**Peer Review and Communication History**

**MS Title**: **What Can We Perceive In Infant Vocalization**

**Author Names**: Alanna A. Beyak, Olivia Cadieux, Matt Cook, Carly S. Cressman, Barbie Jain, Jarod A. Joshi, Spenser L. Martin, Michael Mielniczek, Sara Montazeri, Jolyn Sawatzky, Bradley C. Smith, Jackie Spear, Essence I. Perera, Thomas Thompson, Derek Trudel, Jianjie Zeng, and Melanie Soderstrom

**Submitted:** Jan 26, 2021

**Editor First Decision**: Revise & Resubmit

I have had the opportunity to receive feedback from three expert reviewers and have read the manuscript myself. Overall, the reviewers and I had a positive impression of the work, especially in terms of the novelty of the questions you are asking and the fit for Collabra: Psychology. However, there were a few areas where the reviewers and I agreed there was room for improvement. I would welcome a revision and do view additional data collection as necessary (though note one area where it could be valuable if doing so is feasible).

The reviewers and I found four key areas to focus on:

1. The introduction: Some areas were not well motivated. Reviewer 2 puts it clearly – caregivers are extremely likely to already be aware of the infant’s sex and language and would not need to be able to perceive it. A clearer motivation of the study in general as well as the specific research questions would strengthen the paper. Reviewer 3 also commented that the implicit hypothesis that detecting these variables is important for caregiving is untested. Reviewer 1 gave helpful suggestions for additional literatures to connect to.
2. Individual differences, particularly around the distribution of participant characteristics: All 3 reviewers raised this and I also wondered about it. It would be helpful to know the range of caregiving experience – if it is low overall, it makes sense that this would not be related to task performance. There may also be opportunities to explore variability by including participants with different language experience (rather than excluding those 107 participants) or by collecting data from parents exclusively (this is not necessary, but at least could be worth mentioning in the discussion if additional data collection is not feasible).
3. Infant age: Reviewer 2 in particular was interested in the bin selection process and whether participants were close in their incorrect guesses (for instance, if a participant was judging a 0-7-month-old, it seems different to guess 8-18 than to guess 19-36). I also wondered if you had more detailed information about the age of the infants in the corpus – were participants less accurate for infants near the boundary (e.g., putting a 6-month-old in the 8-18 bin) compared to one in the middle of a bin.
4. Audio selection: More description of the stimuli would be useful. Reviewers raised the possibility that participants could infer age from different properties of the stimuli, such as the presence of words or environmental features in the background. I would not view this as a problem necessarily, but more description of the audio stimuli would be valuable (I also was a little confused about the process of selecting infant clips, the last paragraph of page 9).

See reviews for more detailed feedback from the reviewers.

Thank you for considering Collabra: Psychology as an outlet for your work. I look forward to continuing to work with you.

Sincerely,
Jasmine DeJesus
Section Editor, Developmental Psychology
Collabra: Psychology
jmdejes2@uncg.edu

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Review of “What can we perceive in infant vocalization?”

This manuscript reports on what listeners (in this case, undergrads) understand about the vocalizations of infants, purely from their acoustic forms. Listeners were unable to infer infants’ sex but did identify native language and age; exploratory analyses of listeners’ experiences with infants and other covariates were inconclusive.

I think this is a fascinating topic and a good fit for Collabra: Psychology. The manuscript is concise, analyses are straightforward, and this will make a nice contribution to the developmental science and psychoacoustics literatures. Below are three sets of suggestions for a revision.

First, given the constellation of null and positive results reported, which I think are quite interesting (in particular, failure to detect sex of infant but success at detecting age), I wonder to what extent features of the raw audio that are not the infant’s babble may be at work here. Is there some general difference in the acoustic environment between a 4mo infant and a 24mo infant? I listened to a few of the babbling clips from the BabbleCor Corpus OSF page and they are, of course, very rich, naturalistic recordings; this is good from an ecological validity perspective but also supports my suspicion of a risk of confounding from non-baby-babbling sounds in the audio. I think this concern can be reasonably addressed without collecting new experimental data, but instead by doing some careful analysis of the acoustics of the source audio, or maybe by re-analyzing the main effects in a subset of audio examples where there is minimal non-baby sounds (personally, I would predict that this sort of “preprocessing” would, in fact, increase the size of the main effects, but it’s an empirical question).

