BRIEF REPORT

The role of prosody in infants' preference for speech: A comparison between speech and birdsong

Chiara Santolin[1] (iD) | Sofia Russo[2] | Giulia Calignano[3] |
Jenny R. Saffran[4] | Eloisa Valenza[2]

[1]Centre for Brain and Cognition
& Department of Information and
Communication Technologies, University
Pompeu Fabra, Barcelona, Spain

[2]Department of Developmental Psychology
and Socialisation, University of Padova,
Padova, Italy

[3]Department of Psychology and Cognitive
Sciences, University of Trento, Trento,
Italy

[4]Waisman Center & Department of
Psychology, University of Wisconsin-
Madison, Madison, Wisconsin

Correspondence
Chiara Santolin, Centre for Brain and
Cognition, University Pompeu Fabra, c/
Ramon Trias Fargas 25-27, Barcelona
08005, Spain.
Email: chiara.santolin@upf.edu

Funding information
Fondi per la Ricerca Ordinaria, University
of Padova, Grant/Award Number:
DOR2016

Abstract

Speech preferences emerge very early in infancy, pointing
to a special status for speech in auditory processing and a
crucial role of prosody in driving infant preferences. Recent
theoretical models suggest that infant auditory perception
may initially encompass a broad range of human and non-
human vocalizations, then tune in to relevant sounds for
the acquisition of species-specific communication sounds.
However, little is known about sound properties eliciting
infants' tuning-in to speech. To address this issue, we pre-
sented a group of 4-month-olds with segments of non-na-
tive speech (Mandarin Chinese) and birdsong, a nonhuman
vocalization that shares some prosodic components with
speech. A second group of infants was presented with the
same segment of birdsong paired with Mandarin played in
reverse. Infants showed an overall preference for birdsong
over non-native speech. Moreover, infants in the Backward
condition preferred birdsong over backward speech whereas
infants in the Forward condition did not show clear prefer-
ence. These results confirm the prominent role of prosody
in early auditory processing and suggest that infants' prefer-
ences may privilege communicative vocalizations featured
by certain prosodic dimensions regardless of the biological
source of the sound, human or nonhuman.

1 | INTRODUCTION

Linguistic prosodic features, identified as rhythmic and intonational properties of spoken languages, are processed very early in life by humans. Experience with prosody begins prenatally, when suprasegmental features of the mother's voice are available to the fetus (Gerhardt et al., 1992; May, Byers-Heinlein, Gervain, & Werker, 2011; Moon, Lagercrantz, & Kuhl, 2013; Spence & DeCasper, 1987). At 36 weeks of gestation, indeed, fetuses respond to their mothers' voice while mothers read stories aloud (Hepper, Scott, & Shahidullah, 1993; Voegtline, Costigan, Pater, & DiPietro, 2013) and neonates show physiological responses to the maternal voice (DeCasper & Spence, 1986; Granier-Deferre, Ribeiro, Jacquet, & Bassereau, 2011; Moon & Fifer, 2000). In addition, neonates can tease apart non-native languages based on rhythmical properties, even when the speech is low-pass filtered (Nazzi, Bertoncini, & Mehler, 1998). Early sensitivity to prosodic cues is also supported by decreased neural activity in response to speech stimuli with distorted prosody (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Sambeth, Ruohio, Alku, Fellman, & Huotilainen, 2008). In fact, when speech is presented backwards, eliminating normal prosodic properties (Perani et al., 1998; Ramus, Hauser, Miller, Morris, & Mehler, 2000; Van Lancker, Kreiman, & Emmorey, 1985), 2- to 3-month-old infants fail to discriminate between spoken languages.

Interestingly, it has been suggested that infants' acoustic preferences include nonhuman animal vocalizations. While neonates favor both speech and rhesus monkey calls over synthetic non-speech stimuli, 3-month-olds show selective preference for speech over a variety of sounds (Shultz & Vouloumanos, 2010; Vouloumanos, Hauser, Werker, & Martin, 2010; Vouloumanos & Werker, 2007). However, other studies demonstrated that up until 6 months, infants' listening biases include other primates' sounds. Ferry, Hespos, and Waxman (2013) and Perszyk and Waxman (2016) showed that both speech and lemur calls promote object categorization in 3- to 4- month-olds, but not in 6-month-old infants. Infant auditory perception appears to be initially tuned to a broad range of acoustic communicative sounds including human and nonhuman primate vocalizations and, subsequently, tune in to the relevant sounds for acquiring species-specific communication stimuli (in this case, human language).

