BRIEF REPORT

Violation of non-adjacent rule dependencies elicits greater attention to a talker's mouth in 15-month-old infants

Joan Birulés[1,2] | Anna Martinez-Alvarez[3] |
David J. Lewkowicz[4,5] | Ruth de Diego-Balaguer[1,6,7,8] |
Ferran Pons[1,6]

[1]Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain

[2]Laboratoire de Psychologie et NeuroCognition, University Grenoble Alpes, Grenoble, France

[3]Department of Developmental Psychology and Socialization, University of Padua, Padua, Italy

[4]Haskins Laboratories, New Haven, Connecticut, USA

[5]Child Study Center, Yale School of Medicine, New Haven, Connecticut, USA

[6]Institute of Neurosciences, University of Barcelona, Barcelona, Spain

[7]ICREA, Barcelona, Spain

[8]Cognition and Brain Plasticity Unit, IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain

Correspondence
Joan Birulés, Department of Cognition, Development and Educational Psychology, Universitat de Barcelona, Barcelona 08035, Spain.
Email: joanbirules@gmail.com

Abstract

Infants start tracking auditory-only non-adjacent dependencies (NAD) between 15 and 18 months of age. Given that audiovisual speech, normally available in a talker's mouth, is perceptually more salient than auditory speech and that it facilitates speech processing and language acquisition, we investigated whether 15-month-old infants' NAD learning is modulated by attention to a talker's mouth. Infants performed an audiovisual NAD learning task while we recorded their selective attention to the eyes, mouth, and face of an actress while she spoke an artificial language that followed an AXB structure (tis-X-bun; nal-X-gor) during familiarization. At test, the actress spoke the same language (grammatical trials; tis-X-bun; nal-X-gor) or a

novel one that violated the AXB structure (ungrammatical trials; tis-X-gor; nal-X-bun). Overall, total duration of looking did not differ during the familiar and novel test trials but the time-course of selective attention to the talker's face and mouth revealed that the novel trials maintained infants' attention to the face more than did the familiar trials. Crucially, attention to the mouth increased during the novel test trials while it did not change during the familiar test trials. These results indicate that the multisensory redundancy of audiovisual speech facilitates infants' discrimination of non-adjacent dependencies.

1 | INTRODUCTION

Acquisition of grammar requires infants to learn the sequential relationship between adjacent and non-adjacent linguistic units. Non-adjacent dependencies (NADs) such as, for example, "*Is* [play] *ing*" or "*Is* [listen]*ing*", require infants to learn the temporally distant relationship between the "*Is*" and the "*-ing*" elements, regardless of intermediate [X] elements. Infants can learn NADs by means of statistical learning (for a review see Saffran & Kirkham, 2018) but this ability is constrained by the nature of input, including the similarity of dependent elements, the variability of the intermediate element (Gómez, 2002), the presence of phonological or prosodic cues (Frost & Monaghan, 2016; Mueller, Friederici, & Maïl;nel, 2012; Newport & Aslin, 2004), and attention (de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016; López-Barroso, Cucurell, Rodríguez-Fornells, & de Diego-Balaguer, 2016). Infants can acquire natural-language morpho-syntactic NADs around 18 months of age (Santelmann & Jusczyk, 1998), and artificial-language NADs (AXB-like strings) as early as 15 months of age (Gómez & Maye, 2005).

To date, studies of infants' ability to learn NADs have only investigated this ability in the auditory (A) modality. The fact is, however, that infants usually experience speech as an audiovisual (AV) event. That is, whenever infants find themselves in a social setting and their social partners interact with them *en face*, infants not only hear speech but also see it on their social partner's face. The A and visual (V) speech cues emanating from a talker's vocal track and face are normally spatiotemporally congruent (Chandrasekaran et al., 2009) and, thus, convey highly redundant AV perceptual cues. Such redundant cues are known to increase the perceptual salience of spoken speech and enhance speech processing in adults (Arnold & Hill, 2001; Summerfield, 1979) as well as in infants (Teinonen et al., 2008). Importantly, infants are actually in a position to profit from the multisensory redundancy of AV speech for two reasons. First, they acquire the ability to integrate A and V speech cues in the first months of life (Kuhl & Meltzoff, 1982; Lewkowicz et al., 2010). Second, infants begin lipreading by 8 months of age (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015) and thus acquire the ability to gain direct access to the AV speech cues of their interlocutors. Once they do, it appears that they rely on it for the acquisition of language skills (Tenenbaum et al., 2015; Tsang et al., 2018; Young et al., 2009).