Second, I have, and I suspect that others would share, a strong prior that listener experience should play a role in inferring information about the infant vocalizer, so I am surprised by the mixed results of the exploratory analyses on cargiving experience. How solid are these? Might they be a result of restricted range? (How many undergraduates in the sample have a child of their own, for example? It would be good to provide descriptive statistics for this and other sections to help readers interpret the results). I wonder if an even stronger test of experience might be to compare listeners who have zero experience with babies to everyone else, for example, or better yet, to recruit a small sample of additional participants who are all parents, and see if their performance differs markedly from the ones reported here.

Third, it would be good to connect a bit more to relevant literatures in the introduction and the discussion. For example, I think it would broaden interest in the paper to expand the review to say more about what is detectable in infant cries (e.g., more recent studies, such as Mampe et al. 2009 Current Biology, and citations therein; Soltis 2004 Behavioral and Brain Sciences). This manuscript also connects deeply to the wider, fascinating topic of form and function in vocalization (e.g., in laughter: Bryant et al. 2018 Psych Science; Vouloumanos & Bryant 2019 Scientific Reports; across species: Filippi et al. 2017 Proc B; in infant-directed singing; Mehr et al. 2019 Science; Mehr et al. 2018 Current Biology; and in infant-directed speech: Bryant & Barrett 2007 Psych Science; Moser et al 2020 bioRxiv). These patterns of results would help to situate this paper properly in the literature, especially because they share a common pattern of results: in particular, in form-function tasks, there is often a low level of signal detectable by listeners who are unfamiliar with the signaler, but then on top of that, there is plenty of unexplained variability.

Thanks to the authors for an enjoyable read.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The paper “What Can We Perceive In Infant Vocalization” details a study of 626 undergraduates’ perception of infants’ vocalizations. The authors examined whether participants could accurately identify infants’ sex and age, as well as whether they were learning English or not, and if these perceptions were moderated by participants’ caregiving experience. Participants were somewhat better than chance for language and age, with no effects of moderators.

Overall, this is an interesting and carefully-conducted study, with the potential to make an important contribution to the field. The authors’ efforts to pre-register their study and share their data and code are particularly appreciated. Below are some comments and suggestions to further improve the paper.

Introduction

The authors convincingly argue that adults’ perception of infants’ vocalizations is important for language development. As infants gain linguistic skills, adults adapt their speech to match their infant’s abilities. However, the other research questions are not as well motivated. It is not clear why it would be important for an adult to be able to perceive differences between sex or language. The authors do mention that caregivers interact with female and male infants differently, but it is not clear if this is driven by the infant’s speech or the adult’s socialization. Moreover, caregivers are extremely likely to already be aware of the infant’s sex and language and would not need to be able to perceive it. A clearer motivation of the study in general as well as the specific research questions would strengthen the paper.

Method

The method for this study was pre-registered. The authors followed their pre-registration plan, but the decisions made were not always readily apparent in the manuscript. The authors can consider improving the method section based on the following suggestions:

* Including distributional information on participants’ caregiving experience. Are these variables distributed normally, bimodally, etc. (for example, once could imagine that most undergraduates have little infant caregiving experience, and a few a parents or siblings of infants with a lot of caregiving experience)? This could also limit the interpretation of the analyses with these variables. Even if this entails a deviation from the pre-registration, the authors should ensure that their reported analyses are reasonable given the distributions of these variables.
* Providing example recordings/transcripts for each of the categories (if allowed by BabbleCor agreement) and/or describing the contents of the clips (e.g., do the clips from 19-36 months contain words?)
* Explaining in more detail how the age bins were chosen
* Moving demographic questions about participants (pg. 11-12) into an appendix or table to streamline the body of the manuscript
* Clarifying if participants responded after each infant recording or after batches of 10 recordings. In the pre-registration, this is stated clearly and can easily be brought into the manuscript.
* Expanding and fully justifying exclusion criteria. Again, this information is in the pre-registration, but not clearly described in the manuscript

