Report

Current Biology

Infants respond to alterations in their own voice months before they recognize themselves in the mirror

Highlights

- 5-month-old infants can compensate for upward shifts in their vocal pitch
- Compensation was specific to pitch and did not affect other vocal features
- Vocal self-monitoring emerges months before mirror self-recognition

Authors

Margherita Belia, Fannie Fabre, Camille Borel, Jean-Julien Aucouturier, Nathan Faivre, Louise Goupil

Correspondence

louise.goupil@univ-grenoble-alpes.fr

In brief

Belia et al. show that 5-month-old infants can detect and compensate for upward transpositions of the pitch of their voice, at an age where they show no evidence of recognizing their own face in a mirror. This reveals an earlier developing emergence of self-monitoring in the vocal domain.

Belia et al., 2026, Current Biology *36*, 1–7
December 7, 2026 © 2026 Elsevier Inc. All rights reserved, including those for text and data mining, AI training, and similar technologies.
https://doi.org/10.1016/j.cub.2026.09.089

CellPress

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

Current Biology

CellPress

Report

Infants respond to alterations in their own voice months before they recognize themselves in the mirror

Margherita Belia,[1] Fannie Fabre,[1] Camille Borel,[1] Jean-Julien Aucouturier,[2] Nathan Faivre,[1] and Louise Goupil[1,3,*]
[1]CNRS, Université Grenoble Alpes, LPNC, Grenoble 38000, France
[2]Université de Franche-Comté, SUPMICROTECH, CNRS, institut FEMTO-ST, 25000 Besançon, France
[3]Lead contact
*Correspondence: louise.goupil@univ-grenoble-alpes.fr
https://doi.org/10.1016/j.cub.2026.09.089

SUMMARY

Recognizing our body as our own grounds our perceptions, actions, and social interactions. Even newborns detect cross-modal links between self-related perceptions,[1] considered necessary for body ownership.[2] Yet, clear evidence for a sense of ownership only appears once children not simply detect but also correct mismatches between predicted and perceived self-related sensorimotor inputs. So far, this has mostly been measured using mirror self-recognition, which revealed that children typically do not start recognizing themselves before 18 months of age.[3–5] Infants rarely see their own face, however, and mirror self-recognition has been related to sensorimotor[6] and social experiences.[7] We therefore hypothesized that a sense of ownership over the voice might develop earlier, because most infants routinely perceive themselves vocalizing from birth.[8–10] This might allow them to build sensorimotor models supporting the identification of self-generated signals earlier in this modality. Related research with adults has used the altered auditory feedback paradigm to show that they respond to transpositions of the fundamental frequency (f_0) of their voice by restoring the alignment between predicted and perceived feedback.[11–14] Crucially, these compensatory changes depend on the voice being perceived as self-generated.[15–19] Using an adaptation of this paradigm with 5-month-old infants, we found that they can already perceive and correct upward transpositions of their f_0, at an age when they remain unable to recognize themselves in the mirror. This suggests that hearing ourselves routinely allows us to learn what we sound like months before we recognize what we look like, which might be especially crucial for speech acquisition.

RESULTS AND DISCUSSION

Caregiver-infant dyads engaged in proto-conversations while infants wore custom headphones with an integrated microphone, enabling real-time recording and playback of their voice into their ears (Figure 1A). After a baseline phase lasting until infants produced 10 vocalization bursts, an experimental phase began: infants' fundamental frequency (f_0) was digitally shifted upward (+100 cents, 23 infants) or downwards (−100 cents, 25 infants) during 10 more vocalization bursts, or it was not digitally transposed (24 infants, neutral group). After data collection, infants' vocalizations were extracted from the recordings and concatenated within each phase. The average f_0 and its standard deviation were then calculated for each phase in order to compute the main dependent variable, which was the standardized difference in produced f_0 between the experimental and the baseline phase (hereafter, Δf_0; Figure 1B).

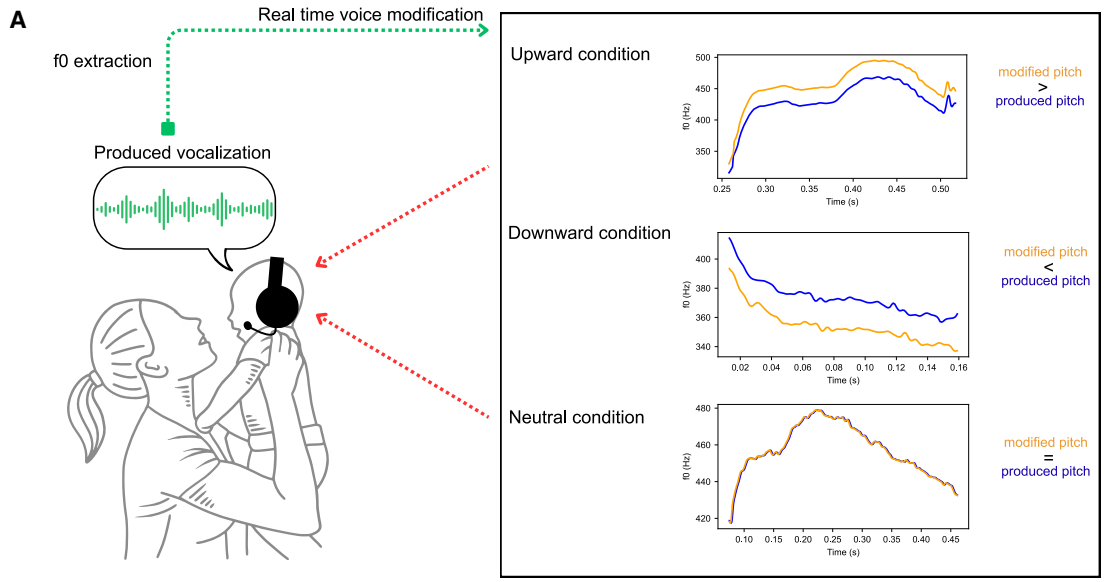
First, we conducted confirmatory analyses of the recordings, which confirmed that the f_0 transposition was successfully applied, with slight variations depending on vocalization types and harmonicity (e.g., for babbling vs. growls, see Figure S1).

