Ms. Ref. No.:  COGNIT-D-19-00069  
Title: The development of infants' responses to mispronunciations: A Meta-Analysis  
Cognition  
  
Dear Dr. Katie Von Holzen,  
  
We have now received three reviews on your manuscript from experts in the field. The reviewers generally commend the value of meta-analyses but also express concerns about the appropriateness of the manuscript for Cognition.   
  
Cognition is primarily concerned with substantial theoretical contributions to the understanding of the mind and manuscripts are expected to sufficiently constrain our theoretical understanding of a phenomenon. None of the reviewers is convinced that this particular analysis makes a sufficiently solid contribution to the field. The theoretical motivation appears oversimplified and the conclusions are not substantially different from what is already known.    
  
I am therefore inclined to recommend that you submit the manuscript to a more specialized journal. I am sorry to be the bearer of such news, but I hope that the comments you have received may be of help in preparing the manuscript for another journal. Cognition can only publish a fraction of all submitted manuscripts, and this usually entails an enthusiastic assessment from reviewers. It might be worth noting that many manuscripts that are declined at Cognition are subsequently published in excellent specialist journals. So, I encourage you to pursue this possibility.   
  
  
For your guidance, I append the reviewers' comments below.  
  
Thank you for giving us the opportunity to consider your work.  
  
Yours sincerely,  
  
Silvia P. Gennari, PhD  
Associate Editor  
Cognition  
  
Journal Policy Statement: Editorial decisions are final, and unsolicited resubmissions cannot be considered for publication.  
  
Reviewers' comments:  
  
  
Reviewer #1: The authors present a competent and thorough review of the experiments  
to date that examine toddlers' knowledge of words' phonological form  
by measuring their fixations to objects upon hearing those objects  
named with a correct pronunciation or a mispronunciation.  The  
manuscript employs up-to-date meta-analytic techniques and generally  
follows the form of prior papers doing the same thing in a range of  
experimental domains.  
  
The authors know what they are doing and the evidentiary base is  
sufficient for such a project to be useful.  I think that if there is  
a question of the paper's suitability for Cognition, the question is  
whether the intended audience is sufficiently large. The techniques  
used here are not new, and the overall conclusions are essentially the  
same as the conclusions drawn by the first study that used this  
technique.  Thus, I have some doubt about the advance that the paper  
represents.  
  
I have a few minor comments to add about the text, given here in  
order.  
  
The introductory paragraph is written awkwardly.  "Acquiring a first  
language means that young learners are solving ..." is very odd  
syntactically.  "word recognition must balance" is strange -- it is  
not the goal that does the balancing.  The "To build robust language  
knowledge" sentence is also a bit of a stumper.  Perhaps the authors  
would be better off simply omitting the first paragraph.  
  
p.9 "As a consequence, our results will be important ..." Opinions  
vary, but my opinion is that it is better to let the importance of the  
results reveal itself.  
  
p.13 Though the scripts are available, I think it would be useful to  
give a bit more detail about the analysis here; the description is  
quite high level.  Some papers give R code (just the formula) which I  
think is usually a good way to manage this.  Ideally readers could  
know better what happened without having to go online.  
  
p,18 What is Q, as a statistic?  
  
p.27 I did not find the description of "Pre vs Post" to be clear:  
"directly compared the post- and pre-naming PTL scores with one  
another."  But this is the same thing that the Difference Score method  
does.  I think what is meant is this: the Difference Score method  
subtracts Pre from Post on every trial, whereas the "Pre vs Post"  
method ... computes Pre target looking across all trials in each  
condition, and likewise Post target looking across all trials in each  
condition, and subtracts the Pre average from the Post average -- not  
even within children, but across children?  So in one case there is  
an early aggregation across trials and then the subtraction, and in  
the other case there is first subtraction and then later aggregation  
across trials?  I hope that this can be described more clearly.  
  
Certainly to me the most surprising result in the paper is the result  
that the "Pre vs Post" analyses do not, on balance, provide evidence  
of mispronunciation effects (Fig. 6, middle violin).  That's really  
weird.  I think it would be useful to drill into this more.  Did those  
studies have something in common about what they were testing?  
  
Considering the figures, I am reviewing this paper based on a  
grayscale printout, which means my interpretation of the graphs relies  
a lot on intuition.  (There are two shades: light gray, light gray,  
and slightly darker gray.) I apologize, sort of, for any errors this  
occasions.  I encourage the authors to make their plots legible when  
appearing in print, by attending to the darkness variation of the  
colors, and labeling the captions according to grayscale rather than  
color.  (Do please still use colors in the plots.  Just make it work  
both ways.)  
  
p.32 On the lack of an age effect on the CP vs MP difference: the  
authors write, "this implies mastery of ... phonological  
distinctiveness earlier than previously thought, which we recommend  
should be further explored experimentally..." The paper presents this  
conclusion and recommendation as if it were news, first reported here.  
But the presence of early effects of mispronunciation on word  
recognition has already inspired several published studies, some of  
which used other methods. (The studies looking at preference for  
infants' own names are examples.)  As it stands this feels a bit like  
doing an analysis of modern communications and proclaiming to Samsung  
that they should think about competing with the iphone.  
  
p.32 why is the start of the analysis window referred to here as the  
"offset time"?  I think it is confusing to call the onset of the  
analysis window the "offset time."  
  
p.33 The authors argue that "increasing age should lead to quicker  
reaction times" and present it as mysterious that experimenters do not  
adjust the beginning of the analysis window according to the age of  
the children.  But at any age, most responses are not at the child's  
absolute minimum.  In an ideal world, researchers would know what  
minimum RT to use at different ages and adjust the window  
accordingly. Realistically, because no-one knows the shape of this  
function, and because the impact on outcomes appears to be minimal in  
existing datasets, experimenters opt not to get into this fight with  
reviewers who worry about possible profusions of free parameters.  
  
  
  
  
  
Reviewer #2: This meta-analysis has much to recommend it. It focuses on a specific question, mispronunciation detection, and includes in the analytic approach a large number of articles, well described techniques, and also explores a number of moderating variables. The write-up is clear and the graphics are helpful. Nonetheless, there are a number of characteristics of the manuscript that decrease my enthusiasm, especially in regard to its appropriateness for Cognition. Perhaps most evident is the skeletal theoretical framework from which the issue is considered. The authors provide a very stripped down, and hence, inaccurate characterization of both Best's PAM and Curtin & Werker's PRIMIR. In fact, these are each set up as 'straw men' with all their nuance and contextualization missing. The essence of PAM has to do with articulatory gestures, and the extent to which articulatory phonetics and articulatory phonology interface, are similar/different and are mutually  
informative. There is one reference to some of Best's work on accent perception, but nothing further that captures the substantial body of work she has done on lexical processing from 6-months to 24 months, as well as straight perception. Similarly, Curtin and Werker's PRIMIR is characterized as making a straight forward prediction of greater use of phonological detail in mispronunciation tasks as a function of age and vocabulary. Yet, PRIMIR's essence is that all detail is available at all ages, but that task and developmental level act as filters that will make such detail more or less available. It is a straight forward prediction from PRIMIR that greater phonological detail will be used in word learning tasks (if there are no additional cues to reference) as a function of age and vocabulary size, but such a prediction is not made wrt mispronunciation detection tasks. Finally, the reason Swingley even began work on mispronunciation detection was to be able to probe the  
phonetics/phonology interface, and to test the assumptions he was building in models of phonological learning. He has given considerable thought to these questions over the years, sees the learning of phonetic categories as informed by lexical acquisition, finds that in many cases infants do have full phonological representations from an early age, but that frequency and acoustic salience also play roles. Thus to give him credit for a seminal study, but treat his work as basically theoretical is not an accurate portrayal. Others, including Thiessen, also have proposed models re the development of phonology. I think that for a journal such as Cognition, it is essential to set up these theoretical stances more comprehensively, or at least accurately.  
  
This leads me to a second major concern. If this meta-analysis was done in order to test the viability of different models of phonological processing, as is apparently the case, then it would seem that different decisions would be made as to which moderator variables to explore. In the paper age and vocabulary are explored, but the number of phonetic feature differences in correctly vs incorrectly pronounced words is not considered. Similarly, the position in the word in which phonetic information is disambiguating (is it at the beginning or end of the word), is not considered. Vowels vs consonants are not considered. And frequency and acoustic salience are also not considered. I realize the number of studies that would fall into some of these categories is likely too small for a conclusive meta-analysis, but given the centrality of those factors to the question of a theory of phonological development, the full set of variable considered seems inadequate to address the  
question posed at the beginning of the article.   
  
  
  
  
Reviewer #3: The authors conducted a meta-analysis of early sensitivity to mispronunciations, focusing on potential changes over development. The conclusion is that children's sensitivity to mispronunciations is robust from an early age, but that there are analysis differences across studies that might be muddying the waters. The conclusion is that children have mature phonological processing from early in development.  
  
This type of meta-analysis is a great service to the field, especially the identification of different analytic practices that may be responsible for skewing results and the identification of possible publication bias. I applaud the authors for undertaking it. This type of meta-analysis also has the potential to be a valuable theoretical contribution towards understanding early word representations (and the factors that shape them).  
  
However, the theoretical contribution of this analysis is not what it could be, for two main reasons. The first is that I am not sure that the authors have correctly characterized the predictions of the main accounts they have covered (PAM and PRIMIR). For example, does the PAM account really predict an increase in tolerance for mispronunciations that have the potential to change the meanings of words? In Best's recent discussions of constancy, she examines cases like accents, where there is some reason to believe that the deviant form maps onto your abstract structure. But is there a prediction in the model that tolerance for mispronunciations should increase for your own accent? As for PRIMIR, my understanding is that it assumes that infants are in fact highly attentive to phonetic detail early on. It is in challenging situations (like word learning situations), with other information overloading the system, that they may not be able to demonstrate this sensitivity (until  
there is a higher-level structure, the phoneme level, that directs them to that information). However, when the situation is simpler, as in the processing of familiar words, PRIMIR does predict detection of mispronunciations, doesn't it? I believe a closer look at these models is warranted, if they are serving as the theoretical meat of this study. Further, an additional possibility, which does not seem compatible with either PRIMIR or PAM (but is stated as being consistent with these models), is raised on p8 - the fact that word representations may not include much detail early on. In contrast, PRIMIR, for example, posits exemplar representations that include considerable acoustic-phonetic detail from the beginning. In other words, it is not clear to me that the vocabulary/age related predictions made in this study fall out from the described models. Perhaps the authors are correct in their characterization, but in that case a more nuanced treatment/more explanation is  
necessary.  
  
Second, I am not convinced that we can draw conclusions about a lack of developmental change because there are a number of factors that are not considered in these analyses that are likely to strongly influence the outcome. These include type of mispronunciation (vowel, consonant, direction, feature type), severity of mispronunciation, type of distractor, whether the mispronunciations are random or part of a new accent, and population (some of the studies include bidialectal or bilingual children, most do not). It is understandable that these factors are not included in the analyses because there would not be enough power to draw any robust conclusions about them. But these factors seem important, and they are also not equally distributed across the factors of interest (age, vocab), which seems additionally problematic. For example, my sense is that most of the studies testing different directions of mispronunciation, varying distractor type, and including  
bilingual/bidialectal children, are in the 18+ month age range. At least one of the populations tested in the Ramon-Casas work is Spanish-dominant bilinguals, whose reduced sensitivity to Catalan vowel changes is expected based on properties of their input. Given that the analysis collapses across all of these factors, it seems premature to accept the conclusion that mispronunciation sensitivity does not change with age (or vocab, see below).   
  
I have a few other comments. First, the authors conclude in multiple places, including the abstract, that the results show that "a mature understanding of native language phonology is present in infants from an early age". This is an overly strong conclusion, even if we accept the absence of an age or vocab effect. The presence of mispronunciation effects does not in and of itself reflect a mature understanding of native language phonology. Mispronunciation effects can reflect the detection of any difference from the typically encountered pronunciation. This is why it is critical to test different types of phonetic changes - both those that are phonologically important in the native language and those that are not (as in Dietrich et al, 2007). A mature understanding would involve selective response to the former. In addition, one might argue that the fact that children accept many mispronunciations as labels for the target object is an indication of immature phonological  
knowledge (see Swingley, 2016). When a phoneme is changed, the label should be mapped to a different object, according to the languages' phonology (at least when a different object is object to map it to).   
  
Distinctiveness vs. constancy: If the authors are equating detection of a mispronunciation with their understanding of phonological distinctiveness and their acceptance of that same mispronunciation with understanding of phonological constancy, this should be laid out explicitly when these terms are first introduced. I don't think that was explicitly done until the discussion. It also needs more justification. As I note above, it is not clear why accepting the types of mispronunciations that are often used in these studies (where a phoneme in the language is changed to another phoneme in the language) is an example of phonological constancy. Does it make sense to think of mispronunciation tolerance as constancy?  
  
The vocabulary effect: There is a weak effect of vocabulary, to be interpreted cautiously because of the small number of studies involved. It seems entirely possible that there would be a stronger vocabulary effect with more data. Many of the studies test children between the ages of 18-24 months (36 out of the 54 entries, if I've counted correctly). Because vocabulary differs wildly at this age, it makes sense that there is no effect of age, but within this age, vocab might matter. The authors are right to lament the fact that so few studies collect vocabulary data. But do they want to come down so hard in favor of the absence of developmental change given that we can't fully evaluate the role of vocabulary?  
  
Minor comments:  
  
There are a number of grammatical errors and typos throughout the manuscript.  
  
P3: "As infants continue to develop into expert language users, their language processing matures and becomes more efficient, including their knowledge of what constitutes a permissible versus word-changing phonological deviation." References should be provided for this. It also seems that the second part of this sentence is exactly what this meta-analysis is testing (and what the authors conclude does not in fact occur, because of the absence of age effects).   
  
P4: In some places - the bottom paragraph here is one example - the word phonological should be phonetic. It should be phonetic when talking about the changes themselves, and phonological when talking about their relevance in the language.   
  
P17: I would suggest that the y axes on Fig3 be the same for the correct and mispronunciation conditions, because otherwise it looks like the effects are the same size  
  
P29: I found the analysis differences to be quite interesting. It does make sense that studies with younger toddlers use larger window sizes. But the fact that there are larger mispronunciation effects for shorter windows, especially as children get older, does raise the possibility that these decisions are made post-hoc and suggests that there is greater need for standardization. I was also intrigued by the fact that the use of individual difference scores, post vs. 50%, and group-level post vs. pre led to different outcomes. In particular, the use of post vs. 50% leads to increased mispronunciation sensitivity with age, but the use of group-level post vs. pre showed a decrease in mispronunciation sensitivity with age. What might be the explanation for these differences (does it have to do with increasing individual differences in familiarity with particular words?), and what is the recommendation going forward?  
  
Table 1 includes a study by Ramon-Casas that has 43-44 month olds. Were these children excluded because the meta-analysis only went up to age 31 months?    
  
Are the ages in months based on the means reported by the authors in the original studies or is the range taken into account? (For example, did the studies listed as involving 18 month olds use only 18 month olds, or did they have, e.g., a range of 17-19 months? Studies differ in how this is reported.)