Results

The analyses conducted are pre-registered and are generally appropriate to address the research questions. The authors can consider improving this section by:

* Explaining the 0.01 alpha level used across analyses. Again, this information is in the pre-registration, but not clearly described in the manuscript
* Clearly labeling the vertical lines in each of the 3 figures as chance and observed value
* Revisiting the interpretation of effect sizes (also in the discussion section). Given the novelty of the research questions, it is difficult to know what effect sizes are meaningful could be considered “small,” “large,” etc. in this context
* If feasible, analyzing the age responses in an additional/exploratory way to consider if participants selected the correct age bin (e.g. 0-7), an adjacent age bin (e.g., 8-18), or the furthest bin (e.g., 19-36). This could potentially reveal that incorrect responses were only one bin off and that participants are maybe better at identifying age than the current analysis reveals.
* Moving the exploratory model results into tables to increase text readability
* Including an exploratory model for sex judgment. Although the pre-registration specified that moderators would only be examined if the main result was significant, I’m not sure how justified this choice is. For example, the main effect could be non-significant, but a subgroup could perform better than chance. This might be a warranted deviation from the pre-registration (and could be identified as such in the manuscript).

Discussion

The discussion includes several possible interpretations and implications of the results. The authors may also consider other interpretations for their results:

* For infant sex, instead of resorting to the interpretation of a Type I error, it is possible that participants were not performing lower than chance but were miscategorizing the infants. That is, they may have been able to discriminate between male and female infant vocalizations but labelled them inversely. Additionally, when interpreting this result, it would also be important to know if there are any characteristics in the speech signal that could differentiate the sex (e.g., an adult’s sex can often be determined by their F0). If there are no perceptual cues available that participants could use, then it would not be surprising that participants were unable to accurately complete this task.
* The manuscript could benefit from a more thorough discussion of the practical implications of these results.
* As these are very new research questions, the authors may consider including some more future directions. Some ideas that come to mind are testing parents who currently have an infant compared to parents with a teenager, or parents verses grandparents, or using a different method that does not require explicit judgment.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Reviewer 3**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This paper reports on an interesting study examining whether adults are able to determine an infant’s sex, age group, and native language from samples of their vocalizations.

The paper is well written and is methodologically sound. It clearly delineates 3 confirmatory analyses, which were pre-registered, and 2 exploratory analyses, all of which are appropriate. The paper appropriately discusses each analysis in terms of statistical significance as well as effect size and direction of effect. The references are also adequate and appropriate, and limitations are adequately discussed.

I have three main comments:

1. It is not clear to me why participants were excluded on the basis of sociodemographic characteristics, particularly speaking other languages. Is it important for the hypothesis that all participants are exclusively English-speaking? Why would knowledge of other languages impact someone’s ability to determine an infant’s age from their vocalizations, for example? While I understand this decision was made a priori and is therefore unlikely to bias results in the favored direction, a rationale for this decision is warranted so that readers can understand the logic for these exclusions.

Relatedly, I wonder if participants could be included for some analyses but not others. For example, participants who selected their gender as “other” cannot be included in the analysis of participant gender, but why can’t they be included in the other analyses?

1. It would be helpful if the paper could provide an explanation for how the researchers chose the age ranges included in the study.
2. The study provides evidence for the paper’s primary hypotheses – that adults can determine some infant characteristics from their vocalizations. Yet there is another implied hypothesis in the paper: that the ability to perceive these characteristics enables caregivers to respond appropriately to infant vocalizations. This implied hypothesis is discussed in the introduction as well as in the paper’s conclusion (e.g., see last sentence in the paper), yet it is not examined (and therefore not supported) by the study. It seems to me an open question whether the kinds of discrimination abilities demonstrated in this paper are associated with caregiver responses to infant vocalizations. Therefore, I would suggest making the speculative nature of this claim clearer in the conclusion.

