Infants prefer to listen to speech: A meta-analysis.

Cécile Issard[1], Sho Tsuji[2], & Alejandrina Cristia[1]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole Normale Supérieure, Département d'Études Cognitives

[2] International Research Center for Neurointelligence, The University of Tokyo

Author Note

Correspondence concerning this article should be addressed to Cécile Issard, Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Études Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France. E-mail: cecile.issard@gmail.com

Abstract

The human auditory system is amazingly efficient at processing speech. Some works suggest that this capacity is present from birth, infants preferring to listen to natural speech than to other types of sounds, enabling them to select the signals that are relevant for communication with conspecifics. However, in experimental studies, a large variety of sounds have been contrasted to speech, with infants of very different ages. Drawing a global picture of how this capacity emerges is therefore difficult. We synthesized the literature by conducting a meta-analysis of studies testing speech preference in infants from birth to one year of age. We found a medium effect size, with infants preferring speech over any other type of sound. Contrary to the results of individual studies, we found no effect of age: infants showed the same amount of preference from birth to one year of age. Still contrary to what individual studies suggested, we found the same amount of preference whether speech was contrasted to other natural sounds or to artificial sounds; as well as whether speech was contrasted to other vocal sounds or to non-vocal sounds. Preference was stronger when the speech stimuli were in the infants' native language. This suggests that the representation of speech as a distinct auditory object comes fromis modulated by the degree of familiarity with the sounds of the language they are exposed to.

Keywords: Meta-analysis, infants, speech preference, auditory development, natural sounds

Word count: X

Infants prefer to listen to speech: A meta-analysis.

Authors' contribution: C.I. reviewed literature. C.I. and A.C. coded data. C.I., S.T. and A.C.analyzed data. C.I. and A.C. wrote the paper.

Introduction

Speech is the main signal for human vocal communication. Adult individuals efficiently detect speech sounds in their environment, and use them to spot conspecifics and build social interactions. Since seminal experimental results suggest that infants react more to speech than to certain other sounds (Ecklund-Flores & Turkewitz, 1996), it is conceivable that newborns are equipped with a preference for speech sounds, and that human infants become further entrenched in this preference as they age and gain additional exposure to speech. Here, we synthesize empirical data on infants' preferences for speech over artificial sounds, natural sounds, as well as human non-speech sounds, with the hope of illuminating the strength of evidence supporting various theories proposed to explain infants' preferential listening and processing studies. [insert Figure 1 here]

Potential dimensions underlying preference patterns

There are three key conceptual explanations for infants preference for speech: (a) Preference for natural over artificial sounds; (b) preference for vocal over non-vocal sounds; and (c) preference for familiar over unfamiliar sounds (see Figure 1 for a representation of these distinctions). These three explanations are mutually compatible, such that one or more may be found to be true. In this section, we explain how these explanations differ, and briefly review some results on each, before turning to the potential effects of age.

Natural versus artificial sounds. Natural sounds are those produced with biological systems, including vocal tracts but also the sound of walking and heart rate. Inspection of their acoustic characteristics reveals that in many cases natural and artificial sounds may not have the same structure. For example, backward speech has unnatural formant transitions and seemingly abrupt closures that are not found in naturally produced sounds. We know that natural sounds are processed more accurately by the auditory system, from the cochlea (Lewicki, 2006) to the auditory cortex (e.g. Gehr et al., 2000; see Mizrahi et al., 2014 for a review). As a result, we can expect a preference for speech over artificial sounds that is present from birth and stable over age, predictions that seem to be validated in a first glance of the literature (Vouloumanos & Werker, 2007; Vouloumanos, 2014).

Familiarity. An alternative to the above explanation could be proposed based on familiarity: Perhaps infants prefer speech to foils when these vary in the level of familiarity, since artificial sounds may be less common (and thus less familiar to the infants) than natural sounds. There are no behavioral results directly testing the prediction that infants show stronger preferences when tested with more familiar speech stimuli (for instance, spoken in their native, as compared to a foreign language), but results from neuroimaging studies do predict indirect evidence for this view. For instance, newborns' brain activation was different for forward than to backward speech when the native language was used as the speech stimuli, but not when a foreign language was used (Sato et al., 2012; May et al., 2018).

Vocal versus non-vocal sounds. Many results summarized above may be accommodated by a third hypothesis, postulating a preference for vocal over non-vocal sounds. Vocal sounds are those made with a mouth, and thus typically a subset of natural sounds. There is some evidence which is easy to explain with this hypothesis but cannot be predicted from the natural sound preference or the familiarity preference: Newborns made more head-turns to speech than to heartbeat (Ecklund-Flores & Turkewitz, 1996), but they listened equivalently to speech and monkey calls (Vouloumanos & Werker, 2010). Arguably, speech and heartbeat are equally natural and they are both familiar to a newborn and thus this result cannot be accommodated by either of the other explanations. Changes as a function of development

When infants are born at term, they have already experienced speech and natural sounds more generally for about three months (and throughout their auditory lifelong experience). While little is known about production experience in utero and how this may affect their perception of sounds as being vocal, it is certain that newborns will have experienced their native speech, which other work suggests newborns prefer to prosodically distinct foreign speech (REF). The latter finding actually suggests a potential inconsistency in findings in this body of literature: While the preference for native over foreign speech is consistent with the familiarity preference discussed above, this familiarity may also predict a preference for native speech over monkey vocalizations, which has not obtained (REF). In fact, 1- to 4-month-old infants responded more to speech than to other human sounds (Shultz & Vouloumanos, 2010). To complicate matters further, one study reports that 9-month-olds listened longer to monkey calls than speech (Sorcinelli et al., 2019), which is in stark contradiction with a preference systematically driven by familiarity. Setting this issue aside, it remains clear that as infants age, they gain experience with stimuli in their environment as well as their own vocal production. In addition to the effects of experience, development of the auditory pathway may affect perception (e.g., Mazuka, XX, & Tsuji, 2018). These age-related changes may affect the preference for speech in various ways. One of them is by infants' developing an increasingly narrow and precise definition of the spoken stimulus they prefer. Indeed, whereas newborns do not prefer speech over monkey calls, three-month-olds do (Vouloumanos et al., 2010). If the definition of speech becomes increasingly narrow, it is conceivable that some of the naturalness, familiarity, and vocal quality effects would change as a function of age, such that very close stimuli (e.g., speech versus another natural sound) initially leads to a weak preference, but, as infants age, this preference may be as strong as that found for very different stimuli (e.g., speech versus an artificial sound). While many articles discuss potential changes in the pattern of preference as a function of age (e.g. Ferry et al., 2013; Shultz & Vouloumanos, 2010; Shultz et al., 2014), only two papers from the same laboratory that we know of includes multiple age groups tested with the exact same stimuli and procedure (Vouloumanos et al., 2010; Vouloumanos & Werker, 2004), such that comparisons are often done using the demonstrably problematic method of stating “there is a difference here but not there, therefore there is an interaction” (REF).

