Infants' preference for speech is stable across the first year of life: Meta-analytic evidence

Abstract

Previous work suggested that humans' sophisticated speech perception abilities stem from an early capacity to pay attention to speech in the auditory environment. What are the roots of this early preference? We assess the extent to which it is due to it being a vocal sound, a natural sound, and a familiar sound, through a meta-analytic approach, classifying experiments as a function of whether they used native or foreign speech, and whether the competitor, against which preference is tested, was vocal or non-vocal, natural or artificial. We also tested for effect of age. Synthesizing data from 775 infants across 38 experiments, we found a medium effect size, confirming at the scale of the literature that infants reliably prefer speech over other sounds. This preference was not significantly moderated by the language used, vocal quality or naturalness of the competitor, nor by infant age. The current body of evidence appears most compatible with the hypothesis that speech is preferred consistently as such, and not just due to its vocal, natural, or familiar nature. We discuss limitations of the extant body of work on speech preference, including evidence consistent with a publication bias and low representation of certain stimuli types and ages.

Keywords: Meta-analysis, Infants, Speech preference, Conspecific detection, Developmental tuning

Word count: 6930

Infants' preference for speech is stable across the first year of life: Meta-analytic evidence

Introduction

Humans acquire their communication skills from early infancy, with specialized speech perception abilities well before they produce their first word. These perceptual capacities manifest in an early preference for speech over other types of sound (Ecklund-Flores & Turkewitz, 1996; Vouloumanos, Hauser, Werker, & Martin, 2010; Vouloumanos & Werker, 2007). Some studies show, at birth, a preference for speech over sine-wave speech (e.g. time-varying sinusoidals tracking the fundamental frequency and the first three formants in Vouloumanos & Werker, 2007), and heartbeat (Ecklund-Flores & Turkewitz, 1996). Studies find no preference between speech and monkey calls at birth (Shultz & Vouloumanos, 2010; Vouloumanos, Hauser, Werker, & Martin, 2010), but a preference for speech over monkey calls at three month old (Vouloumanos, Hauser, Werker, & Martin, 2010). Results like these have led to the theoretical proposal that speech is a privileged signal for humans, whereby newborns have a preference for the vocalizations of humans and non-human primates, and by three months tune in to human speech specifically (Vouloumanos & Waxman, 2014). These studies even found that three-month-olds favor human speech over other human vocal sounds, such as coughing (Shultz & Vouloumanos, 2010). Because these results have been obtained both with the participants' native language (Vouloumanos, Hauser, Werker, & Martin, 2010; Vouloumanos & Werker, 2007), and a foreign language (Shultz & Vouloumanos, 2010), it was claimed that infants tune to the speech signal itself, and not simply the familiar sounds of their native language (Vouloumanos & Waxman, 2014).

These studies also tested factors other than vocal quality which could modulate the preference. Experiments at three months of age showed a preference for speech as compared to environmental sounds, such as wind or water, but in the same paper, other environmental sounds did not yield the same results (Shultz & Vouloumanos, 2010). Regarding

developmental tuning, this study did not test this contrast at other ages, making difficult to determine if infants prefer speech to other natural sounds from birth, or if this preference emerges during the first three months of life, as for monkey calls.

Studies opposing speech to monkey calls or environmental sounds found a preference for speech in both the native language (Vouloumanos, Hauser, Werker, & Martin, 2010; Vouloumanos & Werker, 2007), and foreign languages (Shultz & Vouloumanos, 2010), suggesting that this factor does not determine the preference. However, newborns were only tested in their native language, so the impact of this factor over development has yet to be determined.

A meta-analytic approach

Previous work has thus contributed both important hypotheses and tantalizing results, but the specificity of data collected in any one experiment makes it hard to obtain a bird's eye view. Moreover, some claims are based on juxtaposing a significant result in one condition or age against a non-significant one in another, a practice that is now known to be inappropriate since "the difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman & Stern, 2006). In this paper, we seek to directly test the theoretical predictions presented in the summarized work, namely that the preference for speech may be attributed to a general preference for vocalizations, to a preference for natural sounds, or to familiar sounds. To do so, we employ a meta-analytic approach, which is recommended over narrative reviews (Cristia, Tsuji, & Bergmann, 2021). A meta-analysis can integrate data from experiments that vary in their methodology, as well as test the effect of factors of theoretical importance, by redescribing the stimuli used as a function of those factors. For instance, a study measuring preference for native speech over white noise provides data on a natural versus artificial contrast, as well as a vocal versus non-vocal contrast, thus accounting for how the same stimuli can be both natural and vocal or artificial

and non-vocal.

Also, we can draw a developmental timeline across the age range covered by the literature, beyond age groups tested within papers. This is particularly useful in developmental psychology, which relies on age-related differences that need to be tested statistically (Gelman & Stern, 2006). Meta-analyses offer a powerful statistical approach to directly test for interactions with age across the whole age-range covered by the literature. To give an example from a previous developmental meta-analysis, it had 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, stable familiarity preferences across the first two years have been found for word segmentation in natural speech (Bergmann & Cristia, 2016); and a stable novelty effect ensues for artificial grammars implemented in synthesized speech, whereas those implemented in natural speech led to stable familiarity preferences (Black & Bergmann, 2017). Meta-analyses are therefore important to statistically and systematically test the theoretical predictions proposed in qualitative reviews, and show subtle effects that are difficult to see when reading the literature with a human eye.

Single experiments tell us about what a specific group of participants, presented with a specific set of stimuli, at a specific point in time has done. Meta-analyses are the following step because they provide a principled statistical approach to integrate those individual and specific results into a larger picture. By aggregating the numerous individual studies of a literature, meta-analyses gain statistical power. As a result, meta-analyses can reveal small effects that are difficult to show in individual experiments. By integrating data across different laboratories, they provide evidence for the generalizability of effects, and facilitate comparisons between experimental results. A systematic review also allows us to identify empirical gaps (e.g., age groups that are under-represented but are crucial to tease apart two factors) and conceptual gaps (e.g., use of stimuli that systematically confounds two or more

of those explanations). As a result, meta-analyses are a moment of self-reflection for the field (Cox, Keren-Portnoy, Roepstorff, & Fusaroli, 2022; Nguyen, Versyp, Cox, & Fusaroli, 2021).

Finally, meta-analyses offer tools to detect publication bias in the literature. By aggregating all the available evidence for a phenomenon, we can see if the distribution of effect sizes has an unexpected shape, typically with an excess of positive results due to the difficulty to publish null or negative results. We can further integrate this information, and derive a new estimate of the overall effect size.

Research question and predictions

When an experiment uses native speech as the target and pure tones as the foil, the preference emerging could be due to at least four conceptually separable differences across the two stimuli types: One of them is speech and the other isn't, the former is produced with a mouth (vocal), it is a natural sound, and it is potentially more familiar than the latter. Can preferences observed in the previous body of work be explained by one of these factors?

If infants' (including newborns') preference is attributable to the fact that speech is a **vocal** sound, then effects when the competitor is vocal will be smaller than when it is not vocal. If it is driven by **naturalness**, then effects when the competitor is natural will be smaller than when it is not natural. Finally, if preferences are at least partly due to **familiarity**, then two predictions follow: effects should be larger when the target is the infant's native speech as opposed to foreign speech, and speech preference should increase with age. Of course, it is also possible that the preference for speech cannot be reduced to any of these three factors, in which case effect sizes will not be modulated by such distinctions. In addition, given that many previous empirical studies comment on changes with infant age, we pay special attention to the possibility that age modulates effect size in general, or specifically in conjunction with one of the aforementioned factors.

Methods

This meta-analysis was carried out following PRISMA recommendations (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). In addition, we provide information on all steps (including PRISMA checklist, data, and code) for full transparency and accountability via online supplementary materials; https://osf.io/4stz9/?view_only=d0696591ebf34bfc8430f848cd945ca8.

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 two replies, one of which revealed a formerly undiscovered published study, and communicated unpublished data (Santolin, Zettersten, & Saffran, 2020).

Inclusion criteria

As standard in systematic reviews, after a first screening based on titles and abstracts, 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). If a paper reported results from neurotypical and at-risk infants, we included only the data from the neurotypical group.

Given our key interest in the preference for speech over other sounds, we excluded

studies that contrasted two different speech sounds (e.g., foreign vs. native language, or adult vs. child-directed speech, or mother vs. stranger's voice); or two different non-speech sounds (e.g., backward speech vs. animal vocalizations). In addition, we excluded experiments where the contrast presented to the infants could not be coded according to our three hypotheses (vocal, natural, familiar). This meant the exclusion of experiments where speech was presented in the mother's voice (which confounds speech and individual voice recognition for our familiarity factor). Finally, we excluded neuroimaging experiments to avoid mixing results from different brain regions with different response profiles. 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 1). The full list of the papers that were inspected together with final inclusion decisions are available in a decision spreadsheet (see the online supplementary materials; https://osf.io/4stz9/?view_only=d0696591ebf34bfc8430f848cd945ca8).

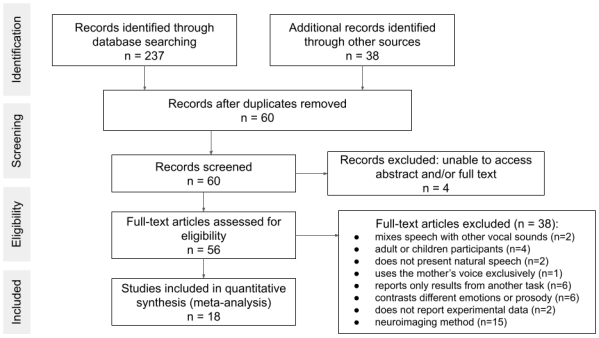

Figure 1. PRISMA flowchart summarizing the literature review and selection process.

Coding

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.

The critical variables for our purpose are key stimuli characteristics, infant age, and testing method (central fixation, high amplitude sucking, head-turn preference procedure). As for key stimuli characteristics, we coded vocal quality, naturalness, and familiarity, as follows.

For **vocal quality**, 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 song, and filtered speech. Non-vocal competitors included backward speech (that has abrupt closures that cannot be produced by a 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).

For **naturalness**, the competitor sound 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. laughter or coughs). If the authors applied acoustic manipulations, the competitor was coded as artificial. Artificial competitors included sine-wave speech, filtered speech[1], 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.

[1]In 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. For **familiarity**, we considered the language in which the speech sounds were recorded (native or foreign).

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.

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 vs. the competitor. If effect sizes 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 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 (Bergmann & Cristia, 2016). We calculated the variance of each effect size using standard formulae (Lipsey & Wilson, 2001). Cohen's d were transformed to Hedges' g by multiplying d by a correction factor for small sample sizes based on the degrees 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).

Analyses use the R (R Core Team, 2018) package Robumeta (Hedges, Tipton, & Johnson, 2010), which allows us 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 (Bergmann et al., 2018).

Results

Database description

We found a total of 19 publications (labeled with an asterisk in the reference list) reporting 39 experiments, for a total of 791 infants, and 55 (not mutually independent) effect sizes. Regarding publication status, 47 effect sizes came from 16 papers that have been 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; Segal, Kligler, & Kishon-Rabin, 2021; Shultz & Vouloumanos, 2010; Sorcinelli, Ference, Curtin, & Vouloumanos, 2019; Spence & DeCasper, 1987; Vanden Bosch der Nederlanden & Vouloumanos, 2021; Vouloumanos & Curtin, 2014; Vouloumanos, Druhen, Hauser, & Huizink, 2009; Vouloumanos, Hauser, Werker, & Martin, 2010, 2010; Vouloumanos & Werker, 2004; Vouloumanos & Werker, 2007; Yamashiro, Curtin, & Vouloumanos, 2020). Additionally, a thesis contributed 5 effect sizes (Ference, 2018), and 3 effect sizes were contributed by authors of unpublished work (Santolin, Zettersten, & Saffran, 2020).

Experiments tended to have small sample sizes, with a median N of 16 children (Range = [4, 60], M = 19.76), which is close to the field standard (Bergmann et al., 2018), but much lower than current recommendations (Oakes, 2017). Infants ranged from 0 to 12 months (1.50 to 380.50 days), although the majority were under 6 months of age (61.54% of the experiments).

There was a fair representation of the factors most relevant to the theoretical hypotheses we aimed to assess. Speech stimuli were recorded in the infant native language in 64.10% of the experiments. The competitor sound was vocal in 46.15% of the experiments. The competitor sound was natural 46.15% of the experiments. Although this means there is some data to assess each of the hypotheses, our systematic review also reveals that these factors are not very well separated in previous work. To start with the most obvious case in which two theoretical explanations are hard to tease apart, there was only 15.38% of effects emerging from stimuli that were natural but not vocal, 10.26% that we classified as vocal but not natural (filtered speech), 46.15% that were neither natural nor vocal, and the rest that were both natural and vocal (35.90%). This means that if we find that both naturalness and vocal quality explain effect sizes, we cannot properly attribute variance to one or the other.

Less obvious is the fact that there are confounds across vocal, natural, and familiarity factors. Figure 2 shows that there are relatively fewer effect sizes in which foreign speech is contrasted against non-vocal sounds, and against non-natural sounds, for which we have fewer than 10 (non-independent) effect sizes. Again, this means that it will be difficult to attribute unique variance to familiarity versus the other factors, if they emerge as significant predictors. In addition, we note that experiments using the infants' native language tested infants from the whole range covered in this meta-analysis (0 to 12 months of age), whereas experiments using a foreign language only tested infants from 3 to 9 months of age, a point to which return.

Although not integral to our study, we report characteristics of discovered data along other dimensions, which are relevant when considering the likely generalizability of conclusions based on previous empirical work. Individual samples comprised 46% 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

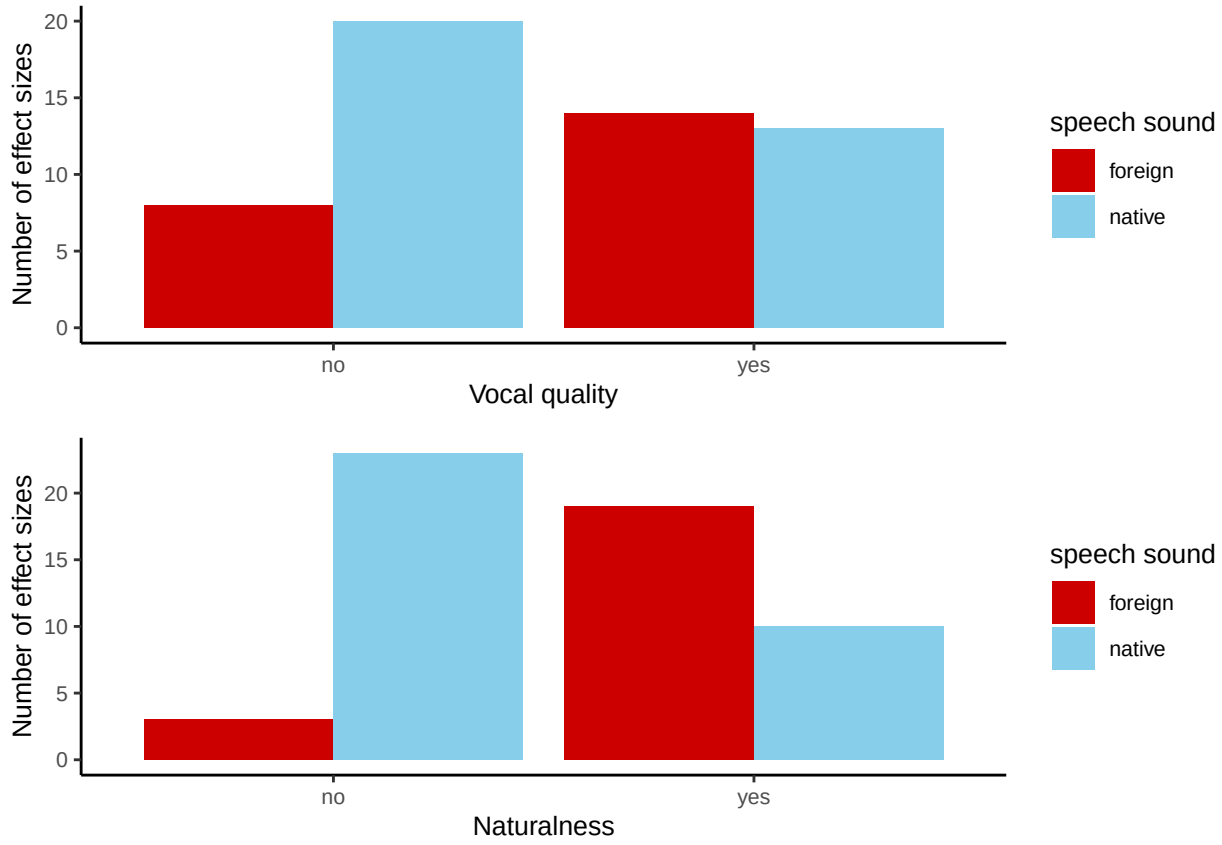

Figure 2. Histograms of the number of effect sizes for each language and moderator status.

experimental methods were used: 79.49% of the experiments used Central Fixation (CF, also called sequential looking preference procedure); 7.69% used High-Amplitude Sucking (HAS); and 12.82% used Head-turn Preference Procedure (HPP). Trial length was fixed in 20.51% of the experiments, and infant-controlled in 76.92% of the experiments.

Speech sounds were spoken by a female in 94.87% of the experiments, with an infant-directed prosody in 53.85% of the experiments. Speech was presented in isolated segments (i.e. words or syllables) in 10.26% of the experiments, and full sentences or passages in 48.72% of the experiments (the remaining were unspecified). The competitors covered a wide range of sounds, as represented in Figure 3.

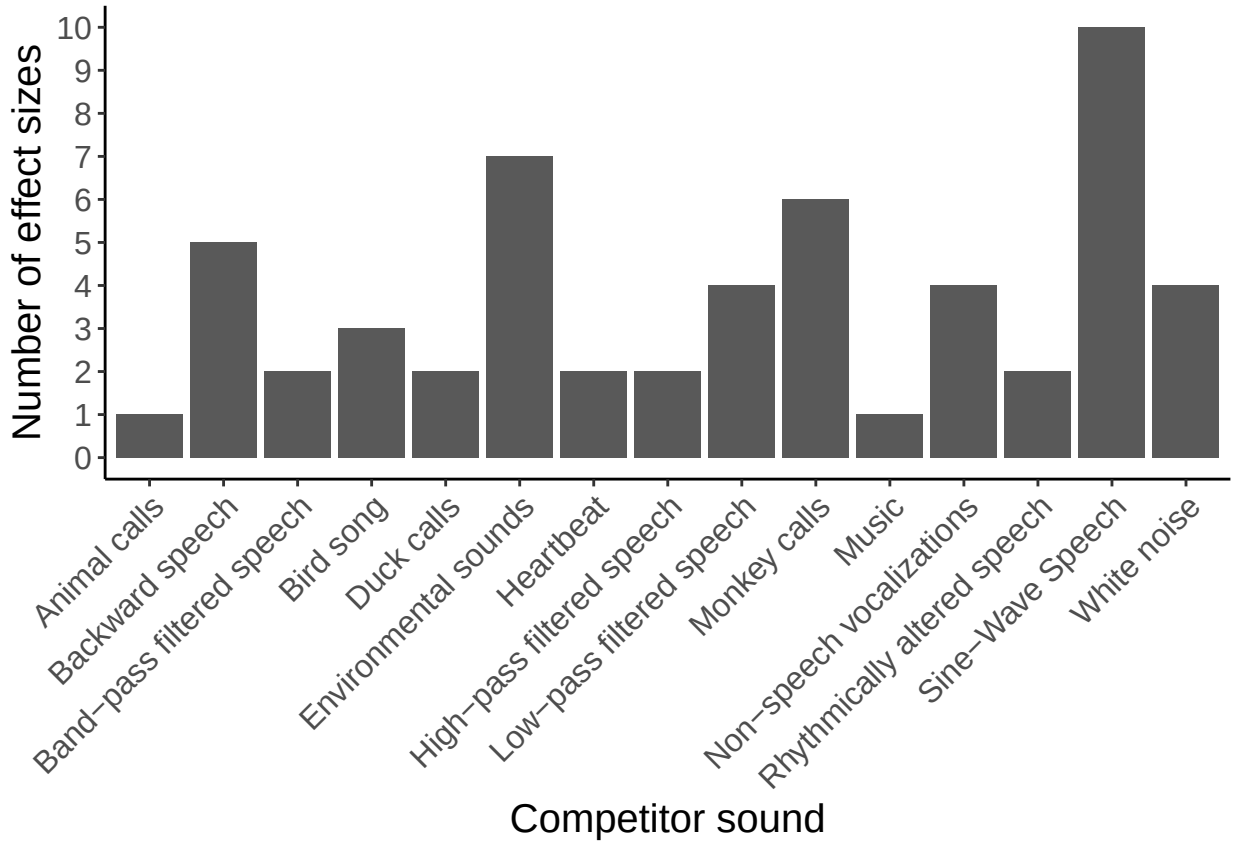

Figure 3. Histogram of the number of effect sizes for each competitor.

Average effect size

Following standard analytic practice, we removed 3 outliers that were more than 3 standard deviations away from the mean. We then integrated all effect sizes in a meta-analytic regression without any moderator, and found an average effect size g of 0.40 (SE = 0.07, CI = [0.26 , 0.55]) (see Figure 4 for the forest plot), corresponding to a medium effect size for the literature as a whole, which was significantly different from zero ($t = 5.59$, $p < .001$). Heterogeneity among effect sizes was estimated at $\tau^2 = 0.16$ ($I^2 = 78.38\%$), which was significant (Q = 206.59, p < 0.01) despite the removal of outliers before running the model. This strongly suggest differences across experiments, and invites analyses using moderators.

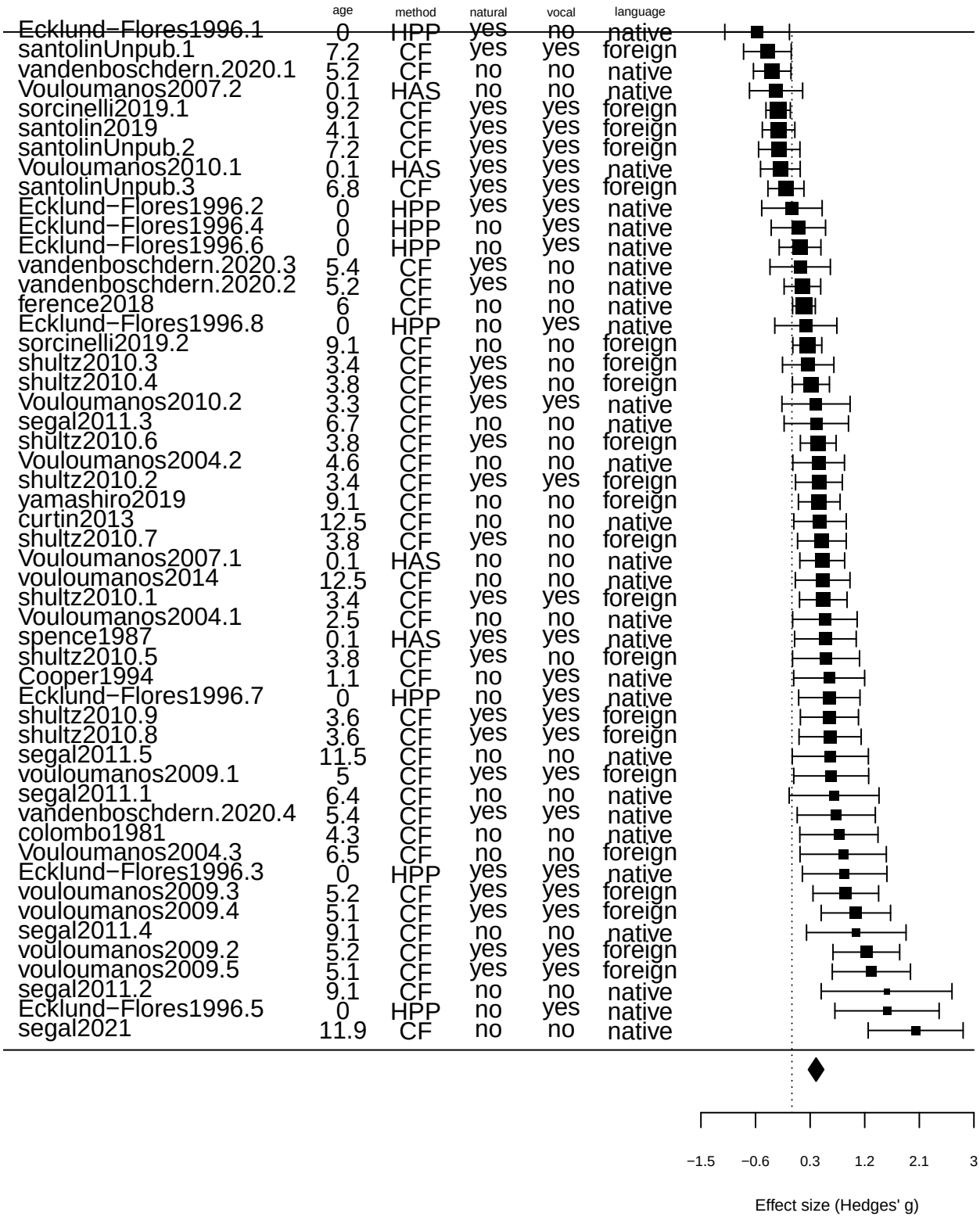

Figure 4. Forest plot of effect sizes available in the literature, along with their respective moderator status. The average effect size is plotted on the bottom line.

Publication bias

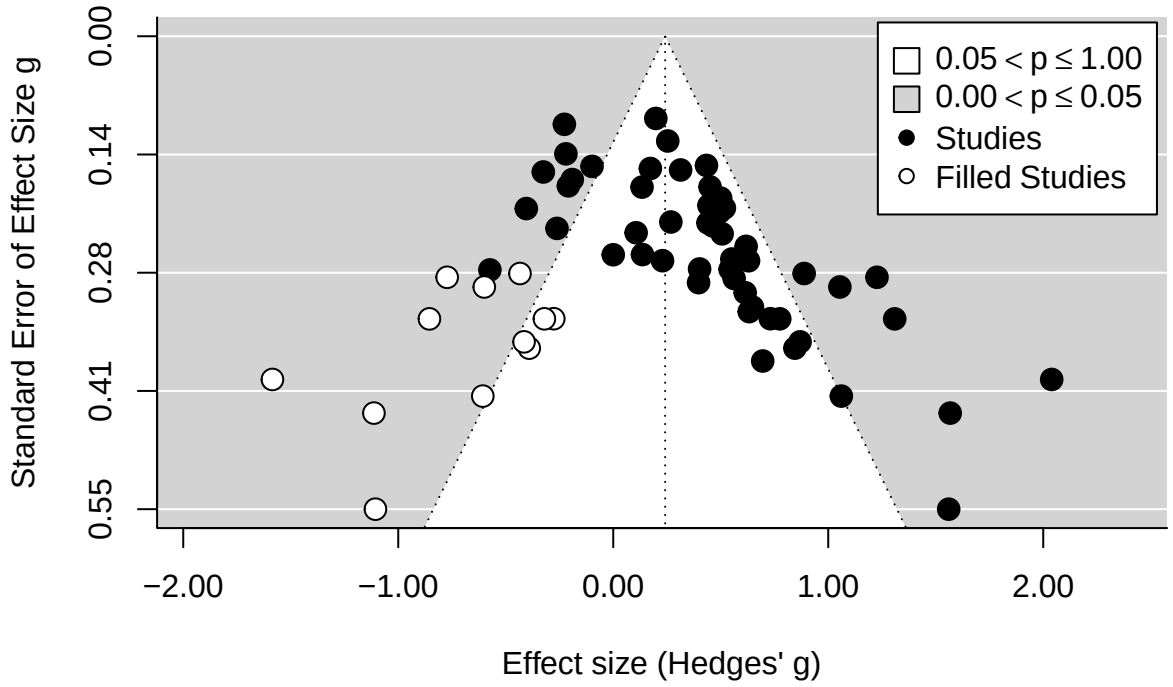

Figure 5. Funnel plot of effect sizes and their respective standard errors. Black dots: effect sizes observed in the literature. White dots: missing effect sizes, suggestive of a publication bias[2]. Vertical line: average effect size after filling the missing effect sizes.

Before proceeding with the moderator analysis, we checked for 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 5)[2]. A regression test on these data was significant ($z = 6.70$, $p < 0.01$), as was the Kendall's tau rank correlation test for funnel plot asymmetry (Kendall's tau $= 0.52$, $p < 0.01$), consistent with a publication bias in the literature.

To check whether this bias fully explains speech preference in the extant body of

[2]If 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.

literature, we symmetrized the funnel plot with the "trim and fill" method (Duval & Tweedie, 2000). To symmetrize the funnel plot, 12 (SE = 4.75) missing experiments were needed on the left side of the plot. The corrected effect size was estimated at 0.24 (SE = 0.07) after filling in the 12 missing experiments, which is still significantly different from zero. Thus, even correcting for a potential publication bias, we still find statistical evidence for infants' preferring speech over competitors.

Moderator analyses

We then tested if heterogeneity could be accounted for by the factors proposed in previous work (vocal, natural, familiar). There are two main ways of answering this question, and they provide converging results. We start with the simplest analysis, in which a single meta-analytic model is fit with the following moderators:

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

These moderators were specified without interactions with each other both to avoid overfitting and because they were sometimes confounded in previous work, as described previously. None of the moderators was significant in this analysis (see Table 1).

Another way of answering our key research question is by fitting a separate meta-regression model for each moderator and its interaction with age. Once again, we observe no significant main effect (nor interaction with age) for any of our predictors (Tables 2, 3, 4). Even though neither main effects or interactions with age were significant, we thought readers would be interested in assessing the extent to which results visually fit expectations from the literature summarized in the Introduction. To begin with the vocal

Table 1

Statistical results of meta-regression with all moderators. The intercept corresponds to the effect size when the competitor is natural, and vocal, and 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); the competitor is artificial (naturalness); and the competitor is non-vocal (vocal quality).

	estimate	SE	t	confidence interval
intercept	0.17	0.17	0.98	-0.19 - 0.54
naturalness	0.00	0.00	0.87	0 - 0
vocal quality	-0.09	0.16	-0.55	-0.45 - 0.27
language	0.27	0.13	2.02	-0.02 - 0.55
age	0.05	0.16	0.34	-0.28 - 0.39

Table 2

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	0.48	0.15	3.17	0.13 - 0.83
vocal quality	-0.40	0.22	-1.86	-0.86 - 0.06
vocal quality*age	0.00	0.00	2.40	0 - 0.01

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	0.36	0.21	1.67	-0.17 - 0.88
naturalness	0.00	0.24	0.00	-0.53 - 0.52
naturalness*age	0.00	0.00	1.05	0 - 0.01

Table 4

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 estimate correspond to changes in the intercept when the target stimuli are in the native language.

	estimate	SE	t	confidence interval
intercept	0.72	0.30	2.44	0.03 - 1.41
language	-0.48	0.32	-1.52	-1.19 - 0.22
language * age	0.00	0.00	2.24	0 - 0.01

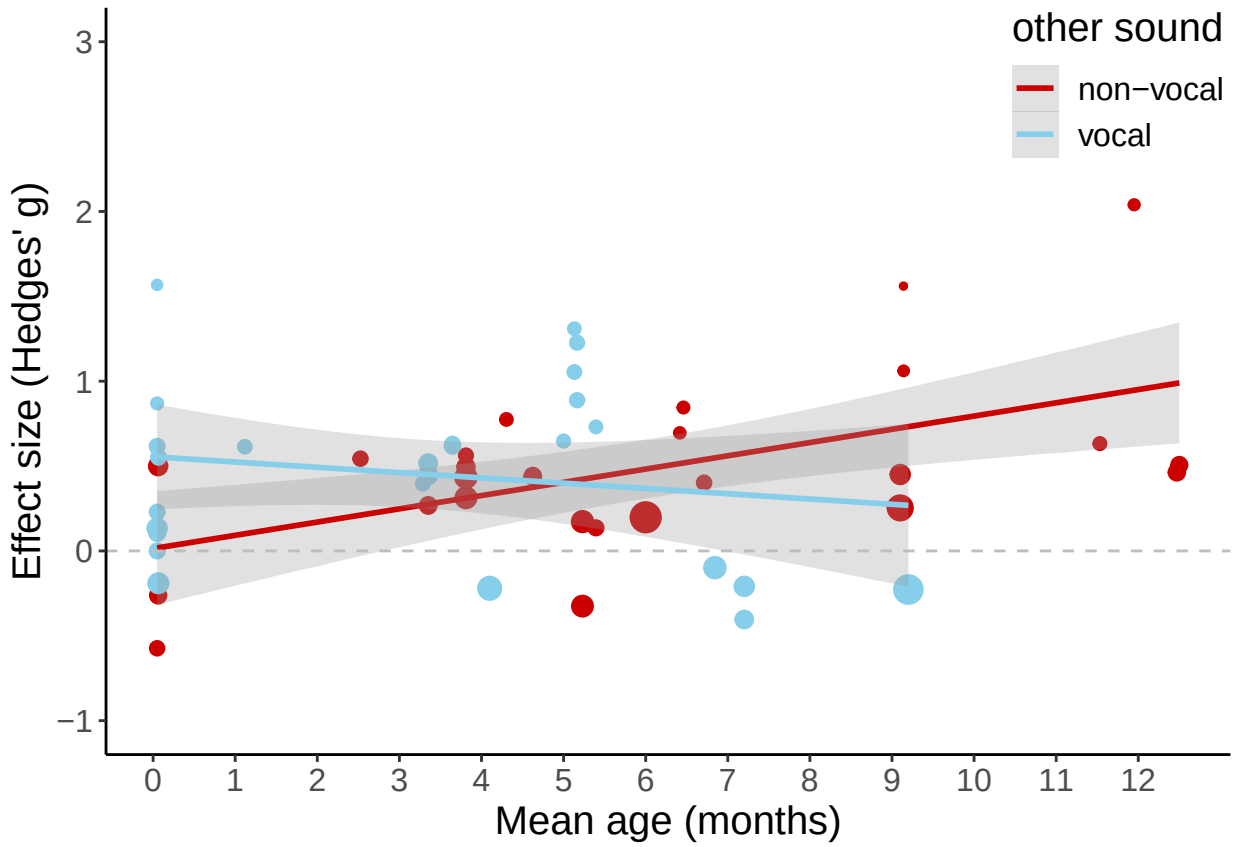

Figure 6. Effect sizes as a function of age and vocal quality of the competitor. The size of each dot is inversely proportional to the variance. Positive effect sizes reflect a preference for the speech sound, negative effect sizes reflect a preference for the competitor sound.

quality of the competitor, the expectation is that effect sizes will be smaller for a contrast between speech and vocal at birth, than speech and non-vocal at the same age; and as infants age, the contrast between speech and vocal should become larger. Neither expectation fits Figure 6, where, if anything, the contrast between speech and non-vocal competitors is larger at birth than that of speech and vocal competitors. As for naturalness of the competitor, if infants prefer speech because it is natural, then we should see smaller effects for the contrast against natural than that for non-natural. That is again not the case: Figure 7 shows overlap across the effect sizes of these two types of experiment. Finally, if infants' speech preference is driven by familiarity, then effect sizes should be larger when native as opposed to foreign speech is used, and this contrast should be larger as infants age.

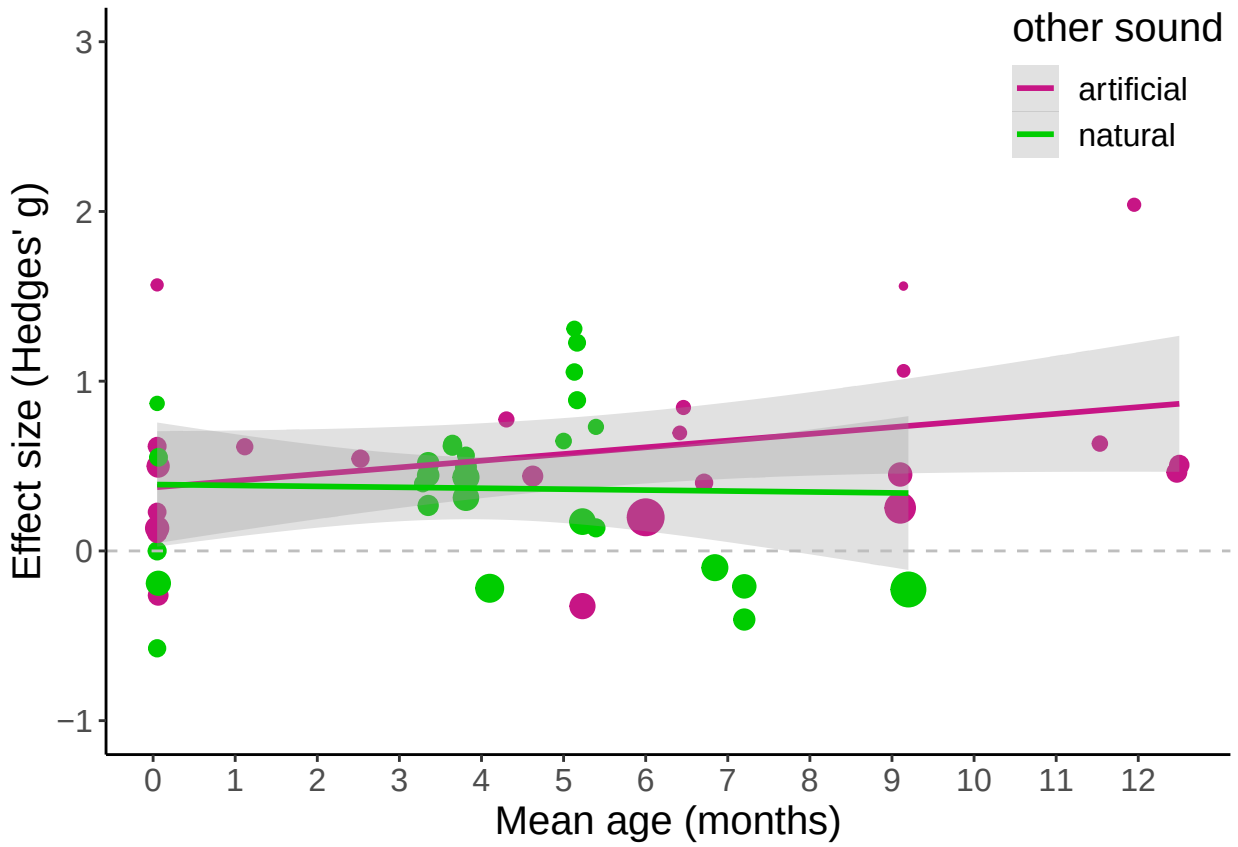

Figure 7. Effect sizes as a function of age and natural quality of the competitor. The size of each dot is inversely proportional to the variance. Positive effect sizes reflect a preference for the speech sound, negative effect sizes reflect a preference for the competitor sound.

The trends in Figure 8 are visually consistent with these predictions, with a positive slope for the native conditions and negative for the foreign conditions. However, it is clear that there is a great deal of overlap between the two curves, and key data (below three and above nine months, particularly gathered with foreign speech) would complete the picture.

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=0.40, g=0.24 when taking the publication bias into account), which was

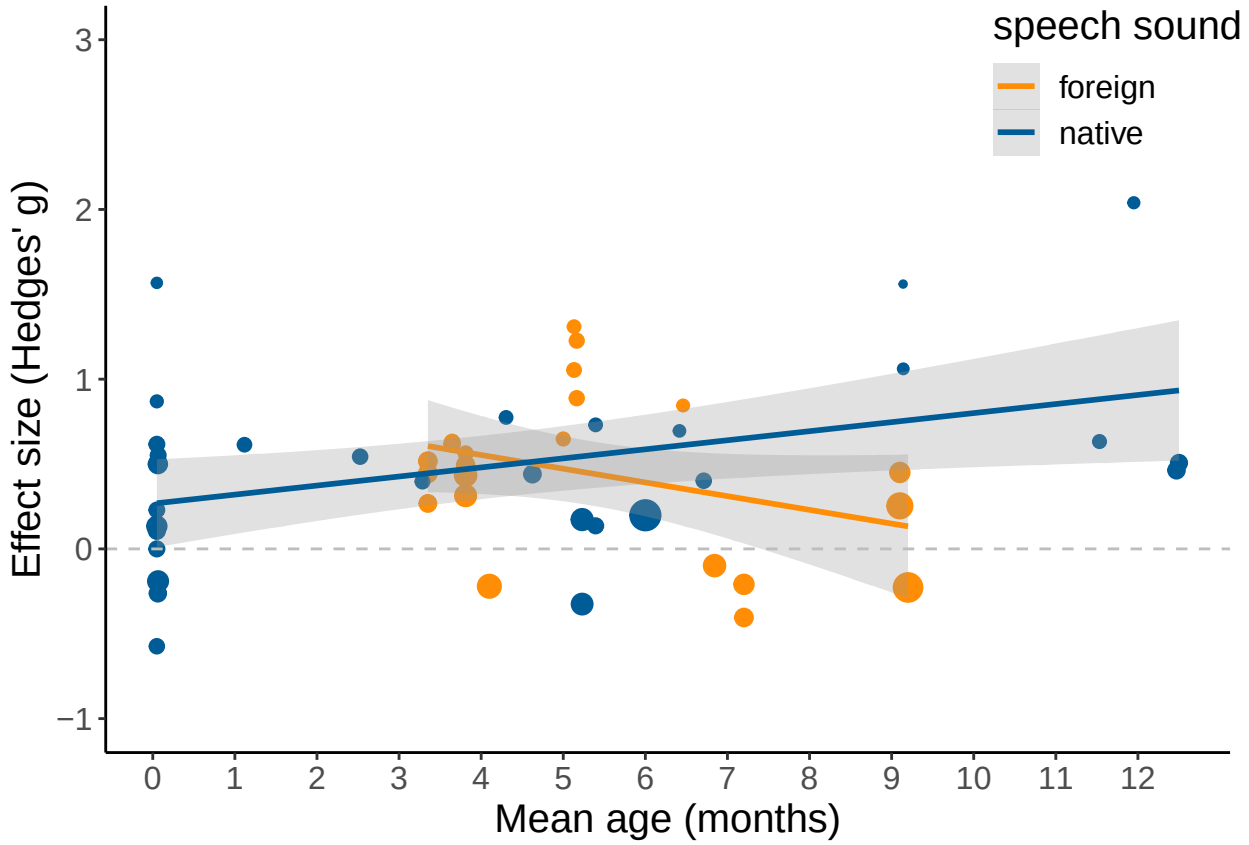

Figure 8. Effect sizes as a function of age and familiarity with the speech sounds. The size of each dot is inversely proportional to the variance. Positive effect sizes reflect a preference for the speech sound, negative effect sizes reflect a preference for the competitor sound.

still significant after correcting for the publication bias. Our meta-analysis shows that this preferential processing of speech sounds is observable from birth on. It is important to stress that this conclusion is not trivial because others have said similar things in the past: One key advantage of meta-analysis over conclusions drawn from individual studies or unsystematic narrative reviews is that we can actually measure the likely presence of publication bias (which *is* present in these data, as we discuss below), and furthermore compensate for this bias statistically, to see if an effect remains significant after doing so (Cristia, Tsuji, & Bergmann, 2021; see also Ioannidis, 2005).

The significant heterogeneity we found among the literature suggests that underlying

factors modulate this effect. Based on previous theoretical claims, one would have predicted infants' speech preference to be larger when the competitor was non-vocal; when the competitor was an artificial sound than when it was a natural one; 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. Thus, although these explanations are conceptually appealing, it does not appear to be the case that we can reduce infants' speech preference to a bias for vocal, natural, or familiar sounds.

Also based on theoretical claims in the literature, we had expected age to play a major role. Indeed, experiments comparing the preference for human speech against human non-speech as well as animal vocalizations more generally (McDonald et al., 2019; Vouloumanos, Hauser, Werker, & Martin, 2010) often discuss 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), nor did it interact with any other moderators (Tables 3, 4, and 2, and corresponding figures). This result was replicated in a separate model declaring age as the only moderator, which showed a significant intercept, similar in size to the intercept found in the meta-regression with no moderator (Supplementary results S1). Crucially, in this supplementary analysis, age was not centered so that 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 S2). This null effect of age, combined with the sizable intercept, confirms that infants reliably prefer speech over other types of sounds stably across the age range covered by previous literature.

For all of our analyses, concerning vocal quality, naturalness, familiarity, and age, it is important to remember that, by aggregating the numerous individual studies of a literature, meta-analyses gain statistical power as compared to individual experiments. As such, meta-analytic results have more cumulative explanatory value than single studies.

Our meta-analysis revealed uneven distributions of experiments across age and stimulus dimensions, and that the distribution of effect sizes in the literature is consistent with publication bias. Moreover, distributions of effect sizes for experiments varying along the three dimensions widely overlap.

This clearly points to the value of meta-analysis: to take stock of a field and inspire follow-up studies. In particular, future experiments should test infants from 1 to 3 months, and older than 9 months. Language production gains in complexity at about 9 months (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. In addition, no experiment in the literature tested newborns with other animal vocalizations or human vocal sounds, such as coughing. It is therefore not possible to know with these individual studies if newborns did not prefer speech from monkey calls because infants have an initial preference for primate vocalizations (as proposed), or for vocal sounds in general (human or not). 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, to address the current publication bias currently affecting this body of work, and potentially revisit the research question that motivated the present study.

Our results suggest that, from birth on, infants show a preference for speech, which cannot be accounted by the three explanations tested here: vocal quality, naturalness, or familiarity with the native language. 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 are sensitive to this specific spectro-temporal structure (though see Norman-Haignere & McDermott, 2018 showing that they only explain neural responses in primary auditory cortex, suggesting that other factors contribute to the behavioral response in later processing stages). 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).

A stable preference for speech in infancy is compatible with the idea that infants are born with the capacity to recognize their conspecifics' communication signals. This parallels what has been proposed for faces: Infants are born with the capacity to orient their attention toward them, even without any prior exposure to faces (Morton & Johnson, 1991; Turati, 2004). This would stem from basic perceptual abilities present at birth, namely that the visual system would be tuned to a spatial structure that correspond to those of faces (Morton & Johnson, 1991; Turati, 2004). As newborns have never been exposed to such visual stimuli before, it would reflect general properties (i.e. filters) of the visual system. Similarly, the auditory system could be tuned to a spectro-temporal structure that speech presents. The combination of this non-specific bias with the systematic variations of the auditory environment (i.e. the fact that an acoustical structure characterizes speech but not other sounds of the environment) would result in preferential responses to speech from birth. However, contrary to faces, fetuses are exposed to speech 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, by tuning the response properties of the auditory system to speech. The fact that familiarity with the language used in the experiment did not modulate infants' preference suggests that this effect is not triggered by familiarity with the sounds of the native language. Infants would therefore form a representation that is specific enough to discriminate speech from

other natural or vocal sounds, but general enough to be independent of the language spoken.

The human species is a gregarious one. Detecting speech signals allows to integrate it with other sensory percepts, such as faces, to form multisensory representations of conspecifics. Five-months old infants were capable to match human faces to speech, as well as monkey faces to monkey vocalizations (Vouloumanos, Druhen, Hauser, & Huizink, 2009). This provides evidence that human infants make correspondences between faces and vocalizations, and that they distinguish their conspecifics from other species. This lays the groundwork for social cognition. Finally, the preference itself may also be a meaningful index of processing that can be used to identify children at risk (Sorcinelli, Ference, Curtin, & Vouloumanos, 2019). Understanding this phenomenon is therefore crucial for both theoretical and clinical advances.

Ultimately, preferential processing of speech may support higher level cognitive tasks. Identifying speech signals and paying attention to them would allow infants to form complex representations of the sensory world, that they can manipulate cognitively. Consistently, 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, 2013; Fulkerson & Waxman, 2007). Interestingly, infants categorized visual stimuli when presented with speech, melodies, monkey (Fulkerson & Haaf, 2003), or lemur vocalizations (Ferry, Hespos, & Waxman, 2013). These results support the idea that infants' cognition is set up to respond to, and manipulate complex sounds (i.e. modulated in time and frequency), especially those of critical ecological importance such as communicative vocalizations.

With a sample size of 791 infants, covering the wide age range of all individual experiments available, our meta-analysis provides evidence for a specialized processing of speech from birth. This parallels infants' attention to faces from birth (Morton & Johnson, 1991; Turati, 2004), and suggests that human cognition is set up to pay attention to signals

from conspecifics specifically. Such systems may be the precursors of humans' advanced social cognition skills.

References

References marked with an asterisk indicate studies included in the meta-analysis.

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 (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

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

Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2022). A bayesian meta-analysis of infants? Ability to perceive audio–visual congruence for speech. *Infancy*, *27*(1), 67–96.

Cristia, A., Tsuji, S., & Bergmann, C. (2021). *A meta-analytic approach to evaluating the explanatory adequacy of theories. MetaPsychology.* OSF Preprints. https://doi.org/10.31219/osf.io/83kg2

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%3C205::AID-DEV2%3E3.0.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.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

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.

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

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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

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

Nguyen, V., Versyp, O., Cox, C. M. M., & Fusaroli, R. (2021). A systematic review and bayesian meta-analysis of the development of turn taking in adult-child vocal interactions.

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

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

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

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

Santolin, C., Zettersten, M., & Saffran, J. R. (2020). Infants' preference for non-native speech versus birdsong. Unpublished data. Retrieved from https://github.com/mzettersten/birdsong

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 Ovid. *Ear and Hearing*, *32*(3), 358–372. https://doi.org/10.1097/AUD.0b013e3182008afc

Segal, O., Kligler, N., & Kishon-Rabin, L. (2021). Infants' Preference for Child-Directed Speech Over Time-Reversed Speech in On-Channel and Off-Channel Masking. *Journal of Speech, Language & Hearing Research*, *64*(7), 2897–2908. https://doi.org/10.1044/2021_JSLHR-20-00279

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

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

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

Turati, C. (2004). Why Faces Are Not Special to Newborns: An Alternative Account of the Face Preference. *Current Directions in Psychological Science*, *13*(1), 5–8.

https://doi.org/10.1111/j.0963-7214.2004.01301002.x

Vanden Bosch der Nederlanden, C. M., & Vouloumanos, A. (2021). Infant biases for detecting speech in complex scenes. *Developmental Psychology*, *57*(9), 1411–1422. https://doi.org/10.1037/dev0000974

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., 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

Vouloumanos, A., & Waxman, S. R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Sciences*, *18*(12), 642–646. https://doi.org/10.1016/j.tics.2014.10.001

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

Yamashiro, A., Curtin, S., & Vouloumanos, A. (2020). 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*, *50*(7), 2475–2490. https://doi.org/10.1007/s10803-019-03924-2