Given that infants selectively attend to a talker's mouth and thus to redundant AV speech cues by the time they begin to acquire their basic speech abilities by the end of the first year of life, it is

possible that selective attention to a talker's mouth may facilitate the learning of NADs. To examine this prediction, we tested 15-month-old infants by first familiarizing them with AV utterances produced by a female actor speaking an artificial language that followed an AXB structure (i.e., tis-X-bun; nal-X-gor). Then, we tested the infants with AV utterances consisting of the same language or ones consisting of an unfamiliar language containing a violation of the previous AXB structure (i.e., tis-X-gor; nal-X-bun). Two outcomes were possible. On the one hand, infants might be expected to devote greater overall attention to the novel than familiar grammar. On the other, given that learning of NADs is a temporally distributed task and that this requires time to manifest itself, infants may exhibit a different time-course of selective attention to the mouth during the familiar as opposed to the novel grammar test trials.

2 | METHOD

Participants. We recruited 29 healthy full-term infants with no history of hearing problems. A bilingual infant questionnaire was filled out by the parents (DeAnda et al., 2016) and only Catalan or Spanish monolingual infants (input language exposure >80%) were included. All families were Causasian and Hispanic. Participant's inclusion required at least one test trial per grammar condition with looking time greater than 2 s and within 2 SD from that trial's mean looking time (as in Frost et al., 2020). Two infants were excluded for not reaching this criteria. The final sample consisted of 27 infants (14 boys; mean age = 15 months; range 14.7–15.4 months). The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Bioethical committee of the University of Barcelona.

Stimuli. We used an A-X-B rule to create two artificial grammars (G1 or G2). G1 strings took the A1-X-B1 and A2-X-B2 form. G2 strings took the form of A1-X-B2 and A2-X-B1 as in Gómez and Maye (2005). The A elements were instantiated as *tis* and *nal* (A1 and A2, respectively), and B elements were instantiated as *bun* and *gor* (B1 and B2, respectively). We used 24 X bisyllabic non-words (e.g., *gadi*, *breso*) to promote learning (Gómez, 2002). Thus, 96 AxB-like strings were created (24 A1-X-B1, 24 A2-X-B2, 24 A1-X-B2, and 24 A2-X-B1). All 96 strings followed Catalan and Spanish phonotactics (the full list is available in the supplementary information section).

The 96 strings were uttered by a Caucasian female actress in an infant-directed manner, recorded from her shoulders up and presented as video clips (see Figure 1). To prevent participants from predicting element onset times, we varied element and stream length and the stream-to-stream silent periods.

Apparatus and procedure. We tested infants in a dimly lit, sound-attenuated room while they were seated in an infant seat 60 cm in front of a 17-inch monitor. The stimuli were presented at approximately 10 degrees of visual angle. The video clips were presented on a Tobii-X120 standalone eye tracker running at 60 Hz using Tobii Studio software while the audio stimuli were presented via two loudspeakers at 65 ± 5 dB, located at the two sides of the monitor. We used the Tobii eye tracker to measure the duration of eye gaze to three areas of interest (AOIs) corresponding to the actor's eyes, mouth, and face (Figure 1).

We used an infant-controlled familiarization/test procedure and the experiment consisted of a calibration phase using Tobii's five-point calibration routine, a 4 m familiarization phase, and a test phase consisting of four 40 s test trials separated by varying inter-trial intervals. During familiarization, each participant was randomly assigned to one of two familiarization groups (i.e., G1 with A1-X-B1 and A2-X-B2 dependencies or G2 with A1-X-B2 and A2-X-B1 dependencies). Infants were
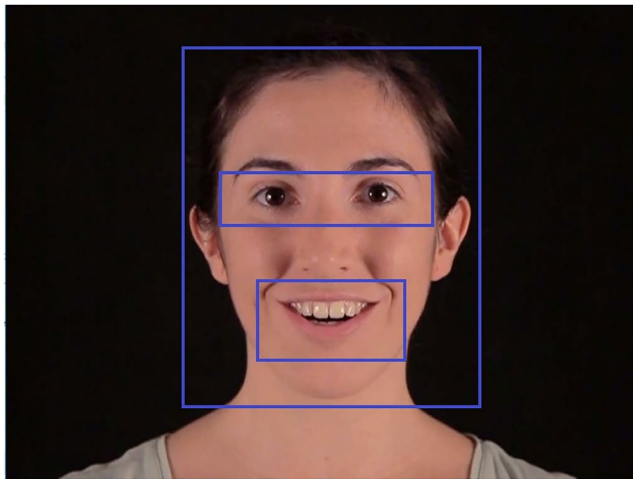
FIGURE 1   Still photo of the actress's face showing the eyes, mouth, and face AOIs used

presented with 48 A-X-B strings, organized into 12 training blocks (i.e., each block consisted of either 3, 4, or 5 AXB strings). Blocks started with an attention-getter and as soon as infants looked at it, a familiarization string was presented. Each string was separated by a 100 ms white fixation point presented in the center of the screen. All familiarization blocks contained at least one string of each dependency and no cue was given related to the upcoming string. The test phase commenced as soon as the familiarization phase ended and consisted of four trials. The test trials were made by using the same syllables presented during familiarization, and consisted of eight A-X-B strings presented pseudorandomly (i.e., four repetitions of each dependency). Two test trials followed G1 dependencies (A1-B1 and A2-B2) and the other two followed G2 dependencies (A1-B2 and A2-B1). The order of the dependencies presented at test was fixed (G1, G2, G1, G2), so for one familiarization group the strings administered first were rule-following "grammatical" trials (G1) while for the other group these first strings were rule-violating "ungrammatical" trials (G2) (see Table S1 in supplementary materials for the complete stimuli description). Each test trial lasted approximately 40 s.

Data Analysis. As in prior NAD-learning studies (Gómez & Maye, 2005), first we compared the total amount of fixation of the face during grammatical and ungrammatical test trials with a mixed analysis of variance (ANOVA). We followed this traditional analysis with one that investigated attentional dynamics by examining changes in eye gaze to the face and mouth as a function of time during the test trials. To do so, we used two generalized linear mixed model (GLMM) analyses and post-hoc likelihood-ratio ($\chi2$) forward model comparisons—comparing each model fit to a reduced one—to assess the predictive power and significance of the included main effects and interactions. The data, code and stimuli are available at https://osf.io/pcuw9/?view_only=de33faf1f0ee402bbbbc15b532392f4b.

3 | RESULTS

To determine whether attention to the grammatical versus ungrammatical test strings differed, we examined the total looking time to the talker's face with a mixed ANOVA, with Familiarization Group (1) as a between-subjects factor and Grammar (2) as a within-subjects factor. Results did not yield any significant effects (all $p's > 0.1$) indicating that total looking duration did not differ in the two grammar conditions. We then investigated the time-course of attention to the talker's face with a GLMM

analysis and post hoc likelihood-ratio (χ2) model comparisons (a full description of analyses and results is provided in the supplementary information section). Results indicated that the full model fitted best, coded in R: Face Look ~ Time × Grammar + (1+Time |Participant)[1].

Figure 2 summarizes the model's results. A significant main effect of Time [$\beta = -0.74$ (0.10), $z = -7.39$, $p < 0.001$] reflected the expected overall decrease of attention across time. The lack of Grammar main effect [$\beta = -0.01$ (0.03), $z = -0.19$, $p > 0.1$] and significant Time x Grammar interaction [$\beta = 0.24$ (0.05), $z = 5.37$, $p < 0.001$] indicate that overall attention was equal across conditions, but that the slope of attention decrease was steeper (i.e., attention decreased faster) in the grammatical than in the ungrammatical test trials (see Figure 2).

If infants exhibited a slower attention drop in the ungrammatical condition because of rule violation, then we would expect increased attention to the source of AV speech, namely the talker's mouth. To test this prediction, we repeated the GLMM analysis with the same factors as described above, but with proportion of looking to the mouth as the dependent measure.[2] Again, forward model comparisons indicated that the full model fitted the best.

Results (see Figure 3) show a main effect of Grammar [$\beta = 0.15$ (0.05), $z = 3.15$, $p < 0.01$] and a significant Time x Grammar interaction [$\beta = 0.23$ (0.07), $z = 3.45$, $p < 0.001$] indicating that infants deployed greater attention to the talker's mouth in the ungrammatical than grammatical trials, and that mouth-looking increased more during the ungrammatical trials than in the grammatical. Further analysis using both the grammatical and ungrammatical trials as base factors for the GLMM model show that the interaction is caused by the fact that mouth attention increased during grammatical trials but that it remained unchanged during the grammatical trials (see S2.2.2 in supplementary materials for full results).
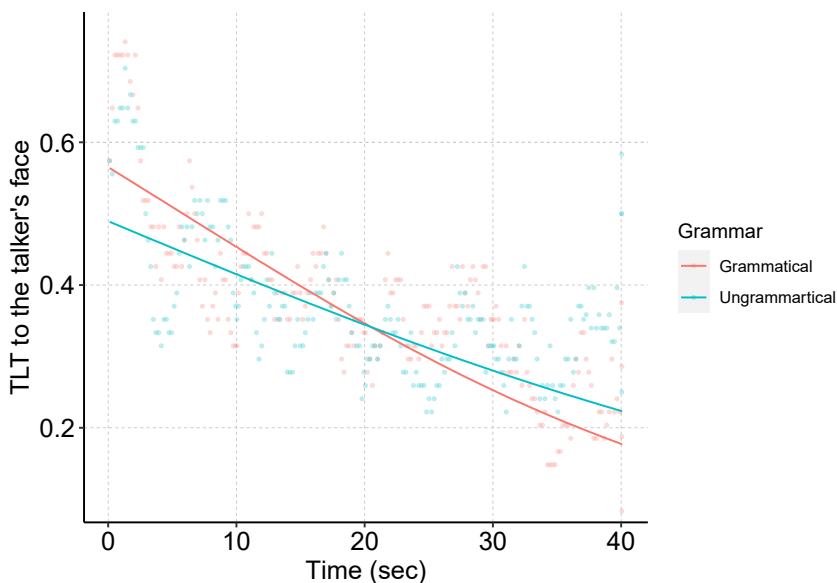

FIGURE 2 Time-course of Proportion-of-Total-Looking-Time (PTLT) to the talker's face during the grammatical (red) and ungrammatical (blue) test trials. The lines represent the fitted GLMM and the dots represent the group PTLT mean for each time point

FIGURE 3 Time-course of Proportion-of-Total-Looking-Time (PTLT) to the talker's mouth during the grammatical (red) and ungrammatical (blue) test trials. The lines represent the fitted GLMM and the dots represent the group PTLT mean for each time point
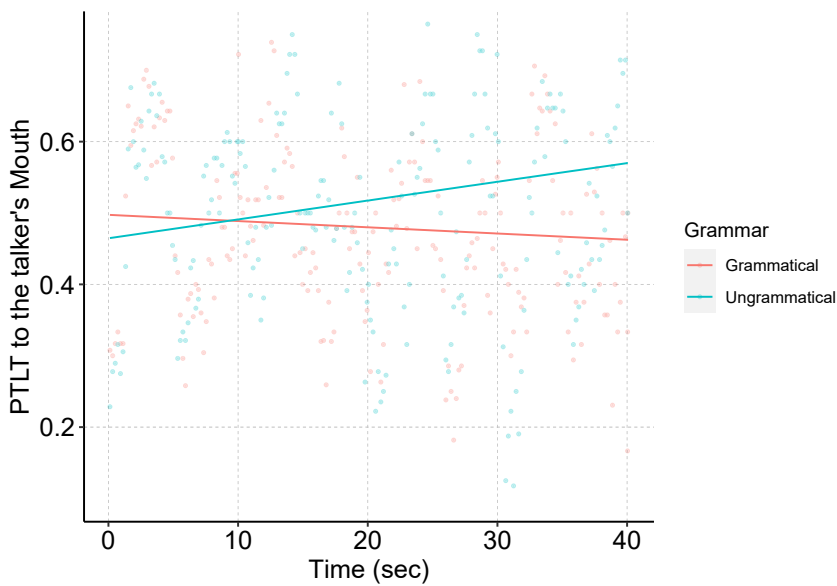
4 | DISCUSSION