A meta-analytic approach

In sum, previous work on infants’ preferences is broadly compatible with preference for natural over artificial, vocal over non-vocal, and familiar over unfamiliar sounds, potentially interacting with infants’ age. Some key predictions are unfortunately missing, including the effect of familiarity and the potential interactions with age. In this paper, we seek to shed light on these gaps by employing a meta-analytic approach.

Meta-analyses offer a way to test the effect of factors that are not part of the original design, by redescribing the stimuli used as a function of those factors and thus combining individual studies that tested different stimuli from a common category (e.g., natural sounds). Meta-analyses may also be able to reveal small effects not obvious in smaller studies because they can be based on combined, and therefore larger, samples. Additionally, by merging across studies carried out by different researchers and in different laboratories, they provide statistical evidence of how much we can trust that the results will replicate to new samples tested elsewhere: If an effect emerges in this combined dataset, it is more likely to be a replicable one that different labs can find. Finally, meta-analyses offers tools to detect publication bias in the literature, providing an index of to what extent global results are trustworthy.

More specifically in the present case, a meta-analysis allows to draw a developmental timeline across the age range covered by the literature, and statistically test how the different factors discussed in different individual studies interact with age in ways that individual studies on speech preference have not yet done. Importantly, previous meta-analyses in infant perception have provided important theoretical and empirical insights that contradicted qualitative reviews. For example, it has been proposed that infants' preference for novel or familiar items related to infants' age such that, all things equal, younger infants showed familiarity preferences whereas older infants exhibited novelty preferences (Hunter & Ames, 1988). However, Bergmann and Cristia (2016?) found stable familiarity preferences for word segmentation in natural speech across the first two years; and Black and Bergmann (2016) found a stable novelty effect for artificial grammars implemented in synthesized speech and a stable familiarity effect for those implemented in natural speech. Meta-analysis are therefore important to statistically and systematically test the theoretical predictions proposed in qualitative reviews. However, meta-analyses have an important limitation we should bear in mind. To begin with, we combine studies with different experimental designs together, which does not allow us to isolate specific variables as well as direct experimentation does. As a result, one can use them to focus on common factors between studies, but may miss subtle effects that require such experimental isolation. Given the scarcity of direct evidence on the potential explanations laid out above (naturalness, vocal quality, and familiarity, as a function of age), we conducted a meta-analysis to test whether infants' preference for speech sounds over other types of sounds is stable in newborns, and how it develops over the first year of life. Based on previous individual studies, we predicted that infants will show (see Figure 2): 1. a preference for speech over artificial sounds that was stable over development; 2. a greater preference for speech over other vocal sounds as a function of age; 3. a greater preference for native speech over foils as a function of age, whereas the opposite will take place for foreign speech.

In addition, since we were engaging in a systematic review for this meta-analysis, we decided to also include neuroimaging and electrophysiological studies. Although these studies can sometimes shed light on functional specialization, they are often used as proxies for the depth of treatment of information, and can thus provide indirect evidence on the perceptual basis for a preference. Put otherwise, if we find a difference in the neural bases, it is possible that a preference would be observed were infants of the same age tested with the same stimuli in a behavioral preference setup. In contrast, if there is no difference in processing at all, it seems unlikely that a preference could emerge.

Methods

To ensure quality, we followed recommendations from the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA; Moher et al., 2009). The PRISMA checklist and flowchart as well as other information can be found on the online supplementary materials (Anonymized, 2019).

Literature search

The information sources used to compose the initial list included suggestions by experts (authors of this work); two google scholar searches (" ("speech preference" OR "own-species vocalization") AND infant", and "("speech preference" OR "own-species vocalization") AND infant - "infant-directed" ") complemented with the same searches in PubMed and PsycInfo; and a google alert, as well as inspection of the reference lists of all full included papers.

Inclusion criteria

After a first screening based on titles and abstracts using more liberal inclusion criteria, we decided on inclusion based on full paper reading as follows. We included studies that tested human infants from birth to 1 year of age, and contrasted speech sounds with any other type of sound, measuring behavioral (e.g., looking times) or neurophysiological responses to the sounds (EEG, fNIRS, fMRI, MEG). We excluded studies that only contrasted foreign against native language, did not present natural speech sounds at all, presented speech in the mother's voice, or intentionally mixed speech with other vocal sounds within the same sound condition. We included published (i.e., journal articles) as well as unpublished works (i.e., doctoral dissertations as long as sufficient information was provided). A PRISMA flow chart summarizes the literature review and selection process (Figure 3). We documented all the studies that we inspected in a decision spreadsheet (available in the online supplementary materials; Anonymized, 2019).

[Insert Figure 3 here]

Coding

The full list of the variables coded is available in the supplementary material. Important to the present project are infant age, methodological variables (testing method: central fixation, head-turn preference procedure, high amplitude sucking/passive listening), and key stimuli characteristics. Specifically, we coded the language in which the speech sounds were recorded (native or foreign language), and natural, vocal for the sound opposed to speech. A sound was coded as natural if it was produced by a biological organism without any further acoustic manipulation. If the authors applied acoustic manipulations it was coded as artificial. A sound was considered as vocal if it was produced by an animal vocal tract, either original or modified.

Data were coded by the first author. In addition, 20% of the papers were selected to be coded by the second author independently, with disagreements resolved by discussion. There were 10 disagreements out of a total of 260 fields filled in indicative of the coders not following the codebook, which led to a revision of all data in four variables.

We coded all the statistical information reported in the included papers. If reported, we coded the mean score and the standard deviation for speech, and the other sound separately. When individual data was provided, we recomputed the respective mean scores and standard deviations based on the reported individual scores. If reported, we also coded the t-statistic between the two sound conditions, or an F-statistic provided this was a two-way comparison. If effect sizes were directly reported as a Cohen's d or a Hedges' g, we also coded this.

It was not feasible to code risk at the level of papers, since in this literature there are no cases of conflict of interest that are as easily discovered as in the medical literature.