Second, we quantified the effect of condition and age on Δf_0 in a linear regression model. Condition (upward vs. downward vs. neutral) was included to assess the effect of the f_0 transposition on vocal production, while age (in months) and the interaction between age and condition was included to assess potential developmental effects. As predicted, adding condition improved the model fit ($F(2) = 4.37$, $p = 0.02$). Δf_0 in the upward condition was significantly lower than Δf_0 in the neutral ($\beta = -0.28$, SE = 0.12, $t = -2.41$, $p = 0.02$, $d = 0.68$) and the downward condition ($\beta = -0.31$, SE = 0.12, $t = -2.7$, $p = 0.008$, $d = 1.08$; Figure 2A). However, Δf_0 in the downward condition was not significantly different from the neutral one ($\beta = -0.03$, SE = 0.11, $t = 0.27$, $p = 0.78$). Adding age did not improve the model fit ($F(1) = 1.77$, $p = 0.18$) nor did the age by condition interaction ($F(2) = 0.66$, $p = 0.52$; the same result was obtained with categorical age, splitting infants into two age groups, above or below 5 months; all p values > 0.2).

Third, we examined compensation in each condition separately (Figure 2A). As predicted, infants in the upward condition lowered their f_0 following the transposition (M = −0.14, SD = 0.19, one-sample t test against zero: $t(22) = -3.27$,

All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089



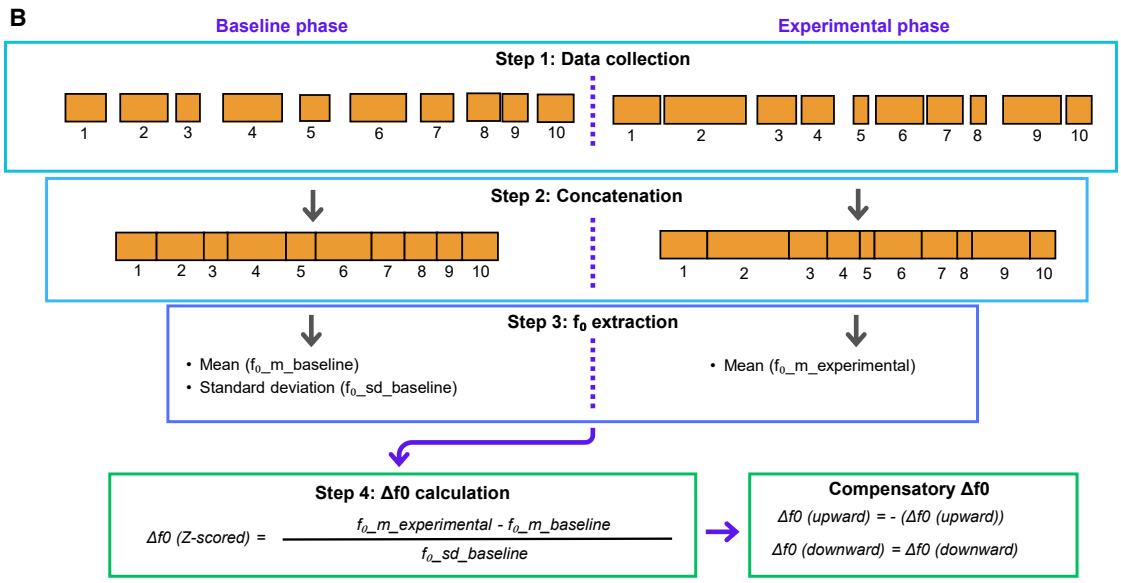


Figure 1. Experimental paradigm and Δf_0 extraction

(A) In the experimental phase, the f_0 of infants' vocalizations (in blue) was manipulated using the DAVID software. The f_0 was shifted upward or downward (in yellow), depending on condition assignment, and the modified voice was played back to the infant in real time.

(B) Analytical approach. Step 1: individual vocalization bursts (orange blocks) were identified from the recordings; Step 2: vocalizations were concatenated within each phase; Step 3: the mean (f_0_m_baseline) and standard deviation (f_0_sd_baseline) of the f_0 across all infant vocalizations in the baseline phase were computed. The same procedure was applied to vocalizations from the experimental phase to obtain the mean f_0 (f_0_m_experimental). Step 4: a standardized Δf_0 for each infant was calculated as the *Z*-scored difference between the experimental and baseline means (Δf_0). A compensatory Δf_0 was also computed by flipping the sign of Δf_0 of infants in the upward condition so it could be directly compared to the downward condition (with a positive value indicating a compensatory response).

See also Figure S1 and Table S1.

p = 0.003, d = 0.7), while infants in the downward condition increased their f_0 (M = 0.18, SD = 0.37, t(24) = 2.34, p = 0.027, d = 0.48). By contrast, no significant change in Δf_0 was observed in the neutral group (M = 0.14, SD = 0.54, t(23) = 1.3, p = 0.21, d = 0.27). To further confirm that these findings could not be attributed to chance variations in f_0 in the two experimental groups, we repeated this analysis after randomly permuting condition assignment across participants, disrupting the link between the f_0 manipulation and the corresponding vocal response. No differences between the experimental groups were found in this case (t(46) = 0.66, p = 0.51), and Δf_0 was not different from chance in either permuted experimental group (upward: M = 0.04, SD = 0.36, t(25) = 0.6, p = 0.6; downward: M = −0.03, SD = 0.36,

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

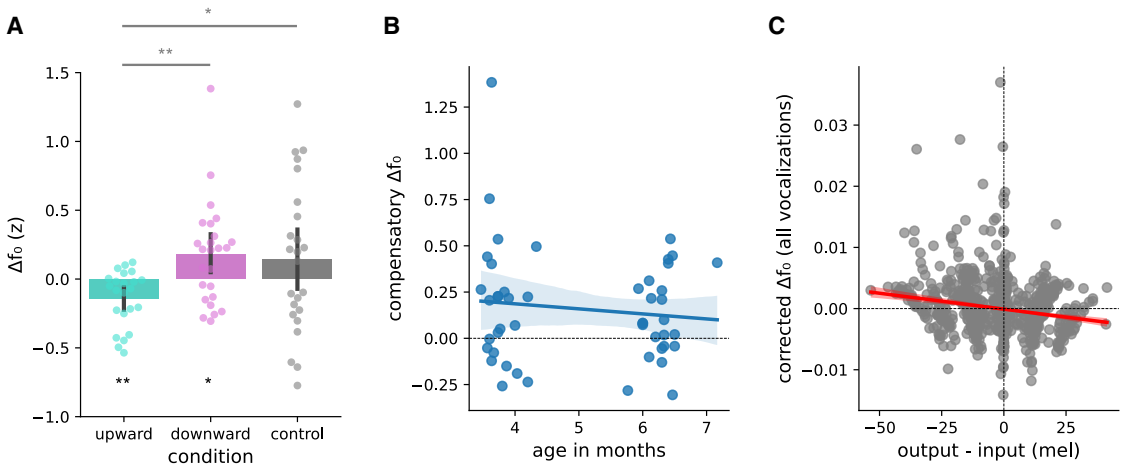


Figure 2. Main results

(A) Bar plots showing the Δf_0 (*Z* scored) separately for the upward (cyan), downward (magenta), and neutral condition (gray). Dots represent individual data, and the black lines show the 95% CI. Gray asterisks show inter-individual comparisons, and black asterisks show comparisons against zero (* $p < 0.05$; ** $p < 0.01$). See Figure S2 for control analyses on other vocal features, and Figure S3A for an analysis of the magnitude of the effect.

(B) Infants' compensatory Δf_0 as a function of age. Positive values indicate that infants changed their f_0 in the compensatory direction during the experimental phase compared to the baseline. See Figure S3B for an analysis of the distribution of this effect. The shaded area shows the 95% CIs.

(C) The *y* axis shows Δf_0 values for all vocalizations, corrected by the unmodified value of the f_0 (in Mel), as a function of the size of the transformation (*x* axis) computed as the difference between the modified f_0 (i.e., output of the software) minus the unmodified f_0 (i.e., input of the software, both in Mel). The shaded area shows the 95% CIs.

$t(21) = -0.34$, $p = 0.74$), which confirms that the compensatory effects cannot be attributed to chance variations of f_0 in these groups.

Next, we tested the size of the compensatory effect by pooling experimental conditions together, flipping the upward condition to make it directly compared to the downward condition, as they were expected to elicit opposite compensatory responses ("compensatory Δf_0" hereafter, where positive values mean that the f_0 changed in the compensatory direction; Figure 1B). Compensatory Δf_0 was highly significantly different from zero ($M = 0.16$, $SD = 0.3$, $t(47) = 3.61$, $p < 0.001$, $d = 0.53$), confirming a medium-size compensatory effect at the group level. We then included this variable in a linear regression to assess the effect of age (Figure 2B). Again, there was no evidence for a developmental effect, be it considered as a continuous or a categorical variable (all *p* values > 0.4).

Individually, 33 out of the 48 participants in the two experimental groups modulated their f_0 in the expected compensatory direction (e.g., lowering their pitch when their f_0 was raised), and 15 modulated their f_0 in the same direction as the manipulation (e.g., rising their pitch when their f_0 was raised). Therefore, infants overall displayed the typical compensatory response reported in adult studies (i.e., the compensatory response was statistically more frequent than what would be expected by chance: $\chi^2 = 6.75$, $p = 0.009$; $\varphi = 0.375$), with some inter-individual variability. By contrast, infants in the neutral group showed no evidence of modulating their f_0 in any direction systematically ($\chi^2 = 0.17$, $p = 0.68$; $\varphi = 0.32$).

As infants produced sounds that varied in voicing (Table S1), the f_0 modification was more accurately applied (and thus more salient) on sounds with larger voiced segments, whereas unvoiced or in-harmonic sounds could not always be transposed throughout, reducing the possibility to perceive the f_0 shift for these sounds (Figures S1B–S1H). Thus, we measured whether infants' f_0 changes scaled with the actual f_0 transposition, predicting a negative relationship between the size of the f_0 transposition for a given vocalization (output-input) and Δf_0 for that same vocalization (Figure 2C). After controlling for the untransformed f_0 and including infant as a random intercept, output-input improved the model fit ($\chi^2(1) = 4.13$, $p = 0.04$) and, as expected if infants were compensating in general, the size of the transposition was negatively related to Δf_0 ($\beta = -0.009$, $SE = 0.004$, $t = -2.03$, $p = 0.04$). This shows that infants' produced f_0 changed linearly with the magnitude of the effective transposition: the higher the infant perceived their f_0 to be, the more they lowered their f_0, and vice versa. Again, this is compatible with a compensatory effect at the group level. In addition, we observed that this relationship held for different vocalization types (speech like: vocants, babbling, and pre-babbling vs. non-speech like: squeals, grunts, and laughs), as there was no interaction between vocalization type and transformation size ($p > 0.44$).

Notably, we found no change in f_0 variability, amplitude (root mean square), voice quality (harmonic-to-noise ratio), or vocalization types across phases and conditions (all $p > 0.7$), consistent with a targeted change of the f_0 (Figures S2A–S2C and S2F). Additionally, caregivers' f_0 did not differ across phases or conditions, and caregiver Δf_0 did not relate to infant Δf_0 (Figure S2D).

Finally, thirty-five infants among the 48 who participated in the altered auditory feedback (AAF) also participated in the mirror test. As expected,[3,4] none of them recognized themselves in the mirror: no infant attempted to touch the mirror or their forehead after seeing the mark, and there were no differences in their looking times to the mirror before and after the mark was introduced ($t(34) = -1.27$, $p = 0.21$).

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

Together, these analyses indicate that infants in the experimental conditions tended to alter their f_0 in the opposite direction of the manipulation more often than would be expected by chance, consistent with our prediction that they should detect and correct f_0 transpositions and with previous child[20,21] and adult[12,14] research. This effect disappeared when condition labels were randomly permuted, and infants in the neutral group showed no significant changes in f_0, suggesting that the observed changes depended specifically on the auditory feedback manipulation.

There was a clear asymmetry, however: evidence for compensation was strong for the upward transposition but equivocal for the downward transposition. In both experimental conditions, at the group level, infants significantly changed their f_0 in the compensatory direction following the transposition compared to the baseline period. However, the extent to which they changed their f_0 following the transposition only differed significantly from what was observed in the neutral condition for the upward condition. The fact that infants significantly lowered their f_0 following the upward transposition is in line with our hypothesis that they can monitor and correct alterations to the f_0 of their own voice. The lack of a significant difference between the neutral and downward conditions was unexpected, and it is therefore difficult to interpret at this point. It might be due to a weaker magnitude of the effect for downward transpositions, as found in older children.[21] Consistent with this hypothesis, the magnitude of f_0 changes was smaller in the experimental conditions compared to the neutral condition, suggesting that the f_0 changes that infants produce spontaneously exceed those required to compensate for the relatively small experimental transposition we introduced (Figure S3A). The results are also compatible with the possibility that the magnitude of the downward transposition (100 cent) was too small for infants to consistently perceive it, and future work could investigate this by increasing the magnitude of the transposition. Finally, it is also possible that infants generally tended to raise their f_0 over the course of the study (e.g., due to increased arousal), making the neutral and downward conditions more similar. This could also be investigated in the future by increasing the magnitude of the downward transposition.

Nevertheless, Figure 2C shows that there is a clear linear relationship between the transposition effectively applied to the infants' voice and their produced f_0, showing a compensatory effect when vocalizations are transposed as intended. The significant group-level effect was accompanied by substantial inter-individual variability (Figures 2B and 2C). This resonates with adult findings: although compensation is typical, fewer responses can be "following"[22–24] ones, especially when transpositions become more predictable.[25] Whether we observed genuine following responses here (which should result in a bimodal distribution of compensatory Δf_0) or the lower tail of a unimodal distribution centered above zero, remains unclear. The latter seems statistically more plausible (Figure S3B), but future studies are needed to further examine these inter-individual differences, and whether they relate to or even predict the development of other abilities such as the onset of canonical babbling.

The present findings show that infants, like adults and children, can monitor and correct features of their own voice based on auditory feedback, with similar compensation magnitudes as adults.[26] Prior developmental work showed that 4-year-olds, but not 2-year-olds, compensated for manipulated vowel formants (F1/F2),[20] whereas even 2-year-olds compensated for f_0 alterations.[21] At this age, children already speak however, so this study left open the possibility that f_0 self-monitoring might be a by-product, rather than a prerequisite, of speech development. The f_0 is primarily modulated using the larynx, which infants start controlling early on, while formant control requires later emerging and finer coordination of supraglottal articulators.[27] Infants may therefore develop reliable audio-motor mappings for f_0 before formants, allowing them to engage in vocal self-monitoring earlier for this simpler dimension. Accordingly, age did not predict f_0 compensation, suggesting that even the youngest infants we tested had sufficient vocal experience for predictive control over their f_0. This shows that at least some aspects of vocal self-monitoring are present before infants' perception becomes finely attuned to native language phonemes,[28] before they can articulate the sounds of their native language[29] or use their voice independently from underlying states of arousal.[30] Although it cannot currently be excluded that predictive sensorimotor models have an innate basis, converging evidence[20,21] suggest that vocal experience is required to build them, much as experience shapes visual self-recognition,[6,7] calling for future research to assess this link more directly.

Our findings are compatible with the hypothesis that vocal self-monitoring is central to oral communicative development. First, it provides a mechanism through which infants form and refine their audio-motor maps.[31,32] As previously argued, this self-generated loop (producing a sound, experiencing its sensory outcomes, comparing it with internal predictions, and adjusting mismatches) could support the production of increasingly complex sounds, such as supraglottal consonants and adult-like syllables underpinning early word production.[33,34] But before this, we suggest that vocal self-monitoring could also underlie infants' discovery of social contingencies in the communicative space, allowing them to detect when their vocalizations elicit a response. This ability is thought to be key for understanding that your own behaviors affect those of others and might thus scaffold the emergence of intentional communication.[35,36]

The current findings also have implications for our understanding of the development of the self, and especially, of body ownership. Indeed, research using the AAF with adult speakers suggests that compensation depends on the voice being perceived as self-generated, as measured via subjective reports of vocal ownership.[15–17] We also know that disturbances in vocal self-recognition and dysfunctions in sensorimotor forward models[37–39] co-occur with auditory-verbal hallucinations.[40,41] Thus, the compensatory adjustments we observed provide indirect evidence that infants can recognize their vocalizations as self-produced, and invites us to envision a more multimodal notion of the developing self.

By focusing on a modality in which infants have extensive opportunities to build predictive sensorimotor models early on, our work opens new venues to further investigate the emergence of self-awareness. For instance, how do sensing the characteristics and limits of one's own body (the sense of "body ownership," e.g., having a sense of what our voice sounds like) and that our actions have consequences (the "sense of agency," e.g., registering that what we do with our voice has specific

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

consequences in the world) interact during development to enable a more integrated sense of self? Our findings also raise the question of why facial recognition is so protracted compared to vocal recognition. The poverty (vs. richness) of the corresponding sensory input might be key: with limited access to their visual image, infants might need to build an abstract representation of themselves to succeed in the mirror task. A protracted passing of this task compared to other self-recognition tasks might therefore indicate that it measures more conceptual aspects of the self,[42] consistent with its association with personal pronouns use,[43] default-mode network activity,[44] and self-memory biases.[45]

To conclude, our findings invite us to embrace a more pluralistic vision of the development of self-awareness that considers how multisensory inputs allow us to build an integrated sense of self. They reveal that young infants can already monitor and modify their own behavior to restore parts of their bodily self (their voice),[41,46] revealing an early-developing, multimodal, predictive representation of the self, grounded in everyday sensorimotor experience.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Louise Goupil (louise.goupil@univ-grenoble-alpes.fr).

Materials availability

The study did not produce new materials.

Data and code availability

- Anonymized data have been deposited in GitHub and are publicly available as of the date of publication at; https://github.com/marghebelia/vocal_self_recognition.
- All code has been deposited in GitHub and is publicly available as of the date of publication at; https://github.com/marghebelia/vocal_self_recognition.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all the families and infants who took part in our study. We are also grateful to the maternity ward of the Couple Enfant Hospital (La Tronche) for facilitating access to families and allowing us to present our research and recruit participants. We also thank Cassandra Miorin and Lucie Vernay, who contributed to data coding and annotation. This study was funded by the Agence National de la Recherche (ANR-RESSAC-23-CE28-0024-01 to L.G.).

AUTHOR CONTRIBUTIONS

Conceptualization, L.G.; methodology, L.G. and J.-J.A.; investigation, M.B., F.F., C.B., and L.G.; visualization, M.B., F.F., and L.G.; funding acquisition, L.G.; project administration, F.F. and L.G.; supervision, L.G.; writing – original draft, M.B.; and writing – review and editing, M.B., L.G., J.-J.A., N.F., and F.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - AAF
  - Mirror self-recognition test
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Software
  - Measures and pre-processing
  - Statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2026.09.089.

Received: April 20, 2026
Revised: August 7, 2026
Accepted: September 30, 2026

REFERENCES

1. Filippetti, M.L., Johnson, M.H., Lloyd-Fox, S., Dragovic, D., and Farroni, T. (2013). Body perception in newborns. Curr. Biol. *23*, 2413–2416. https://doi.org/10.1016/j.cub.2013.10.017.
2. Bahrick, L.E., and Watson, J.S. (1985). Detection of intermodal proprioceptive–visual contingency as a potential basis of self-perception in infancy. Dev. Psychol. *21*, 963–973. https://doi.org/10.1037/0012-1649.21.6.963.
3. Amsterdam, B. (1972). Mirror self-image reactions before age two. Dev. Psychobiol. *5*, 297–305. https://doi.org/10.1002/dev.420050403.
4. Southgate, V. (2024). The origins and emergence of self-representation. Annu. Rev. Dev. Psychol. *6*, 109–131. https://doi.org/10.1146/annurev-devpsych-120621-025747.
5. Suddendorf, T., and Butler, D.L. (2013). The nature of visual self-recognition. Trends Cogn. Sci. *17*, 121–127. https://doi.org/10.1016/j.tics.2013.01.004.
6. Chinn, L.K., Noonan, C.F., Patton, K.S., and Lockman, J.J. (2024). Tactile localization promotes infant self-recognition in the mirror-mark test. Curr. Biol. *34*, 1370–1375.e2. https://doi.org/10.1016/j.cub.2024.02.028.
7. Cebioğlu, S., and Broesch, T. (2021). Explaining cross-cultural variation in mirror self-recognition: New insights into the ontogeny of objective self-awareness. Dev. Psychol. *57*, 625–638. https://doi.org/10.1037/dev0001171.
8. Westermann, G., and Reck Miranda, E.R. (2004). A new model of sensorimotor coupling in the development of speech. Brain Lang. *89*, 393–400. https://doi.org/10.1016/S0093-934X(03)00345-6.
9. Guenther, F.H., and Vladusich, T. (2012). A neural theory of speech acquisition and production. J. Neurolinguist. *25*, 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006.
10. Vihman, M.M., DePaolis, R.A., and Keren-Portnoy, T. (2009). A dynamic systems approach to babbling and words. In The Cambridge Handbook of Child Language, E.L. Bavin, ed. (Cambridge University Press), pp. 163–182.
11. Houde, J.F., and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. Science *279*, 1213–1216. https://doi.org/10.1126/science.279.5354.1213.
12. Burnett, T.A., Senner, J.E., and Larson, C.R. (1997). Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. J. Voice *11*, 202–211. https://doi.org/10.1016/S0892-1997(97)80079-3.
13. Coughler, C., Quinn de Launay, K.L., Purcell, D.W., Oram Cardy, J., and Beal, D.S. (2022). Pediatric responses to fundamental and formant frequency altered auditory feedback: A scoping review. Front. Hum. Neurosci. *16*, 858863. https://doi.org/10.3389/fnhum.2022.858863.

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

14. Jones, J.A., and Munhall, K.G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. J. Acoust. Soc. Am. *108*, 1246–1251. https://doi.org/10.1121/1.1288414.

15. Parkinson, A.L., Korzyukov, O., Larson, C.R., Litvak, V., and Robin, D.A. (2013). Modulation of effective connectivity during vocalization with perturbed auditory feedback. Neuropsychologia *51*, 1471–1480. https://doi.org/10.1016/j.neuropsychologia.2013.05.002.

16. Subramaniam, K., Kothare, H., Mizuiri, D., Nagarajan, S.S., and Houde, J.F. (2018). Reality monitoring and feedback control of speech production are related through self-agency. Front. Hum. Neurosci. *12*, 82. https://doi.org/10.3389/fnhum.2018.00082.

17. Goupil, L., Johansson, P., Hall, L., and Aucouturier, J.-J. (2021). Vocal signals only impact speakers' own emotions when they are self-attributed. Conscious. Cogn. *88*, 103072. https://doi.org/10.1016/j.concog.2020.103072.

18. Moseley, P., Fernyhough, C., and Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. Neurosci. Biobehav. Rev. *37*, 2794–2805. https://doi.org/10.1016/j.neubiorev.2013.10.001.

19. Johns, L.C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., and McGuire, P.K. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. Psychol. Med. *31*, 705–715. https://doi.org/10.1017/S0033291701003774.

20. MacDonald, E.N., Johnson, E.K., Forsythe, J., Plante, P., and Munhall, K.G. (2012). Children's Development of Self-Regulation in Speech Production. Curr. Biol. *22*, 113–117. https://doi.org/10.1016/J.CUB.2011.11.052.

21. Scheerer, N.E., Jacobson, D.S., and Jones, J.A. (2020). Sensorimotor control of vocal production in early childhood. J. Exp. Psychol. Gen. *149*, 1071–1077. https://doi.org/10.1037/xge0000706.

22. Burnett, T.A., Freedland, M.B., Larson, C.R., and Hain, T.C. (1998). Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. *103*, 3153–3161. https://doi.org/10.1121/1.423073.

23. Franken, M.K., Acheson, D.J., McQueen, J.M., Hagoort, P., and Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. Psychon. Bull. Rev. *25*, 1458–1467. https://doi.org/10.3758/s13423-018-1494-x.

24. Larson, C.R., Sun, J., and Hain, T.C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. J. Acoust. Soc. Am. *121*, 2862–2872. https://doi.org/10.1121/1.2715657.

25. Behroozmand, R., Korzyukov, O., Sattler, L., and Larson, C.R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. J. Acoust. Soc. Am. *132*, 2468–2477. https://doi.org/10.1121/1.4746984.

26. Kort, N.S., Nagarajan, S.S., and Houde, J.F. (2014). A bilateral cortical network responds to pitch perturbations in speech feedback. Neuroimage *86*, 525–535. https://doi.org/10.1016/j.neuroimage.2013.09.042.

27. Wermke, K., Robb, M.P., and Schluter, P.J. (2021). Melody complexity of infants' cry and non-cry vocalisations increases across the first six months. Sci. Rep. *11*, 4137. https://doi.org/10.1038/s41598-021-83564-8.

28. Werker, J.F., and Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav. Dev. *7*, 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3.

29. Vihman, M.M. (2014). Phonological Development: the First Two Years (Wiley-Blackwell). https://doi.org/10.1002/9781394260997.

30. Wass, S., Phillips, E., Smith, C., Fatimehin, E.O.O.B., and Goupil, L. (2022). Vocal communication is tied to interpersonal arousal coupling in caregiver-infant dyads. eLife *11*, e77399. https://doi.org/10.7554/eLife.77399.

31. Guenther, F.H. (2016). Neural Control of Speech (MIT Press). https://doi.org/10.7551/mitpress/10471.001.0001.

32. Messum, P., and Howard, I.S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. J. Phon. *53*, 125–140. https://doi.org/10.1016/j.wocn.2015.08.005.

33. McCune, L., and Vihman, M.M. (2001). Early phonetic and lexical development: A productivity approach. J. Speech Lang. Hear. Res. *44*, 670–684. https://doi.org/10.1044/1092-4388(2001/054).

34. Vihman, M. (2015). Perception and production in phonological development. In The Handbook of Language Emergence, pp. 437–457.

35. Ghazanfar, A.A., and Zhang, Y.S. (2016). The autonomic nervous system is the engine for vocal development through social feedback. In Curr. Opin. Neurobiol., *40* (Elsevier Limited), pp. 155–160. https://doi.org/10.1016/j.conb.2016.07.016.

36. Wass, S.V., Phillips, E.A.M., Haresign, I.M., Amadó, M.P., and Goupil, L. (2024). Contingency and Synchrony: Interactional Pathways Toward Attentional Control and Intentional Communication. Annu. Rev. Dev. Psychol. *6*, 63–85. https://doi.org/10.1146/annurev-devpsych-010923-110459.

37. Frith, C.D. (1992). The Cognitive Neuropsychology of Schizophrenia (Lawrence Erlbaum Associates, Inc).

38. Ford, J.M., and Mathalon, D.H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? Int. J. Psychophysiol. *58*, 179–189. https://doi.org/10.1016/j.ijpsycho.2005.01.014.

39. Frith, C.D., Blakemore, S.J., and Wolpert, D.M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res. Brain Res. Rev. *31*, 357–363. https://doi.org/10.1016/S0165-0173(99)00052-1.

40. Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a reappraisal. Exp. Brain Res. *192*, 527–532. https://doi.org/10.1007/s00221-008-1533-3.

41. Vukojević, J., Jelić, L., McCormack, K., Sušac, J., Muselimović, I., Bagarić, M., Brečić, P., Dellwo, V., Cifrek, M., Savić, A., et al. (2025). Self-other voice confusion in patients with auditory-verbal hallucinations and nonclinical hallucination proneness. Preprint at medRxiv. https://doi.org/10.1101/2025.10.06.25337403.

42. Bremner, A.J., and Southgate, V. (2025). The development of self-awareness in infancy. In Perspectives on Early Parenthood and Infancy: a Psychoanalytic, Neuroscientific, Developmental and Cultural Dialogue (Taylor and Francis), pp. 215–221. https://doi.org/10.4324/9781003530473-34.

43. Lewis, M., and Ramsay, D. (2004). Development of self-recognition, personal pronoun use, and pretend play during the 2nd year. Child Dev. *75*, 1821–1831. https://doi.org/10.1111/j.1467-8624.2004.00819.x.

44. Bulgarelli, C., Blasi, A., de Klerk, C.C.J.M., Richards, J.E., Hamilton, A., and Southgate, V. (2019). Fronto-temporoparietal connectivity and self-awareness in 18-month-olds: A resting state fNIRS study. Dev. Cogn. Neurosci. *38*, 100676. https://doi.org/10.1016/j.dcn.2019.100676.

45. Grosse Wiesmann, C., Rothmaler, K., Hasan, E., Habdank, K., Yang, C., Yeung, E., and Southgate, V. (2025). The self-reference memory bias is preceded by an other-reference bias in infancy. Nat. Commun. *16*, 6311. https://doi.org/10.1038/s41467-025-61642-z.

46. Orepic, P., Kannape, O.A., Faivre, N., and Blanke, O. (2023). Bone conduction facilitates self-other voice discrimination. R. Soc. Open Sci. *10*, 221561. https://doi.org/10.1098/rsos.221561.

47. Rachman, L., Liuni, M., Arias, P., Lind, A., Johansson, P., Hall, L., Richardson, D., Watanabe, K., Dubal, S., and Aucouturier, J.J. (2018). DAVID: An open-source platform for real-time transformation of infra-segmental emotional cues in running speech. Behav. Res. Methods *50*, 323–343. https://doi.org/10.3758/s13428-017-0873-y.

48. Cannon, E.N., Simpson, E.A., Fox, N.A., Vanderwert, R.E., Woodward, A.L., and Ferrari, P.F. (2016). Relations between infants' emerging reach-grasp competence and event-related desynchronization in EEG. Dev. Sci. *19*, 50–62. https://doi.org/10.1111/desc.12295.

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

49. Friedrich, M., Mölle, M., Friederici, A.D., and Born, J. (2020). Sleep-dependent memory consolidation in infants protects new episodic memories from existing semantic memories. Nat. Commun. *11*, 1298. https://doi.org/10.1038/s41467-020-14850-8.

50. Buder, E.H., Warlaumont, A.S., Oller, D.K., Peter, B., and MacLeod, A. (2013). An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective. In Comprehensive Perspectives on Child Speech Development and Disorders: Pathways from Linguistic Theory to Clinical Practice, *4*. 103–134. https://digitalcommons.memphis.edu/facpubs/15524.

51. Boersma, P., and Weenink, D. (2019). Praat: doing phonetics by computer. [computer program: Version 6.0.56. http://www.praat.org].

52. Junge, C., Everaert, E., Porto, L., Fikkert, P., de Klerk, M., Keij, B., and Benders, T. (2020). Contrasting behavioral looking procedures: a case study on infant speech segmentation. Infant Behav. Dev. *60*, 101448. https://doi.org/10.1016/j.infbeh.2020.101448.

53. Rose, S.A., Gottfried, A.W., Melloy-Carminar, P., and Bridger, W.H. (1982). Familiarity and Novelty Preferences in Infant Recognition Memory: Implications for Information Processing. Dev. Psychol. *18*, 704–713. https://doi.org/10.1037/0012-1649.18.5.704.

54. Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P., and Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. J. Open Source Softw. *6*, 3139. https://doi.org/10.21105/joss.03139.

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	https://github.com/marghebelia/vocal_self_recognition
Software and algorithms		
DAVID	Rachman et al.[47]	https://doi.org/10.3758/s13428-017-0873-y
Custom code for coding infant gaze direction	This paper	https://github.com/marghebelia/vocal_self_recognition

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

A total of 48 infants (19 females; M_{age} = 5.03 months, age range = 3.47-7.17, SD_{age} = 1.27) were included in the two experimental groups: 23 infants had their f_0 digitally shifted upwards by +100 cents (upward group) during the experimental phase, while 25 had their f_0 shifted downwards by -100 cents (downward group; see the "method details" section for full information on the experimental procedure). A separate group of infants (N = 24, 15 females; M_{age} = 5 months, age range = 3.47-6.6, SD_{age} = 0.18) underwent the exact same procedure but we did not modify their voice during the experimental phase (neutral group). Thus, a total of 72 infants were included in the final analyses. No formal a priori sample size calculation was performed, the target sample size was decided a priori based on published work with adults.[17]

In addition, 53 infants took part in the study but could not be included due to: child fussiness (n = 24); child not vocalizing during the whole duration of the experiment (n = 13); child trying to remove the headset (n = 13); child sucking on a dummy (n = 1); child's tiredness (n = 2). This attrition rate is consistent with that reported in much of the infant literature involving placing body equipment.[20,48,49] It also highlights the challenges of adapting the AAF paradigm to infants, that made it strongly depart from existing AAF literature. Unlike adults and children, who can be instructed to produce specific sounds and are more likely to tolerate recording equipment, infants should be disposed to both wear an unfamiliar headset and vocalize spontaneously at the same time to provide sufficient data, which explains our relatively high attrition rate. Technical error led to the loss of data for an additional 28 infants (either because of software failure, or because the headphones did not appropriately cover the infants' ears).