Relatedly, the paper seems to equate an adult’s ability to determine an infant’s age group (what was tested in the study) with their ability to determine the maturity of an infant’s vocalizations (what would enable appropriate responses from caregivers). It certainly makes sense that perception of an infant’s age would be correlated with maturity, but these are not identical constructs. It would be pertinent to discuss the extent to which judgments of an infant’s age group in the context of this study are likely to be capturing perceptions of vocal maturity that could influence caregiver responses to their infants (whose ages, of course, they know).

The figures are all appropriate and useful. Data and code to ensure reproducibility is being made openly available.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |

**Author Response**
Dec 9,2021

Dear Dr. DeJesus,

We thank you for the very helpful feedback from yourself and the three reviewers, and for the opportunity to address the concerns. Below we outline our response to these comments. We apologize for the lengthy delay in submitting our revised manuscript. We look forward to hearing back from you regarding the status of our manuscript.

*Editor comments*

*The reviewers and I found four key areas to focus on:*

1. *The introduction: Some areas were not well motivated. Reviewer 2 puts it clearly – caregivers are extremely likely to already be aware of the infant’s sex and language and would not need to be able to perceive it. A clearer motivation of the study in general as well as the specific research questions would strengthen the paper. Reviewer 3 also commented that the implicit hypothesis that detecting these variables is important for caregiving is untested. Reviewer 1 gave helpful suggestions for additional literatures to connect to.*

We thank the reviewers for this feedback and for the suggested literature. We have beefed up the literature review as recommended and more clearly and robustly articulated the motivations for the study. See responses to individual reviewers for details.

1. *Individual differences, particularly around the distribution of participant characteristics: All 3 reviewers raised this and I also wondered about it. It would be helpful to know the range of caregiving experience – if it is low overall, it makes sense that this would not be related to task performance. There may also be opportunities to explore variability by including participants with different language experience (rather than excluding those 107 participants) or by collecting data from parents exclusively (this is not necessary, but at least could be worth mentioning in the discussion if additional data collection is not feasible).*

We have added some additional analyses to explore these questions with the existing dataset as recommended, which can be found on our OSF site (https://osf.io/2a6b4/). See responses to individual reviewers for details. Given the challenges of recruiting a large enough parent sample during the pandemic, we have elected not to collect new data on this topic at this time, although we agree it is a question worth further exploration. We have added a comment about this to the Discussion on page 26-27.

1. *Infant age: Reviewer 2 in particular was interested in the bin selection process and whether participants were close in their incorrect guesses (for instance, if a participant was judging a 0-7-month-old, it seems different to guess 8-18 than to guess 19-36). I also wondered if you had more detailed information about the age of the infants in the corpus – were participants less accurate for infants near the boundary (e.g., putting a 6-month-old in the 8-18 bin) compared to one in the middle of a bin.*

We appreciate this point. Our “bins” were somewhat ad hoc (although informed by the literature on infant language development) and the reason for binning was more practical to simplify participant responding and to facilitate analysis, than conceptual in nature. We have explored this further as suggested by examining whether infants closer to the edges of the bins were more likely to be mis-classified, and whether infants in the youngest and oldest bins were more likely to be misclassified to the middle bin. In general the reviewers are correct but the effect is small and our experiment was designed for less nuanced analyses. The findings are reported on our OSF site (see <https://osf.io/3rdmf/>).

1. *Audio selection: More description of the stimuli would be useful. Reviewers raised the possibility that participants could infer age from different properties of the stimuli, such as the presence of words or environmental features in the background. I would not view this as a problem necessarily, but more description of the audio stimuli would be valuable (I also was a little confused about the process of selecting infant clips, the last paragraph of page 9).*

We have revised this section to more clearly articulate the process for selecting the clips and to address the concern regarding ambient sounds and the presence of words. The clips themselves are available via the BabbleCor corpus for additional review to address other specific concerns. The specific clips we used can also now be found in our github repository (we had to remove them in order for the OSF pre-registration to go through without error, but they have now been put back). We are not sure what additional verbal descriptives might be helpful in describing the clips but are happy to add additional detail if there are specific questions about their nature beyond what we have already reported.

*Reviewer 1:*

*This manuscript reports on what listeners (in this case, undergrads) understand about the vocalizations of infants, purely from their acoustic forms. Listeners were unable to infer infants’ sex but did identify native language and age; exploratory analyses of listeners’ experiences with infants and other covariates were inconclusive.*

*I think this is a fascinating topic and a good fit for Collabra: Psychology. The manuscript is concise, analyses are straightforward, and this will make a nice contribution to the developmental science and psychoacoustics literatures. Below are three sets of suggestions for a revision.*

We thank the reviewer for their supportive comments and very useful feedback.

*First, given the constellation of null and positive results reported, which I think are quite interesting (in particular, failure to detect sex of infant but success at detecting age), I wonder to what extent features of the raw audio that are not the infant’s babble may be at work here. Is there some general difference in the acoustic environment between a 4mo infant and a 24mo infant? I listened to a few of the babbling clips from the BabbleCor Corpus OSF page and they are, of course, very rich, naturalistic recordings; this is good from an ecological validity perspective but also supports my suspicion of a risk of confounding from non-baby-babbling sounds in the audio. I think this concern can be reasonably addressed without collecting new experimental data, but instead by doing some careful analysis of the acoustics of the source audio, or maybe by re-analyzing the main effects in a subset of audio examples where there is minimal non-baby sounds (personally, I would predict that this sort of “preprocessing” would, in fact, increase the size of the main effects, but it’s an empirical question).*

We have clarified in the manuscript that part of the objective of the clip selection was to exclude clips for which ambient noise of this type might be a concern. The specific clips we used can be found at https://github.com/melsod/OCSWinter2020/tree/master/audio/selected\_audio\_files for review. We are not sure what kind of acoustic measurement could reasonably rule out this possibility as a confound entirely though. We have added some more explicit language to this effect in the Discussion section on page 27.

*Second, I have, and I suspect that others would share, a strong prior that listener experience should play a role in inferring information about the infant vocalizer, so I am surprised by the mixed results of the exploratory analyses on cargiving experience. How solid are these? Might they be a result of restricted range? (How many undergraduates in the sample have a child of their own, for example? It would be good to provide descriptive statistics for this and other sections to help readers interpret the results). I wonder if an even stronger test of experience might be to compare listeners who have zero experience with babies to everyone else, for example, or better yet, to recruit a small sample of additional participants who are all parents, and see if their performance differs markedly from the ones reported here.*

We were also surprised not to find an effect of caregiving experience. To further probe this null result, we took your suggestion to compare participants who reported no experience against those who reported any experience. Although we found promising results in our exploratory analysis of age discriminations, the pattern of results were inconsistent enough that we are uncomfortable reporting this directly in the manuscript. We have made a brief reference to this analysis and other additional exploratory analyses in the manuscript on page 19 and the interested reader can review it here: <https://osf.io/zx4md/>. (Note: for some reason OSF does not properly parse the pdf directly from github, so we have uploaded it separately to the OSF Storage.)

We have also added some histograms in our supplemental materials regarding participant descriptives to help further situate the findings.

*Third, it would be good to connect a bit more to relevant literatures in the introduction and the discussion. For example, I think it would broaden interest in the paper to expand the review to say more about what is detectable in infant cries (e.g., more recent studies, such as Mampe et al. 2009 Current Biology, and citations therein; Soltis 2004 Behavioral and Brain Sciences). This manuscript also connects deeply to the wider, fascinating topic of form and function in vocalization (e.g., in laughter: Bryant et al. 2018 Psych Science; Vouloumanos & Bryant 2019 Scientific Reports; across species: Filippi et al. 2017 Proc B; in infant-directed singing; Mehr et al. 2019 Science; Mehr et al. 2018 Current Biology; and in infant-directed speech: Bryant & Barrett 2007 Psych Science; Moser et al 2020 bioRxiv). These patterns of results would help to situate this paper properly in the literature, especially because they share a common pattern of results: in particular, in form-function tasks, there is often a low level of signal detectable by listeners who are unfamiliar with the signaler, but then on top of that, there is plenty of unexplained variability.*

We thank the reviewer for the great suggestions and agree that our manuscript is strengthened by a broader grounding in the literature. We have added several of these papers and some others on pages 2-3 of the Introduction and 26 of the Discussion.

*Reviewer 2:*

*The paper “What Can We Perceive In Infant Vocalization” details a study of 626 undergraduates’ perception of infants’ vocalizations. The authors examined whether participants could accurately identify infants’ sex and age, as well as whether they were learning English or not, and if these perceptions were moderated by participants’ caregiving experience. Participants were somewhat better than chance for language and age, with no effects of moderators.*

*Overall, this is an interesting and carefully-conducted study, with the potential to make an important contribution to the field. The authors’ efforts to pre-register their study and share their data and code are particularly appreciated. Below are some comments and suggestions to further improve the paper.*

We thank the reviewer for these comments.

*Introduction*

*The authors convincingly argue that adults’ perception of infants’ vocalizations is important for language development. As infants gain linguistic skills, adults adapt their speech to match their infant’s abilities. However, the other research questions are not as well motivated. It is not clear why it would be important for an adult to be able to perceive differences between sex or language. The authors do mention that caregivers interact with female and male infants differently, but it is not clear if this is driven by the infant’s speech or the adult’s socialization. Moreover, caregivers are extremely likely to already be aware of the infant’s sex and language and would not need to be able to perceive it. A clearer motivation of the study in general as well as the specific research questions would strengthen the paper.*

We have added some additional language to strengthen our motivations on pages 2-3, 5-6, and 7. In particular, we have clarified the nature of our interest in whether infant sex is perceived by adult listeners based on infant vocalization.

Method

*The method for this study was pre-registered. The authors followed their pre-registration plan, but the decisions made were not always readily apparent in the manuscript. The authors can consider improving the method section based on the following suggestions:*

* *Including distributional information on participants’ caregiving experience. Are these variables distributed normally, bimodally, etc. (for example, once could imagine that most undergraduates have little infant caregiving experience, and a few a parents or siblings of infants with a lot of caregiving experience)? This could also limit the interpretation of the analyses with these variables. Even if this entails a deviation from the pre-registration, the authors should ensure that their reported analyses are reasonable given the distributions of these variables.*

We have added this information to our OSF page and referenced them in the Participants section on page 10. We acknowledge that the distributions are not normally distributed and that they may impact our results. However, with our relatively large sample size and more general regression model the impact of non-normality should be limited. We have also conducted some additional analyses using a categorical split on caregiving/childcare experience that are available for review on our OSF page and referenced in the paper.

* *Providing example recordings/transcripts for each of the categories (if allowed by BabbleCor agreement) and/or describing the contents of the clips (e.g., do the clips from 19-36 months contain words?)*

We do not have access to the original transcripts, but only ~500 ms clips. The clips used in our study are now explicitly available on our git repository.

* *Explaining in more detail how the age bins were chosen*

This information has been added on page 10-11.

* *Moving demographic questions about participants (pg. 11-12) into an appendix or table to streamline the body of the manuscript*

Done (see page 35).

* *Clarifying if participants responded after each infant recording or after batches of 10 recordings. In the pre-registration, this is stated clearly and can easily be brought into the manuscript.*

This has been clarified on page 13.

* *Expanding and fully justifying exclusion criteria. Again, this information is in the pre-registration, but not clearly described in the manuscript.*

This has been clarified on page 14-15.

*Results*

*The analyses conducted are pre-registered and are generally appropriate to address the research questions. The authors can consider improving this section by:*

* *Explaining the 0.01 alpha level used across analyses. Again, this information is in the pre-registration, but not clearly described in the manuscript*

Added on page 15.