Statistical analysis

Individual effect sizes. Once the data were coded, we computed individual effect sizes (Cohen's d) that were not directly reported in the papers, along with their respective variance. We first computed the correlation between the two measurements (e.g. looking time during speech and during monkey calls). We computed this correlation based on the t-statistic, the respective means and SDs (Lipsey & Wilson, 2001) if they were all reported. If not, we imputed this correlation randomly. We computed effect size using the respective means and SDs (Lipsey & Wilson, 2001). If they were not reported, we computed effect size based on t- or F-statistic (Dunlap et al., 1996). We finally calculated the variance of each effect size (Lipsey & Wilson, 2001)..

Effect sizes were first computed as Cohen's d, and then transformed to Hedges' g by multiplying d by a correction based on the degree of freedom (REF). The R code used to calculate effect sizes is available in the online supplementary materials (Anonymized, 2019).

Meta-analytic models. We first estimated the global effect size in this literature by fitting a meta-analytic regression without any moderator using the R package Robumeta**(REF)**. Next, we used the R package Robumeta **(REF)** to fit meta-analytic regressions that take into account the hierarchical structure of the data, including effect sizes possibly obtained from the same infants groups within papers. We specified the following moderators:

- mean age of children;
- experimental method (Central fixation, Head-turn Preference Procedure, High Amplitude Sucking/Passive Listening);
- familiarity with the language used (native or foreign);
- naturalness of the contrastive sound (coded as yes if it was natural and no otherwise).
- vocalness of the contrastive sound (coded as yes if it was vocal and no otherwise).
- interactions with age for familiarity, naturalness, and vocalness.

We did not center age. Centering age makes the intercept (global effect size) correspond to the mean age of the dataset, which would be preferable in datasets where age does not vary much or is not a crucial moderator. However, inspection of our predictions in Figure 1 reveals that we have clear expectations regarding main effects and interactions which are best reflected if age is not centered. The intercept (global effect size) in this model therefore corresponds to the preference found when age is zero.

```
## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: g_calc ~ 1
##
## Number of studies = 29
## Number of outcomes = 45 (min = 1 , mean = 1.55 , median = 1 , max = 3 )
## Rho = 0.8
## I.sq = 80.38913
## Tau.sq = 0.1489067
##
```

```
##                 Estimate StdErr t-value  dfs    P(|t|>) 95% CI.L 95% CI.U
## 1 X.Intercept.    0.478 0.0766    6.23 26.9 0.00000116     0.32    0.635
##   Sig
## 1 ***
## ---
## Signif. codes: < .01 *** < .05 ** < .10 *
## ---
## Note: If df < 4, do not trust the results

## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: g_calc ~ test_lang * mean_age_1 + natural * mean_age_1 + vocal * mean_age_1 +
##
## Number of studies = 29
## Number of outcomes = 45 (min = 1 , mean = 1.55 , median = 1 , max = 3 )
## Rho = 0.8
## I.sq = 79.74901
## Tau.sq = 0.1719252
##
##                            Estimate  StdErr t-value   dfs P(|t|>) 95% CI.L
## 1           X.Intercept.  0.61408 0.75668   0.812  6.94   0.444 -1.17827
## 2             test_lang2 -0.52687 0.69383  -0.759 10.02   0.465 -2.07240
## 3             mean_age_1 -0.00340 0.00526  -0.646  2.42   0.574 -0.02263
## 4               natural1 -0.10810 0.42229  -0.256  4.16   0.810 -1.26251
## 5                 vocal1 -0.15907 0.36515  -0.436  4.21   0.685 -1.15328
## 6               methodCF  0.58562 0.53771   1.089  2.83   0.360 -1.18528
## 7              methodHPP  0.73931 0.48707   1.518  2.26   0.254 -1.14292
## 8  test_lang2.mean_age_1  0.00316 0.00287   1.102  8.45   0.301 -0.00339
```

```
## 9    mean_age_1.natural1  0.00306 0.00299   1.026  6.34   0.343 -0.00415
## 10     mean_age_1.vocal1 -0.00227 0.00303  -0.749  6.62   0.479 -0.00951
##    95% CI.U Sig
## 1   2.40644
## 2   1.01866
## 3   0.01584
## 4   1.04631
## 5   0.83515
## 6   2.35652
## 7   2.62154
## 8   0.00971
## 9   0.01028
## 10  0.00497
## ---
## Signif. codes: < .01 *** < .05 ** < .10 *
## ---
## Note: If df < 4, do not trust the results

## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: g_calc ~ natural * mean_age_1
##
## Number of studies = 29
## Number of outcomes = 45 (min = 1 , mean = 1.55 , median = 1 , max = 3 )
## Rho = 0.8
## I.sq = 80.93995
## Tau.sq = 0.1700603
##
```

```
##                        Estimate  StdErr t-value  dfs P(|t|>) 95% CI.L
## 1         X.Intercept.  0.589570 0.36139   1.631 3.96   0.179 -0.41749
## 2             natural1 -0.137200 0.39044  -0.351 8.99   0.733 -1.02061
## 3           mean_age_1 -0.000655 0.00297  -0.221 3.06   0.839 -0.00999
## 4 natural1.mean_age_1  0.000695 0.00304   0.228 4.99   0.828 -0.00713
##   95% CI.U Sig
## 1  1.59662
## 2  0.74621
## 3  0.00868
## 4  0.00852
## ---
## Signif. codes: < .01 *** < .05 ** < .10 *
## ---
## Note: If df < 4, do not trust the results
```