Participants were recruited as part of a larger set of studies run at our laboratory, where 3.5-4-month-olds and 6-6.5-month-olds were recruited separately. Among the 72 infants included in the final sample, 36 were younger infants (downward: N = 12, M_{age} = 3.78 months, SD_{age} = 0.25; upward: N = 12, M_{age} = 3.79, SD_{age} = 0.21; neutral: N = 12, M_{age} = 3.77, SD_{age} = 0.18) and 36 were older infants (downward: N = 13, M_{age} = 6.26 months, SD_{age} = 0.35; upward: N = 11, M_{age} = 6.31, SD_{age} = 0.15; neutral: N = 12, M_{age} = 6.24, SD_{age} = 0.18). All infants had no known hearing, vision or developmental problems. Information on ancestry, race, ethnicity, and socioeconomic background could not be collected under the terms of our ethical approval, limiting our ability to assess the generalizability of the findings across demographic groups. Participants were recruited from the maternity ward of the Couple Enfant Hospital in Grenoble, France. Infants also participated in other unrelated studies ongoing in the laboratory, and the order in which they took part in other studies was counterbalanced across participants. All parents provided written informed consent prior to their participation and received a book and a certificate as a thank you at the end of their visit. The study was approved by the Ethics Committee of Université Grenoble Alpes (CERGA).

METHOD DETAILS

Infants were randomly assigned to the upward (N = 23), downward (N = 25) and neutral (N = 24) conditions. Two paradigms were used: an altered auditory feedback (AAF) paradigm[12,17] and the mirror task.[3]

AAF

Caregiver-infant dyads freely played on the floor with a set of books and toys. Caregivers were invited to interact spontaneously with their infants and to engage in any routines or games that would typically encourage their child to vocalize. To avoid overlapping vocalizations, caregivers were asked to try and respond to their infant's vocalizations by taking turns. During the session, infants wore DT-770 headphones and DPA-4088 microphones, adapted with an elastic band with velcro to fit their head circumference. This setup allowed us to record the infant's voice and play it back to them simultaneously in real time. The volume was adjusted by the experimenters so as to be loud enough to cover the sound of the voice perceptible via bone conduction.

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

After an initial baseline period, the fundamental frequency (f_0) of the infant's voice was digitally manipulated using DAVID, an open-source digital voice manipulation software that allows real-time voice modification while maintaining a low latency (<20ms), including the possibility to shift the f_0 using a time-varying delay chain.[47] For infants in the upward condition, the f_0 was shifted upwards by 100 cents; for those in the downward condition, the f_0 was shifted downwards by 100 cents. The baseline period continued until the infant produced 10 vocalizations, following.[20] The beginning of the f_0 modification marked the beginning of the experimental phase, which directly followed the baseline period without interrupting the dyadic interaction. The f_0 modification continued either until the infant produced 10 additional vocalizations or until a maximum of 10 minutes had elapsed from the beginning of the baseline phase.

Infants assigned to the neutral condition underwent the exact same procedure, including wearing the same headset and microphone system, and hearing their voice played back in real time through the DAVID software, using the exact same signal processing line. However, their voice was played back to them without being transposed throughout the whole session. To this end, we created three settings in DAVID: a "neutral" setting (which played back the child's voice at their normal f_0) and two "modified" settings (which played back the voice at either the increased or the decreased f_0). To prevent f_0 modifications to the caregiver's voice, experimenters manually switched back to "neutral" whenever the caregiver spoke and reverted to the appropriate "modified" setting whenever the infant vocalized. Although this approach minimized the modification of the caregiver's voice, we conducted a control analysis to check for any effects of the modification of the caregiver's voice (See "statistical analyses" section and Figure S2D). A "neutral" DAVID setting was created for the neutral group despite the fact that no transposition occurred in order to mimic the same switching gesture for the experimenter as in the experimental groups, ensuring an exact consistency in the procedure used across all three conditions.

The dyadic interactions were filmed with two cameras (Sony HDR-CX240) to capture infants' gaze directions and facial expressions.

Mirror self-recognition test

A subset of infants (35/48) took part in the mirror self-recognition test. We conducted this test only with the infants in the experimental groups (upward and downward), as the neutral group only served as a control for the AAF paradigm. This was done to verify that infants in our sample did not recognize themselves in the mirror, as predicted from the literature.

In the original version of this emblematic task, a mark is placed on a child's face; if the child touches the mark while looking at their reflection, it is considered evidence that they recognize themselves.[3,4] To adapt this paradigm to pre-verbal infants, we included the following steps: 1) the infant was shown their reflection in a mirror, while the experimenter pointed to the child's reflection and made comments like "Look!" and "Who's that?". This phase lasted until the child looked away from the mirror; 2) a red mark was placed on the infant's forehead, followed by a short play time with the experimenter. This was done to prevent any reactions to the mark in the following steps from being attributed to recognizing a change in a recently shown image, which would not necessarily indicate self-recognition but rather a response to a change in stimuli, as well as to disconnect temporally the touching gesture used to place the mark from their vision of the mark in the mirror; 3) the infant was then shown the mirror again, with the red mark still on their forehead. The experimenter again drew the infant's attention to their reflection by pointing and using attention-getting phrases similar to those used in the first step. This phase lasted until the child looked at the mirror at least three times.

QUANTIFICATION AND STATISTICAL ANALYSIS

Software

Linear and mixed effects models were run in R. Codes for data visualization and the rest of the analyses (e.g., t-test) were conducted in Python.

Measures and pre-processing

AAF

DAVID recorded an audio track with the modified pitch (output of the software), as well as an audio track with the unmodified pitch (input of the software). Using Audacity, an audio editing software, the two were aligned (to compensate for the delay introduced by DAVID's signal processing) and played simultaneously until the onset of the f_0 modification was clearly identified. This timing was noted down and used to infer which vocalizations were produced before and after the f_0 modification. All infant vocalizations produced before the f_0 modification were considered to be produced in the baseline phase (BL), whereas those produced after the alteration of f_0 were considered to be in the experimental phase (EXP).

The audio files were then manually segmented in Audacity to identify the exact start and end times of each vocalization. Two contiguous vocalizations were counted as distinct if they were separated by at least 500 ms of silence or, if separated by less than 500 ms, when they belonged to different vocalization categories; otherwise, vocalizations of the same category separated by less than 500 ms were considered as a single vocalization. Each segment was then manually coded for several features. First, we noted whether the child or their parent was speaking, and whether the segment contained a co-vocalization or turn-taking. Second, we classified the infant vocalization type using an adaptation of the classification scheme in.[50] Every sound produced by the infant was assigned to one of the categories in Table S1. All coding was performed blind to phase, as coders worked on the unmodified audio recordings.

