Psychology of Language and Communication 2015, Vol. 19, No. 2

DE GRUYTER OPEN

DOI: 10.1515/plc-2015-0006

PRZEMYSŁAW TOMALSKI
University of Warsaw

DEVELOPMENTAL TRAJECTORY OF AUDIOVISUAL SPEECH INTEGRATION IN EARLY INFANCY. A REVIEW OF STUDIES USING THE MCGURK PARADIGM

Apart from their remarkable phonological skills young infants prior to their first birthday show ability to match the mouth articulation they see with the speech sounds they hear. They are able to detect the audiovisual conflict of speech and to selectively attend to articulating mouth depending on audiovisual congruency. Early audiovisual speech processing is an important aspect of language development, related not only to phonological knowledge, but also to language production during subsequent years. This article reviews recent experimental work delineating the complex developmental trajectory of audiovisual mismatch detection. The central issue is the role of age-related changes in visual scanning of audiovisual speech and the corresponding changes in neural signatures of audiovisual speech processing in the second half of the first year of life. This phenomenon is discussed in the context of recent theories of perceptual development and existing data on the neural organisation of the infant 'social brain'.

Key words: infant, audiovisual speech perception, McGurk effect, eye-tracking, event-related potentials

Introduction

Human infants successfully deal with multisensory social information from birth (for review see Streri, Hevia, Izard, & Coubart, 2013) and there is some evidence for at least auditory processing of speech even during the prenatal period (Kisilevsky et al., 2009; Partanen et al., 2013). The ability to detect and orient towards faces and towards the human voice is often considered funda-

Address for correspondence: Przemysław Tomalski, Neurocognitive Development Lab, Faculty of Psychology, University of Warsaw, Poland. E-mail: p.tomalski@uw.edu.pl

mental for social and cognitive development (Westermann et al., 2007). Within the research on early language and communicative development, studies of infant phonological knowledge have dominated the landscape for the best part of the last 30 years (Werker & Hensch, 2015). More recently there has been a considerable increase in the number of studies of early audiovisual speech perception, which showed its importance not only for infant phonological repertoire, but also for language comprehension and production throughout subsequent years of life (e.g. Altvater-Mackensen & Grossmann, 2015). Moreover, there is growing evidence that phonological representations may be inherently bi- or polymodal in nature (see Guellai, Streri, & Yeung, 2014), thus highlighting a need for a better understanding of infant processing across modalities. While human speech is a multisensory experience this paper focuses on the two modalities most important for language development: the auditory and the visual modalities, thus by necessity reducing speech to a bimodal phenomenon.

This article reviews some recent work delineating the complex trajectory of development in audiovisual speech integration and audiovisual mismatch detection. The central issue is the role of age-related changes in visual scanning of audiovisual (AV) speech and the corresponding changes in neural signatures of AV speech processing for phonological development in the second half of the first year of life. Some evidence suggests that there are different trajectories of development for distinct groups of speech sounds with the formation of phoneme categories occurring earlier for vowels than for consonants (Patterson & Werker, 2002; Polka & Werker, 1994). However, the vast majority of research on infant AV speech processing has been conducted using consonant stimuli. For this reason this paper discusses developmental trajectories for consonants and it should be noted that not all conclusions can be directly generalized onto vowels.

The review starts with a short survey of infant phonological knowledge prior to the first birthday, followed by a brief account of audiovisual speech processing in the same period. Next, it discusses the adult and infant research employing the McGurk paradigm (McGurk & MacDonald, 1976), and considers these results in the context of recent theoretical accounts of audiovisual processing in infancy: the intersensory redundancy hypothesis (Bahrick & Lickliter, 2014; Lewkowicz, 2003), the intersensory perceptual narrowing hypothesis (Pons, Lewkowicz, Soto-Faraco, & Sebastian-Galles, 2009) in addition to highlighting the social-pragmatic aspects (Tomasello & Carpenter, 2007) of audiovisual speech. The predictive coding theory (Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009) provides a convenient framework to explain the age-related changes in the neural processing of described phenomena. The article concludes with a brief discussion of audiovisual speech processing difficulties in atypical development and suggestions for future research directions.

Infant phonological knowledge before the age of 12 months

Infants typically show dramatic increases in their phonological knowledge throughout the first year of life. By 6 months of age they discriminate many speech sound contrasts: native as well as non-native (Kuhl et al., 2006). While infants initially hear all phonetic contrasts in different languages, this perceptual ability appears to decline before their first birthday (Eimas, 1975), although some data suggest a more selective loss of ability to discriminate non-native contrasts (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Best & McRoberts, 2003). Between 10 and 12 months of age this universal phonetic ability is gradually replaced by speech sound processing attuned to phonemes present in the native language (Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005; Tsao, Liu, & Kuhl, 2006; Werker & Tees, 1984) and this shift occurs several months earlier for vowels than for consonants (Polka & Werker, 1994). Altogether, between 6 and 12 months of age there is a facilitation of processing of native language phonemes at a cost of decreased processing of non-native phonemes (Kuhl et al., 2006; see below).

The quality of auditory language input seems crucial for early phonological learning. Infants early on prefer infant-directed speech ("motherese"; Fernald, 1985), which is produced at higher pitch, with slower articulation and exaggerated intonation contours. Motherese facilitates phonological learning by exaggerating acoustic differences (Kuhl et al., 1997), and there is evidence that greater exposure to infant-directed speech during infant-parent interactions is associated with larger vocabulary size before the age of 12 months (Altvater-Mackensen & Grossmann, 2015), as well as at 24 months (Ramírez-Esparza, García-Sierra, & Kuhl, 2014).

Motherese also helps to engage infant attention to the speaker's face (Cooper & Aslin, 1990), which allows the infant to obtain additional cues that may aid phonological processing. Visual information from the speaker's mouth facilitates speech processing in two ways. Firstly, it complements the auditory stream of speech when parts of it are underspecified, for example in noisy environments (Sumby & Pollack, 1954). Secondly, by providing correlated and possibly redundant information about the speech stream, which may improve the comprehension of difficult-to-understand passages even when hearing conditions are good (see Campbell, 2008). Despite the multisensory nature of human speech, the subject of audiovisual processing in phonological development has been relatively understudied. The following section offers a brief summary of research on audiovisual speech perception in infancy.

Audiovisual speech processing in early infancy

Infants already attend to visual cues during audio-visual (AV) speech perception in the first months of life. Newborns are more likely to imitate facial gestures of an audiovisually congruent rather than visual-only model (Coulon, Hemimou, & Streri, 2013). Very young infants can learn arbitrary face-voice

associations (Brookes et al., 2001), and they attend more to fluent synchronized than desynchronized speech (Dodd, 1979). By 4 months of age they detect AV asynchrony when observing the production of isolated syllables (Lewkowicz, 2010) and in a continuous speech stream (Pons & Lewkowicz, 2014). Infants aged 2 and 4 months prefer watching faces with mouth articulations matching auditory vowels rather than mismatching them (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003). Around this age infants may also match seen and heard consonants (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). Thus, in the first months of life infants are already able to detect corresponding patterns of mouthing and auditory speech in both vowels and consonants.

Phonological development capitalizes on statistical learning to build phonetic categories. Specifically, when a continuum of speech sounds on a phonetic contrast is constructed, e.g. /ba/ - /da/, infants may use the frequency distribution of phonetic variation to learn to discriminate that contrast. Already at 6 months of age infants that are familiarised with a bimodal distribution show enhanced discrimination of non-native phonetic contrasts. Conversely, a unimodal (more frequent sounds in the middle of the continuum, less frequent sounds towards each end) frequency distribution may lead to reduced discrimination of phonetic contrasts that infants normally discriminate at that age (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002; Pegg & Werker, 1997). Importantly, 6-month-olds may also use visual speech cues to enhance their phonetic discrimination of auditory speech in statistical learning tasks. Teinonen, Aslin, Alku and Csibra (2008) exposed infants to speech sounds along a /ba/-/da/ auditory continuum that followed a unimodal frequency distribution, which previously attenuated contrast discrimination (see Maye et al., 2002). To this auditory stimulation infants in the experimental group were presented visually with two types of pairings. Visual articulation /ba/ was paired with sounds from the /ba/ side of the continuum, while visual /da/ was paired with sounds from the /da/ side of the continuum. The control group received the same auditory stimuli but paired with only one visual articulation (for different participants either only visual /ba/ or only visual /da/). Following the exposure only infants, who saw both the /ba/ and the /da/ articulations discriminated the /ba/ - /da/ auditory contrast. Thus, already at 6 months of age infants may use visual speech to enhance their phonetic discrimination when the distributional properties of the auditory input are insufficiently informative.

This work demonstrates that early on infants may benefit from focusing their attention on the articulating face to aid their phonetic learning. However, Desjardins and Werker (2004) showed that at 4.5 months of age the online integration of seen and heard speech is not mandatory, even if possible. In a series of studies they found that when habituated to an incongruent auditory /bi/ - visual /vi/ stimulus, only male infants show evidence of integration, while when habituated to a congruent audiovisual /bi/ only female infants do so. Further, unlike in adults, the audiovisual matching of speech cues in infancy remains unaffected by

the degradation of phonetic information in studies employing sine-wave speech (Baart, Vroomen, Shaw, & Bortfeld, 2013) and may not depend on linguistic experience until after the first birthday (Pons & Lewkowicz, 2014). On the basis of their work Lewkowicz and colleagues (2015) have argued that before the age of 12 months infants process audiovisual speech by relying predominantly on temporal cues rather than on phonetic categories and only around 12-14 months they bein to use phonetic information for audiovisual matching. However, this conclusion is inconsistent with the results of Teinonen et al. (2008), where visual tokens were used by 6-month-olds to learn to discriminate a phonetic contrast in an auditory test. While there is ample evidence to suggest that before their first birthday infants can perform audiovisual matching by detecting temporal synchrony of cues, this does not preclude the use of visual speech for phonetic learning prior to the age of 12 months. What remains unclear is to what extent young infants actually use visual speech cues to construct their phonetic categories and whether this ability is related to visual attention to articulating faces.

Audiovisual speech integration task

One method of measuring capacities for audiovisual speech integration in adults and infants is to present them short videoclips of human faces pronouncing syllables (/ba/ or /ga/), where the visual and auditory speech components of the stimuli do not match (Kushnerenko, Teinonen, Volein, & Csibra, 2008). In these non-matching circumstances two different speech illusions can be perceived: fusions and combinations, known as the McGurk effect (McGurk & MacDonald, 1976). Of particular interest is what happens when a visual /ga/ and auditory /ba/ are presented together (VgaAba) as these are often fused by adults and perceived as the sound /da/ or /θa/. On the other hand, a visual /ba/ dubbed onto auditory /ga/ (VbaAga) leads to an audiovisual mismatch and is often perceived as the combination /bga/ or /baga/.

Early on studies of adult native English speakers demonstrated that the McGurk effect is a robust phenomenon (Massaro, Cohen, & Smeele, 1996; Rosenblum & Saldaña, 1996). A commonly held view was that audiovisual integration is mandatory and automatic, so that the McGurk effect does not require the direction of attentional resources to occur (Soto-Faraco, Navarra, & Alsius, 2004). While some studies indicate that even in the absence of directed attention the mouth movements are difficult to ignore (Buchan & Munhall, 2011, Paré et al., 2003), other research calls the automaticity into question. The presence of visual distractors reduces the incidence of illusory McGurk percepts, which are also vulnerable to auditory distractors (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Schwartz, 2010; Tiippana, Andersen, & Sams, 2004). This suggests that visual fixations on the mouth of incongruent audiovisual stimuli affect the fusion and combination effects. This conclusion is supported by adult eye-tracking data: individuals who

frequently perceive the McGurk illusion spend more time fixating the mouth (Gurler, Doyle, Walker, Magnotti, & Beauchamp, 2015). Given that in adults the integration of hearing and vision is vulnerable to increased attentional load, the successful processing of McGurk stimuli in infancy may require additional attention resources and be related to infants' visual scanning.

Developmental studies of audiovisual speech processing that employed the McGurk effect offer ambiguous results. While there is evidence from behavioural studies that infants as young as 4 months of age can perceive the McGurk fusion illusion similarly to adults (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997), a more careful analysis of this effect using different experimental designs shows that it may occur only for female or male infants depending on the procedure (Desjardins & Werker, 2004). Also, there is ample evidence for a more protracted trajectory of development of audiovisual speech integration throughout childhood. Already in their original study McGurk and MacDonald (1976) reported lower susceptibility to the fusion effect for children aged 3-5 and 7-8 years, a result corroborated by a recent report using the McGurk illusion (Sekiyama & Burnham, 2008). These results are consistent with gradual increases in the influence of visual cues on audiovisual speech perception found in other studies (Baart, Bortfeld, & Vroomen, 2015; Hockley & Polka, 1994; Massaro, Thompson, Barron, & Laren, 1986; Massaro, 1984).