Taken together, these findings point to an early preference for speech, a prominent role of prosody in driving such preference, and an initial openness to nonhuman primate vocalizations. What remains unknown is (a) whether such privilege is limited to our closest phylogenetic relatives, and (b) which properties of the speech signal drive infants' tuning-in to speech. Given that infants' biases are not restricted to native languages, there must be a particular acoustic component, or the combination of multiple components, in the signal attracting infants toward spoken languages. One possibility is that infants prefer harmonically rich biological sounds, characterized by prosodic features such as rhythm and pitch variations. If this is the case, then nonspeech sounds with these properties should also attract infants' preferences.

The present research was designed to test this hypothesis, investigating whether infants would favor non-native speech over a nonprimate vocalization: birdsong. Songs produced by birds are harmonically rich and contain prosody-like components shared with human speech (rhythm, pitch excursions, changes in duration, intonation-like patterns; see Mol, Chen, Kager, & ter Haar, 2017 for a review). For this reason, birdsong represents one of the most complex vocal sounds in nature (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Doupe & Kuhl, 1999; Petkov & Jarvis, 2012; Samuels, 2015; Yip, 2006). We investigated 4-month-olds' preferences for birdsong (sung by an adult male European starling, *Sturnus vulgaris*) versus Mandarin Chinese (Forward condition). To assess the contribution of the prosodic structure of the linguistic materials, we also compared infants' preferences for birdsong versus Mandarin presented backwards (Backward condition). Recall that reversed speech violates the canonical temporal organization of phonemes, dramatically distorting

SANTOLIN ET AL.

INFANCY

THE OFFICIAL JOURNAL OF THE
INTERNATIONAL CONGRESS
OF INFANT STUDIES

WILEY

3

the prosodic features (Ramus et al., 2000). Importantly, although most studies in this literature have employed individual words as speech stimuli, we opted for longer, connected streams in order to preserve normal prosodic components typically conveyed by natural languages, and similar melodic properties of bird songs.

If prosodic structure plays an important role in eliciting acoustic biases in early infancy, we expected to observe a preference for birdsong over backward speech, as the comparison involves a prosodic-like stimulus (birdsong) and a prosodically distorted stimulus (backward speech). In contrast, we expected no preference between birdsong and forward speech because both stimuli include rich natural prosodic structure. This result, if obtained, would expand our understanding of the range of biological stimuli to which infants are attracted, and support the hypothesis that the prosodic characteristics of the signal play a central role in infant auditory preferences.

2 | METHOD

2.1 | Participants

Thirty-five 4-month-olds were randomly assigned to one of two between-subjects conditions; 17 infants to the Forward condition (nine males; mean age 4.1 months) and 18 infants to the Backward condition (11 males; mean age 3.9 months). Inclusion criteria were set prior to data collection and consisted of (a) minimum eight trials with looking time longer than 500 ms, (b) more than 70% of eye-gaze registered by Tobii, (c) no prior exposure to Asian languages, (d) being full term, (e) no history of sensory, neurological, or language disorders. The present study was conducted according to Declaration of Helsinki guidelines, with written informed consent obtained from a parent/guardian before data collection. All procedures were approved by the University of Padova ethics committee (Comitato Etico della Ricerca Psicologica, Area 17).

2.2 | Stimuli

The materials consisted of a stream of birdsong paired with either Mandarin Chinese played forward (Forward condition) or backward (Backward condition). The two conditions thus differed only in the forward versus backward presentation of the speech stream. The speech stimulus was recorded by a female native Mandarin speaker in an infant-directed speech register. Mandarin, a tonal language, was selected because it has very different temporal and prosodic structure from Italian, the participants' native language, to minimize infants' familiarity with the speech stimulus. The birdsong was a natural recording of a male European starling (*S. vulgaris*).[1] Streams were composed of either two different segments of Mandarin or two different segments of birdsong, each formed by 8.63 s-long sentence repeated twice. Overall stream duration was 17.26 s. The four streams were presented in alternation at test. All sounds were scaled for intensity at 60 dB using Praat 6.0.20.[2] A colored checkerboard with a blinking external frame was played in conjunction with the auditory materials.

---

[1]Speech stimuli are available as Supporting Information.

[2]Pitch (average) information. Forward Mandarin Segment 1: 293.4 Hz and Segment 2: 327.5 Hz; Backward Mandarin Segment 1: 293.5 and Segment 2: 328 Hz; Birdsong Segment 1: 3,035.75 Hz and Segment 2: 2,882.61 Hz. Pitch properties were not normalized across stimuli as pitch represents a key prosodic property of human versus songbird vocalizations.

