

Infants' preference for speech decomposed: Meta-analytic evidence

Cécile Issard¹, Sho Tsuji², & Alejandrina Cristia¹

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

² International Research Center for Neurointelligence, The University of Tokyo

This work was supported by an Agence Nationale de la Recherche grant to A.C. (ANR-17-CE28-0007 LangAge, ANR-16-DATA-0004 ACLEW, ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017); and the J. S. McDonnell Foundation Understanding Human Cognition Scholar Award to A.C.

The authors declare no conflict of interest. Funding sources did not take part in study design, data collection or analysis.

Our data is fully available in the corresponding OSF repository: <http://tidy.ws/bqjc4U>

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

Previous experimental work suggested that infants initially prefer a broad range of natural or vocal sounds, and then tune in to speech. To check whether this explanation holds for the entire body of literature, we conducted a meta-analysis of experiments testing speech preference in infants. Infants reliably preferred speech over other sounds, but this preference was not significantly moderated by vocal quality, or naturalness of the competitor. We found no effect of age: infants showed the same strength of preference throughout the first year of life. Speech therefore appears to be preferred from birth, even to other natural or vocal sounds. These results contradict current views of the literature, and call for further investigation of the phenomenon.

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

Word count:

Infants' preference for speech decomposed: Meta-analytic evidence

Highlights

- Infants reliably prefer natural speech over other types of sounds, from birth to the end of the first year of life
- Speech is preferred over both artificial and other natural sounds
- Speech is preferred over both non-vocal and other vocal sounds
- The difference between whether infants are familiar or not with the language used was not significant

Introduction

How do infants build an organized auditory world? Among the numerous auditory objects they have to recognize, speech is probably the most important. Speech as a way to study the development of object recognition.

Acoustic communication is a crucial behavior in humans that support social interactions. Given its significance, previous work argued for an early precursor of such social communication, manifested as an early preference for speech over other types of sound (Ecklund-Flores & Turkewitz, 1996; Vouloumanos & Werker, 2007; Vouloumanos, Hauser, Werker, & Martin, 2010). This auditory bias would initially encompass a broad range of natural or vocal sounds (Figure 1), and then tune in to species-specific communicative vocalizations, namely speech (e.g. Ferry, Hespos, & Waxman, 2013; Shultz & Vouloumanos, 2010; Shultz, Vouloumanos, Bennett, & Pelphrey, 2014; Vouloumanos & Werker, 2004; Vouloumanos et al., 2010). However, this hypothesis has never been directly tested, and other factors might also explain it. there are only few attempts to quantitatively assess the role of different potential factors driving this preference. Here, we synthesize the available

empirical data on infants' preferences for speech over non-speech sounds to assess the explanatory role of the factors cited in the literature.

Potential dimensions underlying preference patterns

Three main factors have been suggested to drive infants' early speech preference: Familiarity, naturalness, and vocal quality. We will review the extant literature on these before turning to proposed changes over development.

Familiarity. A first hypothesis is that infants prefer speech to other sounds since it is a frequent sound in their experience. Newborns prefer their native speech to prosodically distinct foreign speech (e.g., Mehler et al., 1988; Moon, Cooper, & Fifer, 1993), which supports a preference for sound patterns heard frequently in the womb. If this was indeed the case, infants would show a stronger preference for speech over other sounds when tested in their native language, and a weaker preference when tested with a foreign language. Two different experiments testing three-month-old infants with the same stimuli found a similar amount of preference as compared to monkey calls whether speech was in the native language (Vouloumanos et al., 2010) or a foreign language (Shultz & Vouloumanos, 2010), but results from neuroimaging experiments provide contradictory evidence. Newborns' brain activation was different for forward than backward speech when the native language was used as the speech stimuli, but not when a foreign language was used (May, Gervain, Carreiras, & Werker, 2018). Qualitative review does therefore not provide clear evidence for familiarity as a driving factor of infants' preference for speech. As these experiments used very different competitors, it is likely that familiarity with the language used interacts with properties of the competitor to drive infants' preference for speech.

Naturalness. A second hypothesis postulates a preference for natural over artificial sounds. Natural sounds are those produced by biological systems, such as heart beat, the

sound of walking, or animal vocalizations, and environmental/geophysical sounds, such as wind, rain, or the sound of a river. Natural sounds are processed more efficiently by the auditory system, from the cochlea (Smith & Lewicki, 2006) to the auditory cortex (see Mizrahi, Shalev, & Nelken, 2014 for a review). As a result, this should be reflected in infants behavior, with a preference for natural sounds, including speech, over artificial competitors from birth. Consistently, newborns increased their sucking rate more during speech than during an artificial equivalent, sine-wave speech (Vouloumanos & Werker, 2007), and three-month-olds did not listen significantly longer to speech than to another type of natural sound, environmental sounds (Shultz & Vouloumanos, 2010). However, other experiments demonstrate that infants still prefer speech over other natural sounds. Three-month-olds listened longer to speech than to water sounds (Shultz & Vouloumanos, 2010), and newborns made more head-turns to speech than to heartbeat (Ecklund-Flores & Turkewitz, 1996). These results suggest that being a natural sound can be a contributing factor, but not the only explanation for infants' speech preference.

Vocal quality. Perhaps the most widely cited hypothesis across the literature is that of an initial preference for vocal sounds in general, that later restricts to speech (e.g. Vouloumanos et al., 2010). Vocal sounds are characterized by modulations introduced by the vocal tract, with harmonically related energy peaks. In contrast, backward speech has unnatural formant transitions and seemingly abrupt closures that cannot be produced by the vocal tract¹. Accordingly, newborns listened equivalently to speech and monkey calls (Vouloumanos et al., 2010), and neuro-imaging results failed to report differential response to speech, human non-communicative vocalizations, and rhesus calls in 1- to 4-month-old infants (Shultz et al., 2014).

¹In the case of filtered speech, the modulations introduced by the vocal tract are still present at the retained frequencies, and formant transitions are consistent with vocal production constraints. For this reason, filtered speech can be considered as vocal but not natural. Because the womb acts as a low-pass filter, newborn infants are familiar with low-pass filtered speech, but this familiarity fades after birth.

Changes as a function of development

Development may affect the preference for speech in various ways. Whereas newborns do not prefer speech over monkey calls, there is some evidence that three-month-olds do (Vouloumanos et al., 2010, but see Shultz et al. (2014)). This suggests that as they age, infants might develop an increasingly narrow definition of the stimulus they prefer. In this case naturalness, familiarity, and vocal quality effects might differentially change as a function of age: 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). 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). However, statements about the factors driving this phenomenon are often done using the demonstrably problematic method of concluding that there is an interaction without actually testing for it statistically (e.g. Gelman & Stern, 2006). To our knowledge, only two papers from the same laboratory include multiple age groups tested with the exact same stimulus categories and procedure (Vouloumanos & Werker, 2004; Vouloumanos et al., 2010). It is therefore important to directly test these statements with larger datasets, ideally across the whole literature.

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. In this paper, we seek to directly test these interpretations of the literature by employing a meta-analytic approach.

Meta-analyses are quantitative syntheses of experiments testing a comparable phenomenon with comparable approaches. As such, they can reveal small effects not obvious

in individual experiments by combining them to obtain larger samples. By integrating data across different laboratories, they provide evidence for the generalizability of effects across labs. Finally, meta-analyses offer tools to detect publication bias in the literature.

Meta-analyses can also involve combining experiments that may vary in their methodology. One limitation is therefore that one cannot isolate specific variables as well as in direct experimentation, and meta-analyses may miss subtle effects.

Specifically for the present case, a meta-analysis allows to statistically test different explanations. We can test the effect of factors that are not part of the original design, by redescribing the stimuli used as a function of those factors. For instance, a study measuring preference for native speech over native backward speech provides data on a natural versus artificial, as well as a vocal versus non-vocal contrast. We can also draw a developmental timeline across the age range covered by the literature.

Meta-analyses can even provide theoretical and empirical insights that contradict 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 (2017) found a stable novelty effect for artificial grammars implemented in synthesized speech, whereas those implemented in natural speech led to stable familiarity preferences. Meta-analyses are therefore important to statistically and systematically test the theoretical predictions proposed in qualitative reviews.

The present study

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 reliable in newborns, and how it develops over the first year of life. Assuming all three factors are true, and further assuming that the definition of the preferred stimulus narrows with age, we predicted that infants will show (see Figure 2):

1. a greater preference for native speech over non-speech as a function of age, but a smaller preference for foreign speech over non-speech with age;
2. a greater preference for speech over other natural sounds as a function of age, but a preference for speech over artificial sounds that is stable over development;
3. a greater preference for speech over other vocal sounds as a function of age, but a stable preference for speech over non-vocal sounds over development.

Methods

Literature search

We composed the initial list of papers with suggestions by experts (authors of this work); one google scholar search ((*“speech preference” OR “own-species vocalization”*) AND *infant* - *“infant-directed”*), the same search in PubMed and PsycInfo (last searched on 2019-09-24); and a google alert. We also inspected the reference lists of all included papers. Finally, we emailed a major mailing list to ask for missing data. We received only two replies, one of which revealed a formerly undiscovered published study, but no unpublished data was made available to us.

Inclusion criteria

After a first screening based on titles and abstracts using more liberal inclusion criteria, we decided on final inclusion based on full paper reading. We included experiments that tested human infants from birth to one year of age, and contrasted speech sounds with any other type of sound, measuring behavioral preferences to the sounds (e.g., looking times). We excluded experiments that did not contrast speech to another sound type, such as experiments contrasting two different speech sounds (e.g. foreign against native language, or adult vs. child-directed speech), or on the opposite did not present natural speech sounds at all (e.g. backward speech vs animal vocalizations). We excluded experiments that presented speech in the mother's voice to avoid confounds between speech and individual voice recognition. Experiments that intentionally mixed speech with other vocal sounds within the same sound condition were also excluded. Finally, we excluded neuroimaging experiments to avoid mixing results from different brain regions with different response profiles. The full list of the papers that were inspected but excluded is available in the supplementary materials (https://osf.io/4stz9/?view_only=d0696591ebf34bfc8430f848cd945ca8). 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 papers that we inspected in a decision spreadsheet (available in the online supplementary materials; https://osf.io/4stz9/?view_only=d0696591ebf34bfc8430f848cd945ca8).

Coding

The critical variables for our purpose are infant age, methodological variables (testing method: central fixation, high amplitude sucking, head-turn preference procedure), and key

stimuli characteristics. Specifically, we coded the language in which the speech sounds were recorded (native or foreign), and whether the sound opposed to speech was natural or not, as well as vocal or not. This competitor was coded as natural if it was produced by a biological organism without any further acoustic manipulation. Natural competitors included animal calls, environmental sounds (e.g. wind or water sounds), heartbeat, bird song, non-speech vocalizations (e.g. laughs or coughs). If the authors applied acoustic manipulations it was coded as artificial. Artificial competitors included sine-wave speech, filtered speech, white noise, instrumental music, and speech with altered rhythmic structure. The only exception was for newborn experiments presenting low-pass filtered speech mimicking the filtering applied by the womb. Given the recency of the intra-uterine environment to newborns (about 2 days), we coded these as natural. The competitor sound was considered as vocal if it was produced by an animal vocal tract (human or not), either original or modified. Vocal competitors included non-speech vocalizations, animal calls, bird songs, and filtered speech. Non-vocal competitors included backward speech (that has abrupt closures that cannot be produced by the vocal tract), white-noise, environmental sounds, instrumental music, heartbeat, and sine-wave speech (that lacks the harmonic structure introduced by the natural resonance of the vocal tract). If a paper reported results from neurotypical and at-risk infants, we coded only the data from the neurotypical group.

Data were coded by the first author. In addition, 20% of the papers were randomly selected to be coded by the last author independently, with disagreements resolved by discussion. There were 10 disagreements out of a total of 260 fields filled in, and they were 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 infant-level 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.

The PRISMA checklist, data, and code can be found on the online supplementary materials (https://osf.io/4stz9/?view_only=d0696591ebf34bfc8430f848cd945ca8).

Effect sizes

Once the data were coded, we extracted effect sizes, along with their respective variance. Effect sizes were standardized differences (Cohen's d) between response to speech and to the other sound. If they were not directly reported in the papers, we computed them using the respective means and SDs (Lipsey & Wilson, 2001), or a t- or F-statistic (Dunlap, Cortina, Vaslow, & Burke, 1996). As our effect sizes came from within-subject comparisons (e.g. looking time of the same infant during speech and during monkey calls), we needed to take into account the correlation between the two measurements in effect sizes and effect size variances computations. We computed this correlation based on the t-statistic, the respective means and SDs (Lipsey & Wilson, 2001) if they were all reported; or imputed this correlation randomly if not. We finally calculated the variance of each effect size (Lipsey & Wilson, 2001). Cohen's d were transformed to Hedges' g by multiplying d by a correction for small sample sizes based on the degree of freedom (Borenstein, Hedges, Higgins, & Rothstein, 2011).

We did not center age because our hypotheses included a developmental progression from birth to the end of the first year of life. We were therefore interested in the intercept at age 0 (i.e. birth).

All analyses use the R (R Core Team, 2018) package Robumeta (Hedges, Tipton, &

Johnson, 2010), which allows to fit meta-analytic regressions that take into account the correlated structure of the data, when repeated measures are obtained from the same infant groups within papers.

Results

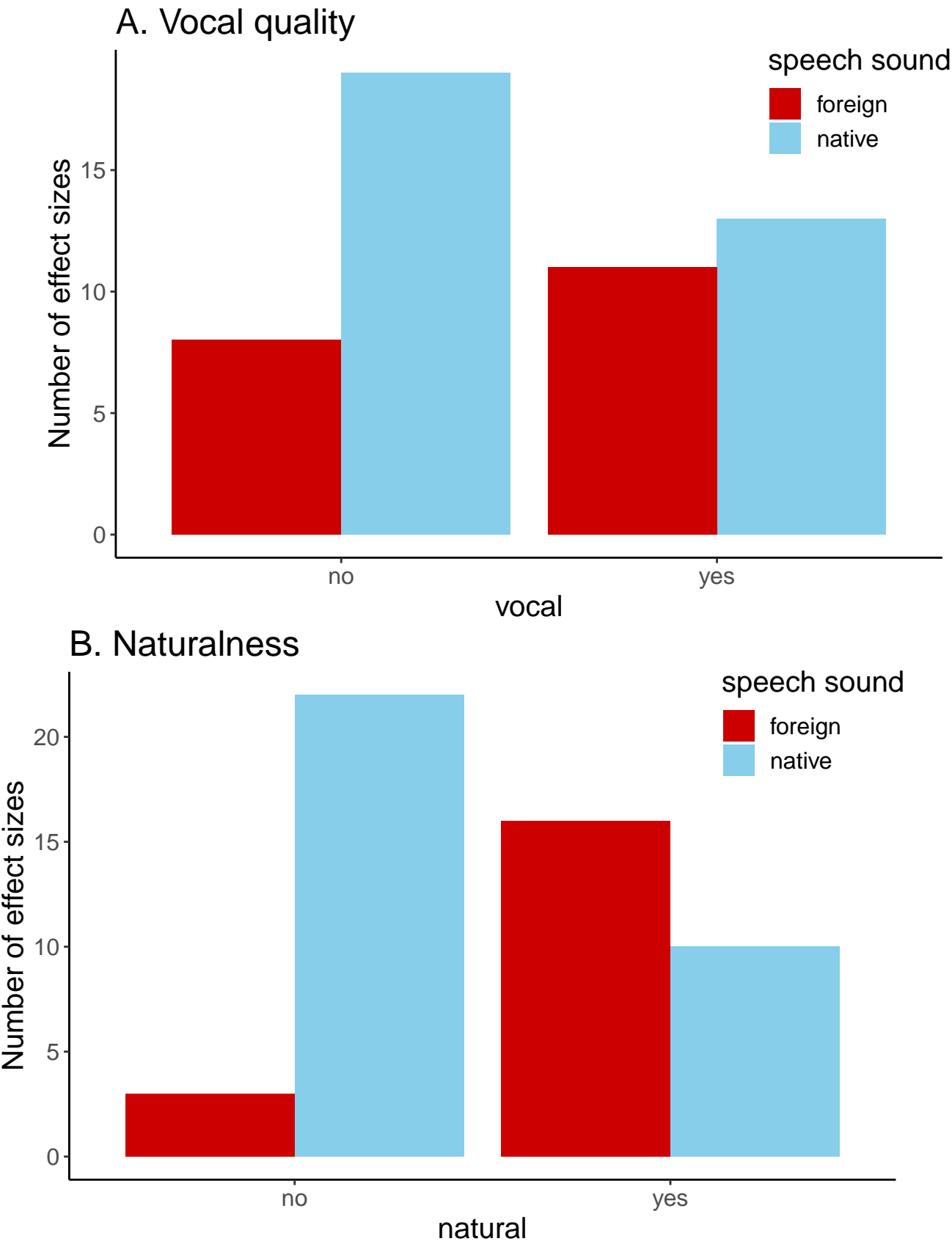
Database description

We found a total of 17 publications (labeled with an asterisk in the reference list) reporting 32 experiments, and 51 (not mutually independent) effect sizes, see Figure 5. 16 papers have been submitted to or published in peer-reviewed journals (Colombo & Bundy, 1981; Cooper & Aslin, 1994; Curtin & Vouloumanos, 2013; Ecklund-Flores & Turkewitz, 1996; Santolin, Russo, Calignano, Saffran, & Valenza, 2019; Segal & Kishon-Rabin, 2011; Shultz & Vouloumanos, 2010; Sorcinelli, Ference, Curtin, & Vouloumanos, 2019; Spence & DeCasper, 1987; Vanden Bosch der Nederlanden & Vouloumanos, 2020; Vouloumanos & Curtin, 2014; Vouloumanos & Werker, 2004, 2007; Vouloumanos, Druhen, Hauser, & Huizink, 2009; Vouloumanos et al., 2010, 2010; Yamashiro, Curtin, & Vouloumanos, 2019). The remaining 1 publication contributing 1 effect size was a thesis (J. D. Ference, 2018).

Experiments tended to have small sample sizes, with a median N of 16 children (Range = [60, 4], M = 19.35, Total: 699). Infants ranged from 0 to 12 months (1.50 to 380.50 days), although the majority were under 6 months of age (75% of the experiments). Individual samples comprised 45% of female participants on average. Infants were native of 6 different languages across the whole database (English, French, Russian, Yiddish, Hebrew, Italian). Experiments were performed in 10 different laboratories from 4 different countries (United States, Canada, Israel, Italy). 3 experimental methods were used: 24 experiments used Central Fixation (CF) (also called sequential looking preference procedure) (Colombo & Bundy, 1981; Cooper & Aslin, 1994; Curtin & Vouloumanos, 2013; J. D. Ference, 2018;

Santolin et al., 2019; Segal & Kishon-Rabin, 2011; Shultz & Vouloumanos, 2010; Sorcinelli et al., 2019; Vanden Bosch der Nederlanden & Vouloumanos, 2020; Vouloumanos & Curtin, 2014; Vouloumanos & Werker, 2004; Vouloumanos et al., 2009, 2010; Yamashiro et al., 2019); 3 used High-Amplitude Sucking (HAS) (Spence & DeCasper, 1987; Vouloumanos & Werker, 2007; Vouloumanos et al., 2010); and 5 used Head-turn Preference Procedure (HPP) (Ecklund-Flores & Turkewitz, 1996). Trial length was fixed in 8 experiments, and infant-controlled in 23 experiments.

Speech sounds were spoken by a female in 30 out of 32 experiments, with an infant-directed prosody in 15 out of the 32 experiments. Speech was presented in isolated segments (i.e. words or syllables) in 5 experiments, and full sentences or passages in 13 experiments. Speech stimuli were recorded in the infant native language in 62.50% of the experiments. Strikingly, experiments using the infants' native language tested infants from 0 to 12 months of age, whereas experiments using a foreign language only tested infants from 3 to 9 months of age (see Figure 9). The competitor sound was vocal in 50% of the experiments. The competitor sound was natural 50% of the experiments. The stimuli characteristics are summarized on Figures ?? and 4.



Average effect size

Integrating across all experiments in a meta-analytic regression without any moderator, we found an average effect size g of 0.36 ($SE = 0.06$, $CI = [0.23, 0.48]$) (Table 1, and Figure 1, diamond), corresponding to a medium effect size. Heterogeneity among effect sizes was estimated at $\tau^2 = 0.11$ ($I^2 = 74.44\%$), which was significant ($Q = 170.72$, $p < 0.001$) despite the removal of outliers before running the model.

Publication bias

We assessed the presence of a potential publication bias in the body of literature by studying the relationship between standard errors of effect sizes as a function of Hedges' g (see funnel plot in Figure 6).² A regression test on these data was significant ($z = 7.31$, $p = 0.00$), as was the Kendall's tau rank correlation test for funnel plot asymmetry (Kendall's $\tau = 0.50$, $p = 0.00$), consistent with a publication bias in the literature. To further investigate this bias, we symmetrized the funnel plot with the "trim and fill" method (Duval & Tweedie, 2000). To symmetrize the funnel plot, 13 ($SE = 4.53$) missing experiments were needed on the left side of the plot. The corrected effect size was estimated at 0.25 ($SE = 0.06$) after filling in the 13 missing experiments.

Moderator analyses

We then tested if the heterogeneity found above could be explained by the dimensions discussed in the literature. Following our hypotheses, we fit a meta-analytic model with the

²If the literature is not biased, effect sizes should be evenly distributed around the mean effect size, with increasing standard error as they go away from the mean effect size (both in the positive and negative directions, white triangle in the funnel plot). This is reflected by a symmetrical funnel plot, with no linear relationship between effect sizes and standard errors.

following moderators:

- mean age of children;
- familiarity with the language used (native or foreign);
- naturalness of the contrastive sound (coded as yes if it was natural and no otherwise);
- vocal quality of the contrastive sound (coded as yes if it was vocal and no otherwise).

None of the moderators was significant (see Figures 7, 8, and 9; and Table 1).

Due to the relatively low number of effect sizes available in the literature, we did not add interactions with age to the model to avoid overfitting. The confidence intervals of all conditions overlap almost exactly across the tested ages, excluding the possibility of such interactions (Table 2, and Figure 7, 8, and 9). The absence of interaction between age and other moderators was confirmed by three separate models for each moderator, that did not yield any significant effect or interaction (Table 3, 4, and 5).

Discussion

Our meta-analysis synthesizes the available literature on infants' preference for speech sounds. When all experiments were considered together with no moderators, we found a sizable intercept ($g=0.36$). For comparison, the average effect for native vowel discrimination using looking time methods is estimated at 0.25 (Tsuji & Cristia, 2014, data inspected in <http://metalab.stanford.edu> on 2019-10-18). These results are in line with adult neuroimaging data showing distinct responses to speech as compared to other natural or environmental sounds, even in low-level auditory regions (Norman-Haignere & McDermott, 2018; Norman-Haignere, Kanwisher, & McDermott, 2015). The significant heterogeneity we found among the literature suggests that underlying factors modulate this effect. We had predicted infants' speech preference to be larger when the competitor was an artificial sound than when it was a natural one; when the competitor was non-vocal; and when the speech

was in the infants' native language. In fact, we were unable to disprove the null hypothesis of no difference for all three factors. Distributions of effect sizes for experiments varying along the three dimensions widely overlap, and the confidence intervals of all conditions overlap almost exactly (Figures 6, 7, and 8). Moreover, Table 1 shows that the estimate for all these factors is close to zero (the maximum being 0.21). Our findings therefore suggest that none of these parameters fully explain infants' preference for speech sounds.

We had also hypothesized age to play a major role, because it may correlate with a reshaping of the category definition for speech itself. Indeed, experiments comparing processing of human speech against human non-speech as well as animal vocalizations more generally (McDonald et al., 2019; Vouloumanos et al., 2010) often discuss these age-related differences in categorization of these sounds. Surprisingly, age did not significantly moderate the overall preference for speech, as shown by the null estimate of this moderator (Table 1). This result was replicated in a separate model for age only, which showed a significant intercept, similar to the intercept found in the meta-regression with no moderator (Supplementary results S5). As age was not centered, this intercept provides an estimate of the effect size at age 0, i.e. at birth. Moreover, the scatterplot of effect sizes as a function of age reveals clearly no change with age even when plotted without other moderators (Supplementary figure S6). This null effect of age, combined with the sizeable intercept, confirms that infants reliably prefer speech over other types of sounds from birth.

In other words, from birth on, infants show a preference for speech, which cannot be reduced to the three simpler explanations tested here: naturalness, vocalness, or familiarity (represented here by comparing effects for native versus competitor against those for foreign versus competitor). This capacity to preferentially listen to speech sounds from birth suggests that infants are born with the capacity to recognize their conspecifics' communication signals. This parallels what has been proposed by the Conspec model for faces: Infants would be born with knowledge about faces, enabling them to orient their

attention toward them, even without any prior exposure to faces (Morton & Johnson, 1991). A capacity to recognize conspecifics, either from their face or their voice, is consistent with the similar processing stages between voice and face perception (Yovel & Belin, 2013). The fact that familiarity with the language used in the experiment did not modulate infants' preference suggests that exposure did not play a crucial role for speech either. However, contrary to faces, fetuses are exposed to speech that is low-pass filtered by the womb throughout the last trimester of gestation (Lecanuet & Granier-Deferre, 1993; Querleu, Renard, Versyp, Paris-Delrue, & Crèpin, 1988). It is therefore possible that prenatal experience with low-pass filtered speech helps infants to form a representation of speech, independently of the language spoken.

basic perceptual abilities are present at birth and become increasingly differentiated and refined over age Turati 2004. Filters, potentially corresponding to neurons response profiles (Moore & Woolley, 2019).

The Conspec model proposes that faces would be detected because of their spatial structure (Morton & Johnson, 1991). Similarly, it is possible that infants prefer speech because of its complex acoustic structure and fast transitions (Rosen & Iverson, 2007). Spectral or temporal modulations taken separately are not sufficient to elicit neural responses similar to the ones elicited by speech (Minagawa-Kawai, Cristià, Vendelin, Cabrol, & Dupoux, 2011). However, speech is characterized by joint spectrotemporal modulations at specific rates (Singh & Theunissen, 2003). It is possible that infants attune to this specific spectro-temporal structure (though see Norman-Haignere & McDermott, 2018 showing that they only explain neural responses in primary auditory cortex). Testing this explanation would require to compute the modulation spectra of the actual stimuli used in the experiments. Thus, we recommend interested researchers to deposit their stimuli in a public archive such as the Open Science Framework (Foster & Deardorff, 2017).

Ultimately, preferential processing of speech may support higher level cognitive tasks.

The human species is a highly social one. Detecting speech signals would allow to integrate it with other sensory percepts, such as faces, to form multisensory representations of conspecifics (Vouloumanos et al., 2009). This would lay the track for social cognition. Identifying speech signals and paying attention to them would allow infants to form complex representations of the sensory world, that they can manipulate cognitively. Infants could categorize visual stimuli (i.e., associate a label to a category of objects) when they were associated to speech, but not pure tones or backward speech (Ferry, Hespos, & Waxman, 2010; Ferry et al., 2013; Fulkerson & Waxman, 2007). Interestingly, infants categorized visual stimuli when presented with speech, melodies, monkey (Fulkerson & Haaf, 2003), or lemur vocalizations (Ferry et al., 2013). These results support the idea that infants may preferentially process complex sounds. Finally, the preference itself may also be a meaningful index of processing that can be used to identify children at risk (Sorcinelli et al., 2019). Understanding this phenomenon is therefore crucial for both theoretical and clinical advances.

The median sample size at present is 16, which is close to the field standard (Bergmann et al., 2018) but much lower than current recommendations (Oakes, 2017). Well-chosen sample sizes are crucial for powerful experiments (Simmons, Nelson, & Simonsohn, 2011). Our meta-analysis provides the average effect size across the literature, which will, in turn, allow researchers to run power analyses to determine the sample size they need in future experiments.

We want to close by drawing the reader's attention to XX limitations of the present study. "Finally, we recognize that readers might have several concerns about the methods used here. First, how can we say that..." (Chen & Wiens, 2020) First, the distribution of effect sizes in the literature is consistent with publication bias, in view of a strong asymmetry of the funnel plot. The trim-and-fill method suggested 13 points may be missing. Given that we have 51 effect sizes in total, a fifth more would be missing. The missing

experiments are in the negative section, i.e., a preference *against* speech, a result that could lead authors to doubt their own data and not submit it to journals, or that would be considered odd by reviewers and editors, who may ask that the data be removed (or who may recommend the paper to be rejected altogether). These missing experiments constitute an important limitation of our results. The literature being biased toward positive effect sizes, the true effect size might be smaller than the one we found (vertical line on Figure 5). To correct this bias, we invite researchers to use registered reports (Kiyonaga & Scimeca, 2019). In this new publication scheme (available for Developmental Science, Infancy, Infant Behavior and Development, and Journal of Child Language at the time of writing, see a full up-to-date list on <https://cos.io/rr/>), manuscripts are submitted before data are collected. Reviewers and editors make publication decisions based solely on the introduction and methods. The paper is then reviewed once more for readability, but it cannot be rejected if the results are surprising or uncomfortable for the field. This would facilitate the publication of experiments failing to report a speech preference, or actually reporting a preference for the competitor, which would in turn help to draw a more accurate picture of the phenomenon. Our dataset can be community-augmented, and we invite researchers investigating this phenomenon to complement it with any data they would have (https://osf.io/4stz9/?view_only=d0696591ebf34bfc8430f848cd945ca8), whatever the results and publication status. [SHOW THAT IT LIKELY DOESN'T CHANGE THE RESULTS. WE WERE ABLE TO SHOW AN OVERALL PREFERENCE, EVEN AFTER CORRECTING FOR PUBLICATION BIAS. REASONING AND SPECULATION FOR MODERATORS WITH MORE NEGATIVE RESULTS?] “Overall, we acknowledge that there are many potential sources of error in our study, but none seem likely to overturn our main conclusions.”(Chen & Wiens, 2020)

Our meta-analysis revealed uneven distributions of experiments across age and stimulus dimensions. One may think that this uneven age distribution could be one potential reason for the lack of an age effect. However, by aggregating the numerous individual studies

of a literature, meta-analyses gain statistical power. As such, meta-analytic results have more cumulative explanatory value than any single study. Previous experiments suggested that infants initially don't discriminate speech from other vocal sounds, and then tune to speech around five months of age. With a sample size of 987 infants, covering a much wider age range than individual experiments, our meta-analysis provide evidence for an alternative view of the phenomenon, namely that infant prefer speech from birth onwards, a preference that cannot be fully explained by stimulus vocalness, or familiarity with the sounds of the language used. This clearly points to the value of meta-analysis: to advance a field and inspire follow-up studies. In particular, future experiments should test infants older than 9 months. Language production gains in complexity at about this age (Oller, Eilers, Neal, & Schwartz, 1999), which could affect infants' speech preference. Experiments using natural vocal stimuli as competitor, and foreign speech as target, would contribute to fill in the gap in the literature that our meta-analysis revealed.

[CONCLUDE REMINDING THE OPEN QUESTIONS OUR MA RAISES]

References

- Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, 19(6), 901–917. <https://doi.org/10.1111/desc.12341>
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting Replicability in Developmental Research Through Meta-analyses: Insights From Language Acquisition Research. *Child Development*, 89(6), 1996–2009. <https://doi.org/10.1111/cdev.13079>
- Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In (pp. 124–129). Cognitive Science Society. Retrieved from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2475527
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.
- * Colombo, J., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*, 4, 219–223. [https://doi.org/10.1016/S0163-6383\(81\)80025-2](https://doi.org/10.1016/S0163-6383(81)80025-2)
- * Cooper, R. P., & Aslin, R. N. (1994). Developmental Differences in Infant Attention to the Spectral Properties of Infant-Directed Speech. *Child Development*, 65(6), 1663–1677. <https://doi.org/10.2307/1131286>
- * Curtin, S., & Vouloumanos, A. (2013). Speech Preference is Associated with Autistic-Like Behavior in 18-Months-Olds at Risk for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 43(9), 2114–2120.

<https://doi.org/10.1007/s10803-013-1759-1>

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1(2), 170–177. <https://doi.org/10.1037/1082-989X.1.2.170>

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>

* Ecklund-Flores, L., & Turkewitz, G. (1996). Asymmetric headturning to speech and nonspeech in human newborns. *Developmental Psychobiology*, 29(3), 205–217. [https://doi.org/10.1002/\(SICI\)1098-2302\(199604\)29:3<205::AID-DEV2>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2302(199604)29:3<205::AID-DEV2>3.0.CO;2-V)

* Ference, J. D. (2018). The Role of Attentional Biases for Conspecific Vocalizations. <https://doi.org/http://dx.doi.org/10.11575/PRISM/31878>

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones. *Child Development*, 81(2), 472–479. <https://doi.org/10.1111/j.1467-8624.2009.01408.x>

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38), 15231–15235. <https://doi.org/10.1073/pnas.1221166110>

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association : JMLA*, 105(2), 203–206.

<https://doi.org/10.5195/jmla.2017.88>

- Fulkerson, A. L., & Haaf, R. A. (2003). The Influence of Labels, Non-Labeling Sounds, and Source of Auditory Input on 9- and 15-Month-Olds' Object Categorization. *Infancy*, 4(3), 349–369. https://doi.org/10.1207/S15327078IN0403_03
- Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not Tones) facilitate object categorization: Evidence from 6- and 12-month-olds. *Cognition*, 105(1), 218–228. <https://doi.org/10.1016/j.cognition.2006.09.005>
- Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, 60(4), 328–331. <https://doi.org/10.1198/000313006X152649>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, 69–95.
- Kiyonaga, A., & Scimeca, J. M. (2019). Practical Considerations for Navigating Registered Reports. *Trends in Neurosciences*, 42(9), 568–572. <https://doi.org/10.1016/j.tins.2019.07.003>
- Lecanuet, J.-P., & Granier-Deferre, C. (1993). Speech Stimuli in the Fetal Environment. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 237–248). Dordrecht: Springer Netherlands.

https://doi.org/10.1007/978-94-015-8234-6_20

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA, US: Sage Publications, Inc.

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2018). The specificity of the neural response to speech at birth. *Developmental Science*, 21(3), n/a–n/a.
<https://doi.org/10.1111/desc.12564>

McDonald, N. M., Perdue, K. L., Eilbott, J., Loyal, J., Shic, F., & Pelphrey, K. A. (2019). Infant brain responses to social sounds: A longitudinal functional near-infrared spectroscopy study. *Developmental Cognitive Neuroscience*, 36, 100638.
<https://doi.org/10.1016/j.dcn.2019.100638>

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoni, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178.
[https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)

Minagawa-Kawai, Y., Cristià, A., Vendelin, I., Cabrol, D., & Dupoux, E. (2011). Assessing Signal-Driven Mechanisms in Neonates: Brain Responses to Temporally and Spectrally Different Sounds. *Frontiers in Psychology*, 2.
<https://doi.org/10.3389/fpsyg.2011.00135>

Mizrahi, A., Shalev, A., & Nelken, I. (2014). Single neuron and population coding of natural sounds in auditory cortex. *Current Opinion in Neurobiology*, 24, 103–110.
<https://doi.org/10.1016/j.conb.2013.09.007>

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16(4), 495–500.

[https://doi.org/10.1016/0163-6383\(93\)80007-U](https://doi.org/10.1016/0163-6383(93)80007-U)

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A Two-Process Theory of Infant Face Recognition. *Psychological Review*, 98(2), 164–181.

Norman-Haignere, S. V., & McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLOS Biology*, 16(12), e2005127.
<https://doi.org/10.1371/journal.pbio.2005127>

Norman-Haignere, S. V., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, 88(6), 1281–1296. <https://doi.org/10.1016/j.neuron.2015.11.035>

Oakes, L. M. (2017). Sample Size, Statistical Power, and False Conclusions in Infant Looking-Time Research. *Infancy*, 22(4), 436–469. <https://doi.org/10.1111/infa.12186>

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders*, 32(4), 223–245. [https://doi.org/10.1016/S0021-9924\(99\)00013-1](https://doi.org/10.1016/S0021-9924(99)00013-1)

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 28(3), 191–212. [https://doi.org/10.1016/0028-2243\(88\)90030-5](https://doi.org/10.1016/0028-2243(88)90030-5)

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>

Rosen, S., & Iverson, P. (2007). Constructing adequate non-speech analogues: What is special about speech anyway? *Developmental Science*, 10(2), 165–168.

<https://doi.org/10.1111/j.1467-7687.2007.00550.x>

- * Santolin, C., Russo, S., Calignano, G., Saffran, J. R., & Valenza, E. (2019). The role of prosody in infants' preference for speech: A comparison between speech and birdsong. *Infancy*, 24(5), 827–833. <https://doi.org/10.1111/infa.12295>
- * Segal, O., & Kishon-Rabin, L. (2011). Listening Preference for Child-Directed Speech Versus Nonspeech Stimuli in Normal-Hearing and Hearing-Impaired Infants After Cochlear Implantation. *Ear and Hearing*, 32(3), 358–372. <https://doi.org/10.1097/AUD.0b013e3182008afc>
- * Shultz, S., & Vouloumanos, A. (2010). Three-Month-Olds Prefer Speech to Other Naturally Occurring Signals. *Language Learning and Development*, 6(4), 241–257. <https://doi.org/10.1080/15475440903507830>
- Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science*, n/a–n/a. <https://doi.org/10.1111/desc.12151>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6), 3394. <https://doi.org/10.1121/1.1624067>
- Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439(7079),

978–982. <https://doi.org/10.1038/nature04485>

- * Sorcinelli, A., Ference, J., Curtin, S., & Vouloumanos, A. (2019). Preference for speech in infancy differentially predicts language skills and autism-like behaviors. *Journal of Experimental Child Psychology*, 178, 295–316.
<https://doi.org/10.1016/j.jecp.2018.09.011>
- * Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development*, 10(2), 133–142.
[https://doi.org/10.1016/0163-6383\(87\)90028-2](https://doi.org/10.1016/0163-6383(87)90028-2)
- Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, 56(2), 179–191. <https://doi.org/10.1002/dev.21179>
- * Vanden Bosch der Nederlanden, C. M., & Vouloumanos, A. (2020). Infant biases for detecting speech in complex scenes. *Under Review*.
- * Vouloumanos, A., & Curtin, S. (2014). Foundational Tuning: How Infants' Attention to Speech Predicts Language Development. *Cognitive Science*, 38(8), 1675–1686.
<https://doi.org/10.1111/cogs.12128>
- * Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, 7(3), 270–276.
<https://doi.org/10.1111/j.1467-7687.2004.00345.x>
- * Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, 10(2), 159–164.
<https://doi.org/10.1111/j.1467-7687.2007.00549.x>
- * Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old

infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, 106(44), 18867–18872.

<https://doi.org/10.1073/pnas.0906049106>

- * Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, 81(2), 517–527.

<https://doi.org/10.1111/j.1467-8624.2009.01412.x>

- * Yamashiro, A., Curtin, S., & Vouloumanos, A. (2019). Does an Early Speech Preference Predict Linguistic and Social-Pragmatic Attention in Infants Displaying and Not Displaying Later ASD Symptoms? *Journal of Autism and Developmental Disorders*.

<https://doi.org/10.1007/s10803-019-03924-2>

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices.

Trends in Cognitive Sciences, 17(6), 263–271.

<https://doi.org/10.1016/j.tics.2013.04.004>

Table 1

Statistical results of meta-regression without any moderator.

	estimate	SE	t	confidence interval
average effect size	0.36	0.06	5.74	0.23 - 0.48

Table 2

Statistical results of meta-regression with all moderators. The intercept corresponds to the effect size when speech is in a foreign language, and the competitor is natural, and vocal, at age 0. The moderator estimates correspond to changes in the intercept when the target stimuli are in the native language (familiarity); the competitor is artificial (naturalness); and the competitor is non-vocal (vocal quality).

	estimate	SE	t	confidence interval
intercept	0.36	0.15	2.42	0.04 - 0.67
familiarity	-0.08	0.12	-0.67	-0.35 - 0.18
naturalness	0.21	0.11	1.89	-0.03 - 0.46
vocal quality	-0.20	0.17	-1.22	-0.56 - 0.16
age	0.00	0.00	0.55	0 - 0

Table 3

Statistical results of meta-regression with naturalness and its interaction with age as moderators. The intercept corresponds to the effect size when the competitor is natural, at age 0. The moderator estimates correspond to changes in the intercept when the competitor is artificial (naturalness).

	estimate	SE	t	confidence interval
intercept	0.29	0.24	1.22	-0.32 - 0.9
naturalness	0.07	0.26	0.27	-0.49 - 0.63
naturalness*age	0.00	0.00	0.06	-0.01 - 0.01

Table 4

Statistical results of meta-regression with vocal quality and its interaction with age as moderators. The intercept corresponds to the effect size when the competitor is vocal, at age 0. The moderator estimates correspond to changes in the intercept when the competitor is non-vocal (vocal quality).

	estimate	SE	t	confidence interval
intercept	0.44	0.17	2.61	0.04 - 0.83
vocal quality	-0.32	0.20	-1.58	-0.75 - 0.11
age	0.00	0.00	0.80	0 - 0.01

Table 5

Statistical results of meta-regression with language of the speech sounds and interaction with age. The intercept corresponds to the effect size when speech is in a foreign language at age 0. The moderator estimates correspond to changes in the intercept when the target stimuli are in the native language (familiarity).

	estimate	SE	t	confidence interval
intercept	0.63	0.31	1.99	-0.11 - 1.37
familiarity	-0.37	0.33	-1.12	-1.1 - 0.37
familiarity * age	0.00	0.00	1.24	0 - 0.01

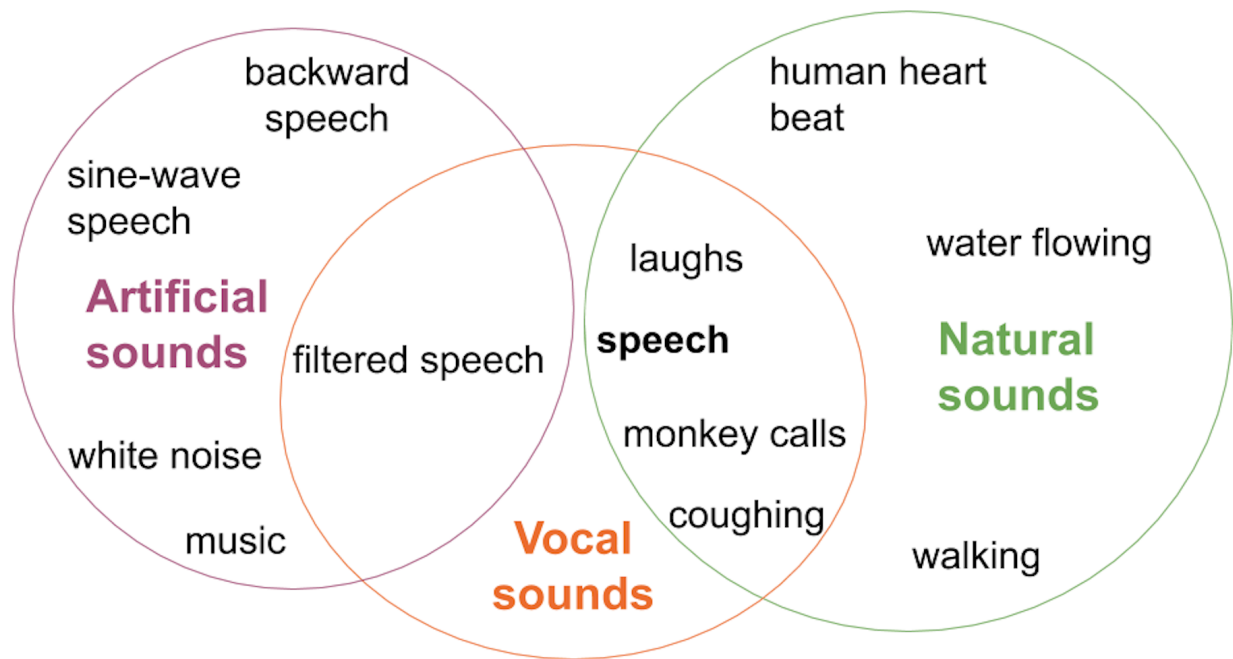


Figure 1. Speech is a natural, vocal sound.

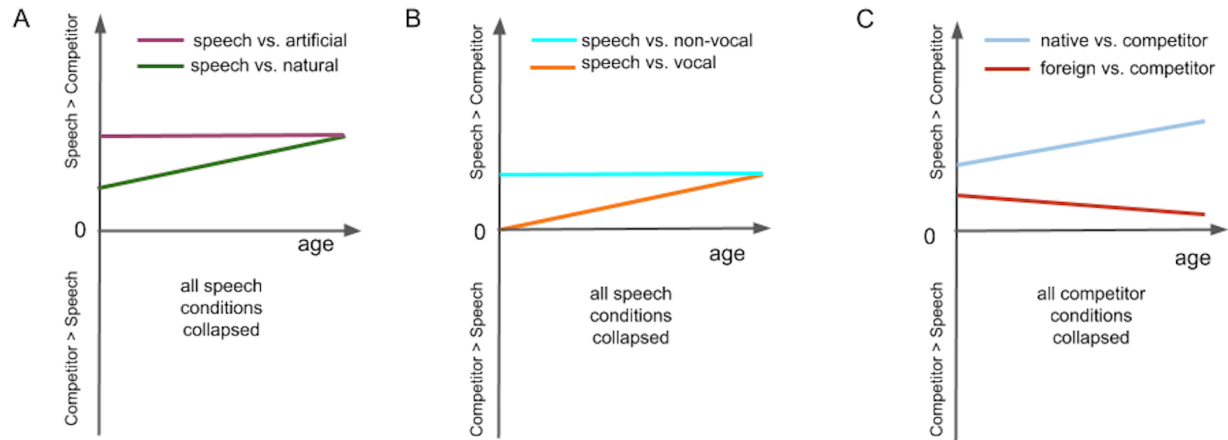


Figure 2. Hypothesized pattern of preference: the x axis shows age, the y axis represents the effect size derived from the contrast between a speech condition and a competitor condition (preference for speech over the competitor is plotted up; the lower quadrants are empty because we do not predict a preference for the competitor over speech). A: Speech contrasted to natural (green) or artificial (purple) competitors. B: Speech contrasted to vocal (orange) or non-vocal (cyan) competitors. C: Collapsing across competitors, separating speech in a foreign language (red); speech in the native language (blue).

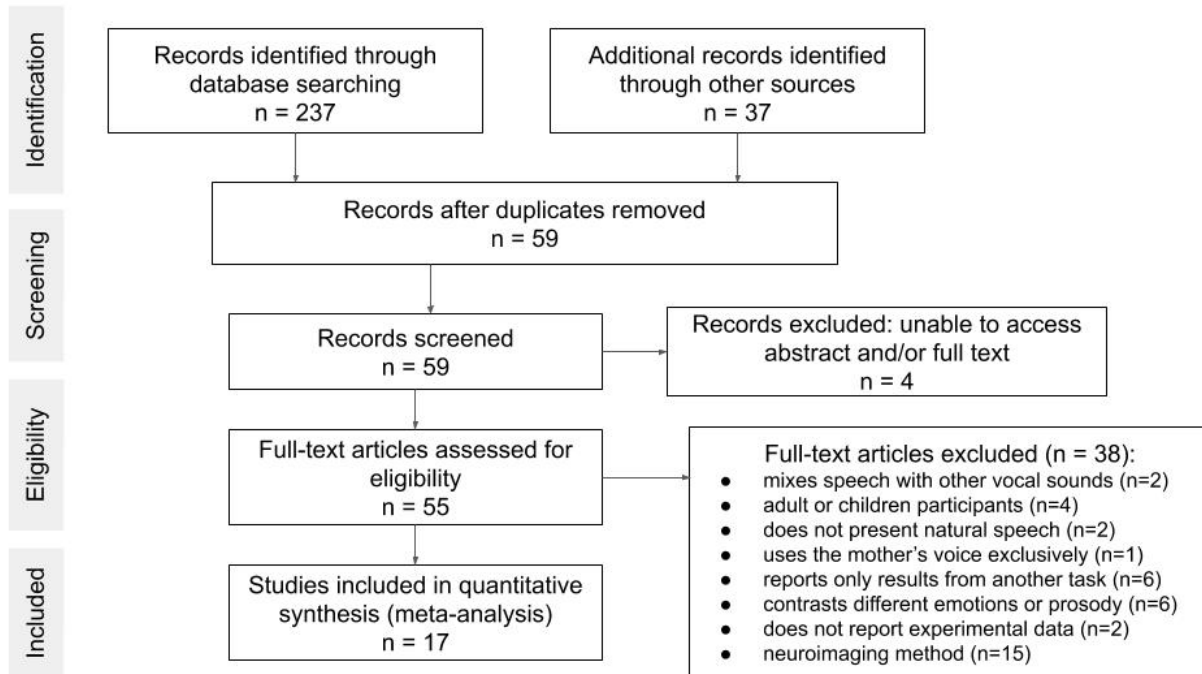


Figure 3. PRISMA flowchart summarizing the literature review and selection process.