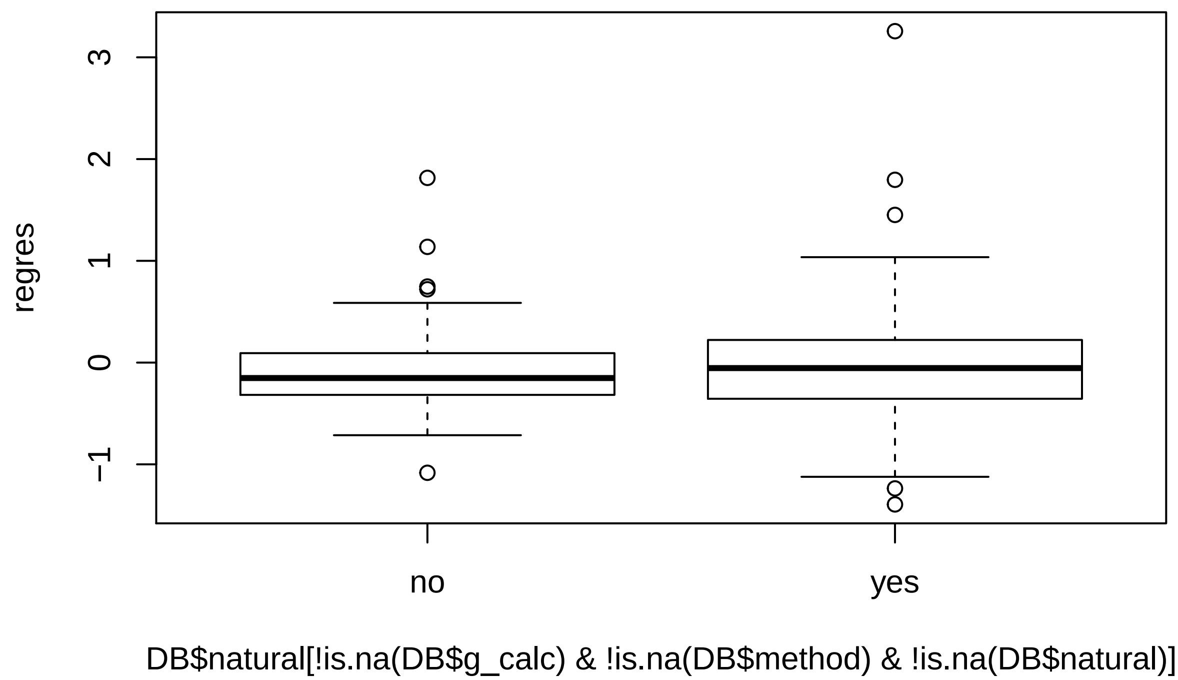
```
## RVE: Correlated Effects Model with Small-Sample Corrections
##
## Model: g_calc ~ test_lang + vocal * mean_age_1
##
## Number of studies = 29
## Number of outcomes = 45 (min = 1 , mean = 1.55 , median = 1 , max = 3 )
## Rho = 0.8
## I.sq = 80.85368
## Tau.sq = 0.1772125
##
##                      Estimate  StdErr t-value   dfs P(|t|>) 95% CI.L
## 1      X.Intercept.  0.639783 0.33516   1.909 16.14  0.0742 -0.07024
## 2        test_lang2  0.035320 0.17727   0.199 14.78  0.8448 -0.34301
```

```
## 3            vocal1 -0.408076 0.31403  -1.299 12.54  0.2172 -1.08902
## 4        mean_age_1 -0.000436 0.00355  -0.123  4.49  0.9076 -0.00989
## 5 vocal1.mean_age_1  0.001172 0.00358   0.327  6.79  0.7533 -0.00735
##   95% CI.U Sig
## 1  1.34980   *
## 2  0.41365
## 3  0.27287
## 4  0.00902
## 5  0.00969
## ---
## Signif. codes: < .01 *** < .05 ** < .10 *
## ---
## Note: If df < 4, do not trust the results
```

```
##
##        animal      backward    BPfiltered       com_voc complex_tones
##             1            22             2             7             1
##    duck_calls      emotions environmental     heartbeat   heartspeech
##             2             2             3             2             2
##    HPfiltered human_walking    LPfiltered  monkey_calls         music
##             2             1             5            11             3
##   non-com_voc     scrambled           SWS         water   white_noise
##             7             2            10             4             2
```

```
##        animal      backward    BPfiltered       com_voc complex_tones
##    1.07499753    0.33642725    0.37521239    0.12226565           NaN
##    duck_calls      emotions environmental     heartbeat   heartspeech
##    1.26795256   -0.16579236    0.27993843    1.02131722    1.43229167
```

```
##    HPfiltered human_walking    LPfiltered  monkey_calls         music
##    0.83689393    0.48158905    1.23310038    0.42686707   -0.30122164
##   non-com_voc     scrambled           SWS         water   white_noise
##    0.38626556    0.02319572    0.33526236    0.36152402    1.08797229
```



Publication bias

We assessed the presence of a potential publication bias in the body of literature by symmetrizing the funnel plot with the “trim and fill” method (Duval, 2005). To do so, we needed to use the simplest regression, without any moderators. We tested the asymmetry of the funnel plot by regressing effect size as a function of effect size standard error and running a Kendall’s tau rank test.

Results

Database description

We found a total of 29 papers reporting 98 (not mutually independent) effect sizes, see Table 1. All of them have been submitted to or published in peer-reviewed journals. Studies tended to have small sample sizes, with a median N of 20 children (Range = 55, M = 21.39, Total: 963). Infants ranged from 0 to 12 months (1.21 to 380.50 days). Individual samples comprised 42 % of female participants on average. Infants were native of 7 different languages across the whole database (English, French, Japanese, Italian, Russian, Yiddish, Hebrew). Studies were performed in 13 different laboratories from 6 different countries (United States, Canada, Israel, France, Japan, Italy). 4 experimental methods were used: 11 studies used Passive Listening (PL) with neuro-imaging, with 8 studies using Near-Infrared Spectroscopy (NIRS), and 3 studies using fMRI; 12 studies used Central Fixation (CF); 3 used High-Amplitude Sucking (HAS); and 1 used Head-turn Preference Procedure (HPP).

Summary effect size

We found a mean weighted effect size g =0.48 (SE = 0.08 CI = [0.32 , 0.63]).

Publication bias

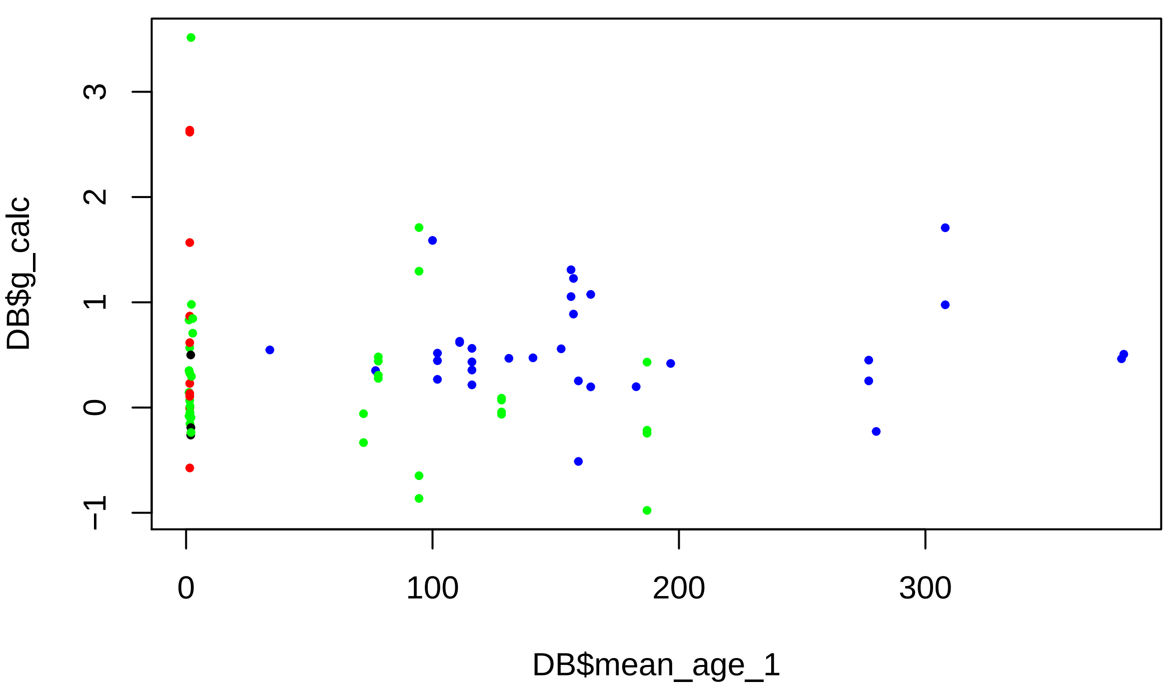
Evidence of bias at level of papers Evidence of bias at level of literature

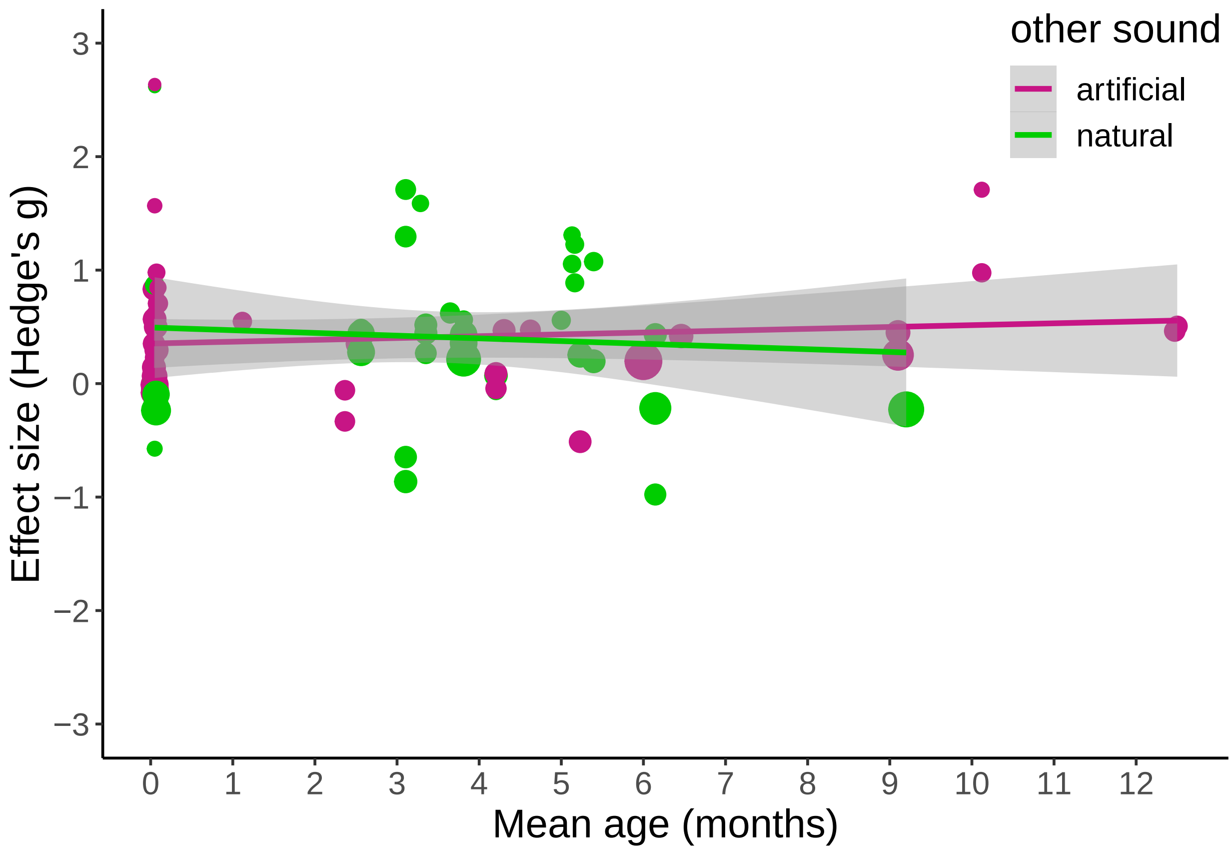
```
##
## Regression Test for Funnel Plot Asymmetry
##
## model:     mixed-effects meta-regression model
```

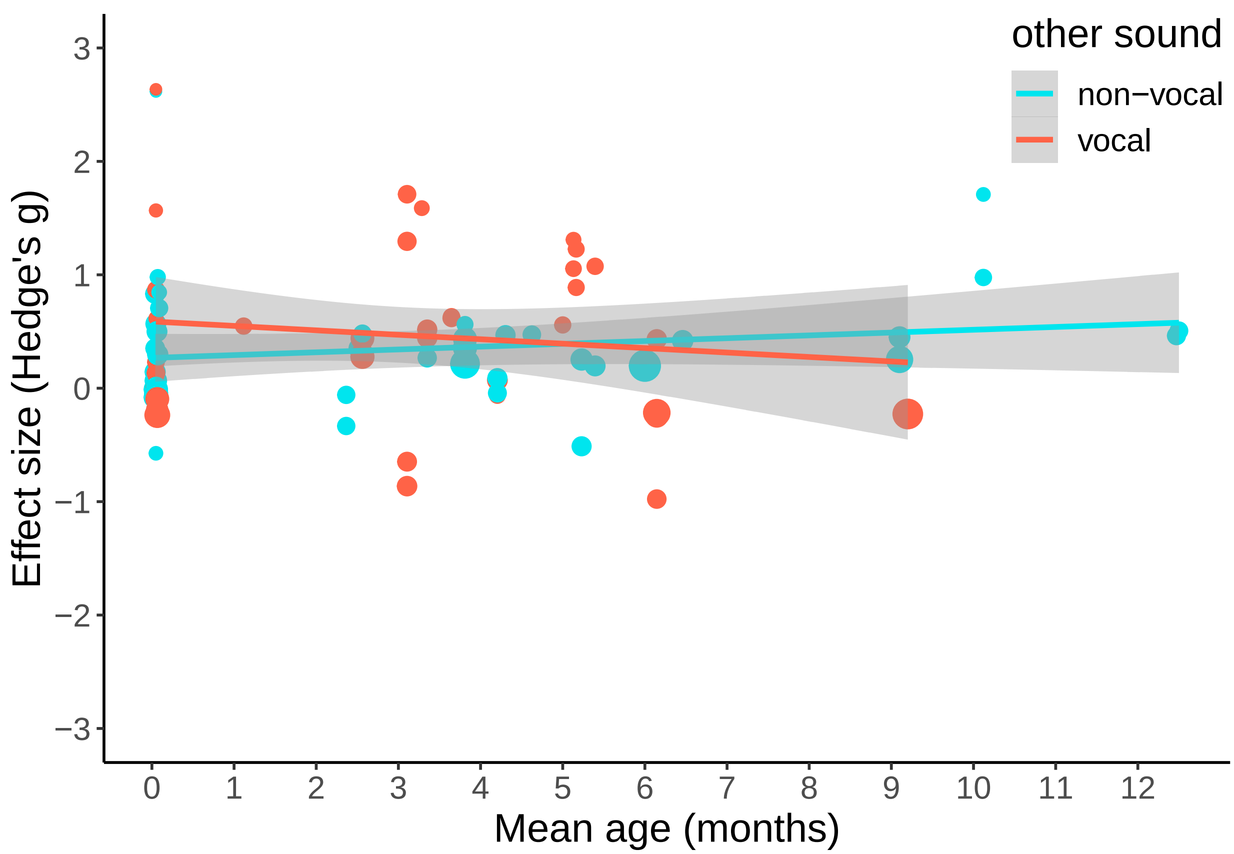
```
## predictor: standard error
##
## test for funnel plot asymmetry: z = 6.8548, p < .0001

##
## Rank Correlation Test for Funnel Plot Asymmetry
##
## Kendall's tau = 0.3550, p < .0001
```