* *Clearly labeling the vertical lines in each of the 3 figures as chance and observed value*

Added.

* *Revisiting the interpretation of effect sizes (also in the discussion section). Given the novelty of the research questions, it is difficult to know what effect sizes are meaningful could be considered “small,” “large,” etc. in this context*

We have added some discussion of this point on page 25. We found it difficult to include discussion of this issue meaningfully within the results section without distracting from the reporting of the results themselves. Our use of the terms within the results section is based on standard practice, although we agree these terms are arbitrary. If the reviewer feels that the comment within the Discussion section is insufficient, we could perhaps add a footnote about the issue.

* *If feasible, analyzing the age responses in an additional/exploratory way to consider if participants selected the correct age bin (e.g. 0-7), an adjacent age bin (e.g., 8-18), or the furthest bin (e.g., 19-36). This could potentially reveal that incorrect responses were only one bin off and that participants are maybe better at identifying age than the current analysis reveals.*

We thank the reviewer for this suggestion. Please see our comment to the editor above for a response.

* *Moving the exploratory model results into tables to increase text readability*

After reviewing the text, we feel the information is more readable in paragraph form as we describe the analytic procedure and implications within the paragraph, so we have left this as-is for now. If the reviewer/editor feel strongly on this point we are willing to make the tables.

* *Including an exploratory model for sex judgment. Although the pre-registration specified that moderators would only be examined if the main result was significant, I’m not sure how justified this choice is. For example, the main effect could be non-significant, but a subgroup could perform better than chance. This might be a warranted deviation from the pre-registration (and could be identified as such in the manuscript).*

At the reviewer’s request we have conducted this analysis and included it in the supplemental materials. Results showed no significant effect of childcare experience, caregiving experience, or gender in judgements of infant sex.

*Discussion*

*The discussion includes several possible interpretations and implications of the results. The authors may also consider other interpretations for their results:*

* *For infant sex, instead of resorting to the interpretation of a Type I error, it is possible that participants were not performing lower than chance but were miscategorizing the infants. That is, they may have been able to discriminate between male and female infant vocalizations but labelled them inversely. Additionally, when interpreting this result, it would also be important to know if there are any characteristics in the speech signal that could differentiate the sex (e.g., an adult’s sex can often be determined by their F0). If there are no perceptual cues available that participants could use, then it would not be surprising that participants were unable to accurately complete this task.*

We have run some basic acoustic measures through Praat and do not see any evidence for clear acoustic differences for sex, but neither do these measures illustrate any differences for the other groups either, where significant effects in the expected direction were found. We expect that a more comprehensive analysis might reveal results, but that would be beyond the scope of our study. We have added some language to this effect in the Discussion section on page 26. We have also added some language from the recent paper by Oller et al. regarding sex differences in vocalizations on page 7.

* *The manuscript could benefit from a more thorough discussion of the practical implications of these results.*

We are not sure specifically what the reviewer is looking for here, but we have added some additional language in the Discussion which hopefully addresses this concern. Note that part of the circumspect nature of our Discussion stems from the fact that our findings, while significant, show far from “perfect” performance. We are therefore hesitant to place too much emphasis on the extent to which these results may be meaningful for caregiver-infant interactions “in the wild”.

*As these are very new research questions, the authors may consider including some more future directions. Some ideas that come to mind are testing parents who currently have an infant compared to parents with a teenager, or parents verses grandparents, or using a different method that does not require explicit judgment.*

We have added some additional discussion on page 26-27.

*Reviewer 3:*

*This paper reports on an interesting study examining whether adults are able to determine an infant’s sex, age group, and native language from samples of their vocalizations.*

*The paper is well written and is methodologically sound. It clearly delineates 3 confirmatory analyses, which were pre-registered, and 2 exploratory analyses, all of which are appropriate. The paper appropriately discusses each analysis in terms of statistical significance as well as effect size and direction of effect. The references are also adequate and appropriate, and limitations are adequately discussed.*

We appreciate these positive comments.

