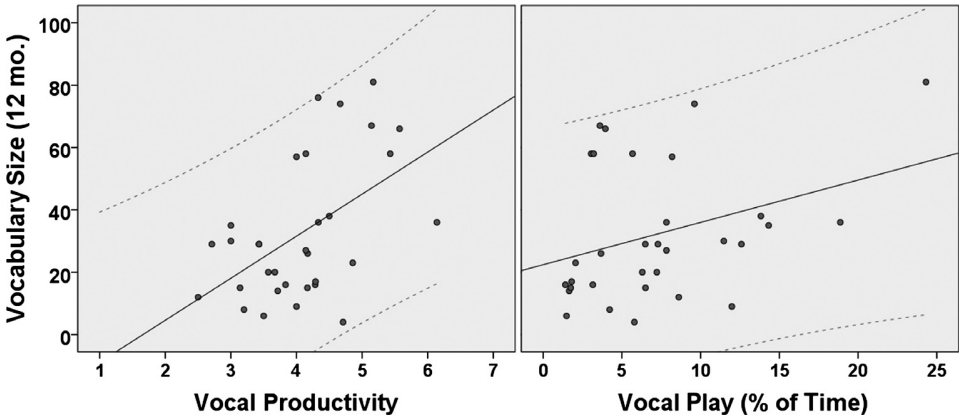

Figure 8. Mean vocabulary size at 12 months of age (in words) plotted against infants' vocal productivity scores and mothers' vocal play behavior. The continuous line depicts the linear regression line (vocal productivity: $R^2 = .263$; vocal play: $R^2 = .114$); the dotted lines indicate the 95% confidence interval.

6-month-olds, we found that mothers' imitation and vocal play behavior as well as infants' vocal productivity and their ability to focus attention were positively related to their sensitivity for mismatches between auditory and visually presented native vowels. Only those infants who often vocalized during daily routines and were well able to focus their attention, as well as those infants who had mothers who imitated them and often used vocal play showed a robust preference for matching audiovisual speech. This suggests (a) that audiovisual perception of native vowels is still developing within the 1st year of life and (b) that infant-specific factors as well as mothers' behavior influences this process. Furthermore, we found that infants' preference for congruent audiovisual speech and their vocal productivity as well as mothers' use of vocal play at 6 months was positively related to infants' vocabulary size at 12 months, suggesting a lasting effect on language development. In the following, we will discuss each of these findings in more detail.

Individual Variation in Infants' Audiovisual Speech Perception

Our results show considerable individual variation, in that some infants showed a robust sensitivity to the congruency of audiovisual speech while other infants seem to be insensitive to the congruency between auditory and visually presented native vowels. This finding challenges the assumption that the ability to match auditory and visual speech cues is innate or develops with little experience early in infancy (e.g., Kuhl & Meltzoff, 1982). Earlier research has shown an astonishing ability to match audiovisual speech cues in infants as young as 2 months of age (Bristow et al., 2003; Patterson & Werker, 2003). Yet, previous studies mainly used the corner vowels /a/, /i/, and /u/, which are farthest apart in the vowel space. Given that some infants had difficulties to detect mismatches in the audiovisual cues associated with the more similar vowels /a/, /e/, and /o/ used in the present study, our results suggest that infants' sensitivity to audiovisual congruency might initially be limited to maximally disparate vowels and further develop in the 1st year of life.

Indeed, more recent work has shown that audiovisual speech perception gets attuned to the native language (Pons et al., 2009) and that integration of auditory and visual speech cues might not be mandatory in infants (Desjardins & Werker, 2004). Our results extend these findings by showing individual variation in the sensitivity to the congruency between auditory and visual speech cues in infants, and by suggesting that the perception of audiovisual *native* sounds still develops in the 1st year of life. This parallels results from auditory speech discrimination finding that infants' ability to discriminate native sound contrasts still improves in the course of development (Kuhl et al., 2006). The finding that mothers' speech style as well as their interactive behavior modulate infants' sensitivity to audiovisual speech contrasts further concurs with the results from auditory discrimination studies (Elsabbagh et al., 2013; Liu et al., 2003).