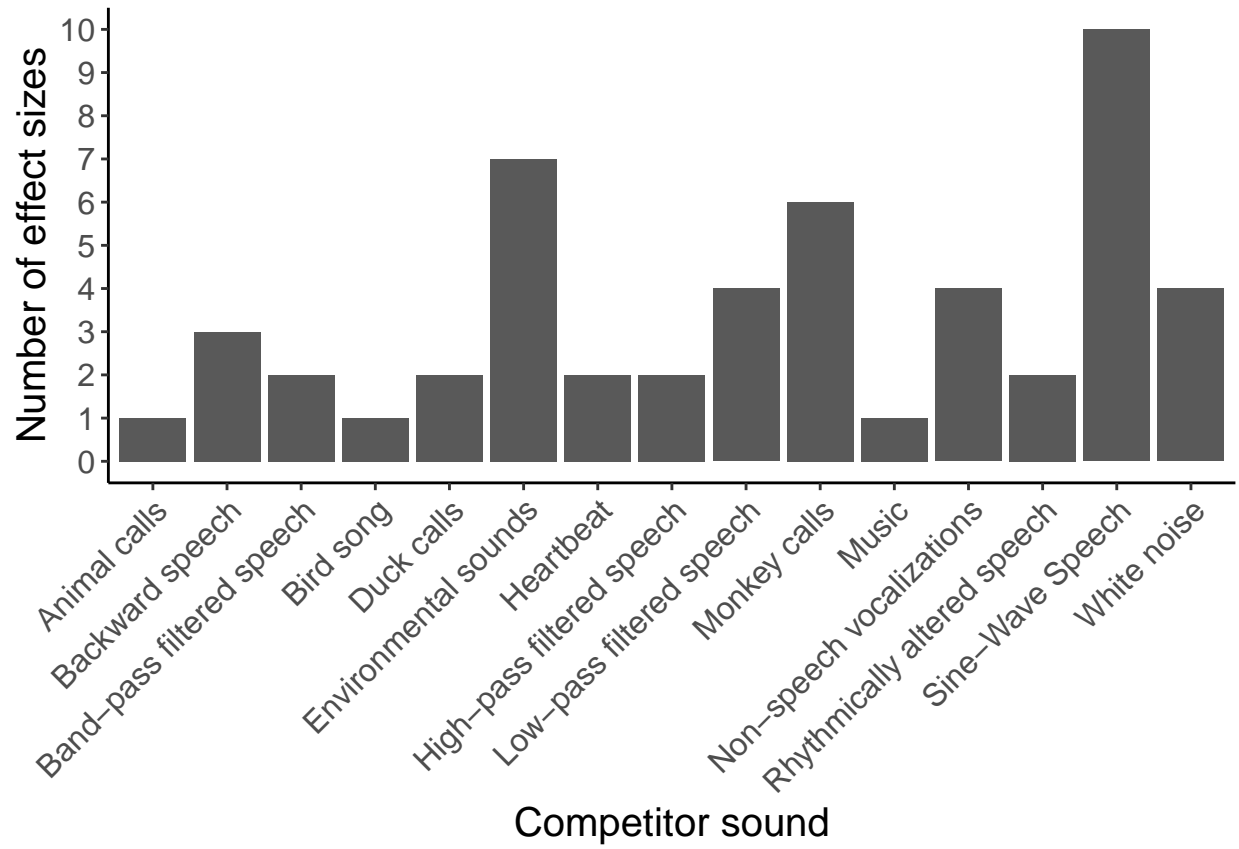


Figure 4. Histogram of the number of effect sizes for each competitor.

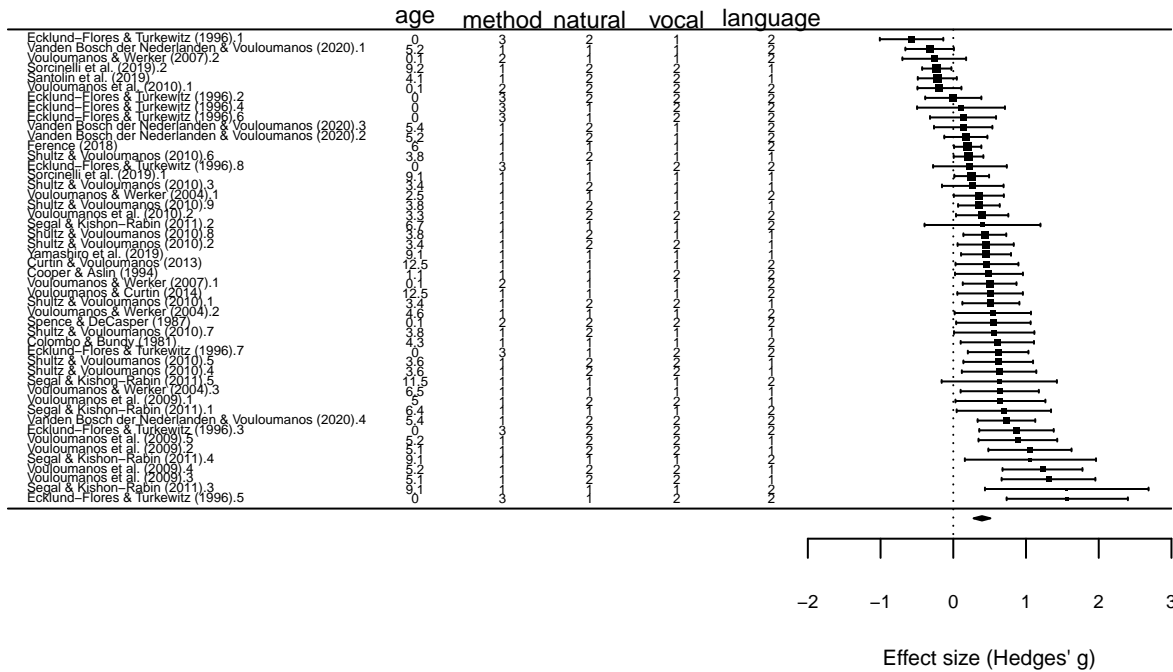


Figure 5. Forest plot of effect sizes available in the literature, along with their respective moderator status.

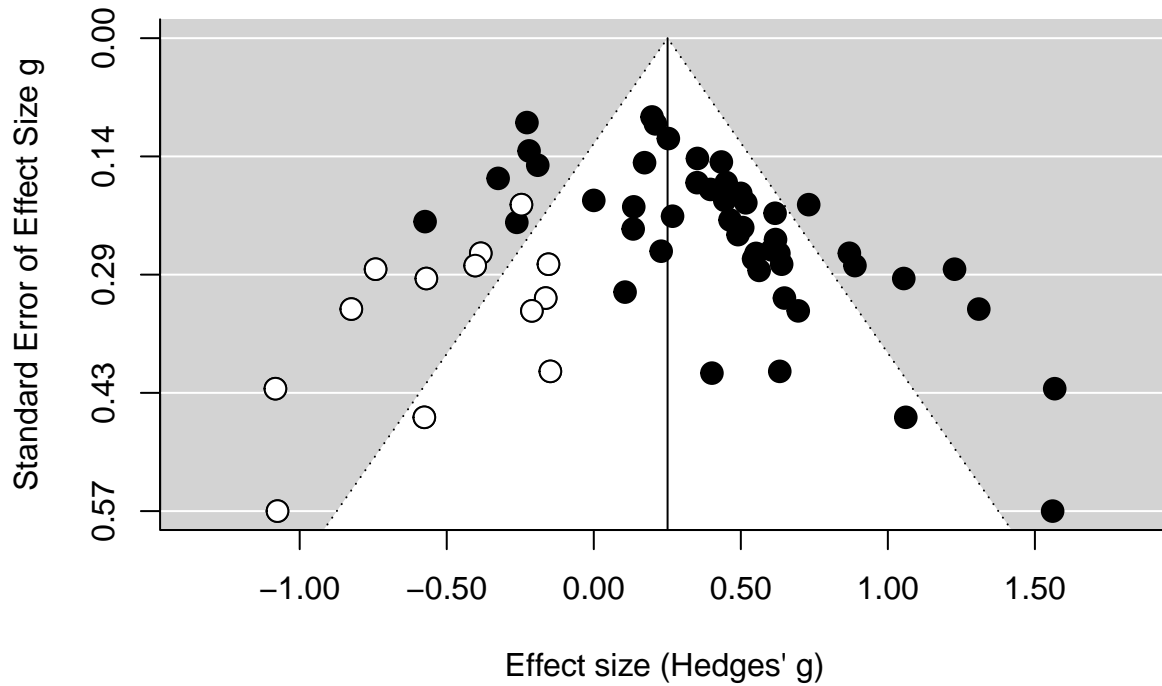


Figure 6. Funnel plot of effect sizes and their respective standard errors. Black dots: effect sizes observed in the literature. White dots: missing effect sizes, suggestive of a publication bias². Vertical line: average effect size after filling the missing effect sizes.

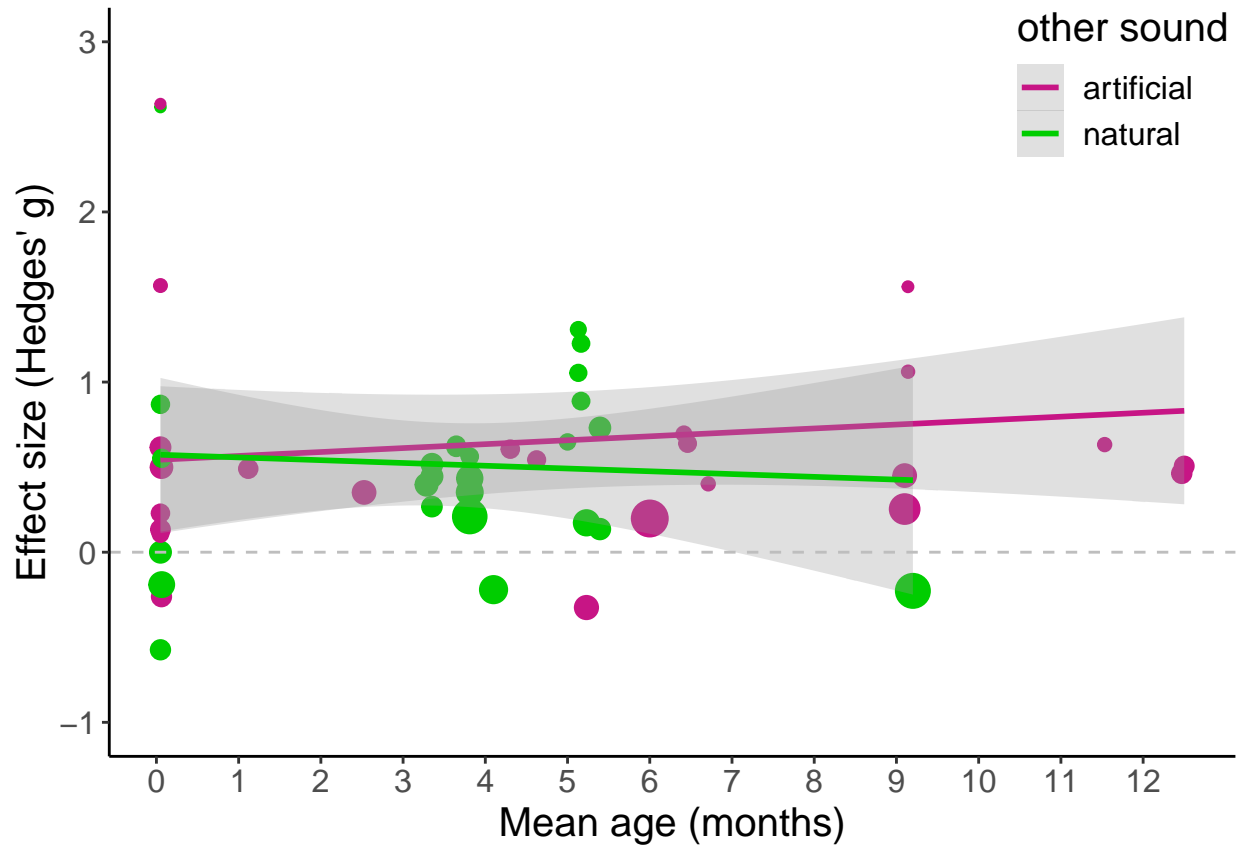


Figure 7. Effect sizes as a function of age and natural quality of the competitor. The size of each dot is inversely proportional to the variance. Positive effect sizes reflect a preference for the speech sound, negative effect sizes reflect a preference for the competitor sound.

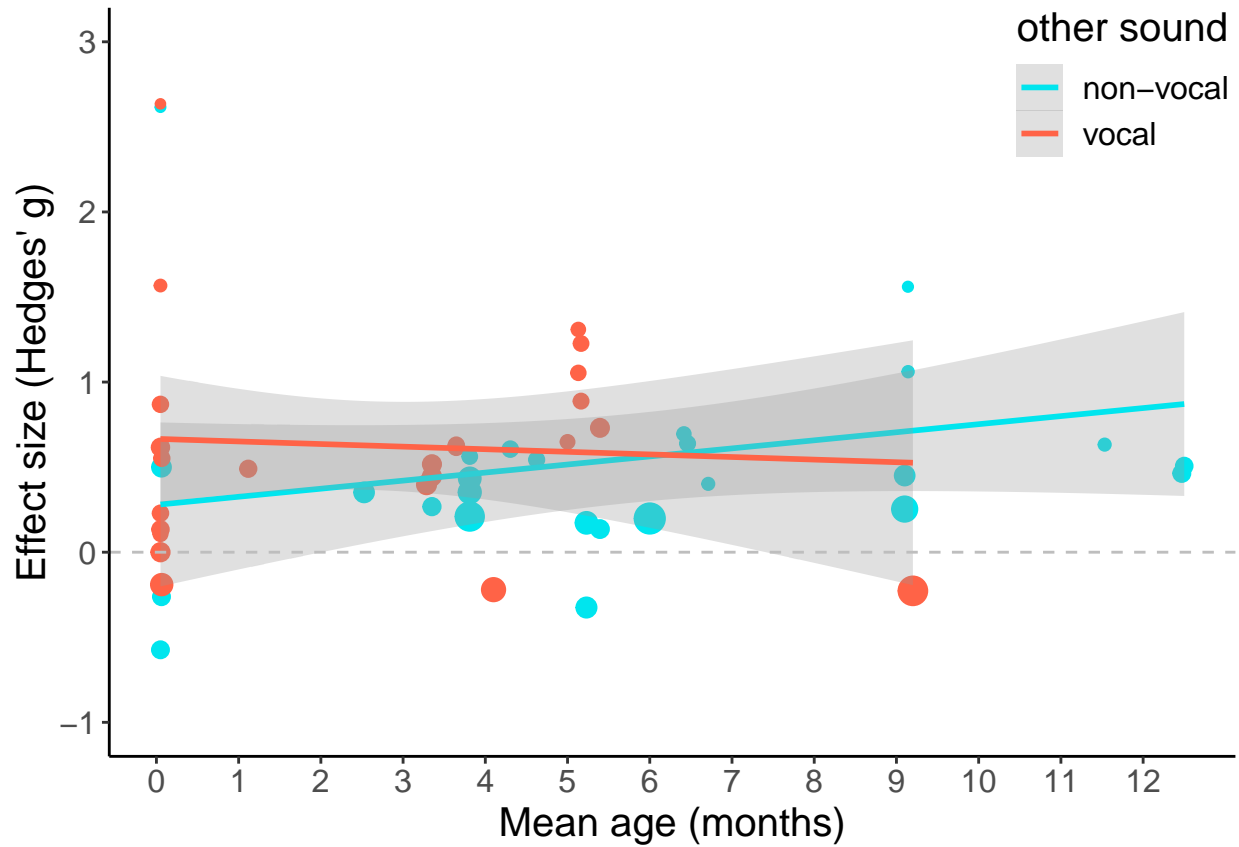


Figure 8. Effect sizes as a function of age and vocal quality of the competitor. The size of each dot is inversely proportional to the variance. Positive effect sizes reflect a preference for the speech sound, negative effect sizes reflect a preference for the competitor sound.