The current study investigated whether 15-month-old infants take advantage of the highly salient redundant AV speech information in a talker's mouth to learn non-adjacent dependencies. Following familiarization to two different AV sequences defined by distinct rules specifying the non-adjacent relationship between their respective initial and terminal elements, we tested infants with familiar (i.e., grammatical) and novel (i.e., ungrammatical) sequences. Infants 1) decreased their attention to the talker's face more slowly when tested with ungrammatical than grammatical sequences, 2) attended more to the talker's mouth when exposed to ungrammatical than grammatical sequences, and 3) increased attention to the mouth in response to the ungrammatical sequences but not in response to grammatical sequences. These results indicate that infants learned and discriminated the NAD rules and that they relied on the high perceptual salience of AV speech to learn them.

Prior studies using A-only speech have found that 15- and 18- but not 12-month-old infants successfully learned morpho-syntactic NAD rules (Gómez & Maye, 2005). Our findings extend these earlier findings by showing that 15-month-old infants also can extract NADs from AV speech. Interestingly, however, we found a novelty preference whereas Gómez and Maye (2005) found a familiarity preference. The different outcomes likely reflect the differential effects of A-only versus AV learning, respectively, and probably a distinct learning strategy when perceptually more salient AV speech cues are available. This interpretation is reasonable on two grounds. First, empirical findings indicate that AV speech is perceptually more salient in that it enhances language learning in infancy (Schure et al., 2016; Teinonen et al., 2008; Weatherhead & White, 2017). Second, infants begin lipreading between 6 and 8 months of age (Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019; Pons et al., 2015). Once they do, they begin to take advantage of the greater salience of AV as opposed to A speech to acquire their native phonology and to overcome the difficulty of processing non-native speech (Lewkowicz & Hansen-Tift, 2012), to learn two close languages (Birulés et al., 2018; Pons

et al., 2015), to learn to imitate their native speech sounds (Imafuku et al., 2019), and to acquire their native language (Tenenbaum et al., 2015; Tsang et al., 2018; Young et al., 2009).

The current study provides novel insights into infant learning of NADs. Through the use of eye tracking combined with innovative time-based data analyses, we were able to evaluate the dynamic changes in selective attention. Here, by default, the task was a temporally distributed one because infants could only determine whether one grammar differed from another by monitoring the unfolding of the stimulus sequence over time. Consistent with the focus on the dynamic aspects of selective attention and with our predictions, we found that attention to the talker's mouth was greater in response to novel than familiar test sequences, and critically, that attention to novel sequences increased over time but that it remained unchanged to familiar sequences. These results indicate that 15-month-old infants can learn NAD rules and that they do so by relying on the highly salient AV speech cues. When the current findings are considered together with previous findings from studies of selective attention to talking faces in infancy, they demonstrate that increased attention to AV speech not only helps infants acquire and disambiguate their native- and non-native-language articulatory and phonological information (Chawarska et al., 2022; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015; Schure et al., 2016) but that it also helps them discover and learn morpho-syntactic rules. If that is the case, then our results suggest that infants also rely on access to the articulatory information visible in a talker's mouth and lips to enhance their extraction of the temporal pattern information inherent in speech. In conclusion, when the published findings to date on the processing of AV speech are considered together with the current findings, they show clearly that AV - as opposed to A-only - speech facilitates speech processing in multiple ways during the initial acquisition of speech and language in infancy. This is in line with findings from a long line of studies with adults demonstrating that access to concurrent visual speech information facilitates speech processing and comprehension (Arnold & Hill, 2001; Savariaux et al., 2004; Sumby & Pollack, 1954; Summerfield, 1979). In the present case, it is reasonable to hypothesize that NAD learning is enhanced through access to the temporally synchronized and, thus, highly correlated dynamic visual motion cues located in a talker's mouth. Presumably, once infants gain access to the correlated auditory and visual dynamic cues representing a particular NAD, their learning is enhanced due to the greater perceptual salience of the audiovisually specified NAD cues. One way for future studies to test this hypothesis is by manipulating the A and V synchrony statistics and investigating the effects of such a manipulation on NAD learning.

The unique contribution of the current study is that an examination of the dynamic changes that occur in a selective attention task during AV speech processing can provide valuable and unique new insights into the relationship between the specific temporal profile of selective attention during the task and learning. Here, we have demonstrated that dynamic changes in selective attention during a grammatical rule-learning task can provide novel insights into the relationship between selective attention to a talker's mouth and the development of speech and language in preverbal infants.

DATA AVAILABILITY STATEMENT

The data, code and stimuli are publicly available at: https://osf.io/pcuw9/?view_only=de33faf1f0ee-402bbbbc15b532392f4b.

ORCID

Joan Birulés https://orcid.org/0000-0001-9708-4922
Anna Martinez-Alvarez https://orcid.org/0000-0001-5551-8171
David J. Lewkowicz https://orcid.org/0000-0001-6685-3178
Ruth de Diego-Balaguer https://orcid.org/0000-0002-2357-5195
Ferran Pons https://orcid.org/0000-0001-5919-8590

ENDNOTES

[1] By the principle of marginality, a factor must be kept if the interaction is significant regardless of the main effect. Moreover, changing the order of the comparisons—maximal and drop1 versus forward—yielded the same results (i.e., keeping the maximal model).

[2] The proportion of looking to the mouth was computed by dividing infants' looking to the mouth by their looking at the face at each time point (coded 1 for mouth, 0 for non-mouth but face, and not including the looks outside of the face).

REFERENCES

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339–355. https://doi.org/10.1348/000712601162220

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2018). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science*, *22*(3), 1–11. https://doi.org/10.1111/desc.12755

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Chawarska, K., Lewkowicz, D., Feiner, H., Macari, S., & Vernetti, A. (2022). Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism. *Journal of Experimental Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13595

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The Language Exposure Assessment Tool: Quantifying exposure for bilingual research in young infants and children. *Journal of Speech, Language, and Hearing Research*, *25*, 1–15. https://doi.org/10.1044/2016

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, *7*, 1–15. https://doi.org/10.3389/fpsyg.2016.00044

Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., Pine, J. M., Rowland, C. F., & Monaghan, P. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology*, *120*, 101291. https://doi.org/10.1016/j.cogpsych.2020.101291

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *4*(4), 527-534. https://doi.org/10.1111/1467-9280.00476

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(918975534), 183–206. https://doi.org/10.1207/s15327078in0702_4

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). *Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze*. Developmental Science. e12825. https://doi.org/10.1111/desc.12825

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

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. https://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy*, *15*(1), 46–60. https://doi.org/10.1111/j.1532-7078.2009.00005.x

López-Barroso, D., Cucurell, D., Rodríguez-Fornells, A., & de Diego-Balaguer, R. (2016). Attentional effects on rule extraction and consolidation from speech. *Cognition*, *152*, 61–69. https://doi.org/10.1016/j.cognition.2016.03.016

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Heinlein, K. B.- (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*, *55*(8), 1–60. https://doi.org/10.1037/dev0000750

Mueller, J. L., Friederici, A. D., & Mañnel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(39), 15953–15958. https://doi.org/10.1073/pnas.1204319109

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

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. https://doi.org/10.1177/0956797614568320

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134. https://doi.org/10.1016/S0010-0277(98)00060-2

Savariaux, C., Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), 69–78. https://doi.org/10.1016/j.cognition.2004.01.006

Schure, S. T., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, *7*(APR), 1–11. https://doi.org/10.3389/fpsyg.2016.00525

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, *36*(4–5), 314–331. https://doi.org/10.1159/000259969

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. https://doi.org/10.1016/j.cognition.2008.05.009

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, *160*, 103–109. https://doi.org/10.1016/j.cognition.2017.01.002

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behaviour and affect at 6 Months: Predicting clinical outcomes and language development in typically developing infants and infants at-risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x.Gaze

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.