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

Vegetative vocalizations (coughs, etc.), whispers, fussy sounds and cries were excluded from further analysis as these sounds are either vegetative, signalling the fact that the infant is off task, and/or unvoiced/non-harmonic sounds, whose f_0 cannot be reliably extracted. Co-vocalizations were also excluded from the analyses because the f_0 extracted from these segments could be a blend of the child's and the adult's f_0. Turn-taking passages, where the two voices did not overlap, were retained, and manually segmented to isolate infants' vocalizations. The segments produced by the child could also contain some environmental noise, such as toys clacking. In cases where the infant's vocalization was covered by the noise, the segment was excluded from the analysis. However, if the noise did not cover the vocalization, the infant's vocalization timings were manually extracted and included in the analysis.

The vocalizations selected in this way were then concatenated within each phase (experimental and baseline), and the resulting sequences were used to extract the f_0 produced in each phase by the infant and the caregiver separately, using Praat[51]; f_0 values were then converted to Mel. This concatenation approach was preferred because spontaneous infant vocalizations are typically brief, and their f_0s are highly variable, making single-vocalization f_0 estimates noisy and unreliable; concatenation yielded a longer, more stable signal, allowing to reduce some of this noise (similar to the procedure to extract ERPs from an EEG signal). The average f_0 and its standard deviation were then calculated for each phase from the concatenated sequences, and were used to calculate our main dependent variable: the average (standardized) produced f_0 shift difference from baseline (Δf_0; see Figure 1B).

Mirror task

The mirror self-recognition test was recorded with two Sony HDR-CX240 cameras. Using a custom Python script based on the package *cv2*, we analysed the infant's gaze direction frame by frame. We identified whether the infant was looking at the mirror, the experimenter, the parent, or away. Gaze information was primarily extracted from camera 1, with camera 2 used only when camera 1 did not capture the infant's gaze during the test, ensuring continuous tracking. As expected, no infant touched the red mark on their forehead or their reflection in the mirror. To confirm that infants showed absolutely no sign of noticing the mark, we compared looking time toward the mirror before and after the red mark was applied to the forehead. The rationale was that, even if infants lacked the motor control to touch the mark, they might still display a novelty preference by looking longer at their reflection once the mark was present.[52,53] No looking time differences were observed ($t(34) = -1.27$, $p = 0.21$).

Statistical analyses

Linear regression assumptions were assessed visually using diagnostic plots of linearity, homogeneity of variance, residual normality, and influential observations, as well as posterior predictive checks using the check_model() function from the performance package[54] in R.

Besides the main and control analyses reported in the main text, we also conducted the following analyses as methodological checks.

Pitch transformation verification

To verify that the f_0 modification was successfully applied, we first extracted the f_0 from the audio played back to the infants through the headset (i.e., the output of the software), and their actual f_0, extracted from the audio recordings, and we computed the difference in Mel units between these two signals (output-input). A linear model was then used to verify the effects of phase, condition and their interaction on output-input in the experimental conditions. Phase alone did not improve the model fit ($F(1) = 0.004$, $p = 0.95$). Adding condition improved the model fit ($F(1) = 131.06$, $p < 0.001$), and the phase*condition interaction enhanced it further ($F(1) = 137.61$, $p < 0.001$). Output-input was significantly lower in the EXP than BL for the downward condition ($\beta = -26.63$, $SE = 2.3$, $t = -11.59$, $p < .001$), higher for the upward condition ($\beta = 28.4$, $SE = 2.39$, $t = 11.86$, $p < .001$), and not significantly different for the neutral condition ($\beta = 0.76$, $SE = 2.34$, $t = -0.32$, $p = 0.74$). It did not differ between upward and downward condition in the baseline condition ($\beta = -0.67$, $SE = 2.34$, $t = -0.28$, $p = .78$). Finally, in the experimental phase the difference was higher in the upward compared to the neutral condition ($\beta = 27.1$, $SE = 2.37$, $t = 11.43$, $p < .001$) and lower in the downward compared to the neutral condition ($\beta = -27.28$, $SE = 2.32$, $t = 11.75$, $p < .001$). Thus, the transposition was effective in shifting the voice upwards in upward condition, downward in the downward condition in the experimental phase compared to the baseline, while in the neutral condition no transposition was observed (see Figure S1 for an overview of the transposition for different types of sounds).

Caregiver pitch

Since the switch between DAVID settings was manual, it was imperfect. Thus, we verified that infants' Δf_0 was not affected by the f_0 modification being briefly applied to the caregiver's voice during switching. Using the same pre-processing steps, we extracted each caregiver's f_0 from their concatenated vocalizations and computed the standardized f_0 difference between EXP and BL (*mum_Δf₀*, Figure S2D). A linear model testing the effect of condition on mum_Δf_0 showed no improvement in model fit ($F(2) = 0.59$, $p = 0.56$), indicating no differences across conditions. We also added the caregiver's average f_0 modification size (output-input mother) as a fixed effect in the model comparison reported in the main text, alongside output-input infant and age, participant as a random effect, and Δf_0 (non-concatenated) as the dependent variable. The only variable that improved the model fit was still output-input infant ($\chi^2(1) = 483.76$, $p < 0.001$). Output-input mother did not improve the model fit ($\chi^2(1) = 0.06$, $p = 0.8$), nor did age ($\chi^2(1) = 0.76$, $p = 0.38$). This further indicates that any accidental caregiver's voice modification did not influence infants' f_0 compensation, which instead was more likely to be driven primarily by the modification applied to the infant's own voice.

Vocalization duration

Previous AAF studies measured f_0 compensation using vocalizations of relatively short, controlled length, which is typically made possible by instructing participants to repeat specific words or syllables (e.g., "bed"[20] or "ba"[21]). This was impossible here because, unlike children or adults, infants cannot be asked to produce specific vocalizations. Instead, they spontaneously vocalize at their own

Please cite this article in press as: Belia et al., Infants respond to alterations in their own voice months before they recognize themselves in the mirror, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.09.089

pace, and produce vocalizations that are very varied in kind and duration. This means that, here, the f_0 transformation was applied to vocalizations that could vary in duration across participants, raising the possibility that any observed differences across conditions might stem from the fact that infants had more or less exposure to the auditory perturbation in one of the conditions (by chance). To rule this out, we tested whether vocalization duration differed by condition and phase using linear mixed effect models, with phase, condition and participant (random effect) as predictors. No predictor improved the model fit (all ps > 0.16), indicating that vocalization duration was comparable across conditions and phases (Figure S2E). It is therefore unlikely that the vocal compensation we observed reflects differences in perturbation exposure rather than genuine f_0 adjustments.