This proposed study will examine the acoustic characteristics of infant-directed speech by fathers, whether infants attend to paternal IDS over paternal ADS, and whether these effects are modulated by the extent of leave taken by the father.

Overall the study is well-designed and the background literature appropriate. I do have some concerns that I outline below to strengthen the study.

Introduction/Literature review:

Page 2-3: The paragraph beginning “Still, the characteristics…” is confusing. It starts out stating that there is variability across cultures but then discusses Norwegian IDS in particular without a comparative lens. Then appears to return to a comparative lens but with quite narrow focus.

In general, I found the review of paternal IDS a bit lean. For example, there are some older studies on e.g. the Father Bridge hypothesis (see work by Tomasello, Berko Gleason) that might provide some relevant bigger-picture theoretical meat. Specifically with respect to acoustic characteristics, some of the seminal work did look at (and compare) both fathers and mothers (see, e.g. Fernald et al. 1989), and there is in fact a fairly broad literature on this topic. In fairness to the authors, this can be a challenging literature to find, because it is often not the primary focus of the study, but rather a secondary comparison. However, more can be found/said than is currently included – a quick google scholar search of “father infant-directed speech” pulls up some key recent studies worth looking at. I don’t think an exhaustive review is necessary, but a little more is warranted. Relatedly:

* With respect to the work on preference for IDS, work by Newman may be worth including.
* The claim on page 7 that “it is unknown whether fathers modulate their IDS” is too strong given that there is indeed existing literature on this topic.
* I believe there IS some literature on the impact of caregiver experience that could be explicitly mentioned – I don’t have any specific suggestions offhand, but I suggest looking into this a little.

Page 8: “The infants… will only differ in….” this is again too strong of a claim. The authors are not (and can not) exhaustively measure all aspects on which the infants/fathers may differ.

Study design/Protocol:

Page 12: Some specific questions about the implementation of the study: What will the timeframe be (and how much might it vary across fathers) between when the father completes the online questionnaire and when they come into the lab? Can the authors clarify “main caregiver” (both for review purposes and to ensure that the question is interpreted consistently by the fathers)? The manuscript suggests that an inclusion criterion is that “the father has not been working at the same time as his paternity leave” but this is not explicitly described as being part of the questionnaire. Perhaps a copy of the actual questionnaire would be helpful? I would also encourage the authors to think about other potential factors that could act as third/nuisance variables – e.g. perhaps the fathers who are more comfortable around their infants are more likely to take longer leaves AND produce stronger IDS?

Page 14: I found the description of the protocol a little confusing as stimulus design/creation was interspersed with discussion of the testing protocol. It would also be helpful to rely less on reference to the ManyBabies protocol, both for clarity for readers who may be less familiar with that study and because there are some key differences with the current approach.

Some specific questions about the protocol:

* What happens if an infant fails to calibrate?
* The decision to use word lists rather than utterances is unusual (and differs from the ManyBabies study). I think this is a reasonable approach but further discussion of this choice is warranted, particularly as it limits direct comparison with many prior studies.
* Why only 8 trials? I am a little concerned about having so few trials, and especially because there is a minimum inclusion threshold of even fewer trials. Emerging work by Krista Byers-Heinlein as a secondary analysis on the original MB1 data, and Melanie Schreiner’s “test-retest” follow-up study to the ManyBabies study suggest that getting reliable data at the individual level for infants on IDS preference is very challenging and 8 trials may be a MINIMUM for reliable data. (Neither of these are published yet, but there are preprints and conference posters etc. available for review.)

Analysis

p.18 The comment at the bottom of the page could use some further fleshing out (how will they be transformed? How will “normally distributed” be assessed?) and this information might be better located where the other transformations are outline, on the following page.

Smaller comments:

p. 3 Bergelson et al. (2018) >> Bergelson et al. (2019).