Recognition of amodal language identity emerges in infancy

David J. Lewkowicz[1] and Ferran Pons[2]

Abstract
Audiovisual speech consists of overlapping and invariant patterns of dynamic acoustic and optic articulatory information. Research has shown that infants can perceive a variety of basic auditory-visual (A-V) relations but no studies have investigated whether and when infants begin to perceive higher order A-V relations inherent in speech. Here, we asked whether and when do infants become capable of recognizing amodal language identity, a critical perceptual skill that is necessary for the development of multisensory communication. Because, at a minimum, such a skill requires the ability to perceive suprasegmental auditory and visual linguistic information, we predicted that this skill would not emerge before higher-level speech processing and multisensory perceptual skills emerge. Consistent with this prediction, we found that recognition of the amodal identity of language emerges at 10–12 months of age but that when it emerges it is restricted to infants' native language.

Keywords
infants, language, multisensory, perceptual development

Interpersonal communication via the speech modality is usually represented by temporally and spatially synchronous auditory and visual streams of information. These streams correspond in terms of their intensity, duration, tempo, and rhythm (Yehia, Rubin, & Vatikiotis-Bateson, 1998) and, as a result, provide a highly salient and redundant communicative signal. Normally, we take advantage of the audiovisual redundancy of everyday speech, and because of this we not only perceive our interlocutors as unified amodal entities (Fowler, 2004; Rosenblum, 2008; Sumby & Pollack, 1954) but even recognize their specific amodal identity (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Lachs & Pisoni, 2004). With specific regard to the latter, this means that adults can link a specific person's face with that person's voice. These facts raise questions about whether we might also be able to recognize other amodal features of talkers such as, for example, the amodal identity of a talker's language and whether such an ability might be present early in development. In other words, can infants perceive the correlation between the vocalizations produced by a talker and the talker's accompanying visual articulatory movements and, in the process, extract the multisensory identity information inherent in such a percept?

The theoretical expectation that the ability to recognize the amodal identity of a talker's language might emerge early in life is consistent with a body of research showing that multisensory perceptual abilities emerge during infancy and that they become relatively sophisticated by the end of the first year of life (Lewkowicz, 2000; Lewkowicz & Ghazanfar, 2009). For example, newborn infants can perceive A-V intensity equivalence (Lewkowicz & Turkewitz, 1980, 1981), learn their mother's face if they have access to her voice in the first hours after birth (Sai, 2005), and can match monkey facial and vocal gestures on the basis of synchrony rather than identity (Lewkowicz, Leo, & Simion, 2010). Two months later, infants begin to detect amodal phonetic information inherent in single syllables (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003) and by 3 months infants can associate specific people's faces and voices (Brookes et al., 2001). By 5 months, infants can

integrate conflicting audible and visible speech information in a manner consistent with the McGurk effect (Rosenblum, Schmuckler, & Johnson, 1997) and can recognize that human faces produce human speech sounds and that monkey faces produce monkey sounds (Vouloumanos, Druhen, Hauser, & Huizink, 2009). By 6 months, infants can perceive illusory spatiotemporal A-V relations (Scheier, Lewkowicz, & Shimojo, 2003) as well as A-V duration equivalence (Lewkowicz, 1986), and by 7 months infants can perceive face-voice affect equivalence (Walker-Andrews, 1986). Finally, by 8 months infants can perceive face-voice gender equivalence (Patterson & Werker, 2002) and can perform adult-like integration of auditory and visual spatial localization cues (Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006).

The rapid improvement in multisensory perception and the emergence of the ability to perceive amodal phonetic information (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003) and amodal species identity (Vouloumanos et al., 2009) suggest that the ability to perceive amodal language identity should emerge sometime during infancy as well. This ability must, however, be preceded by the ability to perceive the amodal attributes of audiovisual speech at the utterance level and this, in turn, requires that infants be able to process speech at the segmental level. Findings show, in fact, that infants begin to detect various segmental speech features during the second half of the first year of life. For example, it is then that infants first begin to detect typical word stress patterns (Jusczyk, Cutler, & Redanz, 1993), recognize

---

[1] Florida Atlantic University, USA
[2] Universitat de Barcelona, Spain

Corresponding author:
David J. Lewkowicz, Department of Psychology & Center for Complex Systems & Brain Science, Florida Atlantic University, Boca Raton, FL 33431, USA.
Email: lewkowic@fau.edu

language-specific sound combinations (Jusczyk, Luce, & Charles-Luce, 1994), use transitional probabilities and/or prosodic cues to identify words in continuous speech (Saffran, Aslin, & Newport, 1996), and begin to recognize word forms and mispronunciations of familiar words (Swingley, 2005; Vihman, Nakai, DePaolis, & Hallé, 2004). In addition, infants begin to understand words at around 10 months of age, and by their first birthday, they begin to acquire their native lexicon and its semantic properties (Benedict, 1979; Fenson et al., 1994; Huttenlocher, 1974). Thus, it is likely that the ability to perceive the amodal character of audiovisual speech at the utterance (i.e., fluent speech) level emerges sometime during the second half of the first year of life. Critically, however, when this ability emerges, it is likely that it is restricted to the infants' native language because multisensory perceptual narrowing produces a decline in responsiveness to non-native audiovisual inputs by the end of the first year of life (Lewkowicz & Ghazanfar, 2009; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009).

To examine this prediction, we tested 6–8- and 10–12-month-old infants and used a paired-preference procedure to investigate their looking at two identical faces speaking silently. One face could be seen speaking in the infants' native language and the other in a non-native language. The experiment consisted of a baseline and a familiarization/test phase. During the baseline phase, infants watched the two faces speaking silently. During the familiarization/test phase, infants heard one of the audible languages, saw the pairs of silently talking faces, heard the same audible language again, and saw the pairs of silently talking faces again but reversed for side of presentation. During both phases, we measured duration of looking and asked whether preferences would change following exposure to audible-only speech. As suggested earlier, our general expectation was that the ability to perceive amodal language identity would emerge sometime in the second half of the first year of life and that this would be reflected in a shift in visual preferences following familiarization. Our more specific expectation was that we would not observe the shift until 10–12 months of age because it is then that multisensory perceptual abilities are sufficiently developed to permit infants to perceive amodal language identity. In addition, we expected that we would only observe the shift following familiarization with the infants' native language because of multisensory narrowing.

Method

Participants

We tested 6–8-month-old ($N = 96$; 44 girls; M age = 7 months, range = 5 months, 25 days–8 months, 29 days) and 10–12-month-old ($N = 96$; 47 girls; M age = 11 months, 6 days, range = 9 months, 24 days–12 months, 19 days) infants. All infants were raised in mostly monolingual English homes. This was established by a detailed questionnaire (Bosch & Sebastián-Gallés, 2001) which determined which language/s were spoken by the parents and other close relatives. As in previous studies (Bosch & Sebastián-Gallés, 2001; Weikum et al., 2007), monolingual infants were defined as those infants who had at least 80–85% exposure to a specific language (in our case, English). All but eight of the infants were Caucasian.

Apparatus & stimuli. Infants were seated in an infant seat or in their parent's lap 50 cm from two side-by-side 17-inch (43.2cm) LCD display monitors that were 6.7 cm apart. The experimenter was outside of the testing chamber and only started the experiment. Infant-looking was recorded via a closed-circuit camera and was later coded off-line by observers who were blind with respect to the experimental hypothesis.

A single and continuous QuickTime-movie was used to present all stimulus events. The visual events consisted of video clips showing the face of the same Caucasian bilingual female silently uttering a script in a highly prosodic style in English on one monitor and in Spanish on the other, and in a temporally similar way in each language (see Figure 1). The auditory events consisted of 20-second clips of audible-only utterances (65 ± 5 dB[A]) that corresponded to the first 20 seconds of one of the visible languages. To ensure that infants did not respond on the basis of some idiosyncratic features of the actor's faces and vocalizations, the audible utterances were not those of the female seen but rather those of two other and different females, each of whom was a native speaker in her respective language.

Procedure. The experiment consisted of six trials: two baseline, two auditory familiarization, and two test trials (Figure 1). Each baseline and test trial lasted 30 seconds, and side of language presentation was switched on the second of each of these two types of trials. Each auditory familiarization trial was followed by a test trial, and during the familiarization trials half the infants at each age heard the English soundtrack while the other half heard the Spanish soundtrack. We presented a multicolored rotating ball in the center prior to each visual test trial as well as during the familiarization trials to give infants something to look at while they listened to the utterance.

Results

We analyzed looking during the first 20 s of the baseline and test trials based on the expectation that infants would map the auditory utterance directly onto the corresponding visual utterance if they were perceiving amodal identity information. In the first analysis, we asked whether infants preferred either silent visual language by examining looking during the baseline trials. Results indicated that, regardless of age, infants looked equally at Spanish vs. English, respectively, in each language-familiarization group (younger, Spanish-familiarized infants: 9.4 seconds vs. 8.5 seconds, $t(47) = 0.94$, *ns*; older, Spanish-familiarized infants: 8.8 seconds vs. 8.2 seconds, $t(47) = 0.92$, *ns*; younger, English-familiarized infants: 8.3 seconds vs. 9.6 seconds, $t(47) = 1.36$, *ns*; older, English-familiarized infants: 8.7 seconds vs. 9.1 seconds, $t(47) = 0.56$, *ns*.

In a second analysis, we addressed the principal question of interest, namely whether infants recognized amodal language identity. To do so, we compared looking prior to auditory familiarization with looking following it based on separate baseline and test-trial proportion-of-total-looking-time (PTLT) scores. To compute each score, we divided the total amount of looking time accorded to the matching visible language by the total amount of time accorded to both languages. We then compared these PTLT scores by way of a mixed analysis of variance (ANOVA), with Age (6–8 and 10–12 months) and Audible-Language (English, Spanish) as the between-subjects factors and Experimental Condition (baseline, test) as the within-subjects factor.

The ANOVA yielded a significant triple interaction, $F(1, 188) = 5.34$, $p < .025$. Follow-up tests indicated that this interaction was due to the younger infants not exhibiting a shift in looking following familiarization, and the older infants
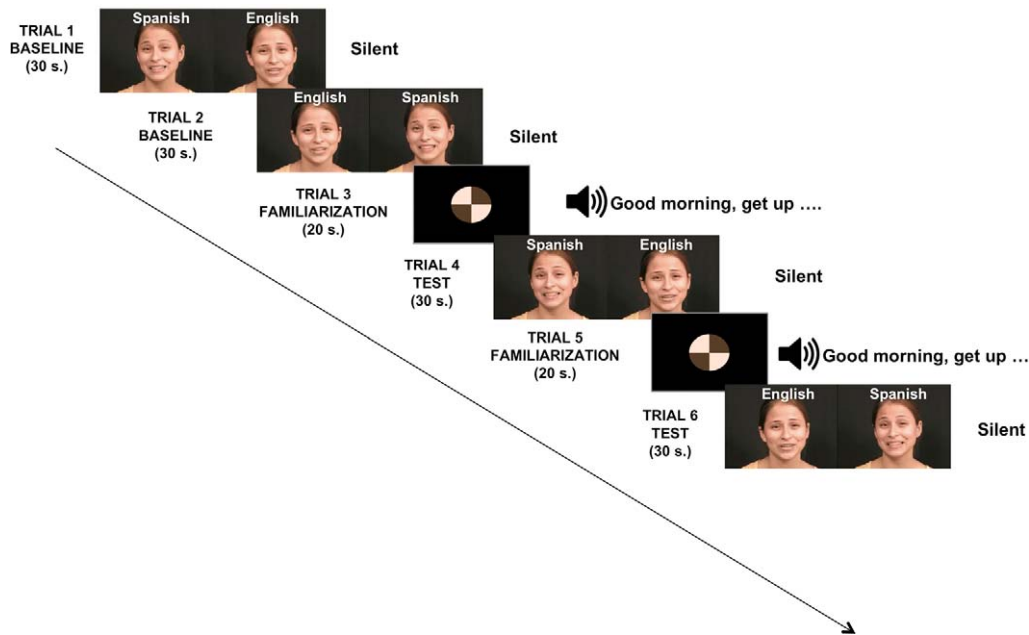
Figure 1. A schematic representation of the types of stimuli presented during the experiment (showing only the English soundtrack condition).
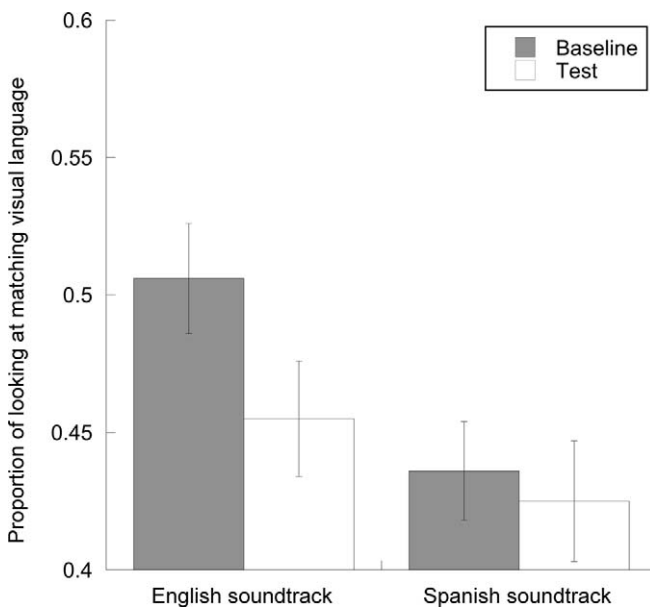

Figure 2. Mean proportion of total looking at the matching visible language during the baseline trials and during the test trials following familiarization with either the English or the Spanish soundtrack in the 10–12 month-old infants. Error bars indicate the standard errors of the mean.

exhibiting a shift following familiarization with the English utterance but not with the Spanish one. That is, no significant effects were observed in the younger infants whereas a significant Audible-Language $\times$ Experimental Condition interaction, $F(1, 94) = 3.83$, $p = .05$, was found in the older infants (see Figure 2). Separate two-tailed t-tests revealed that those older infants who were familiarized with the English audible utterance looked longer at the Spanish face, $t(47) = 2.51$, $p < .025$, but that those who were familiarized with the Spanish audible utterance did not exhibit differential looking.

Discussion

The current findings demonstrate that infants begin to recognize the amodal identity of their native language by the end of the first year of life. Specifically, the findings showed that following familiarization with an auditory-only English speech utterance, older English-learning infants looked longer at a face speaking silently in Spanish than at a face speaking silently in English. This novelty effect is similar to previous reports of novelty effects in both the multisensory perception literature (Gottfried, Rose, & Bridger, 1977) and the visual perception literature (Pascalis, Haan, & Nelson, 2002). It shows that the older infants recognized the correspondence between the previously heard English utterance and the English-speaking face during the test trials. This finding confirms our prediction that infants are likely to begin recognizing native-language, but not non-native language, amodal identity by the end of the first year of life.

The finding that the older infants recognized native-language amodal identity suggests that they extracted, remembered, and matched common visible and audible speech attributes. If so, they did this either by detecting the spatiotemporally correlated patterns of optic and acoustic energy and/or by detecting the invariant optic and acoustic lexical information. Most likely, it was the former, because even 10-month-old infants can only comprehend approximately 10 words (Fenson et al., 1994) and, thus, do not possess a large-enough mental lexicon to detect the specific lexical items in the utterances presented here. If recognition of amodal language identity depends on the detection of the correlated temporal patterns of acoustic and optic energy, then why did the older infants fail to detect the correlated patterns following familiarization with Spanish?

The most likely answer is the fact that responsiveness to speech and language in older infants reflects the operation of two different and opposing developmental processes. The first is the process of

learning and differentiation which leads to the developmental emergence of gradually improving perceptual detection skills (Gibson, 1969). It gradually enables infants to discover native-language segmental, suprasegmental, and lexical features (Johnson & Jusczyk, 2001; Jusczyk et al., 1993, 1994; Saffran et al., 1996; Swingley, 2005; Vihman et al., 2004) and gradually leads to an improvement in multisensory perceptual skills (Lewkowicz, 2002; Lewkowicz & Ghazanfar, 2009). The second and opposing process is perceptual narrowing that occurs at both the unisensory and multisensory processing levels. It is illustrated by findings showing that younger infants exhibit broader unisensory and multisensory perceptual tuning than do older infants. For example, it has been found that younger infants can perceive native as well as non-native audible-only and visible-only speech (Weikum et al., 2007; Werker & Tees, 1984), native and non-native faces (Pascalis et al., 2002), native-race and non-native race faces (Pascalis & Kelly, 2009), and the amodal attributes of native and non-native speech (Pons et al., 2009). In contrast, it has been found that older infants only perceive native audible-only and visible-only speech, faces, and amodal attributes of speech (Mattock & Burnham, 2006; Pascalis et al., 2002; Pons et al., 2009; Skoruppa et al., 2009; Weikum et al., 2007; Werker & Tees, 1984, 2005). Thus, even though infants' perceptual expertise improves during development due to learning and differentiation, the breadth of that expertise narrows due to exposure to native-only auditory, visual, and audiovisual perceptual inputs.

If the older infants recognized amodal language identity by perceiving the correlated temporal patterns of acoustic and optic energy, the perceptual feature that most likely mediated responsiveness was language prosody. It is known that infants can distinguish between languages on the basis of their prosodic (that is, rhythmic) characteristics starting at birth (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998) and that this early and broad sensitivity becomes refined and attuned to the specific prosodic characteristics of the infants' native language (Jusczyk et al., 1993; Nazzi, Jusczyk, & Johnson, 2000; Pons & Bosch, 2010). Because English and Spanish belong to different rhythmic classes (that is, they are stress-timed and syllable-timed languages, respectively), it is likely that the older infants distinguished between these languages based on their unique prosodic characteristics. Consequently, the finding that the older infants successfully perceived their native amodal language identity reflects the emergence of perceptual expertise for multisensory native-language speech attributes.

The finding that the younger infants failed to recognize amodal language identity is consistent with our theoretical predictions offered earlier. Furthermore, the younger infants' failure could not have been due to their inability to perform cross-modal transfer, because prior studies have demonstrated that 6-month-old infants can perform this type of cross-modal transfer in an audiovisual speech perception task (Pons et al., 2009), and that young infants can detect various types of multisensory relations including those inherent in audiovisual speech (Kuhl & Meltzoff, 1982). Therefore, the most likely reason for the younger infants' failure was probably the fact that they find it difficult to perceive the multisensory coherence of audiovisual speech beyond the syllable level.

In conclusion, the current findings demonstrate that the ability to recognize the amodal identity of one's native language emerges by the end of the first year of life. The emergence of this critical perceptual/linguistic skill is parallel with the start of the vocabulary explosion in early development (McMurray, 2007). This is important because multisensory redundancy effects are known to facilitate perception, learning, and memory (Lewkowicz &

Kraebel, 2004) in infancy, and because it has been found that infants take advantage of audiovisual speech redundancy when they begin learning how to talk in the second half of the first year of life (Lewkowicz & Hansen-Tift, 2012). If, by the end of the first year of life, infants take full advantage of the audiovisual redundancy of their native fluent speech, then they can gain access to the redundant auditory and visual attributes of new lexical items at precisely the point when the vocabulary explosion is beginning. These redundant attributes are likely to facilitate the acquisition of new lexical items and, in the process, facilitate the acquisition of language as well.

References

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, *6*, 183–200.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49.

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 & Child Development*, *10*(1–2), 75–82.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994) Variability in early communicative development. *Vol. 59. Monographs of the Society for Research in Child Development*.

Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processing* (pp. 189–201). Cambridge, MA: MIT Press.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton.

Gottfried, A. W., Rose, S. A., & Bridger, W. H. (1977). Cross-modal transfer in human infants. *Child Development*, *48*(1), 118–123.

Huttenlocher, J. (1974). The origins of language comprehension. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 331–368). Potomac, MD: Lawrence Erlbaum.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language*, *44*(4), 548–567.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language*, *33*(5), 630–645.

Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). Putting the face to the voice: Matching identity across modality. *Current Biology*, *13*(19), 1709–1714.

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

Lachs, L., & Pisoni, D. (2004). Crossmodal source identification in speech perception. *Ecological Psychology*, 16(3), 159–187.

Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior & Development*, 9(3), 335–353.

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

Lewkowicz, D. J. (2002). Heterogeneity and heterochrony in the development of intersensory perception. *Cognitive Brain Research*, 14, 41–63.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, 13(11), 470–478.

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*, 109(5), 1431–1436.

Lewkowicz, D. J., & Kraebel, K. (2004). The value of multimodal redundancy in the development of intersensory perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processing* (pp. 655–678). Cambridge, MA: MIT Press.

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match non-human primate faces & voices. *Infancy*, 15(1), 46–60.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, 16, 597–607.

Lewkowicz, D. J., & Turkewitz, G. (1981). Intersensory interaction in newborns: Modification of visual preferences following exposure to sound. *Child Development*, 52(3), 827–832.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10(3), 241–265.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317(5838), 631.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, 24(3), 756–766.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory & Language*, 43(1), 1–19.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, 9(5), 454–464.

Pascalis, O., Haan, M. D., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296(5571), 1321–1323.

Pascalis, O., & Kelly, D. J. (2009). The origins of face processing in humans: Phylogeny and ontogeny. *Perspectives on Psychological Science*, 4(2), 200–209.

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

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 USA*, 106(26), 10598–10602.

Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: The role of syllable weight. *Infancy*, 15(3), 223–245.

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

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17(6), 405.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.

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(1), 29–50.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, 6, 233–244.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, 12(6), 914–919.

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

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8(5), 432–443.

Vihman, M., Nakai, S., DePaolis, R., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3), 336–353.

Vouloumanos, A., Druhen, M., Hauser, M., & Huizink, A. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences*.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, 22, 373–377.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, 316(5828), 1159.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, 7(1), 49–63.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology. Special Issue: Critical Periods Re-examined: Evidence from Human Sensory Development*, 46(3), 233–234.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26(1–2), 23–43.