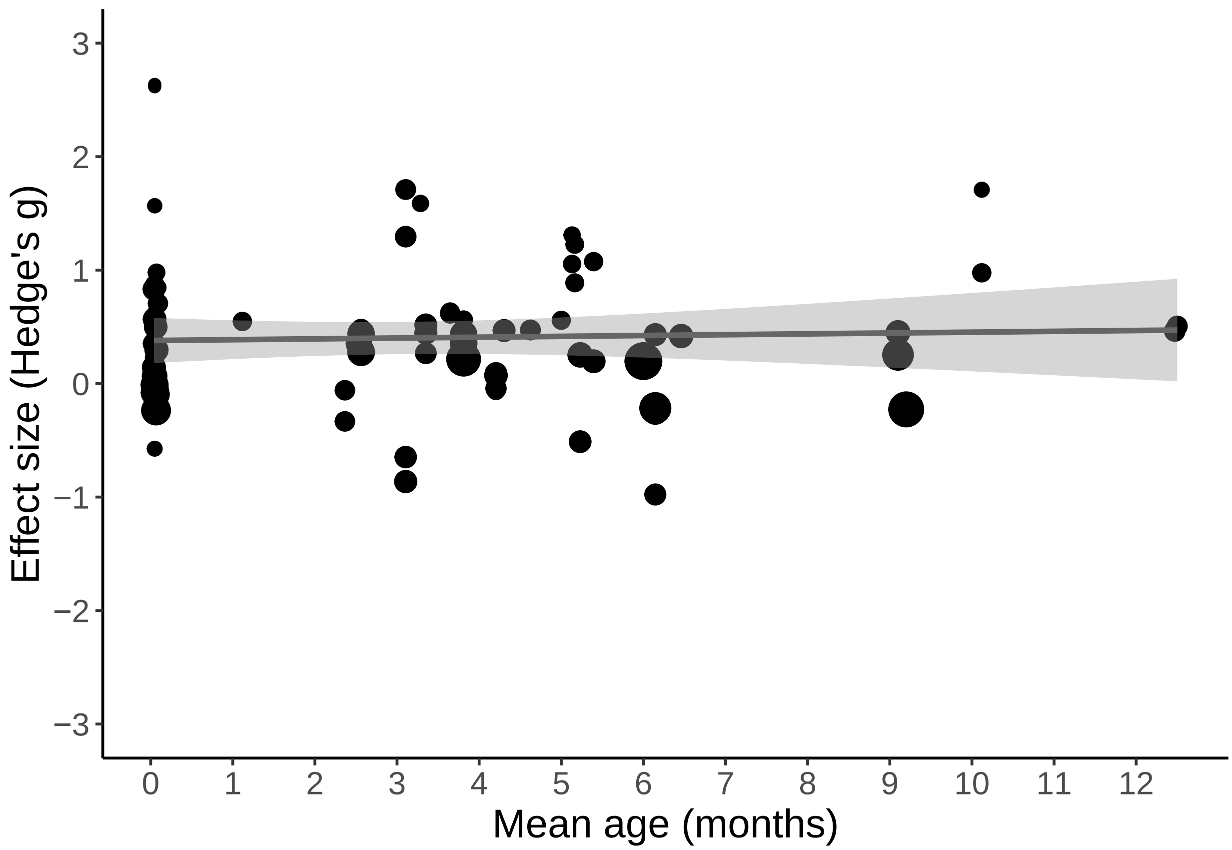
Main effects

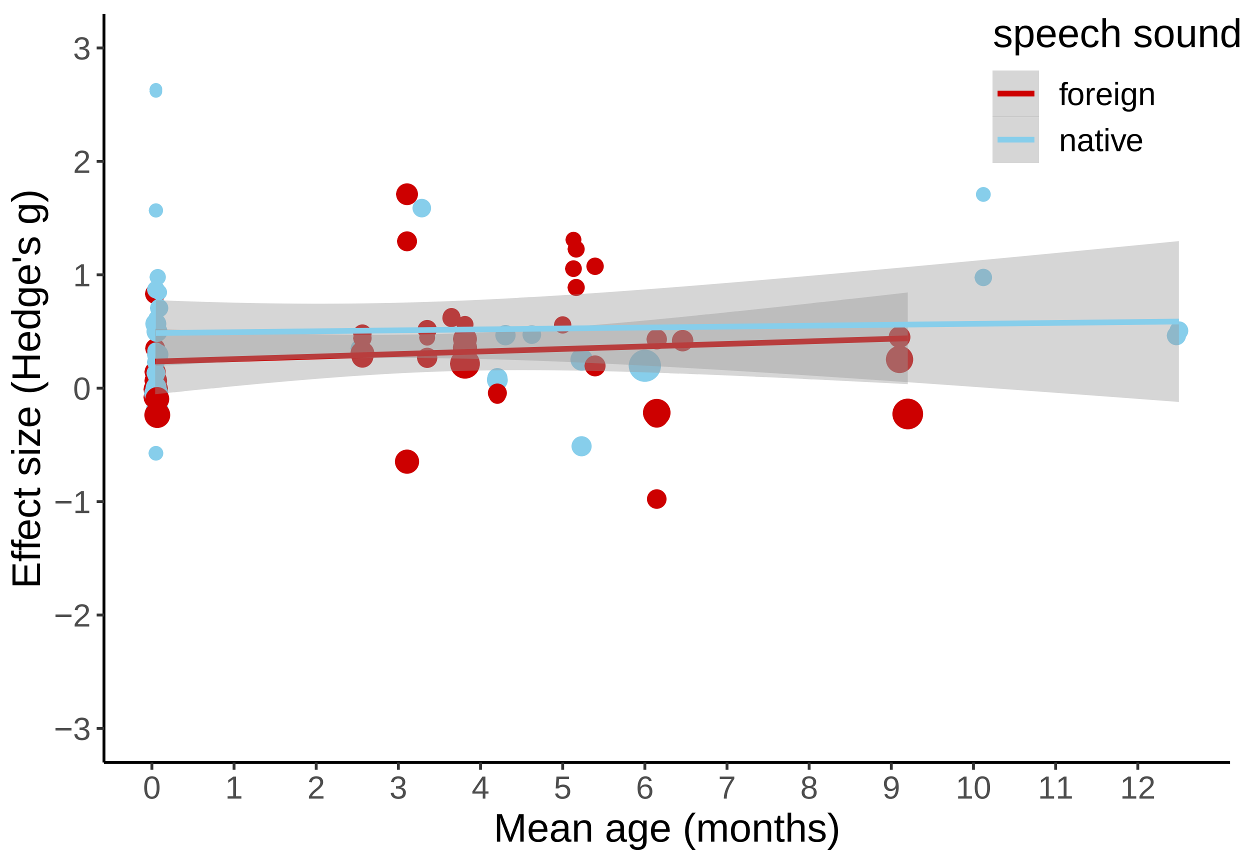
```
##      HAS       CF      HPP       PL
## "black"   "blue"    "red" "green"
```




other sound
artificial
natural
Effect size (Hedge's g)
3
2
1
0
−1
−2
−3
0 1 2 3 4 5 6 7 8 9 10 11 12
Mean age (months)


other sound
non−vocal
vocal
Effect size (Hedge's g)
3
2
1
0
−1
−2
−3
0 1 2 3 4 5 6 7 8 9 10 11 12
Mean age (months)


Effect size (Hedge's g)
Mean age (months)
3
2
1
0
−1
−2
−3
0 1 2 3 4 5 6 7 8 9 10 11 12


speech sound
foreign
native
Effect size (Hedge's g)
Mean age (months)
3
2
1
0
−1
−2
−3
0 1 2 3 4 5 6 7 8 9 10 11 12

Heterogeneity Moderators age, type & interaction

Discussion

Our meta-analysis synthesizes the available literature on infants' preference for speech sound. Our results confirm those of individual studies, showing that infants reliably prefer speech over other types of sounds from birth. When all studies were considered together, we found a sizable intercept at birth. For comparison, the main effect for native vowel discrimination is estimated at XX (Tsuji & Cristia, 2013); that for preference for familiarized words in passages at XX (Bergmann & Cristia,).

A key goal was to interrogate this effect further to assess whether three potential explanations for it hold up, both at birth and with potential interactions with age: A preference for natural over artificial, vocal over non-vocal, and familiar over unfamiliar sounds. We discuss each in turn.

We reviewed evidence that the auditory system encodes better natural than artificial sounds (REF), and that naturalness makes a difference for a higher-level linguistic task, namely word segmentation (Black & Bergmann, 2016). In view of such results, we expected infants' speech preference to be larger when the competitor was an artificial sound than when it was a natural one. In fact, we found that naturalness alone did not significantly moderate infants' preference for speech sounds. Therefore, we can confidently state that a low-level difference in processing of natural versus artificial sounds is unlikely to explain away speech preference. speech triggers different cognitive mechanisms than other sounds. Not incompatible as in this meta-analysis natural speech stimuli when contrasted to only synthetic speech. We don't know if infants would have been able to find "words" with natural sounds other than syllables (i.e. frequent sequences of several natural sounds).