Despite these inconsistencies between different studies, three aspects of the McGurk paradigm make it attractive for infancy research. Firstly, there already exists a body of research using these stimuli in different age groups, providing a necessary developmental context more than for any other audiovisual speech perception paradigm. Secondly, the existing electrophysiological studies of the McGurk effect in infancy (Bristow et al., 2009; Kushnerenko et al., 2008) provide evidence for the processing of AV conflict at the neural level. Thus, behavioural results with the McGurk stimuli can be related to neural responses in different age groups. Finally, the presence of two different incongruent conditions allows to test in a single experiment two separate phenomena: the integration of incongruent but fusible AV tokens (fusion, VgaAba) and the processing of conflict between tokens that cannot be integrated (mismatch, or combination, VbaAga). This means that both conditions can be tested in a short experiment in the same subject, so that individual differences between participants can be quickly assessed with low participant attrition. The following sections focus on research employing both effects to study developmental changes in AV speech processing.

Developmental trajectory of infant attention to articulating faces

Although human newborns show a strong preference for stimuli with eye gaze (Farroni et al., 2005; for review see Johnson, Senju, & Tomalski, 2015), there is limited longitudinal data on the development of scanning patterns of dynamic, articulating faces throughout the first months of life. Jones and Klin (2013) mea-

sured fixation times on the eyes and the mouth of continuously speaking faces. They showed that from the age of 2 months typical infants spend 40% – 60% of time fixating the eyes, but beginning from 6 months of age they fixate more on the mouth relative to the eyes area.

Consistent with this result several other studies revealed an important transition in the visual scanning of articulating faces during the second half of the first year of life. Between the ages of 6 and 10 months, monolingual infants spend progressively more time looking at the mouth and less at the eyes of dynamic articulating faces (Hunnius & Geuze, 2004; Lewkowicz & Hansen-Tift, 2012; Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013; Tomalski et al., 2013), but no such developmental pattern was found for static faces (Wilcox, Stubbs, Wheeler, & Alexander, 2013). This trend then reverses around 12 months of age with a gradual increase in looking to the eyes and a decrease in looking to the mouth, but only for native AV speech (Lewkowicz & Hansen-Tift, 2012), while for non-native speech 12-month-olds continue to exhibit a preference for the mouth (Pons, Bosch, & Lewkowicz, 2015).

Such a pattern of developmental change between 6 and 12 months of age suggests that attention to visual speech cues plays a vital role in the development of speech perception. Several mechanisms have been proposed to explain the role of visual cues in facilitating speech perception development. Firstly, visual cues may enhance auditory speech perception by increasing the saliency of ambiguous or under-specified parts of the speech stream and by providing redundant AV information (Bahrick, Lickliter, & Flom, 2004; Campbell, 2008). Complementary to that view Lewkowicz and colleagues proposed that intersensory redundancy is crucial at the time of perceptual narrowing in phoneme learning (Lewkowicz et al., 2015) before the closure of the sensitive period (Werker & Hensch, 2015). Secondly, in the social-pragmatic approach increased attention to some face parts may help infants to grasp the communicative intent of a speaker (Tomasello & Carpenter, 2007) or to focus attention on what is spoken to the infant (see Tenenbaum, Sobel, Sheinkopf, Malle, & Morgan, 2014). Thirdly, increased attention to articulation at a time when infants engage in canonical babbling may facilitate their own speech production either through imitation, or motor learning (Howard & Messum, 2011) reinforced by a caregiver's feedback (Ramsdell-Hudock, 2014). In summary, attention to the mouth when observing visual speech cues provides typically developing infants with important information that may facilitate their phonological development, while further research is necessary to test both tentative mechanisms.

AV mismatch detection as a test of the intersensory redundancy hypothesis

According to the intersensory redundancy hypothesis infants allocate more attention to stimuli that contain redundant information between modalities

(Bahrick et al., 2004). On that basis Lewkowicz and Hansen-Tift (2012) proposed that between 6 and 12 months of life infants show decreased attention to the eyes and increased attention to the articulating mouth because the mouth provides redundant information about heard speech. However, their study did not provide a critical test of this hypothesis, as it did not compare congruent and incongruent audiovisual speech directly (Lewkowicz & Hansen-Tift, 2012). On the basis of their hypothesis it can be predicted that infants younger than 12 months should prefer audiovisually congruent than incongruent stimuli. To investigate this prediction Tomalski and colleagues (2013) created an eye-tracking task that employed the McGurk stimuli previously used by Kushnerenko et al. (2008) for an ERP study. Infants aged 6-7 and 8-9 months of age were shown short videoclips of articulating faces that were either AV congruent (matching sound and articulation /ba/ and /ga/ – VbaAba, VgaAga) or AV incongruent (fusible VgaAba, and combination/ mismatch VbaAga). Attention to the eyes and the mouth regions of the face was measured with eye-tracking as the total time spent fixating each area out of total time spent fixating the entire face.

They demonstrated that infants as young as 6 months of age differentiate the AV mismatch and AV fusible stimuli and also discriminate them from both congruent AV pairs (Tomalski et al., 2013). Thus they showed that the AV mismatch detection is manifested behaviourally by visual attention allocation at a similar age that it is manifested at the neural level (Kushnerenko et al., 2008). Specifically, they found that infants aged 6-7 months looked significantly less at AV mismatch than AV fusible stimuli and the congruent AV pairs. But infants in the older group, aged 8-9 months looked equally long at AV fusible and AV mismatch (combination) stimuli and longer than during the congruent AV speech. Thus infants in the older group did not show visual discrimination of AV mismatch from AV fusion in terms of looking times, unlike the younger ones.

Tomalski and colleagues (2013) also found that between 6 and 9 months of age there is an age-related increase in attention to the mouth, but only in the AV mismatch condition (VbaAga), when auditory and visual cues are in conflict and cannot be fused to a single percept. No such age-related increase was found for looking at the mouth in either the fusion or the two congruent conditions. Likewise, no age-related change was detected for looking at the eyes in any condition. These results are inconsistent with the intersensory redundancy hypothesis (Bahrick, Flom, & Lickliter, 2002; Lewkowicz, 2000), which would predict increased attention to the mouth only for congruent (redundant) speech cues. Some methodological differences between the Tomalski et al. (2013) study and the Lewkowicz and Hansen-Tift (2012) study may have contributed to these inconsistencies. The first study directly contrasted congruent and incongruent AV tokens, while the second compared infants' native and non-native continuous AV speech that was always congruent. The limited age range in the first study (6-9 months) may have contributed to the discrepancy in findings for the congruent speech. However, the

first study clearly demonstrates that the eyes-to-mouth-to-eyes developmental shift in visual scanning of faces found between 6 and 12 months of age cannot be explained by the intersensory redundancy hypothesis. Thus it is likely that other, unknown mechanisms drive this developmental process.