Interestingly, infants did not show similar individual variation in their ability to discriminate /a/, /e/, and /o/ in auditory speech perception. Discrimination of these sound contrasts was robust across infants and not influenced by any of the assessed mother- or infant-specific variables. This might indicate that the relation between infants' sensitivity to the congruency between auditory and visual speech cues on the one hand and mothers' as well as infants' behavior on the other hand is specific to audiovisual speech perception and does not generalize to auditory speech perception. It is also possible that robust auditory discrimination of the specific contrasts /a/–/e/ and /a/–/o/ used in the current study emerges relatively early in infancy (see Kuhl, 1983, for evidence that English 6-month-old infants correctly categorize /a/ and /ɔ/ across talkers of different age and gender) and can therefore hardly be improved (and influenced by social factors) in the further course of the 1st year of life. In future work, it would thus be useful to assess the influence of social factors by (a) looking at younger infants at an age when auditory discrimination still develops (Polka & Werker, 1994) or (b) by using more difficult native sound contrasts with infants of the same age as used in the current study, such as the fricative contrast /f/–/ʃ/ (e.g., Eilers, Wilson, & Moore, 1977).

The Influence of Mothers' Speech Style: Vocal Play as Exaggerated Input and Motivational Factor

Concerning the influence of mothers' speech style on infants' audiovisual speech perception, we found that the use of vocal play was positively related to infants' preference for congruent audiovisual speech. Why might this rather specific infant-directed style foster the learning of audiovisual speech categories? Mothers' vocal play assessed in the current study very much resembled babbling. Namely, it was more similar to the infants' own

productions than maternal speech and entailed exaggerated acoustic and visual gestures that are typically associated with IDS and actions (Brand, Baldwin, & Ashburn, 2002; Chong, Werker, Russell, & Carroll, 2003; Green et al., 2010; Newport, 1977). Vocal play might thus capture infants' attention and arouse them, and its exaggerated cues might provide an especially rich and salient input for learning to associate acoustic and visual sound patterns. This would comply with previous findings that IDS facilitates language acquisition (e.g., Liu et al., 2003; Ma, Michnick Golinkoff, Houston, & Hirsh-Pasek, 2011; Thiessen, Hill, & Saffran, 2005) and that lacking exposure to IDS might lead to delayed language development (D'Odorico & Jacob, 2006).

Note, however, that another aspect of IDS, that is, the use of high pitch contours and extended pitch ranges, did not influence infants' audiovisual speech perception. This might indicate that pitch characteristics have less impact on audiovisual speech perception. Since audiovisual speech provides additional cues to a sound's identity compared to acoustic speech alone (see also Green et al., 2010), infants might benefit less from exaggerated acoustic cues. However, our pitch measures might also not have been sensitive enough to capture a potential influence of mothers' speech characteristics on infants' audiovisual speech perception. Previous studies measured the infant directedness of mothers' speech by examining the expansion of mothers' vowel space rather than her pitch characteristics (Liu et al., 2003) or they compared infants' performance in language tasks when being presented with infant- versus adult-directed speech (e.g., Thiessen et al., 2005). Absolute measures of mothers' pitch and pitch range when interacting with their infant might be less informative in investigating the influence of IDS.

The Influence of Mothers' Interactive Behavior: Imitation as a Cue to Associate Acoustic, Visual, and Motoric Sound Patterns

Concerning the influence of mothers' interactive style on infants' audiovisual speech perception, we found that mothers' imitation behavior was positively related to infants' preference for audiovisual matches. This agrees with earlier studies finding that contingent social interactions foster language learning (e.g., Elsabbagh et al., 2013; Goldstein et al., 2003; Kuhl et al., 2003). Imitations are indeed very contingent responses that timely mirror the infants' behavior (Ray & Heyes, 2011). This might influence the learning of (audiovisual) speech categories in two ways: It might engage mother and infant, and it might help to establish sensorimotor links.

Previous research has shown that mothers enjoy engaging in imitation games (Lewis, 1979; Pawbly, 1977; Uzgiris, Benson, Kruper, & Vasek, 1989). They tend to perceive their infants' actions as imitating their own behavior even when it only matches at chance level, and reward perceived imitations with smiles and encouragement. Thus, being imitated might render the imitated actions more salient through the contingent mirroring as well as through the positive emotional feedback. In the case of imitated babbles, this might foster the association of articulatory, visual, and auditory characteristics of speech sounds. When an infant produces a sound, the sensorimotor experience allows the coupling of the articulatory gesture and its corresponding acoustic output; when the infant is imitated, she will receive a mirrored response that allows coupling the auditory sound with a visual gesture. This might help to create sensorimotor links and to establish multisensory phoneme categories (see also Westermann & Miranda, 2004). Imitation might thus be a powerful mechanism to boost early language learning (Ray & Heyes, 2011).