Some previous work had found a preference for speech over other natural sounds, which may have been indicative of a preference for vocal sounds. If so, one would have expected greater speech preferences when the competitor was non-vocal than when it was a vocal sound. Our meta-analytic regression, however, revealed no significant difference between these two contrasts.

Do infants show a greater preference for speech when the speech stimuli are in their native language, than when they are in a foreign language? For this factor as well, we could not disprove the null hypothesis of no difference, with widely overlapping distributions of effect sizes for studies using native as opposed to foreign speech stimuli (controlling for the competitor via the other factors discussed above).

We hypothesized age to play a major role, not only because it was correlated with experiences that are crucial for the formation of categories underlying some of the factors above (e.g., vocalness, nativeness), but also because it may correlate with a reshaping of the category definition for speech itself. Indeed, studies comparing processing of human speech against human non-speech as well as animal vocalizations more generally (REF, REF) often discuss these age-related differences in categorization of these sounds. Surprisingly, age did not significantly moderate the overall preference for speech, and it did not interact with the other three factors.

There are at least two potential interpretations of the overall pattern of results. One of them is that, from birth and regardless of changes co-occurring with age, infants show a preference for speech, which cannot be reduced to three simpler explanations: naturalness, vocalness, or familiarity (represented here by the native/foreign contrast). There exists one simple explanation that we could not test here because there were not enough studies with appropriate conditions, and that is the possibility that infants prefer speech because of its complex acoustic structure and fast transitions. To explore this explanation, it would be ideal to carry out acoustic analyses of the actual stimuli used in the studies. Thus, we recommend interested researchers to gather more data in which the competitor is acoustically simple versus complex; and to deposit the actual stimuli in a public archive such as the Open Science Framework (REF). It would also be important to carry out more tests on infants older than 9 months. Language production gains in complexity at about this age REF, and it is possible that this would affect infants' speech preference. We particularly recommend using as competitor natural vocal stimuli, and as target foreign speech, which would help fill in an important gap in our dataset.

More data in general would be helpful also in view of the second potential interpretatio

For interested readers who intend to collect such data, we recommend caution when design

Given the crucial importance of understanding infants' speech preference, we make the fo

Infants' preference for speech is a fascinating phenomenon. Beyond the human species, the capacity to recognize signals from conspecifics might be crucial for survival and it may be present in other social species (e.g. mice, or great apes). To our knowledge, such studies have not been carried out, but we hope they are in the future. More specific to our species, preferential processing of speech may support higher level cognitive tasks, such as categorization (Waxman, 1997; 2007; 2010, Ferry). And the preference itself may be a meaningful index of processing that can be used to identify children at risk REF. For all of these reasons, it is important to take stock of what we know today. Our meta-analysis compiled public results on this key phenomenon, establishing that there is a small to medium effect size associated with it. This preference was not modulated by age, nor three characteristics of the stimuli employed (whether the speech was native or foreign; whether the competitor was natural or artificial, vocal or non-vocal). The analyses also suggested publication bias, for which recommendations were made for researchers, reviewers, and editors.

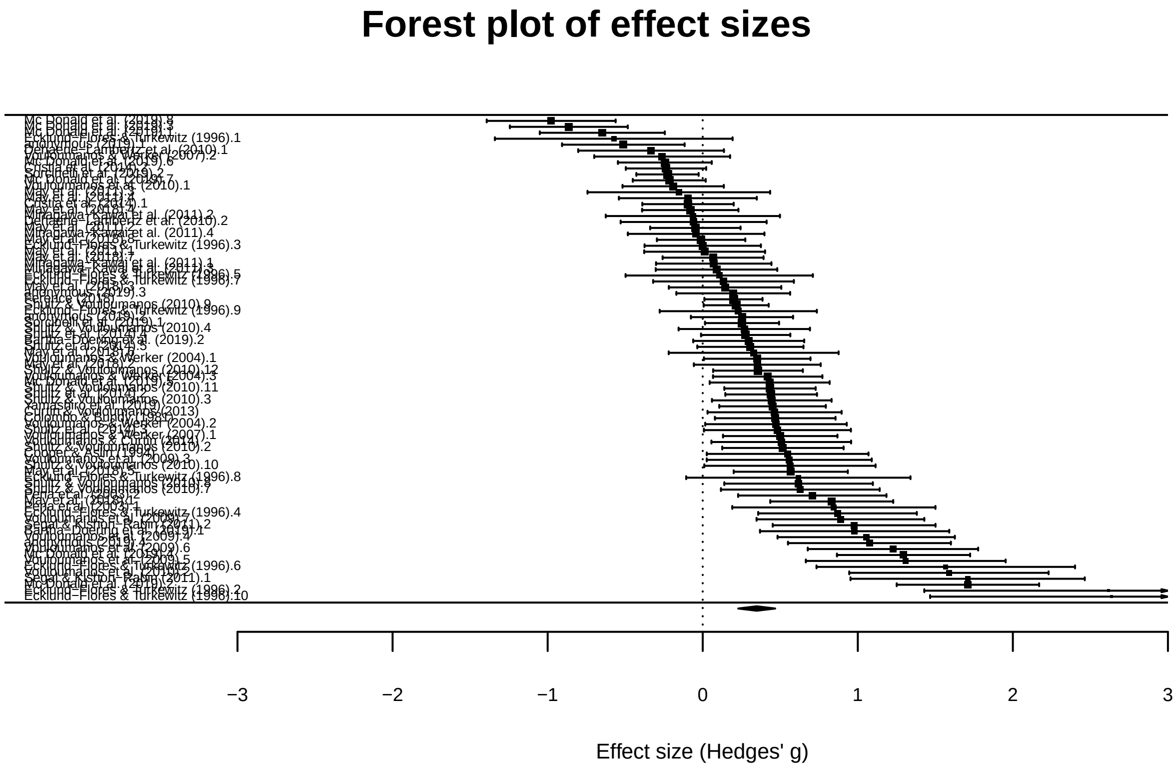


Figure 1

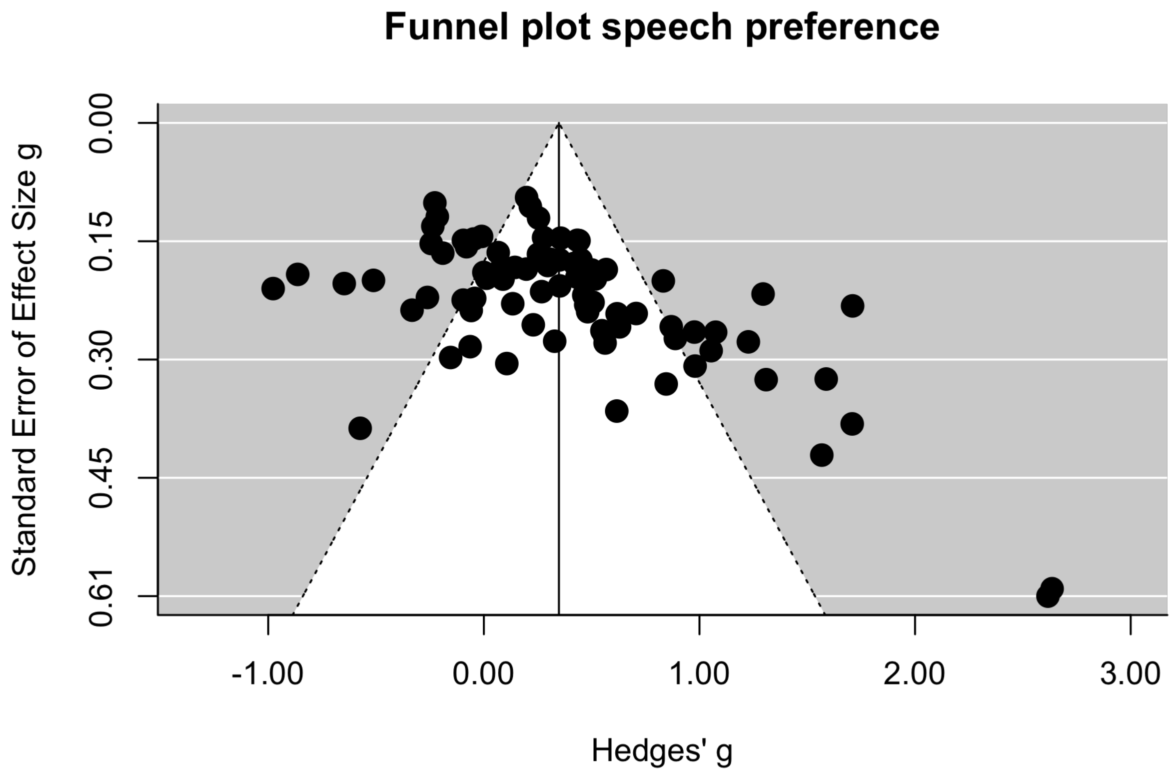


Figure 2

(#fig:publication bias)