Attention to AV mismatch is associated with the neural mismatch response

The detection of mismatch between auditory speech and visual articulation has been documented in several electrophysiological studies, with a mismatch-related ERP response found over frontal-lateral electrodes (Bristow et al., 2009; Kushnerenko et al., 2008; see also Mottonen, Krause, Tiippana, & Sams, 2002; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Crucially, Kushnerenko and colleagues (2008) reported that five-month-olds already show the audiovisual, event-related mismatch response (AVMMR) to a conflicting combination of cues (VbaAga) and that this response is absent in trials where both cues can be fused into a single percept (VgaAba). Importantly, the AVMMR is distinct from potentials evoked by either the auditory or visual components of each stimulus, thus it is most likely a marker of bimodal conflict processing.

These results added to the existing literature by showing that there are active neural mechanisms for mismatch detection in the infant brain not only for subsequent stimuli in the auditory modality (see Kushnerenko, Van den Bergh, & Winkler, 2013), but also for bimodal audiovisual stimuli. Originally, Kushnerenko et al. (2008) suggested that this early capacity for detecting mismatch between auditory and visual speech information is a signature of an important neural mechanism that may assist infants' learning about native speech sounds. However, subsequent results have called this idea into question.

A subsequent study conducted by Kushnerenko, Tomalski, Ballieux, Ribeiro and colleagues (2013) using combined eye-tracking and ERP methods revealed a more complex developmental picture. They tested whether individual differences in attention to the eyes and the mouth of speaking faces is coupled with quantitative (amplitude of event-related potentials) differences in the neural processing of AV mismatch and AV fusion in infants aged 6 to 9 months. They found a very strong negative correlation between the amplitude of the audiovisual mismatch response specifically to VbaAga stimuli (AVMMR) and the proportion of time spent fixating the mouth of articulating faces. While individual infants showed a mouth preference in all conditions, the relationship between looking times and neural responses was strongest for the combination/mismatch condition, with infant looking explaining over 44% of variance in the AVMMR amplitude. Thus infants with strong preference for the mouth compared to the eyes showed less pronounced AVMMR and conversely infants without preference for the mouth showed more pronounced AVMMR. This association could not be explained by infant chronological age.

Given that previous eye-tracking data showed the presence of age-related increase in visual attention to the articulating mouth between the ages of 6 and 9 months (Lewkowicz & Hansen-Tift, 2012; Tomalski et al., 2013), these results can be interpreted as a confirmation that increased interest in visual cues is accompanied by changes in the processing of conflict between audiovisual speech cues at the neural level. Infants that showed the more mature pattern of looking, that is fixating the mouth more than the eyes, had reduced AVMMR. Infants showing the less mature pattern of visual scanning (no clear preference for the mouth over the eyes) had more pronounced AVMMR, characteristic of 2-month-olds (Bristow et al., 2009) and 4- to 5-month-olds (Kushnerenko et al., 2008). This interpretation is consistent with the adult pattern of responses to AV mismatch, where no ERP component equivalent to AVMMR was found (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013), despite the fact that adults have no trouble with detecting AV mismatch (Massaro et al., 1996).

In summary, it is likely that the maturational disappearance of the AVMMR does not indicate that infants lose the ability to detect the AV mismatch, but that the neural networks that support this process undergo a reorganisation. Such a process is described by the Interactive Specialisation (IS) theory (Johnson, 2001; 2011), where early on neural networks that support a process are wider and less defined, therefore may engage more neural resources with larger ERP components. With development the neural network that supports this process becomes more refined and limited in terms of participating brain areas, which may manifest in less pronounced ERPs.

Audiovisual speech integration and later language development

Taken together, the existing results suggest that the audiovisual speech integration task may offer a reliable set of behavioural measures of early individual differences in: 1) visual scanning of articulating faces and 2) the detection of audiovisual mismatch and the integration of AV speech cues. A subsequent longitudinal study demonstrates that this task may also be used as a predictor of language development beyond the first year of life. The cohort assessed at 6-9 months by Tomalski et al. (2013) and Kushnerenko, Tomalski, Ballieux, Ribeiro et al. (2013), was followed up at 14-18 months of age with a standardised measure of vocabulary, the Oxford Communicative Development Inventory (Hamilton, Plunkett, & Schafer, 2000) and a clinical language assessment battery, the Preschool Language Scales, 4th edition (Zimmerman, Steiner, & Evatt Pond, 2002). The follow-up study revealed that both the neural responses and looking times to the mouth and the eyes of faces with incongruent audiovisual information at 6-9 months of age are good predictors of receptive language at 14-18 months (Kushnerenko, Tomalski, Ballieux, Potton, et al., 2013). Longer looking to the eyes and shorter looking to the mouth of faces with incongruent AV speech cues (especially the fusion, VgaAba) predicted higher receptive language in PLS

(explaining ~12% of variance) and productive vocabulary in Oxford CDI (explaining ~18% of variance).

A surprising result was that the amplitude of the AVMMR component was not predictive of language outcomes, while the neural response to the fusion (VgaAba) stimulus – was. The amplitude of the frontal P2 component (140-240 ms from the sound onset; see Kushnerenko et al., 2007) was highly negatively associated with receptive language and explained nearly 38% of variance in PLS Auditory Comprehension scores more than 6 months later. That is, infants who showed less pronounced neural response to fusion at 6-9 months scored higher on a clinical measure of receptive language at 14-18 months of age.

To interpret these results let us comment first on the AVMMR component. This component is not only nearly absent from neural responses of infants that show more mature visual scanning of faces, but it also appears inconsequential for later language development. However, if infant AV mismatch detection is based on temporal synchrony detection rather than actual matching of auditory and visual representations of speech sounds (Baart et al., 2013), then the AVMMR would not index phonological processing, but more basic intermodal matching skills that are less relevant to language development. This view is further supported by studies of older children, indicating protracted development of AV speech integration (Ross et al., 2011). Bart, Bortfeld and Vroomen (2015) extended their work with infants (see above) to the study of children aged 4 to 11 years to reveal that only by around 6.5 years of age do children begin to benefit from previously acquired phonetic knowledge in an AV matching task. That is, children younger than 6.5 years performed the AV matching task using predominantly temporal cues and not their phonetic knowledge. These authors suggest that AV speech processing may follow a U-shaped trajectory of development with early sensitivity followed by a refractory period and later improvements.

The results obtained for the AVMMR and the P2 components are consistent with theories of predictive coding (Friston, 2010), which propose that the human brain generates top-down predictions of sensory events and compares them with bottom-up processes evoked by sensory stimulation. Various kinds of mismatch responses found in event-related potentials would indicate a lack of fit between prediction-driven (top-down) and stimulus-driven (bottom-up) neural representations (see Garrido et al., 2009). This theoretical account can be applied to AV speech processing. Top-down predictions are generated on the basis of prior experience with articulatory patterns and auditory speech and compared with incoming bimodal sensory data. For example, at the beginning of articulation for the /ba/ sound the lips are closed, then moved forward with a sudden opening of the mouth. This pattern of movement begins before the sound onset and it fits only with three sounds in English: /pa/, /ba/ and /ma/. So with prior knowledge of AV speech a precise neural prediction can be generated even before the auditory part can be heard. A mismatch between the prediction and the perceived

sound would lead to a prediction error, which in electrophysiological studies is observed as a neural mismatch response. In human development as infants gather more experience to generate predictions of events, the refinement of neural predictions should lead to the reduction of prediction error, which is indexed by the AVMMR for VbaAga stimuli and the P2 for the fusion – VgaAba stimuli. Thus the negative longitudinal association between the size (amplitude) of the P2 response to fusion and receptive language scores may indicate that infants with more efficient and reliable predictive coding are better at processing AV speech and perhaps better equipped to learn new words.

Altogether the results obtained by Kushnerenko, Tomalski and colleagues showing the association of the reduction in neural response to AV mismatch with preference for fixating the mouth in AV speech reflect an important transition in early development (Kushnerenko, Tomalski, Ballieux, Potton, et al., 2013; Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). They illustrate how developing multimodal and speech processing abilities at the neural level become coupled with infants' active use of strategies for allocating visual attention. Thus they mark the beginning of a period where voluntary attentional control allows preverbal infants to actively select important or desired bits of information from surrounding stimuli.

Social-pragmatic cues in articulating faces

We will now turn to the social-pragmatic interpretation of the audiovisual speech integration task. A long tradition of research has documented the importance of social-communicative cues, whether ostensive (signalling the communicative intent, e.g. establishing eye contact) or referential (indicating an object or an action that is talked about), for the development of all aspects of language (see Csibra & Gergely, 2009). There is electrophysiological evidence of very early influences of a speaker's direct and referential gaze on spoken word processing (Parise, Handl, Palumbo, & Friederici, 2011), while live interactions facilitate infant learning and discrimination of novel, non-native phonetic contrasts (Kuhl et al., 2006). Perhaps one of the accounts most relevant to phonological development is Kuhl's (2014) proposal that language acquisition is „gated by the social brain". Specifically, the complex problem of acquiring language and learning to speak is virtually impossible to solve without the social medium of language learning, for it highlights important perceptual features, provides necessary structuring of learning situations and allows enough practice with immediate feedback. One good illustration of these phenomena is the aforementioned infant-directed speech. Research on referential nature of human eye-gaze, prosody or pointing provides evidence for the role of these social cues in language development: phonetic knowledge (Kuhl, Tsao, & Liu, 2003), word learning (Gliga & Csibra, 2009; Houston-Price, Plunkett, & Duffy, 2006), syntax (Gervain & Werker, 2013).

Does the social-pragmatic account of language development help to explain the age-related changes in visual scanning of articulating faces? Taking the Tomasello and Carpenter (2007) proposal infants' greater attention to the eyes from the age of 12 months could be simply explained by greater focusing on sources of social cues that help to establish shared intentionality. Thus longer looking at the eyes may help to e.g. establish joint attention and cooperative action. However, this theoretical approach is unlikely to explain the findings of greater attention to the mouth in the preceding period of development (Lewkowicz & Hansen-Tift, 2012; Tomalski et al., 2013), as it is difficult to explain how the mouth movements may non-verbally communicate the intent of a speaker.

A recent theoretical proposal has focused on the role of more complex visual behaviour on the part of the infant; namely, alternating fixations on the eyes and on the mouth of dynamic faces interacting with the infant. Ramsdell-Hudock (2014) suggested that such behaviour may be important for receiving feedback from the caregiver when the infant is engaged in her own vocal production. Infants aged 9.5 months produce more speech-like vocalizations when the parent responds contingently to their babbling (Goldstein & Schwade, 2008). Similarly, 10-month-olds that more frequently switch their attention between the actor and the object that is being described in a foreign language show better non-native phoneme discrimination at the neural level in a mismatch response paradigm (Conboy, Brooks, Meltzoff, & Kuhl, 2015). However, this hypothesis has not been directly tested so far, while there is some indirect evidence against it. Infants aged 6-8 months vocalize more when playing independently with toys rather than playing contingently with an experimenter (Harold & Barlow, 2013). More research in naturalistic settings is necessary to establish the trajectory of development of visual scanning and fixation switching between different face parts across the first and second year of life to establish whether it plays a major role in feedback-guided phonetic learning.

Emerging neural organisation of AV speech processing

We will now discuss the data on audiovisual speech processing and mismatch detection in relation to existing studies of changing cortical organisation of auditory and visual processing throughout the first years of life. So far it remains unclear, to what extent early on there is any dedicated neural architecture for the processing of language or phonetic information. Neuroimaging studies provide some evidence for early maturation of left-lateralized areas engaged in speech processing (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002) and for cortical specialisation for the human voice by 5 months of age (Blasi et al., 2011; Lloyd-Fox, Blasi, Mercure, Elwell, & Johnson, 2012).

A major question that requires further investigation is whether the development of AV speech processing is tied with the reorganisation of the dorsal and the ventral visual streams (Johnson, Mareschal, & Csibra, 2001) throughout the

first year of life. Recently, on the basis of adult literature, Berstein and Liebenthal (2014) concluded that lip-reading abilities rely on both configural and dynamic stimulus feature processing. While the former is associated with the ventral visual stream (e.g. ventral temporal cortex), the latter is closely linked with the activity of the dorsal visual stream in the parietal lobe. Infancy work to date has clearly demonstrated that the dorsal stream matures earlier allowing for the visual processing of spatio-temporal information at 4-5 months, while the processing of surface features that relies on the ventral visual stream is more readily available several months later (Kaufman, Mareschal, & Johnson, 2003; Mareschal & Johnson, 2003).

Studies of emerging configural face processing in infancy may help to outline the trajectory of development of phonetic AV speech processing that relies on configural processing dependent on the visual ventral stream. For face processing the behavioural data suggests some degree of specialisation already at 6-10 months of age (e.g. Wheeler et al., 2011), but the electrophysiological evidence indicates that the right hemisphere specialisation for faces does not emerge until 12 months of age (de Haan, Pascalis, & Johnson, 2002). For these reasons it perhaps becomes less surprising that the perceptual narrowing of speech sound contrasts for native language occurs around the same time as narrowing for human faces (but see Watson, Robbins, & Best, 2014). Thus the commitment of neural resources to specific classes of visual articulation stimuli may lead to reduced discrimination of non-native phoneme contrasts for the former (see Pons et al., 2009) and to the face inversion as well as the other-race effects for the latter process. Subsequent infant neuroimaging work may help to ascertain the dorsal and ventral contributions to visual speech processing and the interactions of these networks in the process of arriving at multisensory phoneme representations around the time of perceptual narrowing.

Future directions

The final issue concerns the role of difficulties with audiovisual integration in atypical language development and the emergence of language disorders. There is strong evidence that phonological deficits are central to atypical language development. The learning of phonological information in word learning studies is significantly reduced in children with language impairments compared with same-age peers (Nash & Donaldson, 2005; Steele & Watkins, 2010), while interventions focused on phonological awareness improve their word learning (Zens, Gillon, & Moran, 2009). Difficulties with the integration of auditory and visual speech cues were reported in children with specific language impairment (SLI; Pons, Andreu, Sanz-Torrent, Buil-Legaz, & Lewkowicz, 2013) and autism spectrum disorder (ASD; Megnin et al., 2012). On the one hand children with SLI spend more time fixating the speaking mouth than typical controls, possibly to compensate for auditory processing deficits (Hosozawa, Tanaka, Shimizu,

Nakano, & Kitazawa, 2012). But on the other, the use of visual cues during audiovisual speech integration is known to be less efficient in children and adults with language-learning disabilities (Norrix, Plante, & Vance, 2006; Norrix, Plante, Vance, & Boliek, 2007). Several behavioural studies used the McGurk paradigm to investigate atypical development. In a task where participants reported what they hear older children with ASD showed reduced fusion of incongruent auditory and visual cues in the McGurk illusion (Bebko, Schroeder, & Weiss, 2014) – a divergence from typical development that widens with age (Stevenson et al., 2014). Similar decreased fusion was present in children with developmental language disorder (Meronen, Tiippana, Westerholm, & Ahonen, 2013). In an ERP task adults with ASD showed no differential neural responses to congruent compared with incongruent AV speech, unlike in the IQ-matched control group (Magnée, de Gelder, van Engeland, & Kemner, 2008). This result was found for late ERP components (different from the AVMMR), previously associated with phonological processing, while early sensory components did not differ from the control group. This may suggest a specific difficulty with AV integration of speech cues.

Altogether these studies suggest that audiovisual speech integration deficits are closely related to language difficulties in different atypical trajectories of development. These difficulties may have their origin already in the first years of life. Guiraud and colleagues (2012) found that 8-10 month-old infants at familial risk of autism show reduced detection of salient AV mismatch compared with typically developing controls. What remains unclear is to what extent the difficulties in temporal synchrony detection and later AV phoneme processing contribute to language difficulties and whether they depend on the social-pragmatic difficulties experienced by these children early on.

Conclusions

The aim of this article was to review recent investigations of audiovisual speech integration in early infancy using the McGurk paradigm and to discuss them in the context of new literature on the development of the 'social brain' (Johnson, Grossmann & Cohen Kadosh, 2009; Kuhl, 2007). In particular, this review has focused on a series of studies with 6-9 month-olds conducted by Tomalski, Kushnerenko and colleagues on a group of mono- and bilingual infants from East London, UK. They found an age-related increase in attention to the mouth of articulating faces, but specifically when the auditory and visual speech cues were in apparent conflict. These findings are inconsistent with the existing, intersensory redundancy account of age-related changes in infant attention to the eyes vs. the mouth across the first year of life. However, the existing evidence does not favour the social-pragmatic approach as an alternative explanation of this effect in terms of increasing reliance on face parts that offer salient communicative cues.

Kushnerenko, Tomalski and colleagues also found a strong association between greater attention to articulating mouth and a more mature pattern of neural responses to audiovisual mismatch, i.e. a decreased amplitude of the AVMMR component. This decrease can be explained in terms of predictive coding theories that describe the refinement of top-down neural predictions. However, a subsequent longitudinal language follow-up of infants from their study also possibly suggests that mismatch detection based on temporal cues rather than phonetic information does not constitute an important mechanism for later receptive language development. Altogether these data demonstrate the complexity of developmental trajectories of audiovisual speech processing and the number of issues that warrant further investigation.

Acknowledgments

This work was supported by the FP7 Marie Curie grant (PCIG10-GA-2011-304255). The author acknowledges additional support from the Polish National Science Centre (2011/03/D/HS6/05655, 2012/07/B/HS6/01464), and the Ministry of Science and Higher Education (IP2012 061072). The author wishes to thank Elena Kushnerenko and Robin Panneton for helpful discussions and Marina Zalewska, Joanna Rączaszek-Leonardi and two anonymous reviewers for their comments on an earlier version of the manuscript.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology : CB*, *15* (9), 839-843.

Altvater-Mackensen, N. & Grossmann, T. (2015). Learning to Match Auditory and Visual Speech Cues: Social Influences on Acquisition of Phonological Categories. *Child Development*, *86* (2), 362-378.

Baart, M., Bortfeld, H., & Vroomen, J. (2015). Phonetic matching of auditory and visual speech develops during childhood: Evidence from sine-wave speech. *Journal of Experimental Child Psychology*, *129*, 157-164.

Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2013). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition*, *130* (1), 31-43.

Bahrick, L.E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, *41* (4), 352-363.

Bahrick, L.E., & Lickliter, R. (2014). Learning to Attend Selectively: The Dual Role of Intersensory Redundancy. *Current Directions in Psychological Science*, *23* (6), 414-420.

Bahrick, L.E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13* (3), 99-102.

Bebko, J.M., Schroeder, J.H., & Weiss, J.A. (2014). The McGurk effect in children with autism and Asperger syndrome. *Autism Research*, *7* (1), 50-59.

Bernstein, L.E. & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience*, *8*, 1-18.

Best, C.T. & McRoberts, G.W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46* (2-3), 183-216.

Best, C.T., McRoberts, G.W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, *18* (3), 339-350.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvall, V., Deoni, S., Gasston, D., Williams, S.C.R., Johnson, M.H., Simmons, & A.Murphy, D.G.M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology: CB*, *21* (14), 1220-1224.

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* (5), 905-921.

Brookes, H., Slater, A., Quinn, P.C., Lewkowicz, D.J., Hayes, R., & Brown, E. (2001). Three-Month-Old Infants Learn Arbitrary Auditory-Visual Pairings Between Voices and Faces. *Infant and Child Development*, *10* (1-2), 75-82.

Buchan, J.N., & Munhall, K.G. (2011). The influence of selective attention to auditory and visual speech on the integration of audiovisual speech information. *Perception*, *40* (10), 1164-1182.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45* (4), 204-220.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363* (1493), 1001-1010.

Conboy, B.T., Brooks, R., Meltzoff, A.N., & Kuhl, P.K. (2015). Social interaction in infants' learning of second-language phonetics: An exploration of brain–behavior relations. *Developmental Neuropsychology*, *40* (4), 216-229.

Cooper, R.P. & Aslin, R.N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61* (5), 1584-1595.

Coulon, M., Hemimou, C., & Streri, A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy*, *18* (5), 782-796.

Csibra, G. & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13* (4), 148-153.

De Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14* (2), 199-209.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298* (5600), 2013-2015.

Desjardins, R.N. & Werker, J.F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45* (4), 187-203.

Dodd, B. (1979). Lip reading in infants: attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11* (4), 478-484.

Eimas, P.D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, *18* (5), 341-347.

Farroni, T., Johnson, M.H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: effects of contrast polarity. *Proceedings of the National Academy of Science of the United States of America*, *102* (47), 17245-17250.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8* (2), 181-195.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11* (2), 127-138.

Friston, K.J. (2005). Models of brain function in neuroimaging. *Annual Review of Psychology*, *56* (1), 57-87.

Garrido, M.I., Kilner, J.M., Stephan, K.E., & Friston, K.J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120* (3), 453-463.

Gervain, J. & Werker, J.F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4* (1490), 1-6.

Gliga, T. & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, *20* (3), 347-353.

Goldstein, M.H. & Schwade, J.A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19* (5), 515-523.

Guellai, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: some critical questions. *Frontiers in Psychology*, *5*, 1-7.

Guiraud, J.A., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., Elsabbagh, M., Johnson, M.H., & the BASIS Team. (2012). Atypical audiovisual speech integration in infants at risk for autism. *PloS One*, *7* (5), e36428.

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception & Psychophysics*, *77* (4), 1333-1341.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: Lower scores in the UK than the USA. *Journal of Child Language*, *27* (3), 689-705.

Harold, M.P. & Barlow, S.M. (2013). Effects of environmental stimulation on infant vocalizations and orofacial dynamics at the onset of canonical babbling. *Infant Behavior and Development*, *36* (1), 84-93.

Hockley, N.S. & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America*, *96* (5), 3309.

Hosozawa, M., Tanaka, K., Shimizu, T., Nakano, T., & Kitazawa, S. (2012). How children with specific language impairment view social situations: An eye tracking study. *Pediatrics*, *129* (6), e1453-e1460.

Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, *95* (1), 27-55.

Howard, I. & Messum, P. (2011). Modeling the development of pronunciation in infant speech acquisition. *Motor Control*, *15* (1), 85-117.

Hunnius, S. & Geuze, R.H. (2004). Gaze shifting in infancy: A longitudinal study using dynamic faces and abstract stimuli. *Infant Behavior & Development*, *27* (3), 397-416.

Johnson, M.H. (2001). Functional brain development in humans. *Nature Reviews. Neuroscience*, *2* (7), 475-483.

Johnson, M.H. (2011). Interactive Specialization: A domain-general framework for human functional brain development? *Developmental Cognitive Neuroscience*, *1* (1), 7-21.

Johnson, M.H., Grossmann, T., & Cohen Kadosh, K. (2009). Mapping functional brain development: Building a social brain through interactive specialization. *Developmental Psychology*, *45* (1), 151-159.

Johnson, M.H., Mareschal, D., & Csibra, G. (2001). The functional development and integration of the dorsal and ventral visual pathways: A neurocomputational approach. In C.A. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 339-351). Cambridge, MA: MIT Press.

Johnson, M.H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, *50*, 169-179.

Jones, W. & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504* (7480), 427-431.

Kaufman, J., Mareschal, D., & Johnson, M.H. (2003). Graspability and object processing in infants. *Infant Behavior and Development*, *26* (4), 516-528.

Kisilevsky, B.S., Hains, S.M., Brown, C.A., Lee, C.T., Cowperthwaite, B., Stutzman, S.S., Swansburg, M.L., Lee, K., Xie, X., Huang, H., Ye, H.-H., Zhang, K., & Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32* (1), 59-71.

Kuhl, P.K. (2007). Is speech learning "gated" by the social brain? *Developmental Science*, *10* (1), 110-120.

Kuhl, P.K. (2014). Early language learning and the social brain. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 211-220.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V, Ryskina, V.L., Stolyarova, E.I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277* (5326), 684-686.

Kuhl, P.K. & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science*, *218* (4577), 1138-1141.

Kuhl, P.K. & Meltzoff, A.N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, *7* (3), 361-381.

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* (2), F13-F21.

Kuhl, P.K., Tsao, F.-M., & Liu, H.-M. (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* (15), 9096-9101.

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* (32), 11442–11445.

Kushnerenko, E., Tomalski, P., Ballieux, H., Potton, A., Birtles, D., Frostick, C., & Moore, D.G. (2013). Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*, *4* (432), 1-10.

Kushnerenko, E., Tomalski, P., Ballieux, H., Ribeiro, H., Potton, A., Axelsson, E.L., Murphy, E., Moore, D.G. (2013). Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *The European Journal of Neuroscience*, *38* (9), 3363-3369.

Kushnerenko, E., Winkler, I., Horvath, J., Naatanen, R., Pavlov, I., Fellman, V., & Huotilainen, M. (2007). Processing acoustic change and novelty in newborn infants. *European Journal of Neuroscience, 26* (1), 265-274.

Kushnerenko, E.V., Van den Bergh, B.R.H., & Winkler, I. (2013). Separating acoustic deviance from novelty during the first year of life: A review of event-related potential evidence. *Frontiers in Psychology*, *4* (595), 1-16.

Lewkowicz, D.J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126* (2), 281-308.

Lewkowicz, D.J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, *39* (5), 795-804.

Lewkowicz, D.J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46* (1), 66-77.

Lewkowicz, D.J. & Hansen-Tift, A.M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National -Academy of Sciences of the United States of America*, *109* (5), 1431-1436.

Lewkowicz, D.J., Minar, N.J., Tift, A.H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147-162.

Lloyd-Fox, S., Blasi, A., Mercure, E., Elwell, C.E., & Johnson, M.H. (2012). The emergence of cerebral specialization for the human voice over the first months of life. *Social Neuroscience*, *7* (3), 317-330.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219* (4590), 1347-1349.

Magnée, M.J.C.M., de Gelder, B., van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49* (9), 995-1000.

Mareschal, D. & Johnson, M.H. (2003). The "what" and "where" of object representations in infancy. *Cognition*, *88* (3), 259-276.

Massaro, D.W. (1984). Children's perception of visual and auditory speech. *Child Development*, *55* (5), 1777-1786.

Massaro, D.W., Cohen, M.M., & Smeele, P.M.T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America*, *100* (3), 1777-1786.

Massaro, D.W., Thompson, L.A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, *41* (1), 93-113.

Maye, J., Weiss, D.J., & Aslin, R.N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11* (1), 122-134.

Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82* (3), B101-B111.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264* (5588), 746-748.

Megnin, O., Flitton, A., Jones, C.R.G., de Haan, M., Baldeweg, T., & Charman, T. (2012). Audiovisual speech integration in autism spectrum disorders: ERP evidence for atypicalities in lexical-semantic processing. *Autism Research*, *5* (1), 39-48.

Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, *56* (1), 211-221.

Mottonen, R., Krause, C.M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, *13* (3), 417-425.

Nash, M. & Donaldson, M.L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, *48* (2), 439-458.

Norrix, L.W., Plante, E., & Vance, R. (2006). Auditory–visual speech integration by adults with and without language-learning disabilities. *Journal of Communication Disorders*, *39* (1), 22-36.

Norrix, L.W., Plante, E., Vance, R., & Boliek, C.A. (2007). Auditory-visual integration for speech by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *50* (6), 1639-1651.

Parise, E., Handl, A., Palumbo, L., & Friederici, A.D. (2011). Influence of eye gaze on spoken word processing: an ERP study with infants. *Child Development*, *82* (3), 842-853.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences of the United States of America*, *110* (37), 15145-15150.

Patterson, M.L. & Werker, J.F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22* (2), 237-247.

Patterson, M.L. & Werker, J.F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81* (1), 93-115.

Patterson, M.L. & Werker, J.F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6* (2), 191-196.

Pegg, J.E. & Werker, J.F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, *102* (6), 3742-3753.

Polka, L. & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology. Human Perception and Performance*, *20* (2), 421-435.

Pons, F., Andreu, L., Sanz-Torrent, M., Buil-Legaz, L., & Lewkowicz, D.J. (2013). Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. *Journal of Child Language*, *40* (3), 687-700.

Pons, F., Bosch, L., & Lewkowicz, D.J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, 26 (4), 490-498.

Pons, F. & Lewkowicz, D.J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica*, *149*, 142-147.

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 Sciences of the United States of America*, *106* (26), 10598–10602.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P.K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17* (6), 880-891.

Ramsdell-Hudock, H.L. (2014). Caregiver influence on looking behavior and brain responses in prelinguistic development. *Frontiers in Psychology*, *5* (297), 1-2.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8* (2), 162-172.

Rosenblum, L.D. & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *22* (2), 318-331.

Rosenblum, L.D., Schmuckler, M.A, & Johnson, J.A. (1997). The McGurk effect in infants. *Percept Psychophys*, *59* (3), 347-357.

Ross, L.A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J.J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33* (12), 2329-2337.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., & Foxe, J.J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17* (5), 1147-1153.

Schwartz, J.-L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. *Journal of the Acoustical Society of America*, *127* (3), 1584-1594.

Sekiyama, K. & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, *11* (2), 306-320.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, *92* (3), B13-B23.

Steele, S.C. & Watkins, R.V. (2010). Learning word meanings during reading by children with language learning disability and typically-developing peers. *Clinical Linguistics & Phonetics*, *24* (7), 520-539.

Stevenson, R.A., Siemann, J.K., Woynaroski, T.G., Schneider, B.C., Eberly, H.E., Camarata, S.M., & Wallace, M.T. (2014). Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44* (6), 1470-1477.

Streri, A., Hevia, M., Izard, V., & Coubart, A. (2013). What do We Know about Neonatal Cognition? *Behavioral Sciences*, *3* (1), 154-169.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26* (2), 212-215.

Teinonen, T., Aslin, R.N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108* (3), 850-855.

Tenenbaum, E.J., Shah, R.J., Sobel, D.M., Malle, B.F., & Morgan, J.L. (2013). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy*, *18* (4), 534-553.

Tenenbaum, E.J., Sobel, D.M., Sheinkopf, S.J., Malle, B.F., & Morgan, J.L. (2014). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *18*, 1-18.

Tiippana, K., Andersen, T.S., & Sams, M. (2004, May). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, 16 (3), 457-472.

Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E.L., Murphy, E., Moore, D.G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology*, *10* (5), 611-624.

Tomasello, M. & Carpenter, M. (2007). Shared intentionality. *Dev Sci*, *10* (1), 121-125.

Tsao, F.M., Liu, H.M., & Kuhl, P.K. (2006). Perception of native and non-native affricate-fricative contrasts: cross-language tests on adults and infants. *Journal of the Acoustical Society of America*, *120* (4), 2285-2294.

Watson, T.L., Robbins, R.A., & Best, C.T. (2014). Infant perceptual development for faces and spoken words: An integrated approach. *Developmental Psychobiology*, *56* (7), 1454-1481.

Werker, J.F., & Hensch, T.K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66* (1), 173-196.

Werker, J.F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during thefirst year of life. *Infant Behavior and Development*, *7*, 49-63.

Westermann, G., Mareschal, D., Johnson, M.H., Sirois, S., Spratling, M.W., & Thomas, M.S.C. (2007). Neuroconstructivism. *Developmental Science*, *10* (1), 75-83.

Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian infants scan own- and other-race faces differently. *PloS One*, *6* (4), e18621.

Wilcox, T., Stubbs, J.A., Wheeler, L., & Alexander, G.M. (2013). Infants' scanning of dynamic faces during the first year. *Infant Behavior & Development*, *36* (4), 513-516.

Zens, N.K., Gillon, G.T., & Moran, C. (2009). Effects of phonological awareness and semantic intervention on word-learning in children with SLI. *International Journal of Speech-Language Pathology*, *11* (6), 509-524.

Zimmerman, I.L., Steiner, V.G., & Evatt Pond, R. (2002). Preschool language scales-IV. *Pearson* (http://www.pearsonclinical.co.uk).