2.3 | Procedure

The Infant-Controlled Looking Time Preference procedure (Cooper & Aslin, 1990; Vouloumanos & Werker, 2004) was used to assess infants' auditory preferences. Participants were seated in a high chair in a semi-dark room equipped with a computer screen placed in front of the infant (60 cm distance). Sounds came from audio speakers located on left and right sides of the screen. A remote, infrared eye-tracking camera below the screen recorded eye movements (Tobii X2-60 Eye-Tracker). Testing began after calibration. At the beginning of each test trial, a checkerboard appeared on the screen, providing an area of interest for the eye-tracker (AOI; 19.05 × 19.05 cm). Once the infant fixated the AOI for more than 500 ms, sound presentation begins; one of the stimuli started and kept playing until infant looked away for more than 200 ms. The checkerboard remained on the screen until 25 s elapsed, corresponding to max trial duration, during which infants could accumulate max 17.26 s of looking time (max sound stream duration). There were 12 trials divided in three blocks, presented in random order. The checkerboard changed color at the beginning of each block to keep infants engaged to the task. The dependent measure was infants' looking time, recorded automatically by the eye-tracker. The experiment was programmed in E-Prime 2.0.

2.4 | Statistical methods

A general linear model was used to examine the interaction between condition (*Forward* vs. *Backward*) and sound type (*Mandarin* vs. *birdsong*), applying a 2 × 2 repeated-measures ANOVA. We also ran matched-pairs *t*-tests and Bayesian factor analysis to compare looking time for Mandarin versus birdsong in each condition (Figure 1). Bayes factors were calculated in favor of the null hyp. in the Forward condition (BF01; assuming no difference between birdsong and speech), and in favor of the alternative hyp. in the Backward condition (BF10; assuming a difference), according to Van Doorn et al. (2019). Cohen's *d* is reported as effect size. Analyses had been run with SPSS 19 and JASP 0.9.1.

3 | RESULTS

We obtained a significant main effect of sound type ($F = 9.813$, $p = 0.004$, $\eta_p^2 = 0.229$) suggesting an overall preference for birdsong (13 s) regardless of speech direction (11.7 s), but no significant interaction between conditions ($F = 1.753$, $p = 0.197$, $\eta_p^2 = 0.05$).

In the Backward condition, infants showed a preference for birdsong over backward Mandarin ($t_{(17)} = 2.705$, $p = 0.015$, $d = 0.637$; $BF_{01} = 3.78$); average looking time was 11.2 s ($SD = 3.5$) for backward Mandarin, and 13.1 s ($SD = 3.1$) for birdsong. In the Forward condition, infants showed
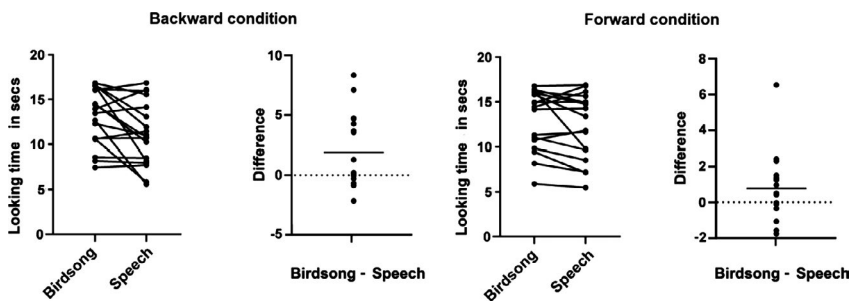

FIGURE 1 Individual looking times and difference score for Backward (left) and Forward (right) conditions are represented. Horizontal lines indicate averages

no preference for forward Mandarin over birdsong ($t_{(16)} = 1.658$, $p = 0.117$, $d = 0.402$; $BF_{01} = 1.28$); average looking time was 12.2 s ($SD = 3.6$) for Mandarin, and 13 s ($SD = 3.3$) for birdsong. While results of the Backward condition reveal a clear preference for birdsong, which is also supported by a medium-to-large effect size, results of the Forward condition are not equally transparent. Although the p-value (>0.05) suggests an absence of preference, results of Bayes factor analysis do not allow us to claim that infants in the Forward group do not show a preference between birdsong and speech.

4 | DISCUSSION

Previous studies support the hypothesis that speech is a special sound for infants even before birth. Interestingly though, this literature leaves open the possibility that infants' early preferences are driven by sound properties shared with other animals' vocalizations (Vouloumanos & Waxman, 2014). To address this issue, we presented 4-month-olds with non-native speech, played either forward or backward, and a birdsong, a nonhuman vocalization that shares some prosodic components with speech.

Infants showed an overall preference for birdsong regardless of whether Mandarin was played forward or backward. Infants also showed a preference for birdsong over backward Mandarin, but no clear preference between birdsong and forward Mandarin. It is likely the case that the prosodically distorted backward speech was not as attractive as its forward-ongoing version, leading infants in the Backward condition to shift their usual preference for speech toward the nonhuman sound. Similarities among speech and birdsong at the level of suprasegmental components may have been sufficient to reduce infants' typical bias for speech: when two prosodically normal sounds are contrasted (birdsong and forward Mandarin) infants no longer show a sharp preference.

Although prosody does not characterize birdsong in exactly the same way it characterizes speech, birdsong *does* possess prosody-like components shared with spoken languages (reviewed in Mol et al., 2017). The connected structure of birdsong, in fact, conveys prosodic information in a way that is similar to spoken languages, and could have been sufficient to render an overall preference for the nonhuman sound. Among prosodic features, birdsong's higher pitch and stronger pitch variations with respect to speech are very likely to have driven the preference for birdsong. Such a preference should not be surprising, though. Previous findings showed that both human and nonhuman vocalizations promote object categorization in 3- to 4- month-olds (Ferry et al., 2013; Perszyk & Waxman, 2016) suggesting that young infants are attracted to naturally produced sounds other than speech, and consider both human and nonhuman vocalizations as reliable sources for learning.

These results broaden our understanding of infants' repertoire of privileged sounds to include an avian song. Infants' acoustic preferences may thus privilege vocalizations featured by certain prosodic properties, even when the biological source of the sounds is not human. What remains unclear though is whether infants are attracted to birdsong's rich prosodic structure (including pitch) or to the novel "content" provided. Further experiments including backward birdsong would help disentangling the role of prosody-like components and the novel aspect of bird vocalizations, and shed light on which properties play a crucial role in driving acoustic biases in early infancy.

CONFLICT OF INTEREST

The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Chiara Santolin 🔟 https://orcid.org/0000-0002-8361-5574

REFERENCES

Berwick, R. C., Okanoya, K., Beckers, G. J., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. https://doi.org/10.1016/j.tics.2011.01.002

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.2307/1130766

DeCasper, A., & Spence, M. (1986). Prenatal maternal speech influences newborn's perception of speech sounds. *Infant Behavior and Development*, *9*, 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013–2015. https://doi.org/10.1126/science.1077066

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*(1), 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15231–15235. https://doi.org/10.1073/pnas.1221166110

Gerhardt, K. J., Otto, R., Abrams, R. M., Colle, J. J., Burchfield, D. J., & Peters, A. J. (1992). Cochlear microphonics recorded from fetal and newborn sheep. *American Journal of Otolaryngology*, *13*(4), 226–233. https://doi.org/10.1016/0196-0709(92)90026-P

Granier-Deferre, C., Ribeiro, A., Jacquet, A., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*, 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Hepper, P., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*, 147–155. https://doi.org/10.1080/02646839308403210

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: Does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, *2*, 222.

Mol, C., Chen, A., Kager, R. W., & ter Haar, S. M. (2017). Prosody in birdsong: A review and perspective. *Neuroscience and Biobehavioral Reviews*, *81*, 167–180. https://doi.org/10.1016/j.neubiorev.2017.02.016

Moon, C., & Fifer, W. (2000). The fetus: Evidence of transnatal auditory learning. *Journal of Perinatology*, *20*, S36–S43.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

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 and Performance*, *24*(3), 756.

Perani, D., Paulesu, E., Sebastián-Gallés, N., Dupoux, E., Dehaene, S., Bertinatti, V., ... Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Perszyk, D. R., & Waxman, S. R. (2016). Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition*, *153*, 175–181. https://doi.org/10.1016/j.cognition.2016.05.004

Petkov, C. I., & Jarvis, E. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*, 12.

Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*(5464), 349–351. https://doi.org/10.1126/science.288.5464.349

Sambeth, A., Ruohio, K., Alku, P., Fellman, V., & Huotilainen, M. (2008). Sleeping newborns extract prosody from continuous speech. *Clinical Neurophysiology*, *119*(2), 332–341. https://doi.org/10.1016/j.clinph.2007.09.144

Samuels, B. D. (2015). Can a bird brain do phonology? *Frontiers in Psychology*, *6*, 1082.

Shultz, S., & Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. *Language Learning and Development*, *6*(4), 241–257. https://doi.org/10.1080/15475440903507830

Spence, M., & DeCasper, A. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development*, *10*, 133–142. https://doi.org/10.1016/0163-6383(87)90028-2

Van Doorn, J., van, den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., … Ly, A. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters: I. Recognition of backward voices. *Journal of Phonetics*, *13*, 39–52.

Voegtline, K. M., Costigan, K. A., Pater, H. A., & DiPietro, J. A. (2013). Near-term fetal response to maternal spoken voice. *Infant Behavior and Development*, *36*(4), 526–533. https://doi.org/10.1016/j.infbeh.2013.05.002

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, *81*(2), 517–527. https://doi.org/10.1111/(ISSN)1467-8624

Vouloumanos, A., & Waxman, S. R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Sciences*, *18*(12), 642–646. https://doi.org/10.1016/j.tics.2014.10.001

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276. https://doi.org/10.1111/j.1467-7687.2004.00345.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Yip, M. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, *10*, 442–446. https://doi.org/10.1016/j.tics.2006.08.001

SUPPORTING INFORMATION

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