

On the Links Among Face Processing, Language Processing, and Narrowing During Development

Olivier Pascalis,¹ H el ene Loevenbruck,^{1,2} Paul C. Quinn,³ Sonia Kandel,¹
James W. Tanaka,⁴ and Kang Lee⁵

¹Universit e Grenoble Alpes, CNRS, ²Grenoble Images Parole Signal Automatique, CNRS,
³University of Delaware, ⁴University of Victoria, and ⁵University of Toronto

ABSTRACT—From the beginning of life, face and language processing are crucial for establishing social communication. Studies on the development of systems for processing faces and language have yielded such similarities as perceptual narrowing across both domains. In this article, we review several functions of human communication, and then describe how the tools used to accomplish those functions are modified by perceptual narrowing. We conclude that narrowing is common to all forms of social communication. We argue that during evolution, social communication engaged different perceptual and cognitive systems—face, facial expression, gesture, vocalization, sound, and oral language—that emerged at different times. These systems are interactive and linked to some extent. In this framework, narrowing can be viewed as a way infants adapt to their native social group.

KEYWORDS—narrowing; face; speech

Social life requires relationships with other group members, acknowledgment of their status, and communication between individuals. Depending on the species studied, communication occurs through vocalization, language, faces and their expressions, or some combination of these. Similarities observed across species may provide insights into the relation between different social communication tools and networks. Based on these observations, we argue here that communicative tools emerged during evolutionary time and that current systems reflect aspects of this evolution.

In humans, faces and language are essential for communication, but they have been studied traditionally as separate areas with little interaction between the two domains, even when their links are acknowledged. In some frameworks, they even have been conceived of as independent cognitive modules. If faces provide an early channel of communication for newborns prior to comprehending gestural or oral language, postnatal exposure to the mother's voice–face combination is required to recognize the mother's face (Sai, 2005). In one study, moving faces were recognized only when sound was present (Coulon, Guellai, & Streri, 2011). Thus, face processing seems to be facilitated by voice processing, even at an early age.

Later, in early childhood, most conversations take place face-to-face. Although auditory information alone is sufficient to understand speech, we rely systematically and unconsciously on visual information provided by a speaker's face. Seeing oro-facial gestures of the speaker accelerates recognition of core words (Fort et al., 2012) and enhances intelligibility in noisy environments (Ben oit, Mohamadi, & Kandel, 1994). Therefore, most human conversations—except when we are on the phone—invoke analyzing facial configurations to locate cues relevant to decode speech. Thus, the integration of audio and facial information is crucial to speech perception.

Olivier Pascalis, H el ene Loevenbruck, and Sonia Kandel, Universit e Grenoble Alpes, CNRS, LPNC UMR5105, France; H el ene Loevenbruck, Grenoble Images Parole Signal Automatique, CNRS, UMR5216, France; Paul C. Quinn, Department of Psychology, University of Delaware, DE, USA; James W. Tanaka, Department of Psychology, University of Victoria, Canada; Kang Lee, Institute of Child Study, University of Toronto, Canada.

This research was supported by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD-46526).

Correspondence concerning this article should be addressed to Olivier Pascalis, Laboratoire de Psychologie et NeuroCognition, Universit e Grenoble Alpes, BP 47 38040, Grenoble, Cedex 9, France; e-mail: olivier.pascalis@upmf-grenoble.fr.

  2014 The Authors. *Child Development Perspectives*   2014 The Society for Research in Child Development

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

DOI: 10.1111/cdep.12064



These observations point to a close link between face and language processing that, we argue, may reflect how social communication evolved and how it develops in infants and children. More specifically, functional links between gestural and oral communication in nonhuman primates as well as infants suggest that social communication is a multimodal system, involving manual and visuo-facial gestures as well as vocalization. This multimodal system is gradually tuned during development, with narrowing occurring in all the different modalities of communication.

FACE PROCESSING, LANGUAGE PROCESSING, AND DEVELOPMENT

Human adults can recognize familiar faces easily and are said to process faces expertly. Faces form a category of stimuli that are homogenous in terms of the positioning of their internal elements, and humans have developed a signature way to discriminate them based on configural (i.e., relational) information, such as the distance between the eyes or between lips and chin. Experience likely plays a critical role in acquiring face expertise (Lee, Anzures, Quinn, Pascalis, & Slater, 2011).

Language is a key tool for social communication because it allows for transmitting complex information that facial expressions cannot. It is a complex cognitive skill requiring recursion and displacement (Chomsky, 1965), yet children acquire it swiftly and without instruction, whereas most adults find learning a second language challenging. Studies of language acquisition have discovered crucial milestones: Vocalizations are observable at birth, babbling emerges at around 6–8 months, children utter their first words at 10–12 months, and they begin to make word combinations and form proto-sentences at around 20–24 months (Vihman, 1996).

Studies of the development of the systems that process faces and language have identified similarities between the two. Face processing develops during the first years of life from a broad nonspecific system to a human-tuned face processor (Nelson, 2001). Faces observed within the infants' visual environment shape and influence the developing face system through a process known as perceptual narrowing: a progression whereby infants maintain the ability to discriminate stimuli to which they are exposed, but lose the ability to discriminate stimuli to which they are not exposed. This course of responsiveness is similar for language development. In the first year, initial discriminatory ability reflecting a universal sensitivity to the sounds of all human languages narrows as a consequence of predominant exposure to one's native language and scarce exposure to other languages (Werker & Tees, 1999). During this time, infants become tuned to their native language and the distribution of phonetic information in the ambient language at the expense of discriminating nonnative contrasts. In other words, infants become experts at processing frequently experienced faces and native sounds.

Narrowing cuts across both visual and auditory modalities, possibly reflecting the development of a common neural architecture (Scott, Pascalis, & Nelson, 2007). Narrowing could be a pan-sensory process; that is, the same phenomenon is observed in various senses during the same period and is part of the development of our multisensory representation of the world (Lewkowicz & Ghazanfar, 2009). This line of thinking raises questions such as: Is perceptual narrowing amodal? Is auditory narrowing linked to visual narrowing?

One argument for the link between the development of face and language processing comes from neuroanatomy. The superior temporal sulcus (STS) is associated with face processing and auditory representation of speech components (Démonet, Thierry, & Cardebat, 2005; Haxby, Hoffman, & Gobbini, 2000). The posterior part of the STS may be considered an amodal convergence zone that plays a key role in integrating face and voice information (Belin, Bestelmeyer, Latinus, & Watson, 2011). These findings suggest similar, interacting, and common brain circuits for processing faces and speech.

Descriptions of narrowing fail to consider the evolution and timing of when face and language processing emerged. What drives or motivates the development of both face and language processing is the urge to communicate. In the rest of the article, we describe several functions of human communication, then explain how perceptual narrowing modifies each of these, and conclude that narrowing is a common characteristic of all social communication.

GESTURAL AND ORAL COMMUNICATION

Human language is described as unique even though some form of communication exists in other species. Understanding the emergence of language during evolution is a challenge, as fossil evidence does not provide much insight into oral language. Two means of communication are seen as potential precursors to human language—vocal calls and gestures—although it is debatable whether language originated in manual gestures or evolved exclusively in the vocal domain. The former hypothesis considers pointing as the initial means to communicate, which later developed into a gestural language. Language may have evolved from manual gestures, and then gradually incorporated vocal elements, so that language involves reciprocity in the actions of partners (Corballis, 2003). The mechanism could be supported by mirror neurons, located in Broca's area in humans (Buccino et al., 2001). This area is involved with vocalization as well as manual action and could have been used as a neural substrate for interspecific communication and then to process speech.

In addition, gestures, and more specifically pointing, are associated closely with language development (Kita, 2003). Ocular pointing (or deictic gaze, at 6–9 months) and later index finger pointing (deictic gesture, at 9–11 months) are key stages in cognitive development that are correlated with stages in speech development. Finger pointing is associated with learning new word

forms and their associated meanings, and when accompanied by word production (at 16–20 months), fosters the emergence of sentences. At later stages, children start using prosodic focus, that is, vocal pointing (Ménard, Løevenbruck, & Savariaux, 2006), or constructions involving a deictic pronoun (Diessel & Tomasello, 2000). Different pointing modalities may share a common cerebral network: Ocular, digital, and prosodic pointing are associated with left parietal activation (Løevenbruck, Dohen, & Vilain, 2009). These findings suggest a link between gesture and language.

However, the referential and combinatorial properties of primate vocal communication suggest that language is also rooted in vocalization (Arnold & Zuberbühler, 2008): Chimpanzees produce and understand functionally referential calls, such as an alarm call for a snake, and monkeys can combine existing calls into higher order meaningful sequences. Furthermore, syllables may derive from cycles of rhythmic opening and closing of the jaw involved in chewing, sucking, and licking, which take on communicative significance as lip smacks, tongue smacks, and teeth chatters (MacNeilage, 1998). These observations suggest a direct evolutionary trajectory from primate vocalizations to human speech rather than a complex route requiring an intermediate stage of gestural communication.

Our view is that functional links between gestural and oral communication, observed in nonhuman primates and infants, suggest that communication is a multimodal system involving manual and visuo-facial gestures as well as vocalization. Human communication may have switched to oral-dominant language for several reasons, including accessibility without seeing the other person (e.g., at night or from a distance) and accessibility while doing something else with the forelimbs (e.g., carrying or using tools; Corballis, 2003). Humans would have gradually used the oro-facial region more than the hand in communicating.

Clearly, different kinds of communication existed before oral language, including vocalizations, facial expressions, and visuo-facial gestures. These findings highlight the strong phylogenetic and ontogenetic links between face and language processing.

NARROWING ACROSS DOMAINS THAT INVOLVE SOCIAL COMMUNICATION

Faces

Although 6-month-olds recognize different races of human faces as well as different monkey faces, 9- to 10-month-olds recognize reliably only faces of their own species and race (for a review, see Lee et al., 2011). Successful social communication relies on our ability to process information that allows us to identify people with whom we interact, such as identity, age, and gender. Specialization for faces of our own race improves our ability to extract such information. Regarding voice recognition, 7-month-olds detected changes in voice only when the language was in their native tongue (Johnson, Westrek, Nazzi, & Cutler, 2011), suggesting that voice recognition develops in pace with increasing competence in language processing. However, younger

infants' ability has not yet been reported and we, therefore, cannot conclude that narrowing has occurred in this domain.

In addition to recognizing faces, infants also learn to recognize facial expressions, which further feeds into their abilities to communicate socially (Quinn et al., 2011). Perceptual narrowing has been found for recognizing emotions in 9-month-old infants, but only for faces of their own race (Vogel, Monesson, & Scott, 2012), suggesting that perceptual narrowing affects stimuli that are important for communication with conspecifics and in-groups.

Audiovisual Speech

By the end of the first year of life, responsiveness to nonnative audiovisual inputs declines both in sound–face matching for other species and in nonnative language (Lewkowicz & Ghazanfar, 2009; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009). In a study that used silent video clips of a bilingual speaker telling a story in two languages, monolingual 4- and 6-month-olds discriminated visually between the two languages, whereas monolingual 8-month-olds did not (Weikum et al., 2007). The link between face and language processing is also illustrated by research in which infants watched and listened to a female speaking their native language or a nonnative language. Four-month-olds looked more at the eyes, 6-month-olds looked equally at the eyes and mouth, and by 8 months, infants shifted their attention to the mouth, regardless of the language spoken. These findings suggest that infants begin to focus on the mouth of a talker precisely when they start babbling (Lewkowicz & Hansen-Tift, 2012). In contrast, 12-month-olds no longer focused on the mouth when exposed to native speech, but continued to look more at the mouth when exposed to nonnative speech (Kubicek et al., 2013; Lewkowicz & Hansen-Tift, 2012).

Music–Rhythm

Music is important for communication and may be involved in comforting, courtship, movement coordination, and social cohesion (Brown, 2003). It requires social skills, such as vocal/gestural imitation, and involves cultural transmission. It may even be considered a form of oral communication that emerged before language (Fitch, 2006). If narrowing happens for any form of communication, it should also occur for music. Indeed, in one study, 6-month-olds were able to discriminate rhythms specific to their culture and those unfamiliar to them; however, 12-month-olds could do so only with a rhythm specific to their culture (Hannon & Trehub, 2005). Furthermore, early and active exposure to culture-specific music rhythms and tonalities may accelerate perceptual narrowing in music (Trainor, Marie, Gerry, Whiskin, & Unrau, 2012).

Auditory Speech

Narrowing of speech perception is also well documented. Infants' speech perception becomes tuned toward their native language at around 10–12 months. Young infants discriminate fine phonetic differences, such as differences in voice onset

time, between consonants such as /pa/ and /ba/ (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Infants are also able to discriminate vowels (e.g., between /a/ and /i/ or /i/ and /u/; Trehub, 1973). Not only can infants younger than 6–8 months discriminate categorically native phonetic contrasts, they can also discriminate those that fall outside their native language. For example, 6- to 8-month-olds who are learning English can discriminate the nonnative dental/retroflex contrasts such as the Hindi /Ta/ versus /ta/ (Werker & Tees, 1999). However, a decline in cross-language consonant perception occurs at 10–12 months. Younger children can discriminate many phonetic differences, whereas older children lose this ability for contrasts that fall outside their native language. Therefore, phonetic discrimination starts as language general but gradually narrows, showing language-specific tuning.

Sign Language

Narrowing has also been observed in perceiving sign language (Palmer, Fais, Golinkoff, & Werker, 2012). Hearing infants are able to discriminate American Sign Language (ASL) signs at 4 months but not at 14 months, whereas infants learning ASL are still able to discriminate signs at the later age. This result suggests that narrowing happens for language regardless of the whether the support is gestural or oral.

NARROWING AS A CATEGORIZATION PROCESS SERVING SOCIAL NEEDS

Our view is that narrowing occurs for different cognitive abilities commonly involved in communication, even though not all evidence uniformly shows that narrowing occurs simultaneously across different domains (see, e.g., Hayden, Bhatt, Kangas, Zieber, & Joseph, 2012, for evidence of own-race specialization several months before language narrowing). Therefore, the underlying mechanism might not be specific to one cognitive ability, but common to all communicative tools. In terms of evolution, it emerged first for processing faces and facial expressions, and therefore, should have been part of primitive language involving rhythm and gestures before becoming part of oral language.

Concomitant occurrence in multiple modalities does not explain why narrowing happens. Our take is that infants are born into a social group that has developed a culture of communication that is unique, opaque (i.e., association between an oral/gestural sign and a referent may be arbitrary), and subject to evolution. The most effective way to integrate within the group may be to adapt rapidly to the group's social habits and communication traditions. During the first 12 months, when infants mainly interact with the mother/caregiver, they have to learn rapidly the appropriate way of communicating when interacting within the social group. The mother/caregiver transmits the basic aspects of communication that are crucial to being part of the community: smiles, language characteristics, and recognition of specific faces.

The child then calibrates its communication systems using learning abilities including imitation. If the child is exposed to several individuals, he or she uses convergence mechanisms to calibrate the system and ends up with finely tuned representations of the faces in the environment as well as detailed representations of the phonemes and prosodic patterns in the ambient language(s).

By this account, narrowing is a categorization process that serves social needs. In the language domain, infants build a broad category including the nonnative contrasts that are lost, and retain tightly tuned categories for native contrasts. In the same way, in the face domain, infants build a large category for other-race faces including multiple other-race face categories (e.g., for infants exposed mainly to Caucasian faces, this category would include Asian and African faces), and build tightly tuned categories organized around subordinate-level identity information for same-race faces (i.e., Olivier vs. Helene vs. Paul). Therefore, narrowing can be conceived of as a system that allows the infant to become more efficient or specialized for the social stimuli at hand in the close environment.

CONCLUSION

In this article, we have argued that perceptual narrowing should be observed for all forms of social communication. During evolution, our social communication used different perceptual and cognitive systems—face, facial expression, gesture, vocalization, sound, and oral language—that emerged at different times. These systems are interactive in adults and their neural mechanisms are linked to some extent. Their development presents similarities as infants adjust to their native social group.

We suggest that the adaptation is accomplished through a specific mechanism dedicated to social cognition, which encompasses the different modalities of communication, including manual and visuo-facial gesture processing, as well as vocalization processing abilities. However, we are uncommitted to whether such a mechanism is part of the core endowment present at birth or is a product of increasing specialization that occurs with development. Behavioral and neuroimaging studies should look at the intertwining of the development of these social abilities. Our suggestion also pertains to the field of neurological or developmental disorders: We predict that deficits in either the development of manual gesture processing, facial gesture processing, or vocalization processing should result in disorders of social communication. This prediction is supported by work on autism spectrum disorders suggesting that social communication strongly relies on the healthy development of these different abilities (Adolphs, Sears, & Piven, 2001; Baron-Cohen, 1989). Although further work is needed to understand this multimodal adaptation process, our account is that the interplay of systems that process faces and language in the development of social communication underlies the occurrences of perceptual narrowing in different domains.

REFERENCES

- Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, *13*, 232–240. doi:10.1162/089892901564289
- Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*, R202–R203. doi:10.1016/j.cub.2008.01.040
- Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology*, *7*, 113–127. doi:10.1111/j.2044-835X.1989.tb00793.x
- Belin, P., Bestelmeyer, P., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*, 711–725. doi:10.1111/j.2044-8295.2011.02041.x
- Benoît, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech, Language and Hearing Research*, *37*, 1195–1203. doi:10.1044/jshr.3705.1195
- Brown, S. (2003). Biomusicology, and three biological paradoxes about music. *Bulletin of Psychology and the Arts*, *4*, 15–17.
- Buccino, G., Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V., . . . Freund, H. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, *13*, 400–404. doi:10.1111/j.1460-9568.2001.01385.x
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, *26*, 199–260. doi:10.1017/S0140525X03000062
- Coulon, M., Guellai, B., & Streri, A. (2011). Recognition of unfamiliar talking faces at birth. *International Journal of Behavioral Development*, *35*, 282–287. doi:10.1177/0165025410396765
- Démonet, J., Thierry, G., & Cardebat, D. (2005). Renewal of the neurophysiology of language: Functional neuroimaging. *Physiological Reviews*, *85*, 49–95. doi:10.1152/physrev.00049.2003
- Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, *11*, 131–152. doi:10.1515/cogl.2001.006
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306. doi:10.1126/science.171.3968.303
- Fitch, W. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*, 173–215. doi:10.1016/j.cognition.2005.11.009
- Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2012). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*, 1207–1223. doi:10.1080/01690965.2012.701758
- Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 12639–12643. doi:10.1073/pnas.0504254102
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233. doi:10.1016/S1364-6613(00)01482-0
- Hayden, A., Bhatt, R. S., Kangas, A., Zieber, N., & Joseph, J. E. (2012). Race-based perceptual asymmetry in face processing is evident early in life. *Infancy*, *17*, 578–590. doi:10.1111/j.1532-7078.2011.00098.x
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*, 1002–1011. doi:10.1111/j.1467-7687.2011.01052.x
- Kita, S. (Ed.). (2003). *Pointing: Where language, culture and cognition meet*. Mahwah, NJ: Erlbaum.
- Kubicek, C., Hillairet de Boisferon, A., Dupierriex, E., Løvenbruck, H., Gervain, J., & Schwarzer, G. (2013). Face-scanning behavior to silently talking faces in 12-month-old infants: The role of pre-exposed auditory speech. *International Journal of Behavioral Development*, *37*, 77–78. doi:10.1177/0165025412473016
- Lee, K., Anzures, G., Quinn, P. C., Pascalis, O., & Slater, A. (2011). Development of face processing expertise. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 753–778). New York, NY: Oxford University Press.
- Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multi-sensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*, 470–478. doi:10.1016/j.tics.2009.08.004
- Lewkowicz, D. J., & Hansen-Tift, A. (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*, 1431–1436. doi:10.1073/pnas.1114783109
- Løvenbruck, H., Dohen, M., & Vilain, C. (2009). Pointing is “special.” In S. Fuchs, H. Løvenbruck, D. Pape, & P. Perrier (Eds.), *Some aspects of speech and the brain* (pp. 211–258). Berlin, Germany: Peter Lang.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*, 499–511.
- Ménard, L., Løvenbruck, H., & Savariaux, C. (2006). Articulatory and acoustic correlates of contrastive focus in French: A developmental study. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 227–251). New York, NY: Psychology Press.
- Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*, 3–18. doi:10.1002/icd.239
- Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the first year of life. *Child Development*, *83*, 543–553. doi:10.1111/j.1467-8624.2011.01715.x
- Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10598–10602. doi:10.1073/pnas.0904134106
- Quinn, P. C., Anzures, G., Izard, C. E., Lee, K., Pascalis, O., Slater, A. M., & Tanaka, J. W. (2011). Looking across domains to understand infant representation of emotion. *Emotion Review*, *3*, 197–206. doi:10.1177/1754073910387941
- Sai, F. Z. (2005). The role of the mother’s voice in developing mother’s face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, *14*, 29–50. doi:10.1002/icd.376
- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*, 197–201. doi:10.1111/j.1467-8721.2007.00503.x
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: Effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences*, *1252*, 129–138. doi:10.1111/j.1749-6632.2012.06462.x
- Trehub, S. E. (1973). Infants’ sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*, 91–96. doi:10.1037/h0034999

- Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford, UK: Blackwell.
- Vogel, M., Monesson, A., & Scott, L. S. (2012). Building biases in infancy: The influence of race on face and voice emotion matching. *Developmental Science, 15*, 359–372. doi:10.1111/j.1467-7687.2012.01138.x
- Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, 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