Note that the effect of mothers' interactive behavior on infants' audiovisual speech perception was specific to imitations. This might indicate a special impact of imitations. It might, however, also indicate that our measure of general responsiveness and attention to the infants' interest was not fine-grained enough to capture their influence. For instance, Elsabbagh et al. (2013) found an impact of individual differences in mothers' behavior on infants' speech perception. Yet, their coding of the interaction was more fine-grained, using a standardized coding manual, and they assessed non-native (rather than native) speech perception. Their measures might thus have been more sensitive to individual differences in mothers' and infants' behavior than the measures in the current study.

The Influence of Infant-Specific Factors: Vocal Productivity and Developing Motor Representations

Next to mothers' use of vocal play and imitation, infants' vocal productivity was positively related to their sensitivity to audiovisual mismatches, suggesting a potential role of articulatory experience on audiovisual speech perception. This is in line with earlier findings reporting a relation between speech

production and perception in language development. Desjardins et al. (1997) found that toddlers' speech production abilities influence their integration of auditory and visual speech cues in perception, and Yeung and Werker (2013) report an influence of concurrently performed mouth movements on audiovisual speech perception in preverbal infants. Recent studies further suggest that infants' preference for specific sound patterns in auditory speech perception is impacted by their productive abilities, such as the production of the corresponding sounds during babbling (DePaolis, Vihman, & Keren-Portnoy, 2011; DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014). Taken together, these studies support the assumption that there is a link between motoric and perceptual knowledge and fit the idea that speech perception and production develop in concert (see also Kuhl, 2000; Vihman, 1996). If infants' own motoric knowledge indeed influences speech perception, then higher vocal productivity might lead to a more robust perception–production link and help to create more mature audiovisual speech categories (for a study relating productive abilities to audiovisual speech perception in infants, see Altvater-Mackensen & Grossmann, 2013).

Another infant-specific factor that influenced sensitivity to audiovisual congruencies was infants' ability to focus their attention, as measured by their general attention span for objects and events. Infants with higher scores were better able to detect audiovisual mismatches, suggesting that the ability to focus attention might serve as a lever for learning. This fits earlier findings on auditory speech discrimination by Conboy, Brooks, Meltzoff, and Kuhl (2008) that infants who are more focused on a language tutor show better language learning. However, the positive effects of focusing attention —or being attentive to learning cues in the environment more generally—are most likely not specific to language learning. Similarly, the other factors that we found to positively influence audiovisual speech perception, that is, mothers' use of vocal play and imitation and infants' vocal productivity, might eventually have the most impact on development because they engage mother–infant interactions and lead to higher arousal and attention to the (speech) input provided (see also Kuhl, 2007).

Impact on Later Language Development

All findings discussed so far provide insights into the factors influencing speech perception at the verge of native language attunement. To further explore which factors potentially have a lasting effect on language development, we related them to infants' vocabulary size at the end of the 1st year of life. Indeed, infants' preference for matching audiovisual speech, their vocal productivity, and mothers' use of vocal play was related to later language growth. The finding that infants who show a robust preference for matching audiovisual speech at 6 months of age have larger vocabularies at 12 months of age resembles the finding of Tsao et al. (2004) that infants who show earlier native language attunement in auditory speech perception have larger vocabularies at toddler age. It suggests that early mastering of the native phoneme system —not only in auditory but also in audiovisual speech perception—fosters further language development.

Similarly, the finding that infants' vocal productivity is related to later vocabulary development concurs with earlier studies showing that infants' development in babbling predicts vocabulary size at 12 months of age (Majorano et al., 2014) and referential word use at 16 months of age (McCune & Vihman, 2001). This might indicate an effect of production experience on word learning. However, it might also be a mediated effect, in that infants who babble more will receive more positive feedback and input from their mothers (see above), which will in turn help phoneme learning. Earlier mastering of phoneme categories might then promote the acquisition of words. In addition, infants who vocalize more might not only be more "talkative" in infancy but also in toddlerhood and might therefore receive more maternal feedback across development.

The effect of maternal vocal play on later vocabulary size is more difficult to explain (note that this correlation was only marginally significant when using two-tailed tests). It might simply be a side effect of the fact that mothers' vocal play enhances the sensitivity to audiovisual congruencies and that audiovisual sensitivity in turn fosters later language learning. However, it might also reflect an effect of mothers' ability to tailor their (speech) input to their infants' capabilities and needs. Mothers who use vocal play to engage with their infant might intuitively provide the most engaging and salient input to their children not only at 6 months of age but also at later stages of language development. This is highly speculative but it points to a need for future studies to investigate how mothers interact with their infants at different stages of development and to see how this influences their infants' language development.

Conclusion

The current study examined audiovisual speech perception in infants who are at the verge of perceptual narrowing. We found that infants differed in their ability to detect congruencies between native auditory and visually presented vowels depending on their vocal productivity and their mothers' interactive (speech) behavior. The findings suggest that learning of audiovisual speech categories is modulated by characteristics of mothers' and infants' mutual behavior, supporting the view that language learning is not only a question of the statistical and acoustic properties of the input provided to infants but also crucially depends on the social interaction in which the learning occurs.

References

Altvater-Mackensen, N., & Grossmann, T. (2013). *The interplay between perception, production and audiovisual integration in phoneme learning*. Poster presented at the International Child Phonology Conference, Nijmegen, Netherlands.

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, *46*, 155–182. doi:10.1177/00238309030460020601

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*, 847–862. doi:10.1016/j.infbeh.2013.09.001

Best, C., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *21*, 295. doi:10.1016/S0163-6383(98)91508-9

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer [Computer software]. Retrieved from http://www.praat.org

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for "motionese" in mothers' infant-directed action. *Developmental Science*, *5*, 72–83. doi:10.1111/1467-7687.00211

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*, 905–921. doi:10.1162/jocn.2009.21076

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*, 204–220. doi:10.1002/dev.20032

Cartmill, E. A., Armstrong, B. F. III, Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Science*, *110*, 11278–11283. doi:10.1073/pnas.1309518110

Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, *12*, 211–232. doi:10.1002/icd.286

Conboy, B. T., Brooks, R., Meltzoff, A. N., & Kuhl, P. K. (2008). *Joint engagement with language tutors predicts brain and behavioural responses to second-language phonetic stimuli*. Paper presented at the annual International Conference on Infant Studies, Vancouver, Canada.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595. doi:10.1111/j.1467-8624.1990.tb02885.x

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, *34*, 590–601. doi:10.1016/j.infbeh.2011.06.005

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*, 642–649. doi:10.1016/j.infbeh.2013.06.007

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*, 85–110. doi:10.1006/jecp.1997.2379

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*, 187–203. doi:10.1002/dev.20033

D'Odorico, L., & Jacob, V. (2006). Prosodic and lexical aspects of maternal linguistic input to late-talking toddlers. *International Journal of Language and Communication Disorders*, *41*, 293–311. doi:10.1080/13682820500342976

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, *20*, 766–780. doi:10.1044/jshr.2004.766

Elsabbagh, M., Hohenberger, A., Campos, R., Van Herwegen, J., Serres, J., de Schonen, S., . . . Karmiloff-Smith, A. (2013). Narrowing perceptual sensitivity to the native language in infancy: Exogenous influences on developmental timing. *Behavioural Sciences*, *3*, 120–132. doi:10.3390/bs3010120

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*, 104–113. doi:10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dun, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*, 477–501. doi:10.1017/S0305000900010679

Gartstein, M., & Rothbart, M. (2003). Studying infant temperament via the revised Infant Behaviour Questionnaire. *Infant Behavior and Development*, *26*, 64–86. doi:10.1016/S0163-6383(02)00169-8

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Science of the United States of America*, *100*, 8030–8035. doi:10.1073/pnas.1332441100

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–523. doi:10.1111/j.1467-9280.2008.02117.x

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip-movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, *53*, 1529–1542. doi:10.1044/1092-4388(2010/09-0005)

Grimm, H., & Doil, H. (2000). *Elternfragebögen für die Früherkennung von Risikokindern* [Parent questionnaire for the early detection of at-risk children]. Göttingen, Germany: Hogrefe.

Henning, A., Striano, T., & Lieven, E. V. M. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, *28*, 519–536. doi:10.1016/j.infbeh.2005.06.001

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263–285. doi:10.1016/S0163-6383(83)80036-8

Kuhl, P. K. (2000). A new view on language acquisition. *Proceedings of the National Academy of Science of the United States of America*, *97*, 11850–11857. doi:10.1073/pnas.97.22.11850

Kuhl, P. K. (2007). Is speech learning "gated" by the social brain? *Developmental Science*, *10*, 110–120. doi:10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioural and electrophysiological measures. *Developmental Science*, *8*, F1–F12. doi:10.1111/j.1467-7687.2004.00384.x

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141. doi:10.1126/science.7146899

Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy: The Minnesota Symposia on Child Psychology* (Vol. 20, pp. 235–266). Hillsdale, NJ: Erlbaum.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*, F13–F21. doi:10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Science of the United States of America*, *100*, 9096–9101. doi:10.1073/pnas.1532872100

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Science of the United States of America*, *105*, 11442–11445. doi:10.1073/pnas.0804275105

Lewis, M. (1979). *Issues in the study of imitation*. Paper presented at the meeting of the Society for Research in Child Development, San Francisco, CA.

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

Ma, W., Michnick Golinkoff, R., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*, 185–201. doi:10.1080/15475441.2011.579839

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219*, 1347–1349. doi:10.1126/science.6828865

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*, 179–204. doi:10.1080/15475441.2013.829740

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111. doi:10.1016/S0010-0277(01)00157-3

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language and Hearing Research*, *44*, 670–684. doi:10.1044/1092-4388(2001/054)

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. doi:10.1038/264746a0

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant-directed speech and the development of speech perception: Enhancing development or unintended consequence? *Cognition*, *129*, 362–378. doi:10.1016/j.cognition.2013.07.015

Mugitani, R., Kobayashi, T., & Hiraki, K. (2008). Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behavior and Development*, *31*, 307–310. doi:10.1016/j.infbeh.2007.12.002

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*, 407–420. doi:10.1111/j.1467-7687.2009.00898.x

Nath, A. R., Fava, E. E., & Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *Journal of Neuroscience*, *31*, 13963–13971. doi:10.1523/JNEUROSCI.2605-11.2011

Newport, E. L. (1977). Motherese: The speech of mothers to young children. In N. J. Castellan, D. B. Pisoni, & G. Pott (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, NJ: Erlbaum.

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voices is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*, 237–247. doi:10.1016/S0163-6383(99)00003-X

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voices. *Developmental Science*, *6*, 191–196. doi:10.1111/1467-7687.00271

Pawbly, S. J. (1977). Imitative interaction. In H. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 203–224). New York, NY: Academic Press.

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, *100*, 577–592. doi:10.1121/1.415884

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421–435. doi:10.1037/0096-1523.20.2.421

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Galles, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Science of the United States of America*, *106*, 10598–10602. doi:10.1073/pnas.0904134106

Ray, E., & Heyes, C. (2011). Imitation in infancy: The wealth of the stimulus. *Developmental Science*, *14*, 92–105. doi:10.1111/j.1467-7687.2010.00961.x

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 58–108). Hoboken, NJ: Wiley.

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-olds. *Developmental Psychology*, *47*, 19–25. doi:10.1037/a0020740

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*, 850–855. doi:10.1016/j.cognition.2008.05.009

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates speech segmentation. *Infancy*, *7*, 53–71. doi:10.1207/s15327078in0701_5

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

Uzgiris, I. C., Benson, J. B., Kruper, J. C., & Vasek, M. E. (1989). Contextual influences on imitative interactions between mothers and infants. In J. Lockman & N. Hazen (Eds.), *Action in social context: Perspectives on early development* (pp. 103–127). New York, NY: Plenum Press.

Vihman, M. M. (1996). *Phonological development. The origins of language in the child*. Oxford, UK: Blackwell.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, *316*, 1159. doi:10.1126/science.1137686

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535. doi:10.1146/annurev.psych.50.1.509

Westermann, G., & Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, *89*, 393–400. doi:10.1016/S0093-934X(03)00345-6

Yeung, H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*, 603–612. doi:10.1177/0956797612458802