*I have three main comments:*

1. *It is not clear to me why participants were excluded on the basis of sociodemographic characteristics, particularly speaking other languages. Is it important for the hypothesis that all participants are exclusively English-speaking? Why would knowledge of other languages impact someone’s ability to determine an infant’s age from their vocalizations, for example? While I understand this decision was made a priori and is therefore unlikely to bias results in the favored direction, a rationale for this decision is warranted so that readers can understand the logic for these exclusions.*

We have added some language on page 14-15 to clarify this decision.

*Relatedly, I wonder if participants could be included for some analyses but not others. For example, participants who selected their gender as “other” cannot be included in the analysis of participant gender, but why can’t they be included in the other analyses?*

Our rationale here was to allow for more direct comparison of the findings across the three analyses and increase homogeneity in our sample to increase clarity of potential results. However, we have run some additional analyses with differing exclusion criteria and found no differences in any of our conclusions. The analyses can be seen here: https://osf.io/65cvt/.

1. *It would be helpful if the paper could provide an explanation for how the researchers chose the age ranges included in the study.*

We have clarified this on page 9-10.

1. *The study provides evidence for the paper’s primary hypotheses – that adults can determine some infant characteristics from their vocalizations. Yet there is another implied hypothesis in the paper: that the ability to perceive these characteristics enables caregivers to respond appropriately to infant vocalizations. This implied hypothesis is discussed in the introduction as well as in the paper’s conclusion (e.g., see last sentence in the paper), yet it is not examined (and therefore not supported) by the study. It seems to me an open question whether the kinds of discrimination abilities demonstrated in this paper are associated with caregiver responses to infant vocalizations. Therefore, I would suggest making the speculative nature of this claim clearer in the conclusion.*

We have added some additional language that we hope makes this more clear on page 25. In the introduction, we feel we are already clear that our research is intended to test a precursor (discriminability) to this “implied hypothesis” (which is well tested by others, in work by Warlaumont, Goldstein, etc.), but have made a minor wording adjustment on page 2 to reinforce this. If the reviewer feels more is warranted, we would welcome some further direction.

*Relatedly, the paper seems to equate an adult’s ability to determine an infant’s age group (what was tested in the study) with their ability to determine the maturity of an infant’s vocalizations (what would enable appropriate responses from caregivers). It certainly makes sense that perception of an infant’s age would be correlated with maturity, but these are not identical constructs. It would be pertinent to discuss the extent to which judgments of an infant’s age group in the context of this study are likely to be capturing perceptions of vocal maturity that could influence caregiver responses to their infants (whose ages, of course, they know).*

We have added some clarification of our approach and rationale on pages 5-6 and 24-25.

*The figures are all appropriate and useful. Data and code to ensure reproducibility is being made openly available.*

**Editor Final Decision: Accept**

Dec 29, 2021

Dear Melanie Soderstrom,

I have now had a chance to read over your manuscript “What Can We Perceive In Infant Vocalization?”, along with the letter describing the changes you made. I have received reviews from two of the original reviewers; both were extremely positive and recommended publication. From my own reading of the manuscript, I appreciate your thoughtful response to the concerns raised on your previous submission. I am pleased to accept your manuscript for publication in Collabra: Psychology.

I had two small final suggestions:

Manuscript (could be applied at proof stage): Pg. 10: Add that modal caregiving was 0 (with n’s with and without caregiving experience if possible).

Supplement (can be addressed any time): I didn’t really follow the logic of the caregiving splits. Does the split at 0 divide between no caregiving and any amount of experience? And then 1 is at least one month caregiving experience vs. not? This could be more clear; given the sample distribution 0 vs. any seems like the most important comparison to me.

I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make at this time, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology, and I wish you the best in the new year!

Sincerely,
Jasmine DeJesus

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The revised version of this manuscript addresses all my prior concerns. The authors have more clearly motivated the research questions and hypotheses, included additional methodological details and justification of methodological choices, and clarified the conclusions from their findings.

The resulting manuscript is an enjoyable read and makes a good contribution to the literature.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Great job on this paper! I look forward to seeing it in print.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |