nature human behaviour

Article

https://doi.org/10.1038/s41562-022-01452-1

A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech

Received: 25 February 2022

Accepted: 23 August 2022

Published online: 03 October 2022

Check for updates

Christopher Cox [1,2,3] ✉, Christina Bergmann [4], Emma Fowler [1,2], Tamar Keren-Portnoy [3], Andreas Roepstorff [2], Greg Bryant [5] and Riccardo Fusaroli [1,2]

When speaking to infants, adults often produce speech that differs systematically from that directed to other adults. To quantify the acoustic properties of this speech style across a wide variety of languages and cultures, we extracted results from empirical studies on the acoustic features of infant-directed speech. We analysed data from 88 unique studies (734 effect sizes) on the following five acoustic parameters that have been systematically examined in the literature: fundamental frequency (f_0), f_0 variability, vowel space area, articulation rate and vowel duration. Moderator analyses were conducted in hierarchical Bayesian robust regression models to examine how these features change with infant age and differ across languages, experimental tasks and recording environments. The moderator analyses indicated that f_0, articulation rate and vowel duration became more similar to adult-directed speech over time, whereas f_0 variability and vowel space area exhibited stability throughout development. These results point the way for future research to disentangle different accounts of the functions and learnability of infant-directed speech by conducting theory-driven comparisons among different languages and using computational models to formulate testable predictions.

Speaking to infants presents caregivers with a substantial challenge. Because infants are not linguistically competent, older individuals modify their speech to them in a variety of ways to communicate. The ways in which caregivers produce infant-directed speech (IDS) have been widely documented, and some clear patterns have emerged across multiple languages. For example, speakers often increase their vocal pitch and pitch variability, slow down their speech and articulate more clearly[1–4]. The discovery of similar acoustic properties of IDS across so many languages and cultures strongly suggests that this speech style plays an important role in linguistic and social development[5].

In the study of signal design in humans and non-human animals, form–function analysis is used to understand how the structural characteristics of signals are shaped by the communicative functions they serve. This approach applies well to the study of IDS[6–8]. For instance, the loud, low-pitched, abrupt onset of a prohibitive yell could be designed to interrupt the behaviour of a baby by exploiting the startle reflex, which quickly re-orients a target infant's attention to the sound source[6]. Similarly, approval vocalizations may induce positive emotions through raised pitch, increased pitch variability, faster speech and modulated loudness reflecting speakers' positive valence and

[1]Department of Linguistics, Cognitive Science & Semiotics, Aarhus University, Aarhus, Denmark. [2]Interacting Minds Centre, Aarhus University, Aarhus, Denmark. [3]Department of Language & Linguistic Science, University of York, York, UK. [4]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [5]Department of Communication, University of California, Los Angeles, CA, USA. ✉e-mail: chris.mm.cox@gmail.com

heightened arousal[6,9,10]. But communicative functions overlap and interact as the cognitive and linguistic skills of the infant develop, and their interactional affordances change[11–13].

One prominent hypothesis holds that the acoustic features of IDS may help infants learn aspects of language[5]. The benefits of IDS to language development are generally attributed to its tendency to increase the clarity of the speech input[14–16]. This hypothesis receives substantial support from longitudinal studies showing positive correlations between parents' tendency to produce acoustically exaggerated vowels and speech discrimination skills[16] as well as expressive vocabulary size[14,17]. Other studies show that acoustically exaggerated vowels induce more mature neural processing of vowel categories in infants[18] and faster word recognition[19]. The cross-linguistic tendency for caregivers to exaggerate the differences between vowel categories might facilitate infants' language development by increasing category separability in the speech stream. An increase in vowel category separability in speech has been shown to co-occur with a greater degree of within-category variability[20–24], which may work in parallel with separability to increase the robustness and generalizability of the categories[25–28].

The functions of IDS have been posited to exhibit change over the course of early infant development, with the speech style initially serving primarily to direct infants' attention and express affect, and later serving more specific linguistic purposes[7]. According to a form–function analysis, these age-related changes in the functions of IDS should manifest themselves in the acoustic properties of caregivers' speech. Despite the implications of unidirectionality in its name, however, IDS also includes feedback from infants—IDS involves reciprocity and interaction where the interdependence of infants' active participation and caregiver responsiveness plays a crucial role[29–33]. The benefits of IDS should be construed as originating in the mutual feedback loops between infant and caregiver, where infants provide an important source of feedback about which signals they prefer to attend to and interact with[29–33].

Many studies have demonstrated that infants prefer to listen to IDS over adult-directed speech (ADS)[1,2,15,34–38]. This preference persists when presented speech is in a foreign language[36,38] or when it is low-pass-filtered and contains only global prosodic information[39]. Even infant-directed songs in a foreign language induce relaxation in babies[40]. A recent large-scale, multi-lab replication study found that infants exhibit linear increases in their IDS preference until at least 15 months of age, the oldest age tested[36,41]. This trajectory was similar to the findings of a meta-analysis reporting a general increase in looking times towards IDS in preverbal infants from 0 to 9 months[42]. In contrast, two studies have reported that infants' IDS preference exhibits a U-shaped pattern. Hayashi et al.[43] found that while groups of both 4- to 6- and 10- to 14-month-old infants paid more attention to IDS than to ADS, 7- to 9-month-old infants did not exhibit a preference. Similarly, Newman and Hussain[44] found a preference for IDS in 5-month-old infants but not in 9- or 13-month-olds.

Infants' shifting preferences for IDS over ADS in the first year of life could reflect dynamic changes in the acoustic features they attend to. For example, Panneton et al.[13] reported that 4-month-old infants listened longer to speech with a higher positive affect (that is, a higher emotion content) and slowed duration, but 8-month-old infants preferred speech with normal duration and lower relative affect. Other studies examining differences in preferences have demonstrated various effects suggesting that infants, even during their first year, might be attending differentially to many aspects of IDS[9–11]. For example, younger infants have been shown to preferentially attend to the intonational variability and positive affect of IDS[45,46]. At this early developmental stage, the tendency for IDS to contain increased pitch variability, modulated loudness contours and rhythmic alterations[35,47] probably serves the function of effectively communicating intentions, including getting an infant's attention, expressing emotions and encouraging behaviour[7]. As infants get older and become more advanced in language development, their attention might shift towards aspects of IDS that provide linguistic information[11,12,48]. If caregivers adapt the acoustic properties of their IDS to suit infants' developmental needs, we may see systematic shifts in acoustic properties over the course of early infancy, such that exaggerated prosodic features associated with communicating intent to young infants should decline, and linguistically relevant properties should be emphasized more for older children, including expansion of the vowel space area[7,15].

The study of IDS across cultures has a long interdisciplinary history. Early linguistic research revealed many regularities in IDS across disparate languages and cultures, as well as language-specific phenomena. In this work, many of the reported features were not acoustic but concerned phenomena such as modified morphemes and grammatical constructions as well as lexical innovations[49]. Naturally, these kinds of features should vary cross-culturally, and variations were noted within villages, including features that were unique to single families or that might spread to a few households at most. Ferguson[49] also discussed cultural variations in attitudes towards baby talk, including its use in public and whether it was more appropriate for men or women to produce it. Other studies have shown that the frequency of speaking to infants in any manner can vary dramatically, with some cultural groups not speaking to infants very much at all[50–52]. A high degree of variability in the rate of IDS use, however, does not preclude universality[53]; rather, IDS may represent a continuum across cultures that exhibits cross-linguistic variability in its rate and acoustic properties. Early rejections of the universality of IDS often conflated the issues of incidence with form; that is, how often IDS occurs during interaction is separate from its acoustic features when it is actually produced. Later analyses focusing on acoustic characteristics of IDS across languages have revealed striking similarities[2,54,55]. Recent large-scale studies have shown that these features occur widely, and the recognition of IDS and infant-directed song is robust[4,36]. Questions regarding within- and between-culture variation are crucial to address when issues of universality are raised[53].

Researchers have now started using day-long recordings of infants[56,57] and open archives of acoustic data[58], allowing for the analysis of more ecological data to investigate infants' linguistic and emotional development through quantitative and computational means[33]. These archives provide data from diverse cultures[4,50] and offer new insights into the role of linguistic input in early language development. For example, US English speakers appear to produce a particularly exaggerated form of IDS relative to other speakers[2,59,60]. Because such a high proportion of studies on IDS examine US English (Fig. 1 and Supplementary Tables 11.1 and 12.1), the field may have a biased view of how IDS manifests itself and how it may affect language development[36,59]. Figure 1 shows the proportion of languages for which IDS has been analysed compared with the total number of languages listed in the World Atlas of Language Structures[61]. Although this world map suggests a considerable bias in the types of languages and cultures investigated, increasing linguistic diversity—while valuable in and of itself—is unlikely to improve our understanding of the cognitive underpinnings of IDS alone. More fine-grained, hypothesis-driven comparisons are also required[62–64], as discussed further in the Discussion. For such comparative approaches to be useful, we need a more careful and theory-driven analysis of the extant IDS literature and how IDS varies across infant ages, languages, experimental tasks and recording environments. It should also be noted that the participants in the studies included in this meta-analysis largely consist of female caregivers residing in Western, educated, industrialized, rich, developed countries[65]. Due to the sparsity of the data on other speaker types and populations, the meta-analysis could not analyse these factors as potential sources of variability in the acoustic measures (for example, kin versus non-kin caregivers), as discussed further in the Discussion.

Many studies have demonstrated that caregivers exhibit age-related changes in the acoustic properties of their IDS. Here we provide an overview of how each of the acoustic features of IDS that

https://doi.org/10.1038/s41562-022-01452-1

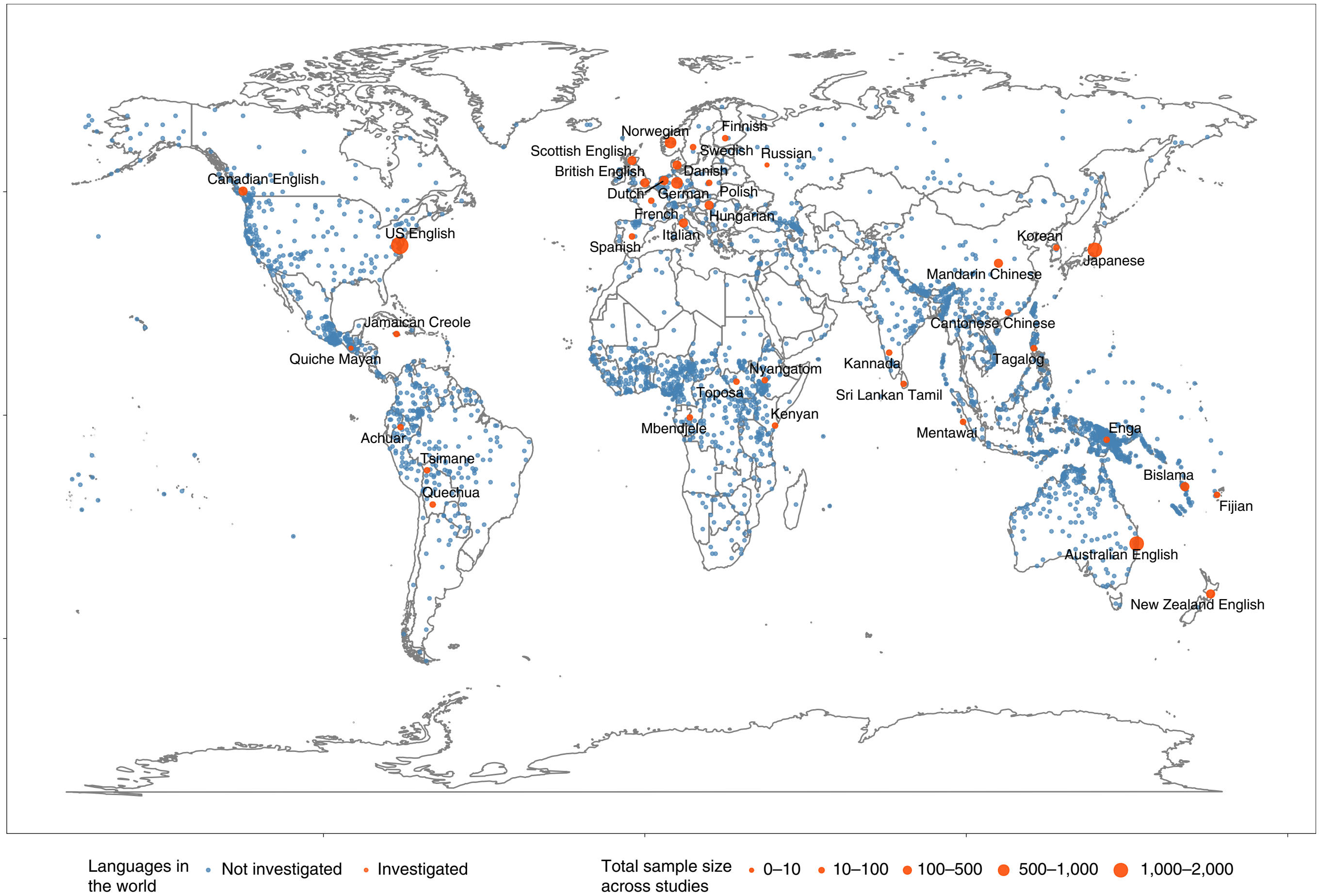


Fig. 1 | World map of IDS data. This plot provides a coarse overview of languages for which IDS has been analysed. It compares the languages included in this meta-analysis with the languages listed in the World Atlas of Language Structures[61]. The data were extracted from https://github.com/cldf-datasets/wals. Each point represents a language; the colour indicates whether the language is included in this meta-analysis, and the point size indicates the cumulative sample size. It should be noted that the exact map positions of the represented languages may not be accurate and that the World Atlas of Language Structures includes languages with very small speaking communities (for example, the vast number of languages in the United States).

we investigated in our meta-analysis have been shown to change as a function of infants' age. See Fig. 2 for a summary visualization.

The most common finding in studies examining the acoustic features of IDS is that IDS utterances, on average, have a higher fundamental frequency (f_0) and f_0 variability than ADS, resulting in the salient perceptual effects of perceived higher pitch and pitch variation[1,3]. Interestingly, many longitudinal studies on f_0 show that caregivers decrease their overall vocal pitch to infants over the course of development[3,66–70], but the findings are mixed, with other studies reporting no change over time[14,71–76]. Variability in f_0 shows a similar pattern. Pitch variation reflects intonational contours that provide information about speakers' expression of affect and intentions[35,77]. Longitudinal studies of f_0 variability in IDS indicate a peak before infants turn 12 months old, with a subsequent decrease over the course of development[3,66–68,70,72,75,76,78].

The tendency for caregivers to expand their vowel space area in IDS represents one of the more subtle adaptations of speech directed to infants. The most common measure calculates the area in acoustic space encompassed by the mean formant values of the three corner vowels: /i/, /a/ and /u/. Because these three vowels represent articulatory extremes and occur in the majority of the world's languages[79], studies focus on how caregivers adapt the acoustic–phonetic characteristics of these vowels in their IDS. Vowel space area is thus used as a measure of how much caregivers clarify their speech to infants[16,17] (but see refs. [20,21,23]). Most studies do not find evidence of any shift in vowel space area at a variety of age ranges[14,20,67,71,80–84]. But some studies have shown changes over time, although there are differences in the direction of the shift[85,86].

Articulation rate measures the speed at which people speak, which can have important consequences for how easily language is processed. This is true not only for young infants but also for adults, including second-language learners and listeners with other impairments[87]. Speaking too fast can prevent proper processing, which could affect phonological perception, emotional communication and other comprehension issues. Several longitudinal studies of articulation rate have shown that caregivers increase their rate of articulation (that is, speed up their speech) over the course of infant development[72,74–76]. Finally, vowel duration plays a crucial role in phonological processing[2], as well as in modulating infant attention and facilitating language development[88]. The exaggeration of the duration of vowels in IDS may make relevant phonological differences more salient to children, thereby facilitating their detection of clause and phrase boundaries[89,90]. Longitudinal studies in several languages indicate that caregivers often decrease relative vowel duration differences in IDS and ADS as infants get older[17,70,91].

In the current meta-analysis, we aimed to investigate the acoustic properties of IDS across infant ages and languages, and to understand these results in relation to the purported functions of IDS. We conducted this investigation by examining the influence of four moderator variables on possible acoustic differences between ADS and IDS: age, language, experimental task and recording environment.

https://doi.org/10.1038/s41562-022-01452-1

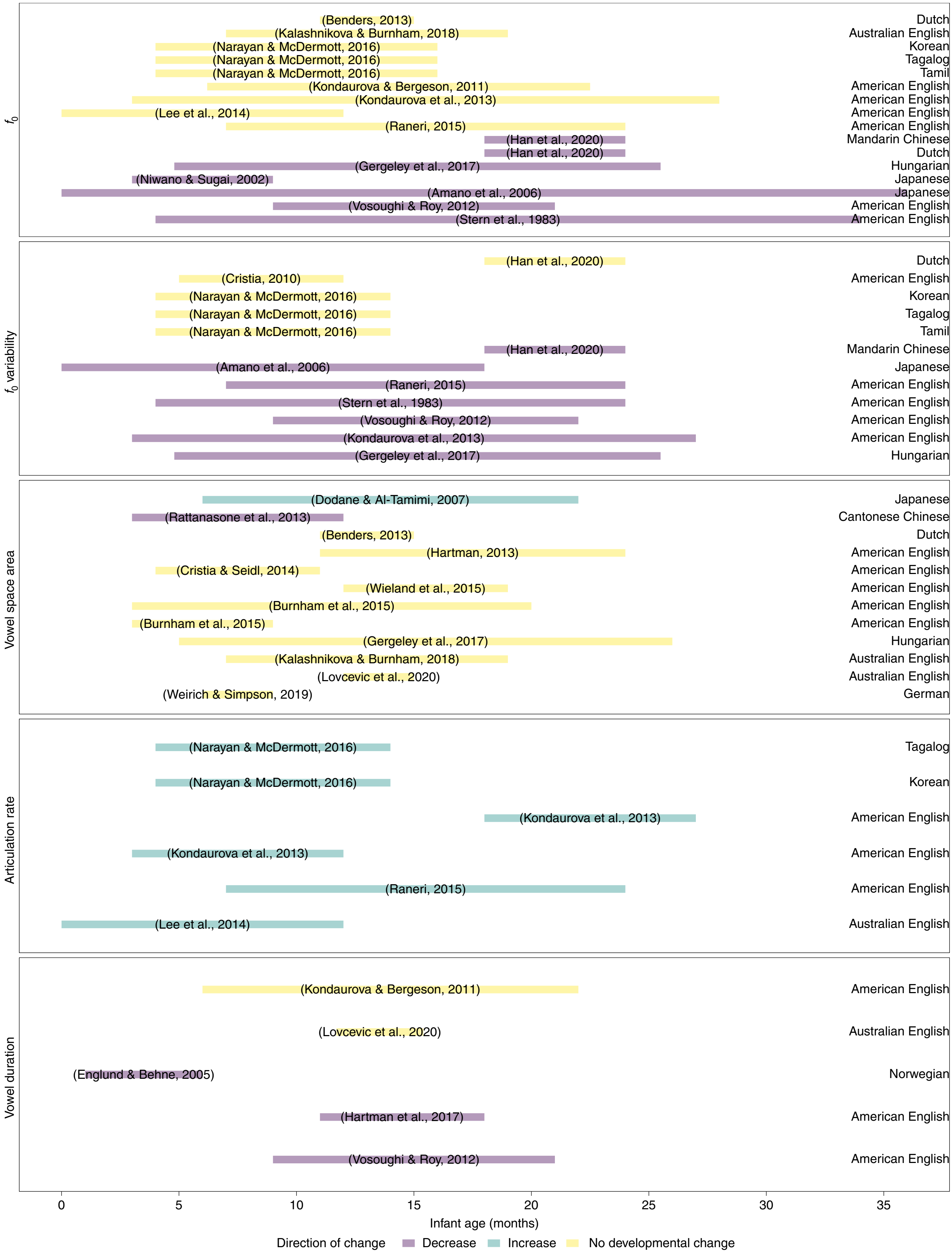


Fig. 2 | An overview of the findings from longitudinal studies for each of the acoustic features. The *x* axis indicates infant age in months, while the colours indicate whether the acoustic feature exhibits an increase (teal), a decrease (purple) or no change (yellow) over the course of the ages investigated. The studies are sorted by the main conclusion regarding any feature change. The data are from refs. [3,14,66–76,78,80–86,91].

https://doi.org/10.1038/s41562-022-01452-1

Table 1 | A summary of the results for the best models for each acoustic variable

Acoustic feature	No. of studies	No. of effect sizes	Average effect size	Evidence ratio	Study s.d.	Robust predictors
f_0	60	262	1.19 (0.81, 1.58)	Inf	0.91 (0.72, 1.14)	Language, age, task, environment
f_0 variability	44	202	0.46 (0.21, 0.71)	817.18	0.76 (0.60, 0.95)	Language, task
Vowel space area	33	84	0.81 (0.44, 1.16)	1,799	0.61 (0.41, 0.86)	Language
Articulation rate	17	60	−1.11 (−1.80, −0.39)	390.3	0.74 (0.42, 1.19)	Language, age, task
Vowel duration	26	81	0.51 (0.16, 0.86)	67.7	0.50 (0.12, 0.92)	Language, age

'Inf' means that all posterior samples are in the direction of the hypothesis. The average effect size refers to the average effect size across infant ages and languages in the best model for the acoustic measure. The numbers in parentheses are 95% CrIs.

The justification for each is briefly described here. First, by pooling data from the studies and quantifying the acoustic changes in IDS as a function of infant age, we can examine which of the acoustic properties of IDS change to become more similar to ADS over early infant development. Specific changes in the acoustic properties of IDS over developmental time would suggest that caregivers exhibit sensitivity to infants' shifting socio-emotional and linguistic needs and adapt their speech accordingly. If IDS in early development serves primarily to convey affect and only later serves a linguistic function, then we might expect to see developmental shifts in the acoustic properties that are primarily associated with linguistic facilitation (for example, vowel space area and vowel duration). Whether these linguistic features are present from birth or become gradually more exaggerated in IDS as infants exhibit linguistic development remains an open empirical question. Over longer timescales (not covered by the studies in this meta-analysis), we would expect all of the acoustic properties of IDS to gradually become indistinguishable from those of ADS. Second, to quantify the amount of cross-linguistic variation that could be observed, we analysed language as a moderator variable. For each acoustic variable, we provided language-specific estimates for each of the languages under investigation, as shown in Supplementary Tables 9.1–9.5. The data were too sparse to allow for an investigation of an interaction between infant age and the language spoken (Supplementary Fig. 7). Last, we analysed experimental task (that is, spontaneous versus read speech) and recording environment (that is, naturalistic versus laboratory) as moderators to examine whether the studies provided commensurable measurements across different conditions.

Whether the acoustic properties of caregivers' IDS change according to experimental task and recording environment remains an open question and an important consideration for future studies of IDS[42]. A cross-tab plot showing how the acoustic measures were distributed across the conditions of task and environment is shown in Supplementary Fig. 8. In addition to these moderator analyses, we conducted sensitivity analyses to quantify the robustness of our findings and to assess the evidentiary strength for each acoustic feature in light of publication bias. We computed the worst-case effect size estimate based only on non-affirmative studies and investigated how sensitive the meta-analytic results were to a potential bias for significant results in the field.

Results

Summary of the results

The overall results indicated a robust cross-linguistic tendency for caregivers to produce IDS with a higher pitch, higher pitch variability, an expanded vowel space area, a slower articulation rate and longer vowel durations. Table 1 provides a summary of the average effect size estimates for each of the acoustic measures as well as the estimated between-study variability. The heatmap in Extended Data Fig. 1 shows that the acoustic properties of IDS and ADS exhibit similar differences across languages, with some language specificity. In the following five sections, we delve deeper into how each of the five acoustic measures are moderated by language, age, experimental task and recording environment, and assess how sensitive the results are to publication bias.

f_0

We combined data from studies reporting either the mean or the median f_0 of utterances, as both measures indicate the central tendency of f_0. The following hierarchical model included 262 individual reported effect size measures from 60 studies. The model with task, environment, age and language as predictors was shown to provide a similar account of the data (stacking weight, 0.481) to the model excluding environment (stacking weight, 0.477), but a better account than the model excluding task (stacking weight, 0.014) and the model excluding task and environment (stacking weight, 0.029).

f_0 across studies. The Bayesian hierarchical intercepts-only model of f_0 revealed an overall estimated effect size of $g = 1.17$ with a 95% credible interval (CrI) of (0.86, 1.45), a between-languages heterogeneity of $g = 0.34$ (0.05, 0.67), a heterogeneity between studies within languages of $g = 0.90$ (0.71, 1.11) and a between-measures heterogeneity of $g = 0.07$ (0.00, 0.21). A standardized mean difference of this size implies that approximately 87.9% of IDS speech samples are expected to exhibit a higher f_0 than ADS speech samples. An overview of how the studies varied with respect to the f_0 estimate is shown in the forest plot in Supplementary Fig. 6.1. All of the studies exhibited effect size estimates on the positive scale, with only 16 of the 60 studies including the null in the lower bound of their CrIs.

f_0 as a function of language. The estimates from the full model are shown in Fig. 3. All of the point estimates for the languages under investigation were in the positive range of effect sizes. The cross-linguistic differences between IDS and ADS in f_0 across languages thus vary mainly according to the extent to which f_0 is higher in IDS than in ADS (see Supplementary Table 9.1 for language-specific estimates and CrIs).

f_0 as a function of age. As shown in the top right of Fig. 3, the model indicated a robust effect of age—as infants' ages increased, the difference in f_0 between IDS and ADS decreased. The estimate for the effect of age is −0.02 (95% CrI, (−0.03, 0.01); evidence ratio, 143.58; credibility, 0.99). This developmental pattern indicates that the cross-sectional data included in this meta-analysis conform to the results reported in most of the longitudinal studies (Fig. 2).

f_0 as a function of task and environment. As shown in the middle-right plot in Fig. 3, caregivers produced a greater f_0 difference between the two speech styles in experimental tasks designed to elicit spontaneous speech (estimate, 0.43; 95% CrI, (0.13, 0.74); evidence ratio, 94.54; credibility, 0.99). As shown in the lower-right plot in Fig. 3, parents recorded in a naturalistic setting as opposed to in the laboratory show a smaller difference between IDS and ADS in terms of f_0 (estimate, −0.48; 95% CrI, (−0.87, −0.07); evidence ratio, 36.54; credibility, 0.97).

Publication bias for f_0. The sensitivity analysis of publication bias for f_0 indicated that no amount of publication bias would be able to attenuate the effect size estimate for the CrI to include null effects, as depicted in Supplementary Fig. 10.1. The worst-case effect size estimate based solely on non-significant studies is 0.60 with a 95% CrI of (0.37, 0.83), as shown in Supplementary Fig. 10.2. This analysis suggests that the effect size estimates might be quite robust to even severe levels of publication bias, assuming that effect size estimates of non-significant studies are representative of those of unpublished studies.

f_0 variability

Some of the studies reported f_0 range ($n = 25$), and others reported the standard deviation of f_0 ($n = 20$). As these measures both capture change in f_0 over the course of the utterance, we grouped them into a single category. If a study reported both measures, we used the standard deviation because range consists of the difference between the highest and the lowest value and is therefore highly sensitive to even one outlier or measurement error. Standard deviation is less sensitive to extreme values and represents the more reliable measure of the two. The effect size distributions of f_0 range and f_0 standard deviation were shown to be strongly correlated and exhibit no notable differences, as shown in Supplementary Figs. 4.1 and 4.2. We extracted 223 effect sizes from 44 of the 88 studies. In this context, a positive Hedges's g value signifies a higher degree of f_0 variability in IDS, and a negative value signifies a lower degree of variability. The model with task, age and language as predictors provided a better account of the data (stacking weight, 0.681) than the model including both task and environment (stacking weight, 0.218), the model excluding task (stacking weight, 0.017), and the model excluding both task and environment (stacking weight, 0.084).

f_0 variability across studies. The Bayesian hierarchical intercepts-only model of f_0 variability showed an overall estimated difference of $g = 0.69$ with a 95% CrI of (0.44, 0.92) and a between-languages heterogeneity of $g = 0.25$ (0.02, 0.52), a heterogeneity between studies within languages of $g = 0.71$ (0.56, 0.88) and a between-measures heterogeneity of $g = 0.11$ (0.01, 0.23). With a standardized mean difference of this size, this implies that approximately 83% of IDS speech samples would show a higher degree of f_0 variability than that of ADS speech samples. An overview of how the studies varied with respect to the f_0 variability estimate is shown in the forest plot in Supplementary Fig. 6.2. The estimated effect sizes were primarily distributed on the positive scale, indicating that the studies provided evidence for greater f_0 variability in IDS than in ADS. Only 1 of the 43 studies on f_0 variability had a negative effect size point estimate, which the authors posit could be a result of caregivers' tendency to produce utterances with a higher minimum f_0 in IDS, thereby reducing the possible f_0 range[92].

f_0 variability as a function of language. As shown in Fig. 4, most of the point estimates for the languages were in the positive range of effect sizes (see Supplementary Table 9.2 for language-specific estimates and CrIs). The cross-linguistic differences between IDS and ADS in f_0 variability mainly related to the degree of exaggeration.

f_0 variability as a function of age. As shown in the top right of Fig. 4, the model indicated no effects of infant age (estimate, 0.00; 95% CrI, (−0.01, 0.01); evidence ratio, 1.33 for no effect; credibility, 0.57). This suggests that f_0 variability in caregivers' IDS remains stable even as infants become older. This is consistent with the results reported in some of the longitudinal studies under investigation (Fig. 2).

f_0 variability as a function of task and environment. The middle-right plot in Fig. 4 shows that caregivers spoke with a higher degree of f_0 variability in spontaneous speech than in read speech (estimate, 0.39; 95% CrI, (0.11, 0.68); evidence ratio, 89.68; credibility, 0.99). The lower-right plot in Fig. 4 indicates that recording the parents in a naturalistic setting as opposed to in the laboratory exerted a weak negative influence on the effect size estimates (estimate, −0.22; 95% CrI, (−0.59, 0.15); evidence ratio, 5.02; credibility, 0.83).

Publication bias for f_0 variability. A sensitivity analysis with a random-effects specification indicates that no amount of publication bias would be able to attenuate the effect size estimate for the CrI to include null effects, as depicted in Supplementary Fig. 10.1. The uncorrected worst-case estimate for the effect size based solely on non-significant studies is 0.33 with a 95% CrI of (0.18, 0.47), as shown in in Supplementary Fig. 10.2.

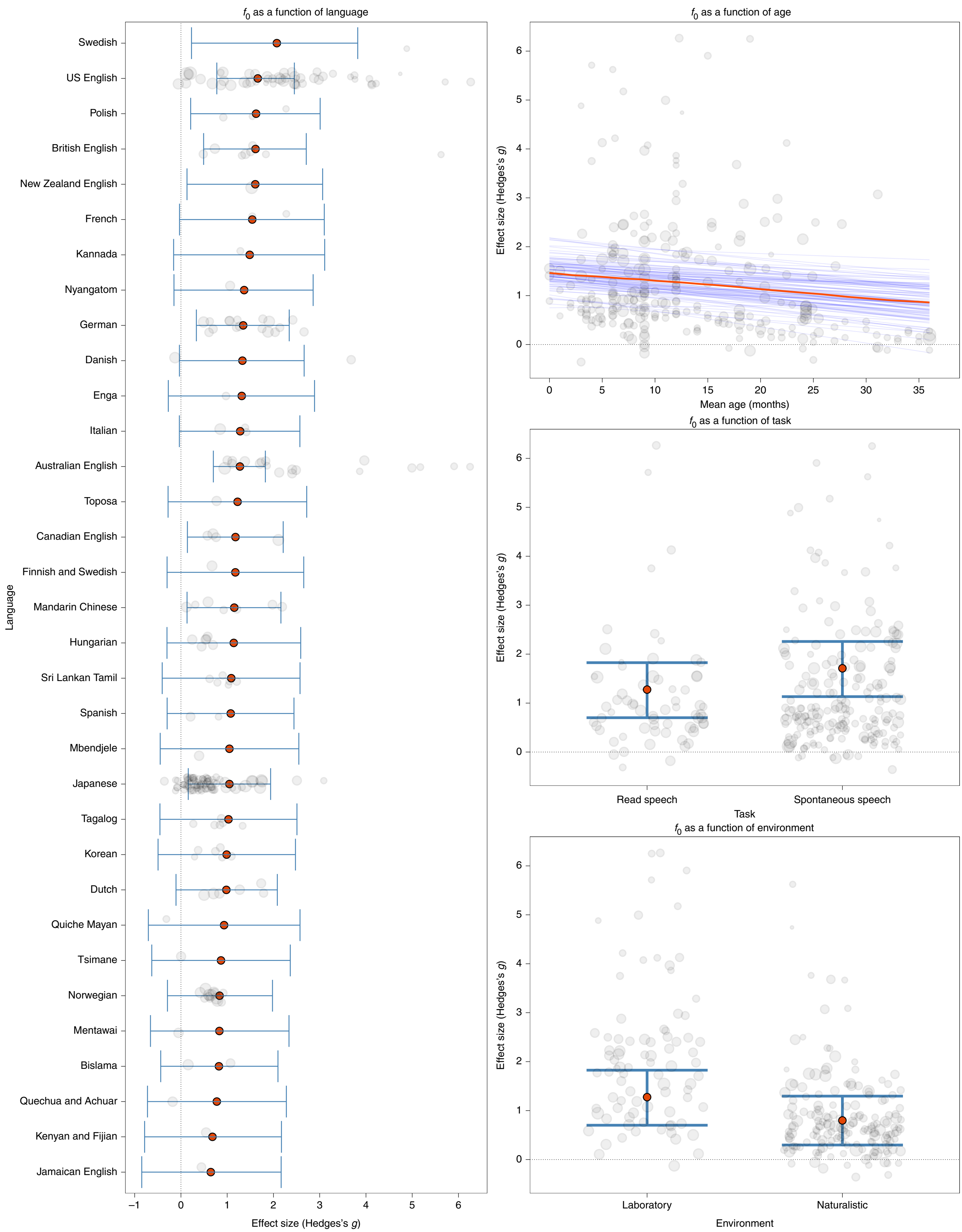
Vowel space area

Thirty-three studies reported vowel space area estimates, for a total of 107 reported effect sizes. In this context, a positive Hedges's g value signifies an expansion of the vowel space area in IDS. The model with age and language as predictors was shown to provide a better account of the data (stacking weight, 0.431) than the model including environment (stacking weight, 0.250), the model including task (stacking weight, 0.193) and the model including both task and environment (stacking weight, 0.127).

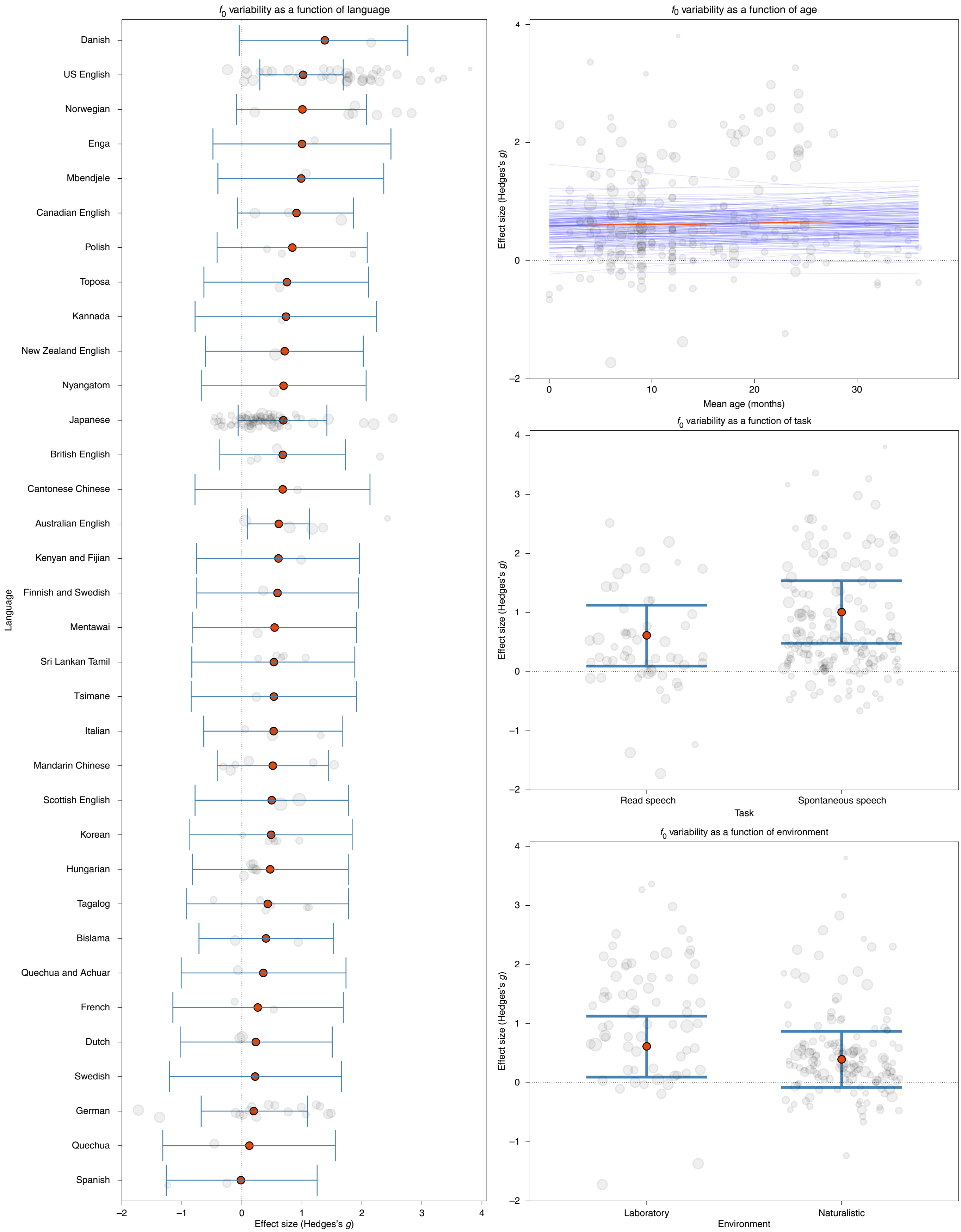
Vowel space area across studies. The Bayesian hierarchical intercepts-only model of vowel space area showed an overall estimated difference in vowel space area of $g = 0.66$ with a 95% CrI of (0.34, 0.98), a between-languages heterogeneity of $g = 0.55$ (0.12, 0.97), a heterogeneity between studies within languages of $g = 0.66$ (0.43, 0.92) and a between-measures heterogeneity of $g = 0.11$ (0.00, 0.28). A standardized mean difference of this size implies that approximately 74% of IDS speech samples overall will show an expanded vowel space area compared with those of ADS speech samples. An overview of how the studies varied with respect to the vowel space area estimate is shown in the forest plot in Supplementary Fig. 6.3. The studies were generally distributed across positive effect sizes; however, 19 of the 33 studies included the null in the lower bound of their CrIs, and 2 of the 33 studies provided evidence for the opposite effect—namely, that ADS exhibited an expanded vowel space area compared with IDS[71,93]. The pooling of data from these studies on vowel space area, then, indicated a moderate effect size, with some of the studies providing conflicting results (possibly due to cross-linguistic differences, as discussed further below and in the Discussion).

Fig. 3 | Model estimates for a total of 3,401 participants across 60 studies investigating 33 distinct languages. Left, effect size estimates for f_0 according to language. The orange points indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Top right, a spaghetti plot showing 100 posterior model predictions for the effect size estimates for f_0 as a function of age. Middle right, the distribution of effect size estimates across experimental tasks. The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Bottom right, the distribution of effect size estimates across recording environments. The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error).

https://doi.org/10.1038/s41562-022-01452-1



https://doi.org/10.1038/s41562-022-01452-1



Vowel space area as a function of language. As shown in Fig. 5, most of the point estimates for the languages were in the positive range of effect sizes (see Supplementary Table 9.3 for language-specific estimates and CrIs). However, there appears to be substantial cross-linguistic variation in the extent to which caregivers expand their vowel space area when speaking to infants.

Fig. 4 | Model estimates for a total of 3,006 participants across 44 studies investigating 34 distinct languages. Left, effect size estimates for f_0 variability according to language. The orange points indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Top right, a spaghetti plot showing 100 posterior model predictions for the effect size estimates for f_0 variability as a function of age. Middle right, the distribution of effect size estimates across experimental tasks. The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Bottom right, the distribution of effect size estimates across recording environments. The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error).

Vowel space area as a function of age. As shown in the top right of Fig. 5, the model indicated no evidence for an effect of infant age. The estimate is −0.00 (95% CrI, (−0.02, 0.01); evidence ratio, 2.04; credibility, 0.66).

Vowel space area as a function of task and environment. As shown in the middle-right plot in Fig. 5, caregivers did not seem to credibly produce a greater vowel space area in the experimental task of producing spontaneous speech (estimate, −0.16; 95% CrI, (−0.49, 0.18); evidence ratio, 3.58; credibility, 0.78). Similarly, as shown in the lower-right plot in Fig. 5, recording caregivers with their infants in a naturalistic setting did not appear to affect the vowel space area of caregivers' IDS (estimate, −0.27; 95% CrI, (−0.76, 0.23); evidence ratio, 4.29; credibility, 0.81).

Publication bias for vowel space area. A sensitivity analysis with a random-effects specification indicated that if moderate publication bias were present in the literature, then the effect size estimate may be closer to null effects. That is, if significant results were fourfold more likely to be published in the literature, the CrI would include an effect size of 0.1, as shown in Supplementary Fig. 10.1. The uncorrected worst-case estimate for the effect size based solely on non-significant studies is 0.20 with a 95% CrI of (−0.01, 0.42), as shown in Supplementary Fig. 10.2.

Articulation rate

Speech production rate is generally measured in one of two ways: articulation rate excludes pause intervals, but speech rate includes them and consequently accounts for speaker-specific ways of conveying information (for example, hesitations and pauses)[94–96]. The majority of studies under investigation here (15 of 17) reported articulation rate as opposed to speech rate. Because both of these measures capture similar acoustic information (that is, the number of output units per unit of time), we have combined the measures in our meta-analysis. But the distinction between them should be made theoretically because a slower speech rate may signify factors in addition to a slower articulation rate (for example, the number and duration of silent pauses)[94]. Here we use articulation rate to refer to this combination of measures.

The acoustic measure of articulation rate was analysed in 17 of the 88 studies and provided 60 separate effect sizes. A negative Hedges's g value in this context signifies a slower production rate in IDS. The model with task, age and language as predictors was shown to provide a better account of the data (stacking weight, 0.999) than the model including environment (stacking weight, 0.000), the model excluding task (stacking weight, 0.001) and the model excluding both task and environment (stacking weight, 0.000).

Articulation rate across studies. The Bayesian hierarchical intercepts-only model of articulation rate showed an overall estimated difference of $g = -1.03$ with a 95% CrI of (−1.53, −0.56) and a between-languages heterogeneity of $g = 0.38$ (0.02, 1.00), a heterogeneity between studies within languages of $g = 0.80$ (0.50, 1.20) and a heterogeneity between measurements of $g = 0.26$ (0.04, 0.47). With a standardized mean difference of this size, this implies that approximately 85% of IDS speech samples will show a slower rate than ADS speech samples. An overview of how the studies varied with respect to the articulation rate estimate is shown in the forest plot in Supplementary Fig. 6.4. The estimated effect sizes of the studies are distributed primarily on the negative scale, indicating that caregivers on average speak slower in IDS than in ADS; however, due to the relative sparsity of data for this acoustic measure, many of the languages include null effects in their CrIs.

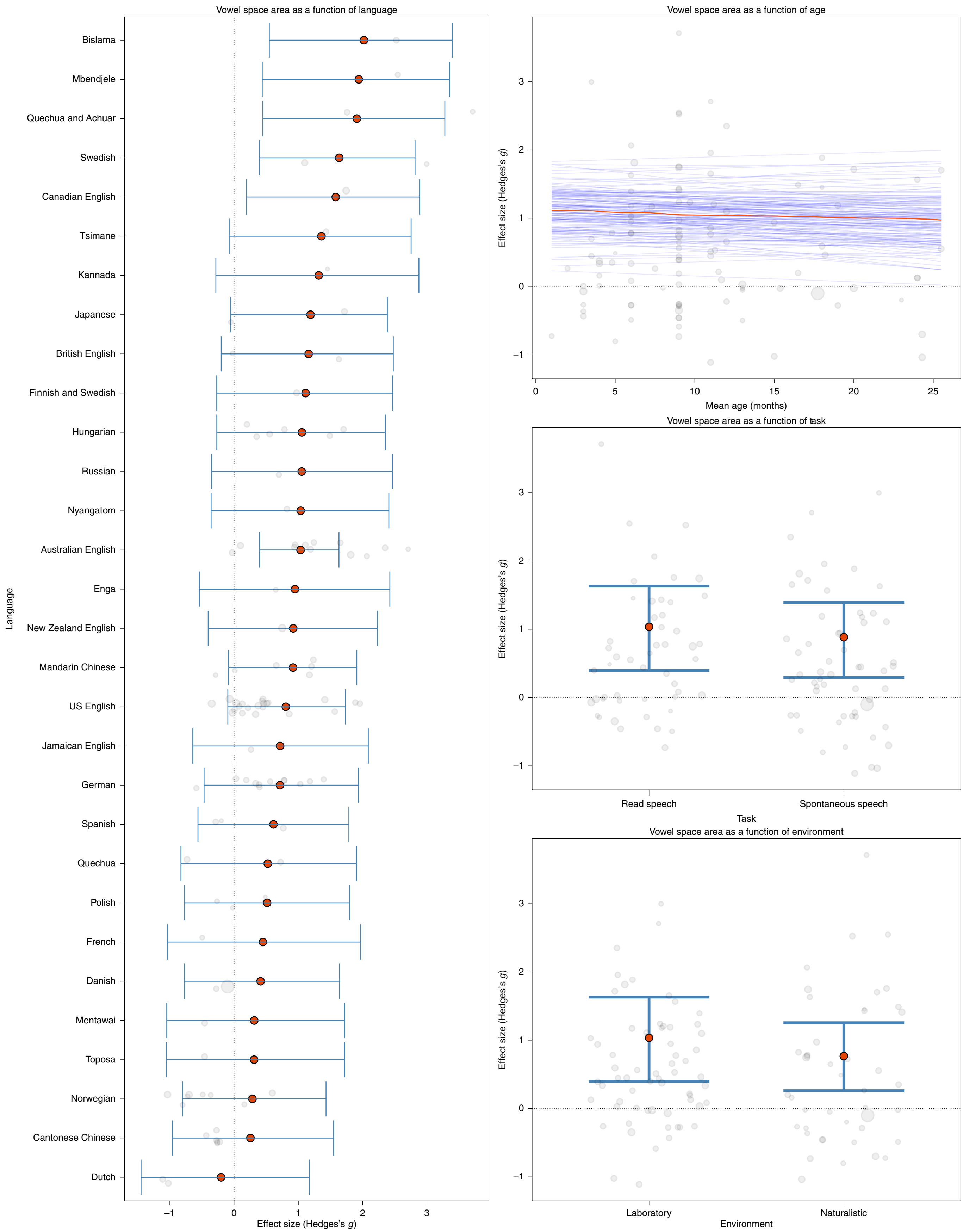
Articulation rate as a function of language. As shown on the left side of Fig. 6, all of the effect size point estimates for the languages under investigation were in the negative range (see Supplementary Table 9.4 for language-specific estimates and CrIs).

Articulation rate as a function of age. As shown in the top right of Fig. 6, the model indicated a reliable effect of infant age. The estimate for the effect of age is 0.02 (95% CrI, (0.00, 0.05); evidence ratio, 33.33; credibility, 0.97). This result shows that caregivers' articulation rate in IDS becomes more similar to that in ADS over the course of infant development from 0 to 30 months.

Articulation rate as a function of task and environment. As shown in the middle-right plot in Fig. 6, caregivers appeared to speak faster to their infants in spontaneous speech than in read speech (estimate, 0.95; 95% CrI, (0.1, 1.73); evidence ratio, 28.34; credibility, 0.97). In contrast, the lower-right plot in Fig. 6 indicates that there is no evidence that recording caregivers outside of the laboratory affects the articulation rate in caregivers' IDS (estimate, 0.15; 95% CrI, (−0.71, 0.96); evidence ratio, 1.66; credibility, 0.62).

Fig. 5 | Model estimates for a total of 1,702 participants across 33 studies investigating 30 distinct languages. Left, effect size estimates for vowel space area according to language. The orange points indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Top right, a spaghetti plot showing 100 posterior model predictions for the effect size estimates for vowel space area as a function of age. Middle right, the distribution of effect size estimates across experimental tasks. The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Bottom right, the distribution of effect size estimates across recording environments. The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error).

https://doi.org/10.1038/s41562-022-01452-1



Publication bias for articulation rate. A sensitivity analysis with a random-effects specification indicated that no amount of publication bias would be able to attenuate the estimate to null, as shown in Supplementary Fig. 10.1. If moderate publication bias were present in the literature, then the effect size estimate may represent a more moderate effect; the uncorrected worst-case estimate for the effect size based solely on non-significant studies is −0.445 with a 95% CrI of (−0.757, −0.133), as shown in Supplementary Fig. 10.2.

Vowel duration

The acoustic measure of vowel duration was analysed in 26 of the 88 studies, and 81 effect sizes were extracted from these studies. We should note that the vowel categories for which data were available differed markedly across studies, with some studies reporting vowel duration only for the articulatory extremes of /i/, /a/ and /u/[82,93], and others reporting vowel duration for the full set of vowel phonemes in their language[97]. In this context, a positive Hedges's *g* value signifies a longer vowel duration in IDS than in ADS, and a negative value signifies a shorter duration. The model with age and language as predictors was shown to provide a better account of the data (stacking weight, 0.393) than the model including task and environment (stacking weight, 0.154), the model including task (stacking weight, 0.242) and the model including environment (stacking weight, 0.211).

Vowel duration across studies. The Bayesian hierarchical intercepts-only model of vowel duration showed an overall estimated difference of $g = 0.48$ with a 95% CrI of (0.08, 0.88), a between-languages heterogeneity of $g = 0.38$ (0.03, 0.92), a heterogeneity between studies within languages of $g = 0.43$ (0.06, 0.85) and a between-measures heterogeneity of $g = 0.17$ (0.01, 0.38). With a standardized mean difference of this size, this implies that approximately 70% of IDS speech samples will show a longer vowel duration than that of ADS speech samples. An overview of how the studies varied with respect to the vowel duration estimate is shown in the forest plot in Supplementary Fig. 6.5. The majority of the effect size estimates were distributed on the positive scale, indicating that caregivers produce vowels with a longer duration in IDS than in ADS.

Vowel duration as a function of language. As shown in Fig. 7, most of the effect size estimates for the languages under investigation were in the positive range (see Supplementary Table 9.5 for language-specific estimates and CrIs). However, there appears to be an influence of language-specific phonological properties, as some languages exhibit substantially longer vowel durations in IDS (for example, Mandarin Chinese), mixed results (for example, US English and Japanese) or no durational differences between the speech styles (for example, Swedish, Norwegian and Danish).

Vowel duration as a function of age. As shown in the top right of Fig. 7, the model indicated a moderate effect of infant age. The estimate for the effect of age is −0.02 (95% CrI, (−0.05, 0.01); evidence ratio, 6.48; credibility, 0.87). This suggests that caregivers' vowel durations in IDS became slightly more similar to those in ADS as infants got older.

Vowel duration as a function of task and environment. As shown in the middle-right plot in Fig. 7, there appeared to be weak evidence that caregivers spoke with a greater vowel duration difference in spontaneous speech (estimate, −0.12; 95% CrI, (−0.97, 0.74); evidence ratio, 1.44; credibility, 0.58), although note that this estimate was based on only three data points for the task of read speech. The lower-right plot in Fig. 7 indicates that recording the infants in a naturalistic setting exerted a weak positive influence on the effect size estimates (estimate, 0.27; 95% CrI, (−0.51, 1.06); evidence ratio, 2.47; credibility, 0.71).

Publication bias for vowel duration. A sensitivity analysis with a random-effects specification indicated that no amount of publication bias can attenuate the estimate to 0.1, as shown in the sensitivity plot in Supplementary Fig. 10.1. The uncorrected worst-case estimate for the effect size based solely on non-significant studies is 0.277 with a 95% CrI of (0.134, 0.417), as shown in Supplementary Fig. 10.2.

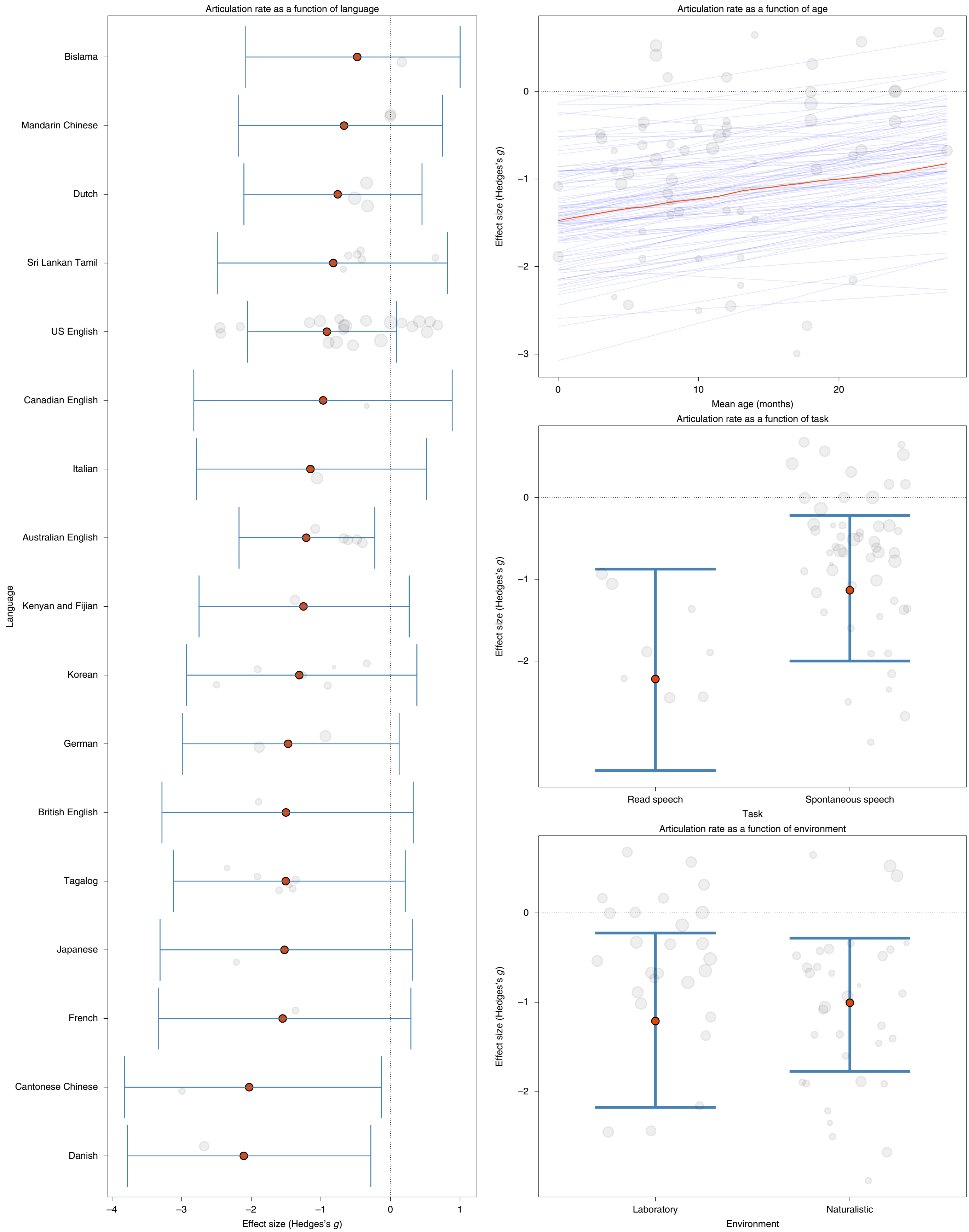
Discussion

The tendency for caregivers to modify their speech to infants represents a widespread cross-cultural and cross-linguistic phenomenon. The aims of this meta-analysis were to examine how the acoustic properties of IDS (1) change over the course of early infant development, (2) vary across languages and (3) differ according to experimental task and recording environment, with an eye towards a better understanding of culturally widespread IDS communicative functions. The results confirmed that across multiple languages and cultures, IDS contains acoustic features that are distinct from ADS, and that different acoustic features operate on varying timescales. Our analysis of publication bias showed that the pattern of acoustic features in IDS would remain reliable even if a strong bias for significant results existed in the literature (although potentially with the exception of vowel space area; Supplementary Figs. 10.1 and 10.2). The findings thus provide reliable evidence that caregivers across multiple languages produce IDS with a higher f_0, a higher degree of f_0 variability, an expanded vowel space area, a slower articulation rate and a longer vowel duration, as summarized in Figs. 3–7 and Table 1 (see also Supplementary Tables 9.1–9.5). The analyses, however, also suggested a high degree of unexplained between-study and between-language heterogeneity. Our analyses of moderators indicated that f_0, articulation rate and vowel duration became more similar to ADS over the course of infants' early development, while vowel space area and f_0 variability remained stable, at least up to 25 and 36 months of age, respectively. Our analysis of the effect of experimental task revealed that spontaneous speech displayed greater differences in f_0, articulation rate and f_0 variability between ADS and IDS, compared with read speech. Recording environment likewise showed a reliable influence on the estimates for f_0.

In the following sections, we discuss our findings in light of the following questions. (1) To what extent do the acoustic features of IDS change over time, and how do these findings speak to the putative functions of IDS? (2) How much do the acoustic properties of IDS vary across languages? (3) What are the sources of variation? We use these questions as opportunities to reflect on the scientific study of IDS and to provide study recommendations that can inform theory building, modelling approaches and future experimental and descriptive investigations.

Fig. 6 | Model estimates for a total of 976 participants across 17 studies investigating 17 distinct languages. Left, effect size estimates for articulation rate according to language. The orange points indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Top right, a spaghetti plot showing 100 posterior model predictions for the effect size estimates for articulation rate as a function of age. Middle right, the distribution of effect size estimates across experimental tasks. The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Bottom right, the distribution of effect size estimates across recording environments. The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error).

https://doi.org/10.1038/s41562-022-01452-1



Changes in IDS features and their relation to functions

The tendency for some of the acoustic features of IDS to change over the course of early development may be due to a form–function relationship between caregivers' acoustic production patterns and infants' attentional allocation to certain aspects of the speech stream[11–13,48]. For example, the increase in articulation rate and parallel decrease in vowel duration during development may reflect caregivers' sensitivity to infants' improved processing of the speech stream. Articulation rate exhibits robustness across languages (Fig. 6), with a universal tendency for caregivers to slow down their speech to infants. Slowed IDS probably eases the cognitive load involved in young infants' speech and language processing[98–100]. Similarly, the decrease in the utterance-global measure of f_0 in IDS may be a consequence of infants' changing preferences to attend to this acoustic feature in the speech stream[13]. Younger infants have been shown to prefer to attend to the positive affect of IDS[45,46], while older infants prefer aspects of the speech stream that provide less positive affect and more linguistically relevant information[11,12,48]. Vocal pitch exhibited a high degree of robustness across languages (Fig. 3), supporting the notion that it is a highly salient property of IDS[1,3] and that caregivers adjust IDS acoustic properties in ways that suit infants' developmental needs[101,102]. Similarly, the cross-linguistic tendency for the acoustic properties of f_0 variability and vowel space area to remain stable throughout early infancy (Supplementary Fig. 6.6) suggests ongoing developmental relevance[18,19]. We should note, however, that vowel space area exhibited cross-linguistic variation (Fig. 5), with some of the studies reporting reduced vowel separability in IDS[71,86,91,93]. Both acoustic features have been implicated in facilitating language development[16,17,39,103], but whether the benefits of IDS derive mainly from its capacity to direct infants' attention or to emphasize linguistic aspects of the speech stream (or both) remains an important open question. We should also note that although infant age appears to affect some of the acoustic measures, the amount of available data across different age ranges varies, ranging from 0–25 months for vowel duration to 0–36 for f_0 and f_0 variability (Supplementary Fig. 6.6). These results highlight the need for an expansion in the availability of data with a high density of observations across many different age ranges.

Computational evidence indicates that vowel space expansion can aid speech intelligibility[25,104–106], but beyond considerations of the information content in the speech signal[5,107], the benefits may simply be a product of the social qualities of IDS, which facilitate learning through increased infant attention[36,38] and social motivation[29,46]. The question of how specific acoustic properties in IDS may facilitate aspects of infant development could be pursued with more detailed theory-driven studies of languages with distinct linguistic systems.

Unexplained variability across studies and languages

Our meta-analytic models revealed a substantial amount of between-study heterogeneity for each of the acoustic features, especially among the studies reporting measures of f_0, f_0 variability and articulation rate (Table 1). Some between-study heterogeneity is expected simply from random sampling error and the mathematics of estimating an effect across a large number of studies[108,109]. But some of this unexplained variance may derive from the inclusion of studies that differ from one another in meaningful ways, such as in study designs, population sample characteristics, cross-linguistic diversity and experimental methodologies[110]. For example, our results indicated larger differences between the speech styles in f_0, f_0 variability and articulation rate for studies recording parents' spontaneous speech as opposed to read speech (Figs. 3, 4 and 6). Without a complete characterization of the sources of this unexplained heterogeneity, factors influencing the generalizability of the effects remain undetermined and therefore constitute an important avenue for future research.

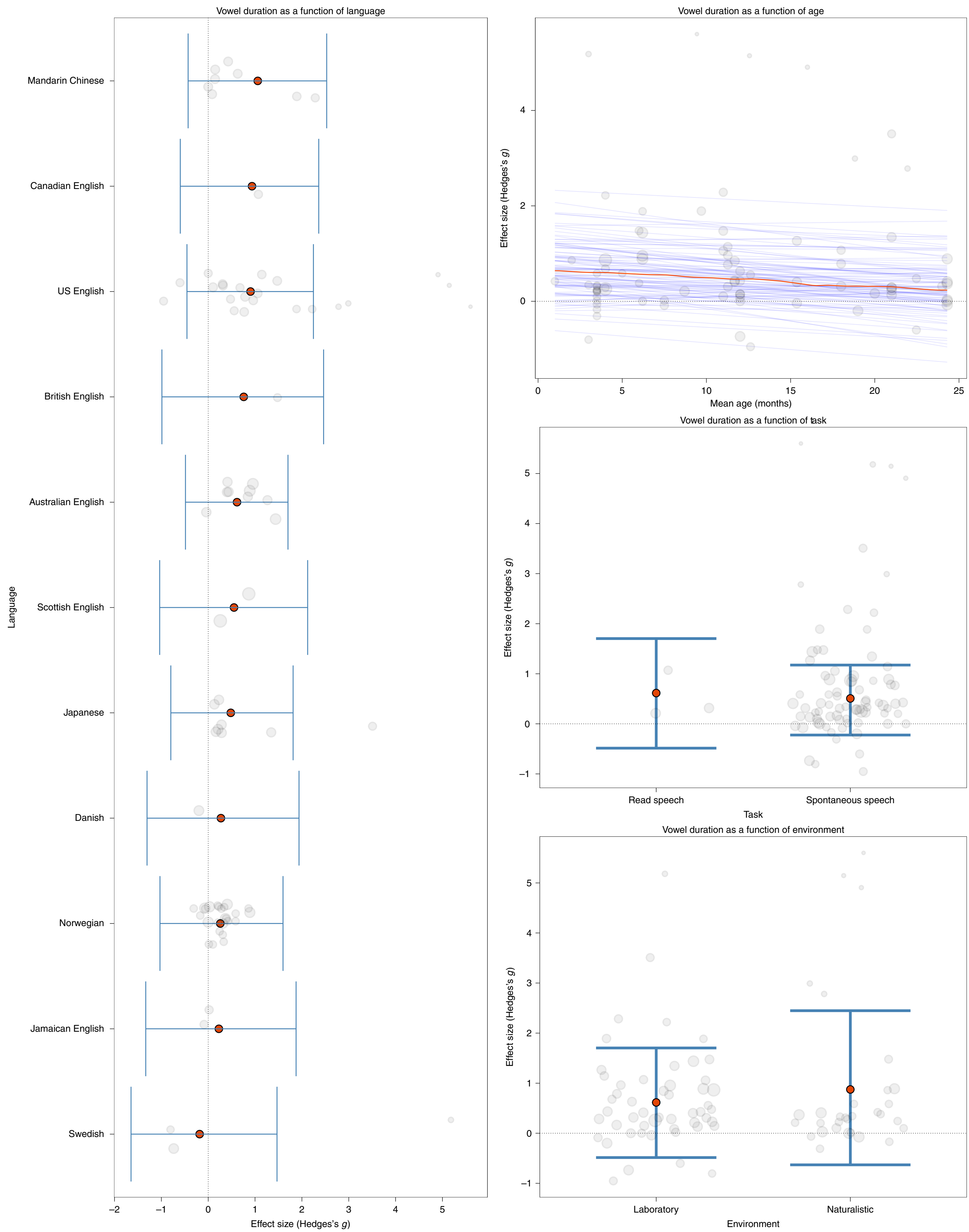
One source of heterogeneity could be the variability induced by cross-linguistic differences in IDS. The acoustic features of IDS were shown to vary across languages, many of which relied on a small number of data points and studies and therefore exhibited substantial uncertainty (Supplementary Tables 9.1–9.5). Part of this heterogeneity and cross-linguistic uncertainty may also depend on the variability caused by subtle differences in phonological systems across languages. For example, although our results suggest a strong cross-linguistic tendency for caregivers to produce IDS with an overall slower articulation rate, Church et al.[111] found that the difference in articulation rate between Canadian English ADS and IDS to 8.5- and 11-month-old infants disappeared when utterance-final syllables were excluded, due to the phonological tendency for utterance-final syllables to be lengthened in Canadian English (see ref. [112] for similar results for Japanese). Similarly, substantial differences in the number and category of vowels included in our analysis of vowel duration may influence the generalizability of results in languages with other types of vowel inventories and phonological systems. Determining the influence of subtle cross-linguistic differences, such as prosodic phonology, as well as vowel inventories and phonemes, will be a fruitful area for future investigations. Although we were unable to accommodate these types of subtle phonological differences between languages in our analyses, these sources of variability highlight the need for fine-grained, theory-driven comparisons of the acoustic properties of IDS across different languages and population characteristics (for example, gender and ethnicity) as well as careful consideration of the causal mechanisms involved[62–64].

Another source of the between-study heterogeneity may be intra-study participant characteristics. Low sample sizes and tight experimental controls characteristic of infant research may result in outcomes that are idiosyncratic to particular study conditions[109]. Between-study differences in participant characteristics, such as gender and kinship, are thus likely to function as potential sources of unexplained heterogeneity. For example, the high prevalence of post-partum depression[113,114] and its attested effects on the prosodic properties of IDS[115–117] may affect the generalizability of the current results to these population samples. The developmental status of the infant, moreover, may also function as a potential source of heterogeneity in IDS properties, as caregivers have been shown to respond differently according to this status[32,101,118]. Future research exploring the effects of diverse speaker characteristics, such as depression, kinship, gender and infants' developmental status, would provide important insights into factors affecting the acoustic properties of IDS.

To allow for more fine-grained temporal analyses of how acoustic features of IDS manifest themselves across early infancy, and to further explore sources of between-study variability, we encourage

Fig. 7 | Model estimates for a total of 1,411 participants across 26 studies investigating 11 distinct languages. Left, effect size estimates for vowel duration according to language. The orange points indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Top right, a spaghetti plot showing 100 posterior model predictions for the effect size estimates for vowel duration as a function of age. Middle right, the distribution of effect size estimates across experimental tasks. The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error). Bottom right, the distribution of effect size estimates across recording environments. The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% CrI, and the grey points are the raw effect size data. The size of each point is proportional to the inverse of the standard error of the effect size (that is, the larger the point, the smaller the standard error).

https://doi.org/10.1038/s41562-022-01452-1



researchers to share participant-level data in open repositories. A cumulative approach to improving the external validity of studies can also be carried out by conducting experiments across multiple laboratories[4,36], affording the exploration of within-lab and between-lab variability. Because logistical constraints may hinder multi-laboratory approaches, we argue that providing access to participant-level data may represent the easiest, most practical alternative.

Despite the finding of substantial between-study heterogeneity, we should emphasize that the studies exhibited consistency with each other; that is, the CrIs for the results of individual studies showed substantial overlap (Supplementary Figs. 6.1–6.5). Moreover, our meta-analytic models included random effects by study to address the dependency among effect sizes as well as predictor variables to explain the heterogeneity between studies. In the following section, we provide a series of recommendations that will enable a better understanding of the factors moderating the acoustic properties of caregivers' IDS.

Recommendations for future research

While solid progress has been made towards examining a wide variety of relevant aspects of IDS, we have identified various shortcomings that should be addressed in future investigations. First, with the continued rise of day-long recordings[57] and open archives of acoustic and phonetically transcribed data[58], as well as the continued development of techniques to automatically assess and code large amounts of audio data[56,119], future research can expand the availability of cross-linguistic data and provide a high density of observations for each participant[4,120]. These technological developments will allow for a more fine-grained resolution and comparison of how IDS differs across individuals, languages and infant ages. Second, as noted above, to further explore the functions and learnability afforded by IDS, more theory-driven comparisons across distinct linguistic systems are needed[62,64], as well as testable predictions from computational models disentangling different theoretical accounts. For example, computational models that explore the supposed learnability afforded by the acoustic properties of IDS constitute fruitful future avenues of research[106,121,122], as do computational models of stimulus-driven attention and prominence of IDS[123] and other sensory inputs more generally[124,125]. Assessing these models on data from a broad range of cultural, linguistic and sociodemographic settings would provide a more robust assessment of theoretical limitations and provide fuel for further theoretical development. Finally, adapting speech to a listener is not a unilateral phenomenon. We want to highlight the importance of considering the mutual feedback loops between infant and caregiver, with infants being an important source of information regarding which sort of signal would be most beneficial for their developmental progress[29–33]. This is especially important given the substantial variability in developmental trajectories across individuals. Studies investigating the importance of the bidirectional process of adaptation between infants' communicative signals and caregiver responsiveness on a turn-by-turn basis comprise another fruitful avenue of future work that can deliver new accounts, predictions and data from both interactants' viewpoints[5,14,32,91,101,126,127].

The current meta-analysis investigated the acoustic features of IDS across a variety of languages and cultures by aggregating data from three decades of research on this speech style. We found robust evidence that adults worldwide often speak to infants in ways that differ systematically from how they speak to other adults (that is, they alter a range of acoustic features). Moreover, how caregivers speak to infants changes as a function of infants' ages. We propose that the observed modifications in acoustic features over the course of early infancy may reflect caregivers' dynamic sensitivity to changes in infants' attention to specific acoustic properties in the speech stream.

Our results provide support for several findings in the literature, including the robust effects of cross-linguistic differences, infant ages and experimental tasks. However, the precise nature of these differences remains elusive. We therefore recommend that future studies (1) share participant-level data to enable the analysis of individual differences and intra-study variability, (2) conduct theory-driven comparative studies of cross-linguistic differences, (3) formulate computational models on the functions and learnability afforded by IDS, and (4) conduct longitudinal studies on the importance of dynamic adaptation to the developmental process.

Methods

To obtain a comprehensive sample of the available literature on acoustic properties of IDS, we conducted a systematic literature search on PubMed and Web of Science, in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guidelines[128] (Supplementary Fig. 1.1 and Supplementary Tables 1.2 and 1.3). The search terms used were "motherese" OR "baby talk" OR "child-directed speech" OR "infant-directed speech" OR "caretaker speech" OR "parentese", with no search limits in the query to target studies broadly. The first systematic search was conducted independently by two of the authors (R.F. and E.F.) in June 2017 and updated by a third author (C.C.) in December 2021; C.C. screened for missed studies from before and after the date of the first systematic search. Disagreements in the screening of papers were resolved with discussions in the first phase between E.F. and R.F. and in the second phase between C.C. and R.F.; if the paper was thought to contain relevant data for the meta-analysis (see below), the paper was included in the successive phase of the review. Disagreements were therefore rare and mainly motivated by studies where relevant information was reported only in the Supplementary Information. As of December 2021, the search strategy yielded a total of 602 papers, which were manually screened for inclusion according to the following criteria: (1) the infants had to be typically developing, (2) the studies had to include the quantification of an acoustic feature, (3) the studies had to include a comparison condition with ADS and (4) the speech had to be spoken to an infant by one or both of their primary caregivers.

On the basis of the initial set of 602 papers, we used Connected Papers and Research Rabbit to find an additional 48 relevant studies. After excluding 54 duplicate studies, we screened the titles of 596 studies and excluded a further 302 studies that were unrelated to the current investigation. We read the abstracts of the remaining 294 studies and evaluated each with reference to the above exclusion criteria. Of the 294 papers, 175 studies had no relation to IDS, 17 studies had no comparison condition with ADS and 15 studies examined atypical populations and had no relevant control sample of typically developing infants to extract data on. We further discuss the importance of future studies investigating more diverse speaker characteristics in the Discussion. To the best of our knowledge, the present review of a total of 88 studies represents a comprehensive sample of the literature on IDS.

To assess the state of the literature and to explore the extent to which the studies build a common discourse with reciprocal references, we used the R package bibliometrix[129] to build coupling and direct-citation networks of the studies, as shown in Supplementary Fig. 2.1. The studies cluster into three main groups and exhibit considerable overlap in the studies they cite. Furthermore, they cite each other somewhat independently of the acoustic measure reported. The collection of studies investigated here thus represents a coherent intersection of papers that build a common discourse on a variety of relevant aspects of IDS.

Data extraction

The following meta-analyses allowed us to explore how each acoustic measure differed across infant ages, languages, experimental tasks and recording environments. We classed the 88 relevant papers into five clusters on the basis of the acoustic measure reported: f_0, f_0 variability, vowel space area, articulation rate and vowel duration. If an individual study reported multiple acoustic measures, the study was included in all of the relevant clusters. It should be noted that other acoustic measures of IDS were reported in some of the studies under investigation (for example, syllable duration (three studies), pause duration (five studies) and intensity (five studies)); however, the studies provided insufficient data for meta-analysis.

To standardize the measures and to allow for comparison among the studies, we calculated Hedges's g, an effect size variant that is preferred for small sample sizes[130,131]. For our purposes, this effect size represents the standardized mean difference between ADS and IDS—that is, the bigger the effect size, the larger the difference between the speech styles. A positive effect size indicates that the value for IDS is greater than that for ADS, and vice versa. This implies that an acoustic property of IDS that becomes more similar to ADS over the course of development would manifest as a shift towards an effect size of zero.

When the raw means and standard deviations were reported in the papers, we calculated the effect sizes with standard formulae for Hedges's g (that is, $g = \frac{\text{mean}_1 - \text{mean}_2}{\text{s.d.}_{\text{pooled}}}$, where $\text{s.d.}_{\text{pooled}} = \frac{(n_1-1)\text{s.d.}_1^2 + (n_2-1)\text{s.d.}_2^2}{n_1+n_2-2}$), as formulated in ref. [132], where the standard deviation of each group is weighted by its sample size, using the R package esc[133]. For the remaining studies that did not report the raw data, the effect sizes were calculated either by using the reported d values or one-sample t values or by digitally extracting the raw data from published plots using the WebPlotDigitizer application[134]. In certain cases, the standard deviation of the effect size could not be calculated from the reported data or plots. To include these effect sizes in the meta-analysis, we imputed these missing standard deviation values ($n = 110$) by using multivariate imputation by chained equations based on a Bayesian linear regression model in the R package mice[135], as described further in Supplementary Section 3. We checked that this process of multiple imputation did not bias the estimation of the overall effect size for each acoustic measure by comparing the estimates of the intercepts-only models for the imputed and non-imputed datasets. The results of these analyses are shown in Supplementary Table 3.1. All hierarchical Bayesian models in this paper pool the results of analyses performed on the imputed datasets. In Supplementary Tables 11.1 and 12.1 (see also Fig. 1), we provide more information about the size of the sample investigated for each language.

Hierarchical Bayesian model

In the following meta-analyses of the five acoustic features, we combined the weighted results of comparable studies and provided pooled estimates of the overall effect sizes. We estimated and adjusted for heterogeneity in population samples and methodologies by allowing the estimate to vary by study. The hierarchical structure of the random-effects model posits that the true effect size may be study-specific and thereby accounts for repeated measures[136–138]. The CrI of the pooled estimate thus aggregates information from both within-study sampling error and between-study variance[139]. The hierarchical Bayesian robust regression models were fitted to the meta-analytic data using a Student's t likelihood. With this type of robust regression model, longer-tailed distributions are implemented to reduce the influence of outliers. This method incorporates outliers without allowing them to dominate non-outlier data[140]. See Supplementary Section 5 for a detailed account of the models and choice of priors (Supplementary Table 5.1), prior and posterior predictive checks (Supplementary Fig. 5.2.1), prior–posterior update plots (Supplementary Fig. 5.3.1), and prior sensitivity analyses for the model estimates (Supplementary Fig. 5.4.1) and evidence ratios (Supplementary Fig. 5.4.2) of intercept and age.

Moderator analyses

We began by building intercepts-only models to condition the data for each of the acoustic measures on the variance associated with individual studies. With these models, we posited that effect sizes were nested within languages and within studies. To quantify the within-language variability due to different studies reporting data on the same language and repeated measures within these studies, we included nested effects of study and measures within the random-effects term (that is, (1 | Language/StudySite/measurement)). We used these three-level intercepts-only models to assess the within-language, between-study heterogeneity and report how the effect size estimates of each study deviate from the pooled effect size estimate (Supplementary Figs. 6.1–6.5).

We then constructed a second model to analyse the influence of potential moderators on the variation of effect sizes across studies. This second model allowed us to explore the effects of the following predictors on each of the acoustic measures: infant age, language, experimental task and recording environment (the justifications for these predictors are described in the Introduction). We refer to this second model as the full model for the remainder of this paper.

We performed pairwise leave-one-out information-criterion-based model comparison[141] between the full model and models without each of the predictor variables. We report leave-one-out stacking weights[142] in favour of the model. Stacking weights indicate the probability that the model including the variables is better than the model without the predictor variables. All computations were performed in R v.4.2.0 (ref. [143]) using brms v.2.17 (ref. [144]) and Stan v.2.21 (ref. [145]) in RStudio v.1.4 (ref. [146]).

For each acoustic measure, we provide the estimates from the full model and report 95% CrIs, evidence ratios, credibility scores and leave-one-out stacking weights for each of the models. CrIs indicate the range of values within which there is a 95% probability that the true value of the parameter is included given the assumptions of the model. The evidence ratio provides the ratio of likelihood in favour of a hypothesis; that is, an evidence ratio of 5 indicates that the hypothesis is 5 times more likely than the alternative, while an evidence ratio of 'Inf' (infinite) suggests that all of the posterior samples are compatible with the hypothesis and not with the alternatives[144,147,148]. The credibility score refers to the percentage of posterior samples in the direction of the hypothesis under investigation[144]. Lastly, stacking weight refers to the probability that the model including a predictor provides a better model of the data than the model without the predictor[141]. The estimates from the best model for each acoustic variable are reported in Supplementary Tables 9.1–9.5.

We chose to assess publication bias by conducting quantitative sensitivity analyses and estimating the severity of the publication bias required to attenuate the CrI of the pooled effect size to include null values[149]. Traditional assessments of publication bias rely on Spearman rank correlations between effect size and standard error and exhibit certain limitations[150]. These traditional methods, for example, provide binary decisions either rejecting the null hypothesis of no publication bias or not and fail to control for type I error rates when used with standardized mean difference effect sizes and conventional variance estimates[151,152]. This is especially the case when within-study sample sizes are relatively small or between-study heterogeneity is high[153]. We therefore chose to assess how robust the meta-analytic estimates would be to varying assumptions of publication bias[149]. These methods assume that meta-analytic studies represent samples from an underlying (possible) population of published and unpublished studies, where the probability of selection for significant studies is higher. The potential presence of publication bias is thereby assessed (1) by varying assumptions as to how much more likely significant studies are to be published than non-significant studies and (2) by calculating the amount of publication bias required to attenuate the estimates so that the evidence in favour of an effect becomes negligible. This method has limitations, such as relaxing certain distributional assumptions on the population effects and assuming that the non-significant findings available are representative of the whole population of unpublished studies[149]. However, the method still offers substantial benefits over classical funnel plot methods and selection models (see refs. [151–153] for reviews). It should be noted that this method of analysing publication bias sensitivity cannot comment on the severity of publication bias in practice or the opposite; rather, this analysis provides results that allow us to assess the extent to which an effect would be present even if publication bias were a severe issue in the literature. For each acoustic measure, we report the worst-case effect size estimate based solely on the non-significant studies and make sensitivity plots and significance funnel plots (Supplementary Figs. 10.1 and 10.2).

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All data were accessed on PubMed and Web of Science and are available and permanently archived in the following open repository: https://osf.io/hc7me/. Source data are provided with this paper.

Code availability
The analysis and visualization code and a reproducible R Markdown manuscript are available and permanently archived in the following open repository: https://osf.io/hc7me/.

References

1. Fernald, A. Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* https://doi.org/10.2307/1130938 (1989).
2. Fernald, A. et al. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* **16**, 477–501 (1989).
3. Stern, D. N., Spieker, S., Barnett, R. & MacKain, K. The prosody of maternal speech: infant age and context related changes. *J. Child Lang.* **10**, 1–15 (1983).
4. Hilton, C. B. et al. Acoustic regularities in infant-directed speech and song across cultures. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-022-01410-x (2022).
5. Golinkoff, R. M., Can, D. D., Soderstrom, M. & Hirsh-Pasek, K. Baby talk to me: the social context of infant-directed speech and its effects on early language acquisition. *Curr. Dir. Psychol. Sci.* **24**, 339–344 (2015).
6. Bryant, G. A. & Barrett, H. C. Recognizing intentions in infant-directed speech: evidence for universals. *Psychol. Sci.* **18**, 746–751 (2007).
7. Fernald, A. Meaningful melodies in mothers' speech to infants. in *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (eds Papousek, H., Jürgens, U. & Papoušek, M.) 262–282 (Cambridge Univ. Press, 1992).
8. Owren, M. J. & Rendall, D. Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol. Anthropol.* **10**, 58–71 (2001).
9. Kitamura, C. & Lam, C. Age-specific preferences for infant-directed affective intent. *Infancy* **14**, 77–100 (2009).
10. Lam, C. & Kitamura, C. Developmental trends in infant preferences for affective intent in mothers' speech. In *Proc. 11th Australian International Conference on Speech Science & Technology* (eds Warren, P. & Watson, C. I.). 100–105 (2006).
11. Kitamura, C. & Notley, A. The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Dev. Sci.* **12**, 706–714 (2009).
12. McRoberts, G. W., McDonough, C. & Lakusta, L. The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy* **14**, 162–194 (2009).
13. Panneton, R., Kitamura, C., Mattock, K. & Burnham, D. Slow speech enhances younger but not older infants' perception of vocal emotion. *Res. Hum. Dev.* **3**, 7–19 (2006).
14. Kalashnikova, M. & Burnham, D. Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *J. Child Lang.* **45**, 1035–1053 (2018).
15. Kuhl, P. K. et al. Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**, 684–686 (1997).
16. Liu, H., Kuhl, P. K. & Tsao, F. An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* **6**, F1–F10 (2003).
17. Hartman, K. M., Ratner, N. B. & Newman, R. S. Infant-directed speech (IDS) vowel clarity and child language outcomes. *J. Child Lang.* **44**, 1140–1162 (2017).
18. Peter, V., Kalashnikova, M., Santos, A. & Burnham, D. Mature neural responses to infant-directed speech but not adult-directed speech in pre-verbal infants. *Sci. Rep.* **6**, 34273 (2016).
19. Song, Demuth, K. & Morgan, J. Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* **128**, 389–400 (2010).
20. Cristia, A. & Seidl, A. The hyperarticulation hypothesis of infant-directed speech. *J. Child Lang.* **41**, 913–934 (2014).
21. Martin, A. et al. Mothers speak less clearly to infants than to adults: a comprehensive test of the hyperarticulation hypothesis. *Psychol. Sci.* **26**, 341–347 (2015).
22. McMurray, B., Kovack-Lesh, K. A., Goodwin, D. & McEchron, W. Infant directed speech and the development of speech perception: enhancing development or an unintended consequence? *Cognition* **129**, 362–378 (2013).
23. Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H. & Mazuka, R. Vowels in infant-directed speech: more breathy and more variable, but not clearer. *Cognition* **166**, 84–93 (2017).
24. Rosslund, A., Mayor, J., Óturai, G. & Kartushina, N. Parents' hyper-pitch and vowel category compactness in infant-directed speech are associated with 18-month-old toddlers' expressive vocabulary. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/wrku5 (2021).
25. Eaves, B. S., Feldman, N. H., Griffiths, T. L. & Shafto, P. Infant-directed speech is consistent with teaching. *Psychol. Rev.* **123**, 758–771 (2016).
26. Perry, L. K., Samuelson, L. K., Malloy, L. M. & Schiffer, R. N. Learn locally, think globally: exemplar variability supports higher-order generalization and word learning. *Psychol. Sci.* **21**, 1894–1902 (2010).
27. Rost, G. C. & McMurray, B. Speaker variability augments phonological processing in early word learning. *Dev. Sci.* **12**, 339–349 (2009).
28. Rost, G. C. & McMurray, B. Finding the signal by adding noise: the role of noncontrastive phonetic variability in early word learning. *Infancy* **15**, 608–635 (2010).
29. Goldstein, M. H. & Schwade, J. A. Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* **19**, 515–523 (2008).
30. Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M. & Soderstrom, M. Entrainment of prosody in the interaction of mothers with their young children. *J. Child Lang.* **43**, 284–309 (2016).
31. Murray, L. & Trevarthen, C. The infant's role in mother–infant communications. *J. Child Lang.* **13**, 15–29 (1986).
32. Nguyen, V., Versyp, O., Cox, C. M. M. & Fusaroli, R. A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Dev.* https://doi.org/10.1111/cdev.13754 (2022).
33. Warlaumont, A. S., Richards, J. A., Gilkerson, J. & Oller, D. K. A social feedback loop for speech development and its reduction in autism. *Psychol. Sci.* **25**, 1314–1324 (2014).
34. Cooper, R. P. & Aslin, R. N. Preference for infant-directed speech in the first month after birth. *Child Dev.* **61**, 1584–1595 (1990).
35. Fernald, A. & Simon, T. Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* **20**, 104–113 (1984).

36. ManyBabies Consortium. Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Adv. Methods Pract. Psychol. Sci.* **3**, 24–52 (2020).
37. Pegg, J. E., Werker, J. F. & McLeod, P. J. Preference for infant-directed over adult-directed speech: evidence from 7-week-old infants. *Infant Behav. Dev.* **15**, 325–345 (1992).
38. Werker, J. F. & McLeod, P. J. Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Can. J. Psychol.* **43**, 230–246 (1989).
39. Fernald, A. & Kuhl, P. Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* **10**, 279–293 (1987).
40. Bainbridge, C. M. et al. Infants relax in response to unfamiliar foreign lullabies. *Nat. Hum. Behav.* **5**, 256–264 (2021).
41. Tsuji, S. et al. MetaLab: a repository for meta-analyses on language development, and more. In *Proc. Interspeech 2017* 2038–2039 (International Speech Communication Association, 2017).
42. Dunst, C., Gorman, E. & Hamby, D. Preference for infant-directed speech in preverbal young children. *Cent. Early Lit. Learn.* **5**, 1–13 (2012).
43. Hayashi, A., Tamekawa, Y. & Kiritani, S. Developmental change in auditory preferences for speech stimuli in Japanese infants. *J. Speech Lang. Hear. Res.* **44**, 1189–1200 (2001).
44. Newman, R. S. & Hussain, I. Changes in preference for infant-directed speech in low and moderate noise by 4.5- to 13-month-olds. *Infancy* **10**, 61–76 (2006).
45. Kitamura, C. & Burnham, D. Acoustic and affective qualities of IDS in English. In *Proc. 5th International Conference on Spoken Language Processing* 0909 (International Speech Communication Association, 1998).
46. Singh, L., Morgan, J. L. & Best, C. T. Infants' listening preferences: baby talk or happy talk? *Infancy* **3**, 365–394 (2002).
47. Fernald, A. & Mazzie, C. Prosody and focus in speech to infants and adults. *Dev. Psychol.* **27**, 209–221 (1991).
48. Segal, J. & Newman, R. S. Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy* **20**, 339–351 (2015).
49. Ferguson, C. A. Baby talk in six languages. *Am. Anthropol.* **66**, 103–114 (1964).
50. Casillas, M., Brown, P. & Levinson, S. C. Early language experience in a Tseltal Mayan village. *Child Dev.* **91**, 1819–1835 (2020).
51. Cristia, A., Dupoux, E., Gurven, M. & Stieglitz, J. Child-directed speech is infrequent in a forager–farmer population: a time allocation study. *Child Dev.* **90**, 759–773 (2019).
52. Shneidman, L. A. & Goldin-Meadow, S. Language input and acquisition in a Mayan village: how important is directed speech? *Dev. Sci.* **15**, 659–673 (2012).
53. Bryant, G. A. Vocal communication across cultures: theoretical and methodological issues. *Phil. Trans. R. Soc. B* **377**, 20200387 (2022).
54. Grieser, D. L. & Kuhl, P. K. Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* **24**, 14–20 (1988).
55. Papoušek, M., Papoušek, H. & Symmes, D. The meanings of melodies in motherese in tone and stress languages. *Infant Behav. Dev.* **14**, 415–440 (1991).
56. Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A. & Casillas, M. An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behav. Res. Methods* **53**, 818–835 (2021).
57. Xu, D., Yapanel, U. & Gray, S. *Reliability of the LENA Language Environment Analysis System in Young Children's Natural Home Environment* (LENA Foundation, 2009).
58. MacWhinney, B. *The CHILDES Project: Tools for Analyzing Talk* Vol. 2 (Psychology Press, 2014).
59. Floccia, C. et al. British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition* **148**, 1–9 (2016).
60. Shute, B. & Wheldall, K. Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educ. Psychol.* **19**, 221–233 (1999).
61. Haspelmath, M. The typological database of the World Atlas of Language Structures. in *The Use of Databases in Cross-Linguistic Studies* (eds Everaert M., Musgrave S. & Dimitriadis A.) 283–300 (De Gruyter Mouton, 2009).
62. Christiansen, M. H., Kallens, P. C. & Trecca, F. Towards a comparative approach to language acquisition. *Curr. Dir. Psychol. Sci.* https://doi.org/10.1177/09637214211049229 (2022).
63. Deffner, D., Rohrer, J. M. & McElreath, R. A causal framework for cross-cultural generalizability. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/fqukp (2021).
64. Trecca, F., Tylén, K., Højen, A. & Christiansen, M. H. Danish as a window onto language processing and learning. *Lang. Learn.* **71**, 799–833 (2021).
65. Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* **162**, 31–38 (2017).
66. Amano, S., Nakatani, T. & Kondo, T. Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *J. Acoust. Soc. Am.* **119**, 1636–1647 (2006).
67. Gergely, A., Faragó, T., Galambos, Á. & Topál, J. Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: an acoustic analysis. *Sci. Rep.* **7**, 1–10 (2017).
68. Han, M., De Jong, N. & Kager, R. Pitch properties of infant-directed speech specific to word-learning contexts: a cross-linguistic investigation of Mandarin Chinese and Dutch. *J. Child Lang.* **47**, 85–111 (2020).
69. Niwano, K. & Sugai, K. Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *Jpn. J. Spec. Educ.* **39**, 59–68 (2002).
70. Vosoughi, S. & Roy, D. K. A longitudinal study of prosodic exaggeration in child-directed speech. In *Proc. Speech Prosody 6th International Conference* (Speech Prosody Special Interest Group, 2012); http://hdl.handle.net/1721.1/80790
71. Benders, T. Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behav. Dev.* **36**, 847–862 (2013).
72. Kondaurova, M. V., Bergeson, T. R. & Xu, H. Age-related changes in prosodic features of maternal speech to prelingually deaf infants with cochlear implants. *Infancy* **18**, 825–848 (2013).
73. Kondaurova, M. V. & Bergeson, T. R. The effects of age and infant hearing status on maternal use of prosodic cues for clause boundaries in speech. *J. Speech Lang. Hear. Res.* **54**, 740–754 (2011).
74. Lee, C. S., Kitamura, C., Burnham, D. & McAngus Todd, N. P. On the rhythm of infant- versus adult-directed speech in Australian English. *J. Acoust. Soc. Am.* **136**, 357–365 (2014).
75. Narayan, C. R. & McDermott, L. C. Speech rate and pitch characteristics of infant-directed speech: longitudinal and cross-linguistic observations. *J. Acoust. Soc. Am.* **139**, 1272–1281 (2016).
76. Raneri, D. P. *Infant-Directed Speech: Maternal Pitch Variability, Rate of Speech, and Child Language Outcomes*. Doctoral dissertation, Univ. Maryland (2015).
77. Knoll, M. A. & Costall, A. Characterising F (0) contour shape in infant- and foreigner-directed speech. *Speech Commun.* **66**, 231–243 (2015).
78. Cristià, A. Phonetic enhancement of sibilants in infant-directed speech. *J. Acoust. Soc. Am.* **128**, 424–434 (2010).

79. Liljencrants, J. & Lindblom, B. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* https://doi.org/10.2307/411991 (1972).
80. Burnham, E. B. et al. Phonetic modification of vowel space in storybook speech to infants up to 2 years of age. *J. Speech Lang. Hear. Res.* **58**, 241–253 (2015).
81. Hartman, K. M. *Patterns and Possible Influences of Maternal Vowel Clarification on Child Language Development*. Doctoral dissertation, Univ. Maryland (2013).
82. Lovcevic, I., Kalashnikova, M. & Burnham, D. Acoustic features of infant-directed speech to infants with hearing loss. *J. Acoust. Soc. Am.* **148**, 3399–3416 (2020).
83. Weirich, M. & Simpson, A. Effects of gender, parental role, and time on infant- and adult-directed read and spontaneous speech. *J. Speech Lang. Hear. Res.* **62**, 4001–4014 (2019).
84. Wieland, E. A., Burnham, E. B., Kondaurova, M., Bergeson, T. R. & Dilley, L. C. Vowel space characteristics of speech directed to children with and without hearing loss. *J. Speech Lang. Hear. Res.* **58**, 254–267 (2015).
85. Dodane, C. & Al-Tamimi, J. *An Acoustic Comparison of Vowel Systems in Adult-Directed-Speech and Child-Directed-Speech: Evidence from French, English & Japanese* (Newcastle Univ., 2007).
86. Rattanasone, N. X., Burnham, D. & Reilly, R. G. Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *J. Phon.* **41**, 332–343 (2013).
87. Huettig, F. & Guerra, E. Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Res.* **1706**, 196–208 (2019).
88. Gleitman, L. R., Newport, E. L. & Gleitman, H. The current status of the motherese hypothesis. *J. Child Lang.* **11**, 43–79 (1984).
89. Seidl, A. & Cristià, A. Developmental changes in the weighting of prosodic cues. *Dev. Sci.* **11**, 596–606 (2008).
90. Soderstrom, M., Seidl, A., Nelson, D. G. K. & Jusczyk, P. W. The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *J. Mem. Lang.* **49**, 249–267 (2003).
91. Englund, K. T. & Behne, D. M. Infant directed speech in natural interaction—Norwegian vowel quantity and quality. *J. Psycholinguist. Res.* **34**, 259–280 (2005).
92. Outters, V., Schreiner, M. S., Behne, T. & Mani, N. Maternal input and infants' response to infant-directed speech. *Infancy* **25**, 478–499 (2020).
93. Steen, V. B. & Englund, N. Child-directed speech in a Norwegian kindergarten setting. *Scand. J. Educ. Res.* https://doi.org/10.1080/00313831.2021.1897873 (2021).
94. Laver, J. *Principles of Phonetics* (Cambridge Univ. Press, 1994).
95. Tsao, Y.-C., Weismer, G. & Iqbal, K. Interspeaker variation in habitual speaking rate: additional evidence. *J. Speech Lang. Hear. Res.* **49**, 1156–1164 (2006).
96. Tsao, Y.-C., Weismer, G. & Iqbal, K. The effect of intertalker speech rate variation on acoustic vowel space. *J. Acoust. Soc. Am.* **119**, 1074–1082 (2006).
97. Englund, K. T. Hypoarticulation in infant-directed speech. *Appl. Psycholinguist.* **39**, 67–87 (2018).
98. Christiansen, M. H. & Chater, N. The now-or-never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* **39**, E62 (2016).
99. Saffran, J. R. & Kirkham, N. Z. Infant statistical learning. *Annu. Rev. Psychol.* **69**, 181–203 (2018).
100. Werker, J. F. & Tees, R. C. Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* **50**, 509–535 (1999).
101. Fusaroli, R., Weed, E., Fein, D. & Naigles, L. Hearing me hearing you: reciprocal effects between child and parent language in autism and typical development. *Cognition* **183**, 1–18 (2019).
102. Smith, N. A. & Trainor, L. J. Infant-directed speech is modulated by infant feedback. *Infancy* **13**, 410–420 (2008).
103. Spinelli, M., Fasolo, M. & Mesman, J. Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Dev. Rev.* **44**, 1–18 (2017).
104. De Boer, B. & Kuhl, P. K. Investigating the role of infant-directed speech with a computer model. *Acoust. Res. Lett. Online* **4**, 129–134 (2003).
105. McMurray, B., Aslin, R. N. & Toscano, J. C. Statistical learning of phonetic categories: insights from a computational approach. *Dev. Sci.* **12**, 369–378 (2009).
106. Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F. & Amano, S. Unsupervised learning of vowel categories from infant-directed speech. *Proc. Natl Acad. Sci. USA* **104**, 13273–13278 (2007).
107. Kalashnikova, M., Carignan, C. & Burnham, D. The origins of babytalk: smiling, teaching or social convergence? *R. Soc. Open Sci.* **4**, 170306 (2017).
108. Mikolajewicz, N. & Komarova, S. V. Meta-analytic methodology for basic research: a practical guide. *Front. Physiol.* **10**, 203 (2019).
109. Song, F., Sheldon, T. A., Sutton, A. J., Abrams, K. R. & Jones, D. R. Methods for exploring heterogeneity in meta-analysis. *Eval. Health Prof.* **24**, 126–151 (2001).
110. Ruppar, T. Meta-analysis: how to quantify and explain heterogeneity? *Eur. J. Cardiovasc. Nurs.* **19**, 646–652 (2020).
111. Church, R., Bernhardt, B., Pichora-Fuller, K. & Shi, R. Infant-directed speech: final syllable lengthening and rate of speech. *Can. Acoust.* **33**, 13–19 (2005).
112. Martin, A., Igarashi, Y., Jincho, N. & Mazuka, R. Utterances in infant-directed speech are shorter, not slower. *Cognition* **156**, 52–59 (2016).
113. Gavin, N. I. et al. Perinatal depression: a systematic review of prevalence and incidence. *Obstet. Gynecol.* **106**, 1071–1083 (2005).
114. Gelaye, B., Rondon, M. B., Araya, R. & Williams, M. A. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. *Lancet Psychiatry* **3**, 973–982 (2016).
115. Kaplan, P. S., Bachorowski, J., Smoski, M. J. & Zinser, M. Role of clinical diagnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy* **2**, 537–548 (2001).
116. Lam-Cassettari, C. & Kohlhoff, J. Effect of maternal depression on infant-directed speech to prelinguistic infants: implications for language development. *PLoS ONE* **15**, e0236787 (2020).
117. Porritt, L. L., Zinser, M. C., Bachorowski, J.-A. & Kaplan, P. S. Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Lang. Learn. Dev.* **10**, 51–67 (2014).
118. Woolard, A. et al. Infant and child-directed speech used with infants and children at risk or diagnosed with autism spectrum disorder: a scoping review. *Rev. J. Autism Dev. Disord.* **9**, 290–306 (2022).
119. Cychosz, M. et al. Vocal development in a large-scale crosslinguistic corpus. *Dev. Sci*. https://doi.org/10.1111/desc.13090 (2021).
120. Le Franc, A. et al. The ACLEW DiViMe: an easy-to-use diarization tool. In *Proc. Interspeech 2018* 1383–1387 (International Speech Communication Association, 2018).
121. Adriaans, F. & Swingley, D. Distributional learning of vowel categories is supported by prosody in infant-directed speech. *Proc. Annu. Meet. Cogn. Sci. Soc.* **34**, 72–77 (2012).
122. Ludusan, B., Mazuka, R. & Dupoux, E. Does infant-directed speech help phonetic learning? A machine learning investigation. *Cogn. Sci.* **45**, e12946 (2021).
123. Räsänen, O., Kakouros, S. & Soderstrom, M. Is infant-directed speech interesting because it is surprising? Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition* **178**, 193–206 (2018).

124. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* **7**, e36399 (2012).
125. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks effect in infant auditory attention. *Child Dev.* **85**, 1795–1804 (2014).
126. Lam, C. & Kitamura, C. Maternal interactions with a hearing and hearing-impaired twin: similarities and differences in speech input, interaction quality, and word production. *J. Speech Lang. Hear. Res.* **53**, 543–555 (2010).
127. Lam, C. & Kitamura, C. Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation. *Dev. Sci.* **15**, 212–221 (2012).
128. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int. J. Surg.* **88**, 105918 (2021).
129. Aria, M. & Cuccurullo, C. bibliometrix: an R-tool for comprehensive science mapping analysis. *J. Informetr.* **11**, 959–975 (2017).
130. Morris, S. B. Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *Br. J. Math. Stat. Psychol.* **53**, 17–29 (2000).
131. Parola, A., Simonsen, A., Bliksted, V. & Fusaroli, R. Voice patterns in schizophrenia: a systematic review and Bayesian meta-analysis. *Schizophr. Res.* **216**, 24–40 (2020).
132. Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Stat.* **6**, 107–128 (1981).
133. Lüdecke, D. esc: Effect size computation for meta analysis. R package version 05 (2019).
134. Rohatgi, A. *WebPlotDigitizer User Manual Version 3.4* (2014).
135. Van Buuren, S. & Groothuis-Oudshoorn, K. mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
136. Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-analysis* (Academic Press, 1985).
137. Raudenbush, S. W. & Bryk, A. S. Empirical Bayes meta-analysis. *J. Educ. Stat.* **10**, 75–98 (1985).
138. Fernández-Castilla, B. et al. The application of meta-analytic (multi-level) models with multiple random effects: a systematic review. *Behav. Res. Methods* **52**, 2031–2052 (2020).
139. Hedges, L. V. & Vevea, J. L. Fixed- and random-effects models in meta-analysis. *Psychol. Methods* **3**, 486–504 (1998).
140. Jylänki, P., Vanhatalo, J. & Vehtari, A. Robust Gaussian process regression with a Student-*t* likelihood. *J. Mach. Learn. Res.* **12**, 3227–3257 (2011).
141. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
142. Yao, Y., Vehtari, A., Simpson, D. & Gelman, A. Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Anal.* **13**, 917–1007 (2018).
143. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
144. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
145. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
146. RStudio: Integrated Development for R v.1.4 (RStudio, PBC, 2020).
147. Lemoine, N. P. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* **128**, 912–928 (2019).
148. Gelman, A., Simpson, D. & Betancourt, M. The prior can often only be understood in the context of the likelihood. *Entropy* **19**, 555 (2017).
149. Mathur, M. B. & VanderWeele, T. J. Sensitivity analysis for publication bias in meta-analyses. *J. R. Stat. Soc. C* **69**, 1091–1119 (2020).
150. Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M. & Gaigg, S. B. Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Res.* **10**, 384–407 (2017).
151. Jin, Z., Zhou, X. & He, J. Statistical methods for dealing with publication bias in meta-analysis. *Stat. Med.* **34**, 343–360 (2015).
152. McShane, B. B., Böckenholt, U. & Hansen, K. T. Adjusting for publication bias in meta-analysis: an evaluation of selection methods and some cautionary notes. *Perspect. Psychol. Sci.* **11**, 730–749 (2016).
153. Pustejovsky, J. E. & Rodgers, M. A. Testing for funnel plot asymmetry of standardized mean differences. *Res. Synth. Methods* **10**, 57–71 (2019).

Acknowledgements

R.F. has been a paid consultant for F. Hoffman La Roche on related but non-overlapping topics. This project has been supported by seed funding from the Interacting Minds Centre at Aarhus University, awarded to R.F. and E.F. All of the computation done for this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

C.C., R.F. and E.F. conceived the research and extracted the data from studies. C.C., G.B., R.F. and C.B. wrote the initial manuscript with contributions from T.K.-P., A.R. and E.F. C.C. wrote the first and second revisions with contributions from R.F., G.B. and T.K.-P. C.C. and R.F. led the statistical analyses with contributions from C.B. C.C. wrote the computer code with contributions from R.F. C.C. designed the figures.

Competing interests

The authors declare no competing interests.

Additional information

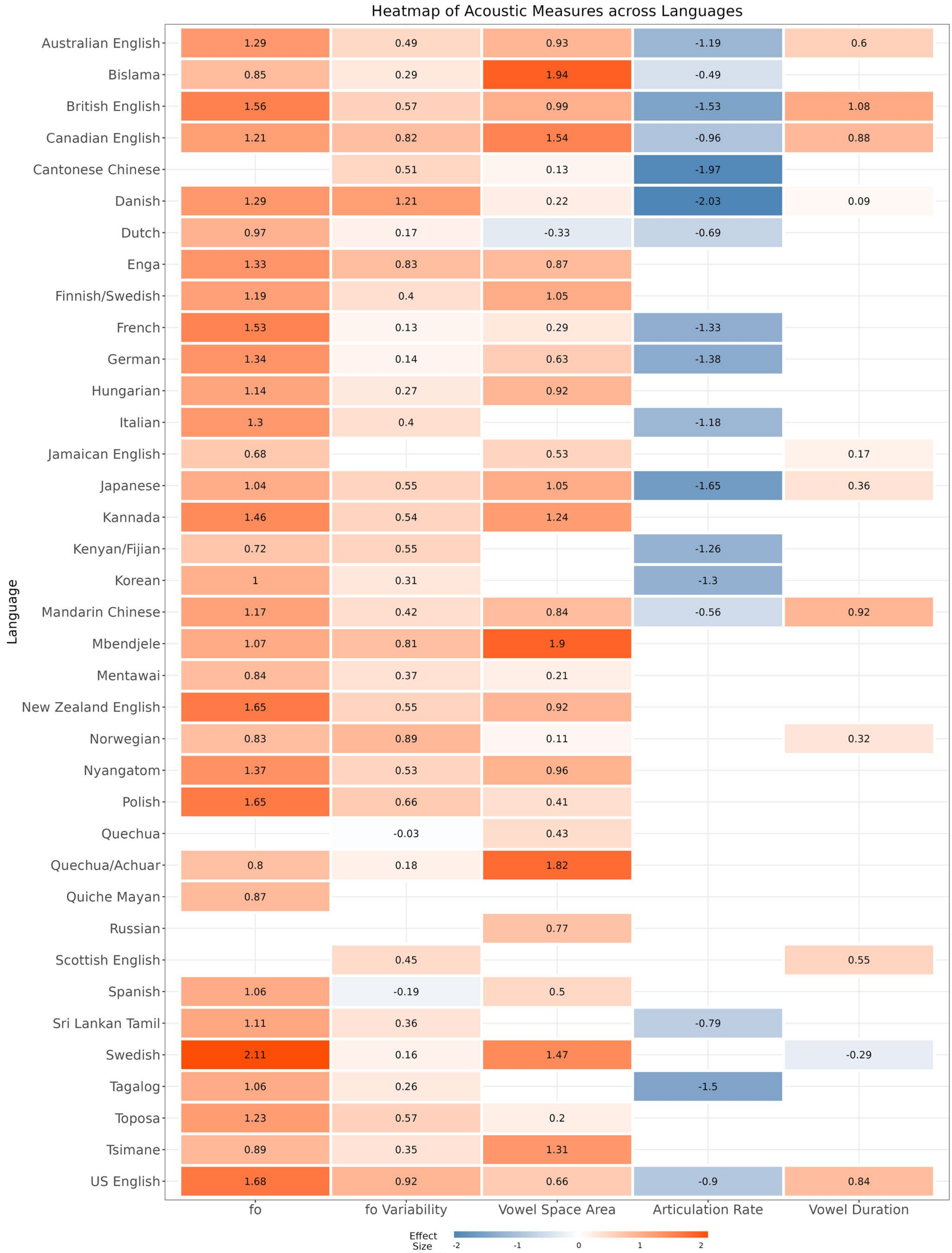
Extended data is available for this paper at https://doi.org/10.1038/s41562-022-01452-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-022-01452-1.

Correspondence and requests for materials should be addressed to Christopher Cox.

Peer review information *Nature Human Behaviour* thanks Christa Lam-Cassettari, Shravan Vasishth and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41562-022-01452-1



Extended Data Fig. 1 | Heatmap of Acoustic Measures across Languages. A heatmap providing an overview of the effect size estimates for each of the acoustic variables and languages. Dark orange shading indicates a strong effect size value on the positive scale. Dark blue shading indicates a strong effect on the negative scale.

nature portfolio

Corresponding author(s): Christopher Cox

Last updated by author(s): Aug 12, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	EndNote, WebPlotDigitizer
Data analysis	Data processing and analysis were carried out in R 4.2.0. All code can be found in the open repository: https://osf.io/hc7me/ R Packages used: ggmap_3.0.0; ggrepel_0.9.1; viridis_0.6.2; viridisLite_0.4.0;loo_2.5.1; mice_3.14.0; boot_1.3-28; brms_2.17.0; rstan_2.21.5; ggalluvial_0.12.3; RColorBrewer_1.1-3; cowplot_1.1.1; lattice_0.20-45; ggridges_0.5.3; PublicationBias_2.2.0; effectsize_0.7.0; esc_0.5.1; robumeta_2.0; metafor_3.4-0; metadat_1.2-0; Matrix_1.4-1; job_0.3.0; clickR_0.8.0; Counterfactual_1.2; glue_1.6.2; moments_0.14.1; here_1.0.1; readxl_1.4.0; forcats_0.5.1; stringr_1.4.0; dplyr_1.0.9; purrr_0.3.4; readr_2.1.2; tidyr_1.2.0; tibble_3.1.7; tidyverse_1.3.1; tidybayes_3.0.2 ggplot2_3.3.6; Rcpp_1.0.9; bibliometrix_3.0.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data were accessed on PubMed and Web of Science and are available and permanantly archived in the following open repository: https://osf.io/hc7me/.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Gender-based analyses of the extracted data were not performed in this study. The participants in this meta-analysis largely consisted of female caregivers residing in Western, educated, industrialized, rich, developed countries. Due to the sparsity of the data on additional dimensions that are likely to impact the acoustics of infant-directed speech (e.g., speaker types, kinship, gender, socio-economic status, fine-grained details of the interaction), the meta-analysis could not systematically analyze these factors as potential sources of variability. We encourage researchers who are interested in these questions to contribute to the openly available dataset and to integrate and update our selection of studies.
Population characteristics	Age of infants who are being addressed by caregivers.
Recruitment	As a meta-analysis, we did not ourselves recruit participants, but instead analyzed data from the included studies. Systematic searches and meta-analyses, however, cannot completely avoid bias, as discussed in S1.1 in the Supplementary Information. Here we discuss i) how our choice of search terms may select a biased subset of the literature, ii) how the published literature itself may represent a biased subset of the literature available, iii) how we counteracted bias in the study selection process, iv) how bias might arise as a function of the reporting of estimates in the included studies.
Ethics oversight	The manuscript relies on publicly available data (published articles) and has been deemed exempt from the need of ethical approval by the local ethical committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	A quantitative meta-analysis of studies on the acoustic properties of infant-directed speech.
Research sample	Research samples are based on the samples of each of the studies included in the meta-analysis. All data were accessed on PubMed and Web of Science and are available and permanantly archived at: https://osf.io/hc7me/ . The samples involve a broad range of languages, cultures and infant age ranges, as our rationale for this study was to synthesise all available evidence on the acoustics of infant-directed speech. It should be noted, however, that the majority of participants included in this meta-analysis consisted of female caregivers residing in Western, educated, industrialized, rich, developed countries. To the best of our knowledge, this sample of 88 studies is a representative sample of the literature on IDS.
Sampling strategy	In order to obtain a comprehensive sample of the available literature on the acoustic properties of IDS, we conducted a systematic literature search on PubMed and Web of Science, in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guidelines. We performed forward and backward (i.e., snowball) literature searches based on this initial search. No sample size calculation was performed, as we relied on publicly available data and studies that have already been published.
Data collection	We extracted data from studies that conformed to our inclusion criteria. We used these data to calculate Hedges' g effect sizes, either with standard formulae for Hedges' g (when the raw means and standard deviations were reported), d-values, one-sample t-values, or by digitally extracting the raw data from published plots using the WebPlotDigitizer application. In certain cases, the standard deviation of the effect size could not be calculated from the reported data or plots. In order to include these effect sizes in the meta-analysis, these missing standard deviation values were imputed by using multivariate imputation by chained equations

	based on a Bayesian linear regression model. The first systematic search was conducted independently by two of the authors (RF & EF) in June 2017 and updated by a third author (CC) in December 2021; CC screened for missed studies from before and after the date of the first systematic search. Disagreements in the screening of papers were resolved with discussions in the first phase between EF and RF and in the second phase between CC and RF; if the paper was thought to contain relevant data for the meta-analysis, the paper was included for the successive phase of the review. Disagreements were therefore rare and mainly motivated by studies where relevant information was reported only in the supplementary materials. None of the authors were blinded to the study hypothesis.
Timing	Systematic searches were performed in June 2017 and updated in December 2021.
Data exclusions	Each of the 602 papers were manually screened by three of the authors for inclusion according to the following pre-established inclusion criteria: i) infants had to be typically-developing, ii) studies had to include quantification of an acoustic feature, iii) studies had to include a comparison condition with adult-directed speech, iv) the speech had to be spoken to an infant by one or both of their caregivers. Based on the initial set of 602 papers, we used Connected Papers and Research Rabbit to find an additional 48 relevant studies. After excluding 54 duplicate studies, we screened the titles of 596 studies and excluded a further 302 studies that were unrelated to the current investigation. We read the abstracts of the remaining 294 studies and evaluated each with reference to the above exclusion criteria. Of the 294 papers, 174 studies had no relation to IDS, 17 studies had no comparison condition with ADS, and 15 studies examined atypical populations and had no relevant control sample of typically-developing infants to extract data on. To the best of our knowledge, the present review of a total of 88 studies represents a comprehensive sample of the literature on IDS.
Non-participation	No participants were involved in our study, as our meta-analysis aggregates data from already conducted studies.
Randomization	Randomization is not applicable to our study, as our meta-analysis aggregates data from already conducted studies. We therefore had no control over how participants were allocated to groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging