Interspeech IDS-ADS Reviews

These reviews correspond to submitted version here:

[https://www.overleaf.com/8592580wqvchqwzgmvp - /30837382/](https://www.overleaf.com/8592580wqvchqwzgmvp#/30837382/)  
**REVIEWER #1**  
---------------------------------------------------------------------------  
Comments  
---------------------------------------------------------------------------  
- Strengths:  
This is a good topic, worth investigating.  
  
- Weaknesses:  
The paper focuses partly on the planned results but partly on methodological  
issues (I mean LENA). This is a bit confusing. It would be better to separate  
the two goals.  
  
- General Discussion:  
The findings seem to be relevant and add new insights into the topic of  
babytalk.  
  
  
**REVIEWER #2**---------------------------------------------------------------------------  
Comments  
---------------------------------------------------------------------------  
  
- Strengths:  
It is an important work partly comparing LENA-classification with human  
annotators, and partly analysing the vocal input during long-recordings using  
children from different corpora. The study is done, the analyses done are well  
chosen, and the text is clear and well written.  
  
- Weaknesses:  
A problamtisation of the parameters chosen would be in place (for example, SES.  
which here is based on mother's education - out of necessity as it appears, but  
still). Everyone using LENA ends up commenting that "further studies are  
needed" since the conclusions based on LENA are so inconclusive and vague. The results from this study adds to this. Why not skip LENA?  
  
- General Discussion:  
It is overall a well done and important study. To merge corpora is the new  
black and we need to see a lot more of it.

**REVIEWER #3**  
---------------------------------------------------------------------------  
Comments  
---------------------------------------------------------------------------  
- Summary:  
This paper compares human annotations against automated LENA analyses to  
evaluate adult speech utterances. Specifically, the authors investigate the  
proportion and quantity of CDS that children hear in relation to the child’s  
age, child’s gender, the caregiver’s gender, and SES. Results show that  
under these naturalistic conditions babies hear about two-thirds of their input  
in a CDS speech register, with male infants hearing more speech overall than  
female infants. Additionally, SES affects the overall quantity of speech that  
an infant hears (more speech with higher SES), but not the proportion.  
  
- Strengths:  
1. Addresses the issue of how well LENA analyses can or cannot capture  
differences in speech.  
2. Very clear analyses of how CDS varies across development due to age, gender,  
speaker gender, and SES. Statistics were clear.  
3. Clear organization and presentation. Very nice discussion/conclusion  
section.  
  
- Weaknesses:  
None.  
  
- General Discussion:  
Overall, very clear and well organized paper addressing an important topic as  
LENA becomes more widely used. Minor editorial comment for final version:  
misspelling of “similar” on line 9 of Discussion/Conclusions section.  
  
**REVIEWER #4**---------------------------------------------------------------------------  
Comments  
---------------------------------------------------------------------------  
- Strengths:  
This well written paper presents on the basis of a large database of a total of  
61 daylong recordings of a child's language environment from four different  
labs in the US a comparison between manually-coded and LENA-classed speaker identification of child-directed and adult-directed (female/male) speech. The amount of annotation work is previously unseen (around 11 000 addressee and speaker gender tags were set by 23 annotators) and the comparison to automatic addressee and speaker gender tagging provided by the LENA system is extremely useful given the broad use of the LENA system in the research community. With the now common daylong recordings featuring in language acquisition research design it is imperative to test the validity and reliability of the LENA system. The authors have undertaken this huge task by contributing their data corpora and their annotation teams, laying the foundation to build an automatic classification system with a more reliable output than the LENA system  
provides. The study is related to the ComParE challenge on speaker identification at Interspeech this year and provides the dataset for the challenge. Methodologically meticulously carried out, the paper uses well-fitted analyses and illustrative graphs depicting the results.  
  
- Weaknesses:  
There are only a few minor comments to improve the paper further:

1. It would be important to specify the annotation guidelines that the coders were  
   using more, at least with a reference to another paper's method section in  
   which the scheme is explained in greater detail than as "unified" and "with an  
   eye towards developing CDS and ADS classifiers".
2. The colour coding of all figures renders the figures uninterpretable, please  
   change to symbols suitable for gray scale printing. Apart from that, the result  
   presentation in the chosen graph types is well-suited and informative.
3. Footnote 2 ended up above Figure 1 instead of below, please adjust.
4. In the discussion, remove residues of ambiguity in the first sentence by  
   writing "than from male ones" or equivalent.
5. 2nd paragraph, line 2: "similar"
6. The last paragraph of the discussion comes quite abrupt and stands isolated,  
   understandably as a function of space issues. Maybe there is something that can still be done?  
     
   - General Discussion:  
   The paper has the potential to become a landmark contribution to the speech  
   science community as it tackles the simple distinction between child- and  
   adult-directed speech and adult speaker gender. This distinction holds a strong informative value for all child language acquisition research, but it still  
   costs an incredible amount of human resources when coded by hand. Thanks to this work together with the ComParE challenge, it has the potential to be  
   solved reliably automatically in the future.