Psychological Bulletin

© 2025 American Psychological Association
ISSN: 0033-2909

2025, Vol. 151, No. 6, 669–695
https://doi.org/10.1037/bul0000479

Acoustic Exaggeration of Vowels in Infant-Directed Speech: A Multimethod Meta-Analytic Review

Irena Lovčević[1], Titia Benders[2], Sho Tsuji[1, 3], and Riccardo Fusaroli[4, 5]
[1] The World Premier International Research Center for Neurointelligence, Institutes for Advanced Studies, The University of Tokyo
[2] Amsterdam Center for Language and Communication, Faculty of Humanities, University of Amsterdam
[3] Département d'Études Cognitives, Laboratoire de Sciences Cognitives et de Psycholinguistique, École Normale Supérieure, École des Hautes Études en Sciences Sociales, Centre National de la Recherche Scientifique, Paris Sciences et Lettres University
[4] Department of Linguistics, Cognitive Science and Semiotics, Aarhus University
[5] Interacting Minds Centre, Aarhus University

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

There is a long-standing debate about the extent to which vowel hyperarticulation, the production of acoustically exaggerated vowels, occurs in infant-directed speech (IDS). This exaggeration has been argued to result in clearer speech sounds that are easier for infants to process and might be positively related to infants' linguistic outcomes. However, previous findings regarding the presence of vowel hyperarticulation in IDS are seemingly inconsistent and contradictory, making it hard to advance our understanding of the role of hyperarticulation in language development and, consequently, the potential functions/roles of IDS in language acquisition. Thus, we adopted a systematic review and meta-analytic approach to investigate the robustness of vowel hyperarticulation in IDS and identify sources of heterogeneity in the literature. We employed four complementary meta-analytic approaches and evaluated the robustness of results across the different choices. We performed both traditional (Frequentist) and Bayesian meta-analyses first on methodologically consistent studies (20 studies, 42 effect sizes) and then on all studies of vowel hyperarticulation in IDS irrespective of the method (35 studies, 80 effect sizes). Findings indicate the presence of vowel hyperarticulation in IDS compared to adult-directed speech (effect sizes ranging from 0.41 to 0.69), as well as systematic and unsystematic variability due to, for example, cross-linguistic variability and methods employed, making it difficult to identify specific factors associated with stronger vowel hyperarticulation. The quantitative results combined with a systematic review of the literature also enable important methodological insights, which we summarize into recommended practices such as enlarging sample sizes and explicitly incorporating sources of heterogeneity in analyses.

Public Significance Statement
This meta-analysis confirms previous findings that when talking to infants, adults modify features of their speech compared to talking to other adults. Specifically, our study demonstrates that mothers exaggerate vowels in their speech to infants in at least 10 different languages. Although language spoken and various methodological choices seem to matter, specific factors could not be pinpointed due to factors co-occurring in the data. These findings have implications on our understanding of the infants' social-communicative environment, having broader implications on theories of language development.

Keywords: vowel hyperarticulation, infant-directed speech, meta-analysis, Frequentist approach, Bayesian approach

Supplemental materials: https://doi.org/10.1037/bul0000479.supp

This article was published Online First June 2, 2025.
Lucy Bowes served as action editor.
Irena Lovčević https://orcid.org/0000-0002-4528-185X
All analyses reported in this article were preregistered via Open Science Framework prior to data analysis. These preregistrations can be accessed at https://osf.io/n29mr and https://osf.io/t4bzv. Furthermore, analysis scripts and data are available at https://osf.io/n29mr.

The authors have no conflicts of interest to disclose. This research was funded by The Ministry of Education, Culture, Sports, Science and Technology World Premier International Research Center Initiative startup grant to Irena Lovčević, The Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research Grants 20H05617 and 20H05919 to Sho Tsuji, and Seed funding from the Interacting Minds Center to Riccardo Fusaroli.

The authors thank Christina Dideriksen and Emma Fowler for their help with data collection. Also, the authors thank all authors of the included meta-analyses who contributed additional data and answered their questions.

Irena Lovčević played a lead role in data curation, investigation, writing–original draft, and writing–review and editing, a supporting role in visualization, and an equal role in conceptualization, formal analysis, and methodology.
continued

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Infants' social-communicative environment plays an important role in their language development, raising questions about which environmental factors scaffold language learning and how. A potential candidate is infant-directed speech (IDS), a speaking style employed in interaction with infants. In comparison to adult-directed speech (ADS), IDS is usually characterized by a number of enhanced visual (e.g., Chong et al., 2003), linguistic (e.g., Kavanaugh & Jirkovsky, 1982), and acoustic (e.g., D. Burnham et al., 2002) features. Here, we focus on one of these acoustic features, vowel hyperarticulation. Vowel hyperarticulation has been suggested to scaffold language acquisition by facilitating infants' speech processing (Lovcevic et al., 2024; Peter et al., 2016; Song et al., 2010; Zhang et al., 2011) and vocabulary development (Dilley et al., 2020; Hartman et al., 2017; Kalashnikova & Burnham, 2018; Lovcevic, Kalashnikova, & Burnham, 2020; U. Marklund et al., 2021). However, there is a lack of consistency across studies as well as the absence of vowel hyperarticulation in some languages, challenging the robustness of this IDS feature. Given these suggested functions of vowel hyperarticulation, it is important to know whether it is a robust phenomenon. Hence, we conducted a systematic review of the literature and meta-analytically assessed the robustness of vowel hyperarticulation in IDS and the factors, both theoretically and methodologically important, that might affect it. This will give insight into the sources of variability in hyperarticulation as well as diversity in current research practices.

Vowel Hyperarticulation

One of the key acoustic IDS features is phonetic exaggeration, most often measured as vowel space hyperarticulation. Vowel space hyperarticulation, also called vowel space expansion, refers to an increase in the space defined by the center of peripheric vowel categories (Smiljanić & Bradlow, 2009). Specifically, when the vowel category centers (usually the means of formant frequencies) are projected on a two-dimensional space defined by the first and second formant frequencies ($F1 \times F2$ with $F1$ inversely related to vowel and tongue height and $F2$ inversely related to vowel and tongue-body backness; Ladefoged, 2006), the resulting space is greater in IDS compared to ADS (see Figure 1 for an example of vowel triangles in IDS and ADS). Such vowel hyperarticulation thus reflects more dispersed vowel categories.

Vowel hyperarticulation is not a feature unique to IDS. It has been observed when speakers' goal is to increase their speech clarity (Bradlow et al., 2000; Ferguson & Kewley-Port, 2002, 2007; Krause & Braida, 2004; Moon & Lindblom, 1994; Picheny et al., 1986), for example, in noisy conditions (Lombard speech; Castellanos et al., 1996), and when speakers engage listeners with real or perceived communicative needs: adults with hearing loss (Ferguson & Kewley-Port, 2002, 2007; Hazan & Baker, 2011), nonnative listeners (foreigner-directed speech; Lorge & Katsos, 2019; Piazza et al., 2022; Uther et al., 2007), computers (computer-directed speech; D. Burnham et al., 2002), and pets that have speech-like vocalization ability (e.g., parrots, Xu et al., 2013). Hence, vowel hyperarticulation might result from speakers' adjustments to the listeners' needs.

From the listener's perspective, hyperarticulated vowels might augment speech intelligibility by providing easily perceptible acoustic distinctions between vowels (Bradlow et al., 1996). Two sets of evidence support this proposition. On the one hand, speakers perceived as more intelligible tend to produce more expanded vowel spaces compared to less intelligible speakers (Bond & Moore, 1994; Bradlow et al., 1996; Byrd, 1994; Hazan & Markham, 2004; Moon & Lindblom, 1994). Conversely, the vowel space expansion of clear speech has a positive impact on its speech intelligibility in noise (Ferguson & Kewley-Port, 2007; Moon & Lindblom, 1994; Smiljanić & Bradlow, 2005), although not all studies find this (Krause & Braida, 2004).

Vowel Hyperarticulation in IDS

The first empirical evidence on the presence of a greater vowel space in IDS compared to ADS comes from a U.S. English study by Bernstein Ratner (1984). This study demonstrated an increase in IDS vowel space expansion over language development stages (preverbal, verbal, 3–4 word utterances), suggesting modifications according to infants' language ability.

Kuhl et al. (1997) followed up on Bernstein Ratner (1984) but with greater methodological impact and suggested vowel hyperarticulation as a language-general feature of IDS by demonstrating its presence in IDS of U.S. English, Russian, and Swedish mothers. These three languages were chosen for their differences in vowel system: Russian has five vowels, U.S. English nine, and Swedish 16. Yet, the point vowels /a, i, u/ occur in each of them (and in the majority of world languages, Ladefoged & Maddieson, 1996). As the methodological choices made in this study have continued to influence the field, and consequently our meta-analytic endeavor, we report the methods in some detail here. Mothers were recorded while speaking to their 2- to 5-month-old infant (IDS) and while speaking to an adult experimenter (ADS). In order to obtain maternal production of the 3 point vowels, the mothers in U.S. English and Russian samples were provided with toys whose names contained each of the target vowels (one toy per vowel for U.S. English, two toys per vowel for Russian). In the Swedish sample, all content words containing the target vowels were analyzed. To calculate the vowel space, the averaged $F1$ and $F2$ values of the three corner vowels were used, projected as Cartesian coordinates on the x–y plane. The vowel space was calculated using the following formula (Equation 1, Kuhl et al., 1997):

Titia Benders played a supporting role in investigation, writing–original draft, and writing–review and editing and an equal role in conceptualization. Sho Tsuji played a supporting role in investigation, supervision, and writing–review and editing and an equal role in conceptualization. Riccardo Fusaroli played a lead role in supervision and visualization, a supporting role in data curation, writing–original draft, and writing–review and editing, and an equal role in conceptualization, formal analysis, and methodology.

Correspondence concerning this article should be addressed to Irena Lovčević, The World Premier International Research Center for Neurointelligence, Institutes for Advanced Studies, The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan. Email: irena.lovcevic@ircn.jp

$$\text{Area} = \frac{|(F1/\text{a}/ \times (F2/\text{i}/ - F2/\text{u}/) + F1/\text{i}/ \times (F2/\text{u}/ - F2/\text{a}/) + F1/\text{u}/ \times (F2/\text{a}/ - F2/\text{i}/)|}{2}. \quad (1)$$

(see Equation 1 above)

Here, $F1$/a/ refers to the average value of the first formant frequency for vowel /a/, whereas $F2$/i/ refers to the average value of the second formant frequency for vowel /i/ and so forth.

A larger vowel triangle area in IDS compared to ADS, as calculated using this formula, would be evidence for vowel hyperarticulation in IDS. Indeed, the results of this study demonstrated a much larger vowel space in IDS compared to ADS (Hedges' $g = 1.61$) across all three languages.

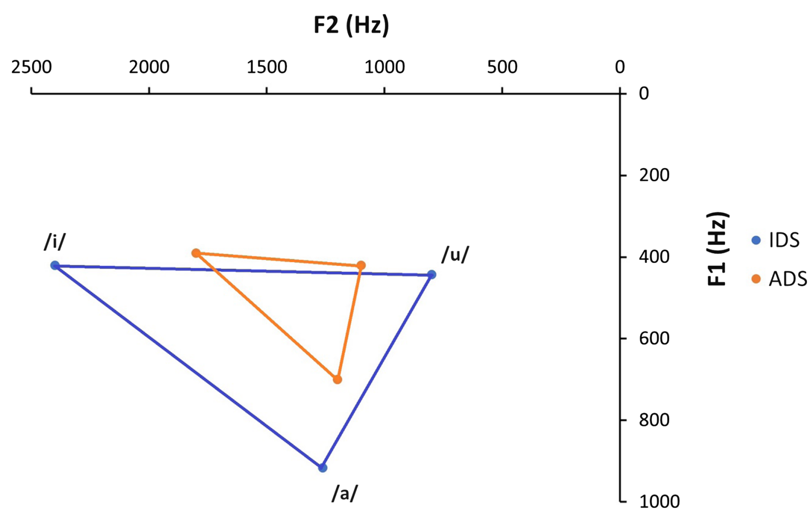
Numerous follow-up studies assessed vowel hyperarticulation in IDS using the same protocol as Kuhl et al. (1997): IDS recorded during mother–infant play sessions, typically developing infants, spontaneous speech, three "corner" vowels (a mid-low vowel, e.g., /a/; a front-high vowel, e.g., /i/; and a back-high vowel, e.g., /u/), and the same formula for computing vowel spaces. A majority of these studies demonstrated the presence of vowel hyperarticulation when studying languages such as U.S. English (Cristia & Seidl, 2014), British English (Uther et al., 2007), Australian English (D. Burnham et al., 2002; Kalashnikova & Burnham, 2018; Kalashnikova et al., 2017, 2018; Xu et al., 2013), Mandarin (Liu et al., 2003; Tang et al., 2017), and Japanese (Miyazawa et al., 2017). On the contrary, a few studies demonstrated a smaller vowel space in IDS compared to ADS; thus, vowel *hypo*articulation (e.g., Cantonese, Rattanasone et al., 2013; Norwegian, K. Englund & Behne, 2006) or no difference in vowel space areas between IDS and ADS (e.g., Japanese, Shochi et al., 2009; Australian English, Lovcevic, Kalashnikova, & Burnham, 2020). As all the study methods were comparable, these varied findings suggest that vowel hyperarticulation in IDS might not be a universal phenomenon. Further studies employing different methods than Kuhl et al. (1997) (e.g., inclusion of more than three vowels, inclusion of vowels other than corner vowels, IDS recorded during reading sessions, Benders, 2013; Cox, Dideriksen, et al., 2023; K. Englund & Behne, 2006) also provide conflicting findings regarding the presence of vowel hyperarticulation in IDS.

Potential Moderators of Vowel Hyperarticulation in IDS

To better understand the conflicting findings, the current meta-analysis assessed the effects of several moderator variables on vowel hyperarticulation in IDS. We had three groups of moderators: *infant-related moderators* (native language, age, risk status), *design-related methodological moderators* (methodological decisions related to experimental design in a broader sense: measurement scale, recording context, recording order of IDS and ADS, type of elicitation stimulus, interlocutor, method of formant measurement, recording duration), and *vowel-related methodological moderators* (methodological decisions related only to the analyzed vowels: vowel type, number of vowels, number of word types, stress-carrying vowels). Table 1 presents a detailed list of moderators included in this meta-analysis. In the following paragraphs, we will review these moderators in more detail. We identified a majority of these moderators a priori based on expert knowledge, and some additional moderators were

Figure 1

Hypothetical Vowel Space Triangles for /a/, /i/, and /u/ in ADS (Orange) and IDS (Blue)



Note. The x-axis represents the average value of each vowel on the first formant ($F1$), and the y-axis represents the same on the second formant ($F2$). The depicted difference between ADS and IDS is due to $F1$ raising for /a/, $F2$ raising for /i/, and $F2$ lowering for /u/. There might be other patterns leading to a greater vowel space triangle in IDS than ADS (e.g., vowel-specific raising of $F1$ and $F2$, Dodane & Al-Tamimi, 2007). ADS = adult-directed speech; IDS = infant-directed speech. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 1

Moderators Included in Meta-Analysis 1 (Narrow-Search, Only Including Articles Whose Methods Follow Kuhl et al., 1997) and Meta-Analysis 2 (Broad-Search)

Moderator	Variable type	Meta-analysis
Infant-related factors		
Infants' age	Continuous	1 and 2
Native language	Categorical (each language separate factor)	1 and 2
Risk status	Categorical (two categories: at risk, not at risk)	2
Design-related methodological factors		
Measurement scale	Categorical (two categories: Hertz and logarithmic)	1 and 2
Recording order (indicates whether IDS or ADS was recorded first)	Categorical (six categories: IDS first, ADS first, not fixed-counterbalanced, not fixed-randomized, not fixed-ordering principle unspecified, ordering principle unspecified)	2
Elicitation stimulus in IDS (type of stimulus used to elicit target words)	Categorical (eight categories: none, object without written label, object with written label, picture without written label, object and picture without written label, written verbiage, memorized verbiage, unspecified)	2
Elicitation stimulus in ADS	Categorial (nine categories: none, object without written label, object with written label, picture without written label, object and picture without written label, written verbiage, memorized verbiage, not specified whether IDS stimuli were retained in ADS, unspecified)	2
Recording context	Categorical (four categories: unconstrained, reading, singing, repetition of sentence/phrase/text provided by experimenter)	2
Interlocutor in IDS	Categorical (four categories: own child, other child, pictured child, absent/imaginary child)	2
Interlocutor in ADS	Categorical (four categories: adult experimenter, unspecified adult, absent interlocutor, unspecified)	2
Method of formant measurements	Categorical (eight categories: automatic only, automatic—outliers excluded, automatic—outliers human checked, automatic—human reliability, automatic—human verification, automatic—human verification and reliability, human—primary coder, unspecified)	2
Session duration	Categorical (five categories: 0–4 min, 5–10 min, 11–30 min, more than 30 min, unspecified)	2
Vowel-related methodological factors		
Vowel type	Categorical (two categories: only corner vowels, corner vowels and other vowel[s])	2
Vowel number	Numeric (number of vowels used for vowel area calculations)	2
Target-word types (how many word types were analyzed for each target vowel)	Categorical (six categories: one target, several targets, all content words, all words analyzed in focus utterance, all words, unspecified)	2
Stress (whether the analyzed vowels carry word stress)	Categorical (three categories: stressed, both stressed and unstressed, unspecified)	2

Note. Meta-analysis 1 is narrow-search meta-analysis that included studies following Kuhl et al. (1997) methodology. Meta-analysis 2 is broad-search meta-analysis that included all studies on vowel hyperarticulation in IDS regardless of the methods they used. IDS = infant-directed speech; ADS = adult-directed speech.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

added during the literature review. Moderators that were identified during the literature review were recording order of IDS and ADS, type of elicitation stimulus, interlocutor in IDS, interlocutor in ADS, session duration, and target-word types.

Infant-Related Moderators

Native Language. The demonstrated presence of vowel hyperarticulation in certain languages/dialects (e.g., U.S. English, Australian English) but not in others (e.g., Cantonese, Norwegian) might suggest cross-linguistic differences. Such differences could stem from structural linguistic factors (such as the size of the vowel space or the presence of tones) as well as cultural factors (e.g., acceptability of using "baby language" or parenting practices), but such variability is complex to measure and due to data sparsity beyond the scope of this meta-analysis. For the present study, we thus simply test for the presence of cross-linguistic differences in the degree of hyperarticulation in IDS. Given the findings on cross-linguistic differences in several IDS features (Han et al., 2018; Kitamura et al., 2001; Narayan & McDermott, 2016; Bernstein Ratner & Pye, 1984), we hypothesize cross-linguistic differences in the presence and degree of vowel hyperarticulation in IDS. Although cross-linguistic differences might be confounded with certain laboratory factors, our inclusion of design-related methodological moderators aims to account for cross-laboratory differences in the execution of the study.

Age. Although age-related modifications have been observed for several other IDS features (e.g., pitch height and range; Kitamura & Burnham, 2003), there is no clear evidence for age-related modifications in the degree or presence of vowel hyperarticulation in IDS. Limited longitudinal evidence is conflicting, with reports of age-related modifications on the one hand (Bernstein Ratner, 1984) and no evidence for (or against) modifications on the other (Kalashnikova & Burnham, 2018; Kalashnikova et al., 2017). Similarly, cross-sectional evidence provides mixed findings regarding the presence of vowel hyperarticulation at different ages. For example, vowel hyperarticulation has been found in IDS in infants between two and 12 months in some studies (e.g., D. Burnham et al., 2002), but not in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

others (e.g., Rattanasone et al., 2013). It has also been found in IDS to 18- to 24-month-olds (e.g., Miyazawa et al., 2017), but not in IDS to 15-month-olds (e.g., Lovcevic, Kalashnikova, & Burnham, 2020). Both longitudinal and cross-sectional studies also differ in native language (e.g., American English in Bernstein Ratner, 1984; Australian English in Kalashnikova & Burnham, 2018, Japanese in Miyazawa et al., 2017), making it hard to disentangle the effect of age from the effect of the native language. Hence, the current meta-analysis will assess the effect of infants' age on vowel hyperarticulation in IDS to infants from birth up to 24 months of age (age range encompassing the majority of IDS studies). Due to demonstrated nonlinear changes in other IDS features across development (Kitamura & Burnham, 2003; Kitamura et al., 2001; Stern et al., 1983), we will explore the linear and nonlinear effects of infants' age on vowel hyperarticulation. Age-related changes in vowel hyperarticulation are also relevant to the question of whether it reflects infants' perceived needs and thus varies between stages of language acquisition.

Risk Status. A relatively recent set of studies otherwise following Kuhl et al.'s (1997) method except in focusing on typically developing infants was conducted with different populations, namely infants with hearing loss or at risk for a developmental disorder (i.e., dyslexia). These also demonstrated diverse findings from the presence of vowel *hyper*articulation in IDS (infants with hearing loss, Wieland et al., 2015), vowel *hypo*articulation (also infants with hearing loss, Lam & Kitamura, 2010), and no detected difference between IDS and ADS (infants at risk for dyslexia, Kalashnikova et al., 2018; infants with hearing loss, Lovcevic, Kalashnikova, & Burnham, 2020). These divergent findings were reported both across dialects (American and Australian English) and within the same language dialect (e.g., Australian English: Lam & Kitamura, 2010; Lovcevic, Kalashnikova, & Burnham, 2020). Although factors inherent to these disorders may lead to heterogeneity in findings, the factors responsible for divergent findings in typically developing infants could also lead to spurious group differences. Regardless, if vowel hyperarticulation in maternal speech is associated with infants' risk status, we expect a smaller degree of vowel hyperarticulation in IDS to at-risk infants compared to typically developing infants (Kalashnikova et al., 2018, 2020; Kondaurova et al., 2012; Lam & Kitamura, 2010). On the other hand, if infants' risk status is not related to vowel hyperarticulation in IDS, we expect no difference in the degree of vowel hyperarticulation between the two groups of infants (Lovcevic, Kalashnikova, & Burnham, 2020; Wieland et al., 2015). It should be noted that as many risk factors have a significant heritable component, it should be hard to determine if any effect of risk status is due to the caregiver's characteristics or the infant's characteristics. It is not the focus of this study to ascertain whether the presence of vowel hyperarticulation in IDS to at-risk children is due to caregiver's or infant's characteristics but to examine the association between at-risk-ness and degree of vowel hyperarticulation. Furthermore, although the caregiver's diagnostic status might also affect IDS production (e.g., postnatal depression, Brookman et al., 2020; Kaplan et al., 2015), this is beyond the scope of the present study.

Design-Related Methodological Moderators

Measurement Scale. IDS studies differ in whether they used linear (e.g., Hertz) or logarithmic scales (e.g., Bark, Mel) for vowel space assessment. Hertz is a unit of frequency established by the International Standard of Units (SI) and defined as one cycle per second, thus representing a physical measure of frequency. Psychoacoustic scales, in contrast, are logarithmic in nature, and they better capture human perception of acoustic differences and human perception's higher acuity for lower than for higher frequencies (Volkmann et al., 1937; Zwicker, 1961). Because much of the discussion of hyperarticulation in IDS vowels revolves around perceptual clarity of the input, expressing hyperarticulation in perceptually based scales appears to be a reasonable choice.

Although Kuhl et al. (1997) found comparable results with both Hertz and Mel scales, the effect sizes for hyperarticulation may be impacted by the scale on which formant frequencies are measured in conjunction with the exact changes in the $F1$–$F2$ plane leading to hyperarticulation. Compared to the linear Hertz scale, logarithmic scales emphasize changes in lower frequencies and deemphasize changes in high frequencies. Thus, hyperarticulation that is (primarily) due to the lowering of inherently low formants will appear to be larger on a logarithmic scale (e.g., the low $F1$ of /i/ and /u/ or the low $F2$ of /u/ lowering further in IDS). Conversely, hyperarticulation that is (primarily) due to the raising of inherently high formants will appear larger on the Hertz scale (e.g., the high $F1$ of /a/ or the high $F2$ of /i/ raising further in IDS). Both these patterns of shift have been observed in the literature but have not yet been systematically connected to vowel hyperarticulation outcomes. Here, we expect that hyperarticulation will be associated with larger effects in studies employing the Hertz scale rather than a logarithmic scale, as formant raising might be more prevalent in IDS (Benders, 2013; K. Englund & Behne, 2006).

Method of Formant Measurement. The majority of hyperarticulation studies employed some form of automatic formant measurement (Nearey et al., 2002). Such automatic measurements are not error-free, as testified by algorithm evaluations (e.g., Nearey et al., 2002) and continued efforts to improve formant tracking (Dissen et al., 2019). Potential measurement errors caused by automatic measurement can be reduced by manually checking formant measures or by eliminating formant values that are statistical outliers. If measurement errors are not all caught, the remaining noise in the data may lead to unreliably estimated effects. Hence, in the present study, we will investigate whether the size of the vowel hyperarticulation effect is associated with the rigor with which measurement errors were prevented. Specifically, we will examine if different levels of automatic versus manual formant measurements affect the observed size of the vowel hyperarticulation effect.

Recording Context. Besides the mother–infant play session, some studies employed additional recording contexts (reading/singing/repetition of a sentence, phrase, or text that the experimenter has provided). Such differences in recording context can result in segmental and suprasegmental speech differences (Cox, Dideriksen, et al., 2023; Dellwo et al., 2015; de Ruiter, 2015; Ernestus et al., 2015). Vowel hyperarticulation in IDS has been found during the repetition of a sentence, phrase, or text provided by the experimenter (Andruski et al., 1999; Gergely et al., 2017), but not during singing (Audibert & Falk, 2018). Findings regarding reading are conflicting, with observed vowel hyperarticulation in some studies (Weirich & Simpson, 2019) and no attested vowel hyperarticulation in other studies (E. B. Burnham et al., 2015; Dodane & Al-Tamimi, 2007). Although these studies concern different native languages (American/British English, German, French, and Japanese), they did not differ in the infant's age range and the measurement scale used (psychoacoustic scales). Based on observations from other acoustic aspects of IDS, we expect a lower degree of vowel hyperarticulation when the recording context is constrained (reading, repetition) compared to an unconstrained context (Cox, Bergmann, et al., 2023).

Order of IDS and ADS Recordings. Previous evidence suggests that the acoustic reduction associated with the repeated production of a word might continue when interlocutors are changed (Bard et al., 2000, 2004). Hence, if a caregiver speaks about a small set of toys to two interlocutors, their speech might be clearer in the first compared to the second conversation, all else being equal. Therefore, vowel hyperarticulation effect sizes may be larger in studies in which caregivers always speak first to their infant than in studies that started with ADS, with effect sizes for counterbalanced order of speech register being somewhere in the middle.

Interlocutor in IDS. Besides assessing vowels in IDS to real infants, some studies asked parents to imagine that they are speaking to an infant. The evidence suggests that vowels in speech to imaginary infants are less clearly articulated than IDS to real infants (Knoll & Scharrer, 2007). Thus, we expect more pronounced vowel hyperarticulation for studies in which a caregiver spoke to her own child compared to a pictured or imagined child.

Additional design-related methodological moderators that we assessed are *type of elicitation stimulus*, *interlocutor in ADS*, and *session duration*. Although no extensive literature on the impact of these potential moderators exists, these factors introduce variability in assessment methods and could therefore affect study results.

Vowel-Related Methodological Moderators

Number of Vowels and Type of Vowels (Corner vs. Noncorner). Some studies diverging from Kuhl et al.'s (1997) method included a fourth and fifth vowel in addition to the canonical three vowels. In the majority of studies (if not all), these additional vowels are noncorner vowels. Limited evidence suggests the absence of vowel hyperarticulation for noncorner vowels even in data sets that display hyperarticulation of the corner vowels (e.g., McMurray et al., 2013). Findings of studies that included a fourth and fifth vowel are conflicting despite some of them being conducted in the same native language and within a similar age range, with evidence either for vowel *hyper*articulation in IDS (Weirich & Simpson, 2019), vowel *hypo*articulation in IDS (Audibert & Falk, 2018; Benders, 2013; Cox, Dideriksen, et al., 2023; K. T. Englund, 2018), or no detected difference between registers (Bohn, 2013; Dodane & Al-Tamimi, 2007). This raises the question of whether hyperarticulation, if attested, is a feature of the corner vowels rather than a more extensive vowel inventory. In the current meta-analysis, we expect more pronounced vowel hyperarticulation in studies that analyzed exactly the three standard corner vowels compared to more vowels (McMurray et al., 2013).

Stress-Carrying Vowels. Vowels in stressed syllables tend to be produced with more peripheral vowel qualities than vowels in unstressed syllables. This is observed in both languages with strong vowel reduction in unstressed syllables (e.g., English) and in languages with no unstressed vowel reduction (e.g., Polish; Rojczyk, 2019). IDS is often characterized by a slower speaking rate and longer vowels, which might be particularly true for vowels in stressed syllables and content words, as the duration of unstressed syllables and function words may not differ between IDS and ADS (Bernstein Ratner, 1984; Swanson et al., 1992). Thus, vowel hyperarticulation might be more pronounced in studies that assessed only stress-carrying vowels.

Type of Target Words. As content words are more often stressed than function words (Berry, 1953), they are also produced with more peripheral vowel qualities. For the same reasons outlined above in the context of stress-carrying vowels, hyperarticulation in IDS might be more evident in studies that considered only vowels in content words compared to studies that included all vowels regardless of word type.

Number of Target Words. Speakers produce words more clearly in the presence of similar-sounding words to avoid (potential) confusion. For example, speakers produce /p/ in *pat* more clearly if they want to disambiguate it from *bat* (e.g., Baese-Berk & Goldrick, 2009). Whereas only very few IDS vowel studies have elicited minimal-pair words (e.g., *pea, bee*; Burkinshaw, 2020), it is typical to elicit a small set of words that are phonologically similar (e.g., *bead*, *po*t, *boo*t, Kuhl et al., 1997; *sheep*, *shoe*, *shark*, D. Burnham et al., 2002). Wang et al. (2018) hypothesized that having few toys or objects might result in increased parental awareness of these words being the study focus, leading to emphasis on their differences, especially for infant listeners who may need that. Hence, vowel hyperarticulation in IDS might be more pronounced in studies using a smaller number of target words.

The Present Study

The current literature seems to present contradictory and unsystematic findings regarding the presence of vowel hyperarticulation in IDS. We therefore set out to systematically review and meta-analyze the literature to identify underlying sources of heterogeneity (moderators), considering only methodologically homogeneous studies (following Kuhl et al., 1997, Meta-analysis 1 [narrow-search]) as well as a mix of these homogeneous and more heterogeneous ones (Meta-analysis 2 [broad-search]). Beyond estimating the current evidence for vowel hyperarticulation in IDS, we also aim to identify good research practices and issues and to provide guidelines for future studies.

Meta-Analysis 1 (Narrow-Search Meta-Analysis)

In the Meta-analysis 1 (narrow-search), we only included studies following Kuhl et al.'s (1997) method (see the Meta-Analysis 1 Inclusion Criteria section). Regarding the moderators, given the restricted set of studies, we could only meaningfully assess the effect of two infant-related (infants' age and native language) and one design-related methodological moderator (measurement scale).

Meta-Analysis 2 (Broad-Search Meta-Analysis)

In the Meta-analysis 2 (broad-search), we extend our inclusion criteria to methodologically diverse studies on vowel hyperarticulation in IDS. Accordingly, in addition to research questions from the Meta-analysis 1 (narrow-search), we will assess the effects of infants' risk status, all design-related methodological factors as well as the vowel-related methodological factors.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Method

Data Collection

Transparency and Openness

The analyses reported in this study are all preregistered on the Open Science Framework (OSF; Lovcevic, Benders, et al., 2020, https://osf.io/n29mr). The preregistration was performed during data collection, but before data inspection and analysis. The original preregistration was for the Meta-analysis 1 (narrow-search) and was successively amended to include the Meta-analysis 2 (broad-search; additional 16 studies; Lovcevic et al., 2021; preregistration link: https://osf.io/t4bzv). These amendments were made prior to data analysis and were motivated by literature review and assessment of whether the studies available warranted the analysis of specific design-related and vowel-related methodological moderators. The OSF project also contains analysis scripts and data. We followed the review protocol that was based on the framework of the MetaLab platform (Gasparini et al., 2022; https://langcog.github.io/metalab/).

Systematic Literature Search

The search for relevant studies was conducted following the Preferred Reporting Items for Systematic reviews and Meta-Analyses protocol (Page et al., 2021) and included the search for relevant reports, abstract screening, and full-text retrieval and screening (see Preferred Reporting Items for Systematic reviews and Meta-Analyses checklist in the Supplemental Material S1). All document types (journal articles, theses, conference proceedings, and unpublished data) were included in Meta-analysis 1 (narrow-search) and/or Meta-analysis 2 (broad-search) if they fulfilled the inclusion criteria below.

Meta-Analysis 1 Inclusion Criteria

- *Participants:* Full-term typically developing infants from birth to 24 months of age and their mothers with no diagnostic status (e.g., clinical depression).
- *Method:* Studies with both IDS and ADS conditions; IDS recorded during mother–infant play session; ADS recorded during semistructured interviews with an adult experimenter; vowel space measured between the "corner" mid-low vowel (e.g., /a/), front-high vowel (e.g., /i/), and back-high vowel (e.g., /u/)[1]; vowel hyperarticulation calculated using the measure used by Kuhl et al. (1997; see Equation 1 in Introduction).

Meta-Analysis 2 Inclusion Criteria

- *Participants*: Full-term typically and nontypically (preterm, at risk for dyslexia, infants with hearing loss) developing infants from birth to 24 months of age and their mothers with no diagnostic status (e.g., clinical depression).
- *Method*: Studies with both IDS and ADS conditions; IDS recorded during mother–infant play session eliciting spontaneous speech, but also during reading or singing sessions or sessions where speakers were instructed to say predefined sentences to interlocutors; any vowels assessed; vowel hyperarticulation calculated either using the measure used by Kuhl et al. (1997) or any other measure (e.g., Euclidean distances or Quadrilateral measure).

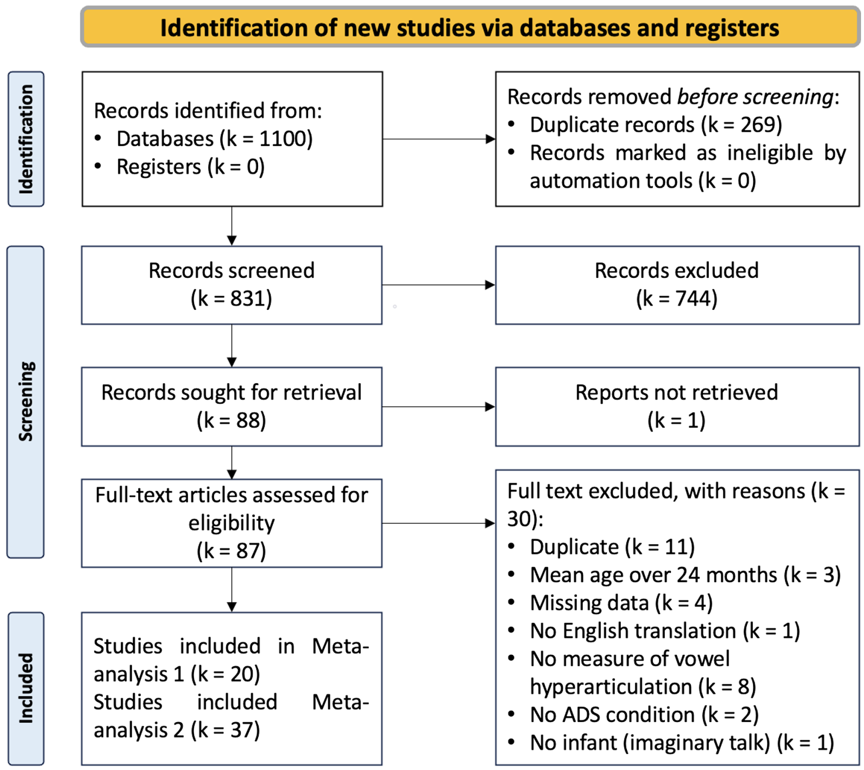
The literature search according to these criteria was conducted between August and December 2020. First, a systematic database search was conducted in August 2020 using Google Scholar with the keywords *vowel hyperarticulation in infant-directed speech*. The search was complemented by a seed search of reports citing the seminal study on vowel hyperarticulation in IDS (Kuhl et al., 1997) as listed in Google Scholar. We retrieved a total of 1,080 reports, 19 of which were deemed eligible for the Meta-analysis 1 (narrow-search), and 16 of which were deemed eligible for the Meta-analysis 2 (broad-search; see Figure 2 for the flow of study reports and Table 2 for the overview of included studies). Three authors (I.L., T.B., and R.F.) independently conducted the first search for relevant reports and considered a report as eligible based on full-text screening. The first author assessed the agreement between the chosen reports. Disagreements, which were almost nonexistent, were resolved via discussions between all authors. If the full text could not be retrieved or data necessary for the meta-analysis were missing from any of these eligible reports, the study authors were contacted. Two reports were excluded because we were not able to retrieve missing data. No report fulfilling the inclusion criteria included bilingual or multilingual infants. An updated search was conducted in February 2024 using Google Scholar using the keyword phrases *vowel hyperarticulation in infant-directed speech*, *acoustic features of infant-directed speech*, *vowel hyperarticulation in child-directed speec*h, and *acoustic features of child-directed speech*. This yielded 20 additional reports, but after full-text screening based on our criteria, only four of those reports were included in the meta-analysis: two were included in Meta-analysis 1 (E. Marklund & Gustavsson, 2020; Panneton et al., 2024), and two were included in Meta-analysis 2 (Rosslund et al., 2022, 2024). Seventeen reports were excluded for various reasons, including focusing on kindergarten teachers rather than mothers (1), including both parents without separate measurements (1), data overlap with another included study (1), reporting only IDS (2), lacking a measure of vowel space area (7), being review reports (2), and including children older than 24 months (2).

Effect Size

For the estimates of effect size, we calculated Hedges' g, which is a variant of *Cohen's d* effect size adjusted for small samples (Cumming, 2013; Hedges, 1981; Kline, 2004). Hedges' g was calculated as the mean vowel space size in IDS minus the mean vowel space size in ADS divided by the pooled standard deviation. Thus, positive values would indicate a larger vowel space in IDS compared to ADS (i.e., *hyper*articulation), whereas negative values would indicate a smaller vowel space in IDS compared to ADS (i.e., *hypo*articulation). Supplemental Material S1 details the effect size calculations, including dealing with missing data. Figure 3 depicts effect sizes for individual experiments.

[1] Note that /u/ is not always back-high, as it is centralized in, for example, several dialects of English.

Figure 2
Flow of Study Reports Into the Research Synthesis



Note. ADS = adult-directed speech. See the online article for the color version of this figure.

Coding of Moderator Variables

Meta-Analysis 1. Here, we included three moderator variables: infants' age, native language, and measurement scale. Infants' age was coded in days, and in cases where it was reported in months, it was converted to days by multiplying the number of months by 30.42. The effect of age on vowel hyperarticulation was assessed in the Frequentist analysis as a polynomial function (linear, quadratic, and cubic effects in three separate models) and in the Bayesian analysis as a monotonic function (i.e., assuming that increases or decreases in effects over time might change nonlinearly, slowing down or accelerating, but never changing direction). Age was not centered because that would make monotonic growth models unfeasible (as they consider age an ordinal variable, starting from "first") and would make comparison across models more complicated. The studies reviewed involved 10 different native languages belonging to four language families (Indo-European, Sino-Tibetan, Japonic/Japanese-Ryukyuan, Nilo-Saharan). Each language was entered as a separate factor level into the analysis. In the case of different dialects of a language, country-wide dialects were each treated as a separate language (U.S./Australian/British English). For one of the studies, data were only available for the data set collapsed over languages (Swedish, Russian, and U.S. English participants) and not for separate language groups; hence, we coded the native language for this study as mixed. Regarding the measurement scale, we grouped the logarithmic scales (Mel, Bark) and contrasted them with the Hertz scale. The moderator variables native language and measurement scale were sum-coded; that is, the mean of a dependent variable for a given level was compared to the mean of the dependent variable across levels.

Meta-Analysis 2. The Meta-analysis 2 included additional moderators (see Table 1 for details). Design-related methodological moderators included recording order, elicitation stimulus in IDS, elicitation stimulus in ADS, recording context, interlocutor in IDS, interlocutor in ADS, method of formant measurements, and session duration. Vowel-related methodological moderators included vowel type, vowel number, target-word types, and stress. All categorical variables were sum-coded.

Analysis Plan

In Meta-analyses 1 and 2, both Frequentist and Bayesian methods were implemented. The use of Frequentist methods is motivated by common practices, thus allowing continuity/interpretability with classical approaches. Bayesian methods were chosen due to their increased modeling flexibility, enabling a more precise implementation of the theoretical expectations (e.g., decrease of hyperarticulation with age, but not necessarily in a linear fashion). All computations were performed in R 4.4.1 (R Core Team, 2020) using *metafor* (Viechtbauer, 2010), *lme4* (Bates et al., 2015), *emmeans* (Lenth, 2016), *brms* (Bürkner, 2017), and *Stan* (Carpenter et al., 2017) packages.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 2
Overview of Studies Included in Narrow-Search and Broad-Search Meta-Analysis

Study	Native language	Country	At risk status	Infant age (months)	Number of participants
Meta-analyses 1 and 2					
D. Burnham et al. (2002)	Australian English	Australia	Typical	6.00	12
Cristia and Seidl (2014, Experiment 1)	U.S. English	United States	Typical	4.15	28
Cristia and Seidl (2014, Experiment 2)	U.S. English	United States	Typical	11.30	18
Donnellan et al. (2020, Experiment 1)	British English	United Kingdom	Typical	5.95	24
Donnellan et al. (2020, Experiment 2)	Alur (Uganda)	Uganda	Typical	3.58	8
Donnellan et al. (2020, Experiment 3)	Kiswahili (Uganda)	Uganda	Typical	5.88	6
Donnellan et al. (2020, Experiment 4)	Lugbara (Uganda)	Uganda	Typical	3.58	3
Hartman et al. (2017, Experiment 1)	U.S. English	United States	Typical	11.00	15
Hartman et al. (2017, Experiment 2)	U.S. English	United States	Typical	18.00	15
Hartman et al. (2017, Experiment 3)	U.S. English	United States	Typical	24.00	15
Kalashnikova and Burnham (2018, Experiment 1)	Australian English	Australia	Typical	7.29	18
Kalashnikova and Burnham (2018, Experiment 2)	Australian English	Australia	Typical	9.15	18
Kalashnikova and Burnham (2018, Experiment 3)	Australian English	Australia	Typical	11.11	18
Kalashnikova and Burnham (2018, Experiment 4)	Australian English	Australia	Typical	15.22	18
Kalashnikova and Burnham (2018, Experiment 5)	Australian English	Australia	Typical	19.27	18
Kalashnikova et al. (2017)	Australian English	Australia	Typical	11.20	8
Kalashnikova et al. (2018)	Australian English	Australia	Typical	10.26	18
Kuhl et al. (1997)	mixed	Sweden/Russia/United States	Typical	3.32	30
Liu et al. (2003, Experiment 1)	Mandarin Chinese	Taiwan	Typical	7.30	16
Liu et al. (2003, Experiment 2)	Mandarin Chinese	Taiwan	Typical	11.20	16
Liu et al. (2009)	Mandarin Chinese	Taiwan	Typical	9.70	17
Lovcevic, Kalashnikova, and Burnham (2020, Experiment 1)	Australian English	Australia	Typical	15.38	15
Lovcevic, Kalashnikova, and Burnham (2020, Experiment 2)	Australian English	Australia	Typical	11.68	15
Lovcevic, Kalashnikova, and Burnham (2020, Experiment 3)	Australian English	Australia	Typical	8.09	11
Lovcevic, Kalashnikova, and Burnham (2020, Experiment 4)	Australian English	Australia	Typical	14.68	9
E. Marklund and Gustavsson (2020)	Swedish	Sweden	Typical	12.00	12
Miyazawa et al. (2017)	Japanese	Japan	Typical	20.40	20
Panneton et al. (2024)	U.S. English	United States	Typical	6.39	10
Rattanasone et al. (2013, Experiment 1)	Cantonese	Not reported	Typical	3.02	11
Rattanasone et al. (2013, Experiment 2)	Cantonese	Not reported	Typical	6.08	11
Rattanasone et al. (2013, Experiment 3)	Cantonese	Not reported	Typical	9.30	11
Rattanasone et al. (2013, Experiment 4)	Cantonese	Not reported	Typical	6.08	11
Rattanasone et al. (2013, Experiment 5)	Cantonese	Not reported	Typical	9.30	11
Rattanasone et al. (2013, Experiment 6)	Cantonese	Not reported	Typical	11.79	11
Starling (2019)	U.S. English	United States	Typical	20.25	4
Tang et al. (2017)	Mandarin Chinese	Australia	Typical	12.00	15
Uther et al. (2007)	British English	Not reported	Typical	8.51	10
Wieland et al. (2015, Experiment 1)	U.S. English	United States	Typical	19.90	20
Wieland et al. (2015, Experiment 2)	U.S. English	United States	Typical	4.95	20
Wieland et al. (2015, Experiment 3)	U.S. English	United States	Typical	13.10	11
Wieland et al. (2015, Experiment 4)	U.S. English	United States	Typical	4.80	11
Xu et al. (2013)	Australian English	Australia	Typical	6.00	11
Meta-analysis 2					
Andruski et al. (1999)	Japanese	Japan	Typical	7.00	14
Ashby (2004)	Australian English	Australia	Typical		6

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

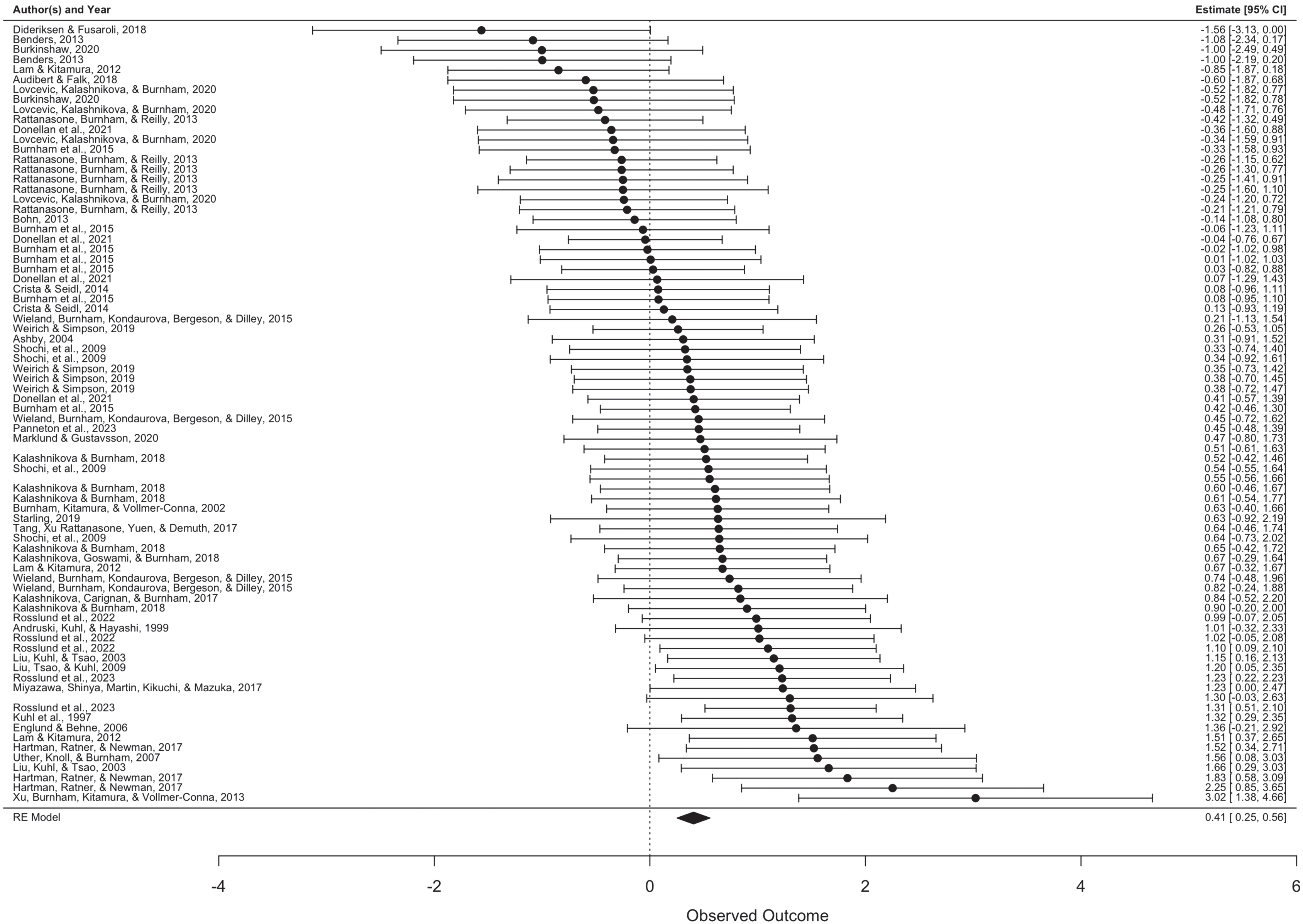
Table 2 (*continued*)

Study	Native language	Country	At risk status	Infant age (months)	Number of participants
Audibert and Falk (2018, Experiment 1)	German	Not reported	Typical	5.80	14
Audibert and Falk (2018, Experiment 2)	German	Not reported	Typical	5.80	14
Benders (2013, Experiment 1)	Dutch	Netherlands	Typical	11.00	18
Benders (2013, Experiment 2)	Dutch	Netherlands	Typical	15.00	18
Bohn (2013, Experiment 1)	Danish	Not reported	Typical	19.00	26
Bohn (2013, Experiment 2)	Danish	Not reported	Typical	20.20	13
Burkinshaw (2020, Experiment 1)	Canadian English	Canada	Typical	7.24	5
Burkinshaw (2020, Experiment 2)	Canadian English	Canada	Typical	14.52	4
E. B. Burnham et al. (2015, Experiment 1)	U.S. English	United States	Typical	3.00	11
E. B. Burnham et al. (2015, Experiment 2)	U.S. English	United States	Typical	5.90	11
E. B. Burnham et al. (2015, Experiment 3)	U.S. English	United States	Typical	9.00	11
E. B. Burnham et al. (2015, Experiment 4)	U.S. English	United States	Typical	3.10	12
E. B. Burnham et al. (2015, Experiment 5)	U.S. English	United States	Typical	9.00	12
E. B. Burnham et al. (2015, Experiment 6)	U.S. English	United States	Typical	12.80	12
E. B. Burnham et al. (2015, Experiment 7)	U.S. English	United States	Typical	20.40	12
Cox, Dideriksen, et al. (2023)	Danish	Denmark	Typical	17.73	5
Dilley et al. (2015)	U.S. English	United States	With hearing loss	39	
K. Englund and Behne (2006)	Norwegian	Norway	Typical	5.52	6
Gergely et al. (2017, Experiment 1)	Hungarian	Hungary	Typical	4.80	5
Gergely et al. (2017, Experiment 2)	Hungarian	Hungary	Typical	16.50	6
Gergely et al. (2017, Experiment 3)	Hungarian	Hungary	Typical	25.50	8
Kalashnikova et al. (2018)	Australian English	Australia	At risk for dyslexia	10.26	18
Kalashnikova et al. (2020, Experiment 1)	Australian English	Australia	Typical	11.77	7
Kalashnikova et al. (2020, Experiment 2)	Australian English	Australia	Typical	11.77	6
Kalashnikova et al. (2020, Experiment 3)	Australian English	Australia	At risk for dyslexia	10.79	8
Kalashnikova et al. (2020, Experiment 4)	Australian English	Australia	At risk for dyslexia	10.79	8
Kalashnikova et al. (2020, Experiment 5)	Australian English	Australia	At risk for dyslexia	10.79	8
Lam and Kitamura (2012, Experiment 1)	Australian English	Australia	Typical	6.19	16
Lam and Kitamura (2012, Experiment 2)	Australian English	Australia	Typical	6.19	16
Lam and Kitamura (2012, Experiment 3)	Australian English	Australia	Typical	6.19	16
Lovcevic, Kalashnikova, and Burnham (2020, Experiment 1)	Australian English	Australia	With hearing loss	15.09	20
Lovcevic, Kalashnikova, and Burnham (2020, Experiment 2)	Australian English	Australia	With hearing loss	17.64	11
Shochi et al. (2009, Experiment 1)	Japanese	Japan	Typical		10
Shochi et al. (2009, Experiment 2)	Australian English	Australia	Typical		9
Shochi et al. (2009, Experiment 3)	Japanese	Japan	Typical		10
Shochi et al. (2009, Experiment 4)	Australian English	Australia	Typical		9
Rosslund et al. (2024, Experiment 1)	Norwegian	Norway	Typical	8.22	45
Rosslund et al. (2024, Experiment 2)	Norwegian	Norway	Typical	8.22	45
Rosslund et al. (2022, Experiment 1)	Norwegian	Norway	Typical	17.88	21
Rosslund et al. (2022, Experiment 2)	Norwegian	Norway	Typical	17.88	21
Rosslund et al. (2022, Experiment 3)	Norwegian	Norway	Typical	17.88	21
Weirich and Simpson (2019, Experiment 1)	German	Germany	Typical	6.00	15
Weirich and Simpson (2019, Experiment 2)	German	Germany	Typical	6.00	15
Weirich and Simpson (2019, Experiment 3)	German	Germany	Typical	9.00	15
Weirich and Simpson (2019, Experiment 4)	German	Germany	Typical	9.00	15
Wieland et al. (2015, Experiment 1)	U.S. English	United States	With hearing loss	18.20	20
Wieland et al. (2015, Experiment 2)	U.S. English	United States	With hearing loss	11.20	11

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 3

Forest Plot of Individual Experiments and Associated Effect Sizes



Note. For each experiment, we represent the estimated mean effect sizes (Hedge's g) and upper and lower 95% confidence intervals (CIs), while numeric estimates are presented on the right side; RE = random-effects.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Analysis Steps

Publication Bias. We calculated the sensitivity of the meta-analytic results to different levels of publication bias (Mathur & VanderWeele, 2020). We assumed that meta-analytic studies represent samples from an underlying population of published and unpublished studies, where the probability of selection (i.e., publication) for affirmative studies (with a significant effect size in the expected direction) is higher than for nonaffirmative ones (with no significant effect size or a significant effect size in the unexpected direction). Sensitivity to publication bias is thus assessed by artificially increasing the proportion of studies from the nonaffirmative group, increasingly attenuating a potential publication bias. At each increment, the overall meta-analytic effect is estimated. We can assess at which level of publication bias the credible intervals of the estimate first cross 0. This measure thus allows us to estimate how much more likely it would need to be for affirmative than for nonaffirmative studies to be published in order for publication bias to explain away observed effects. Next, we explored possible cues of actual publication bias in the relation between effect size estimate and uncertainty. We represented the data in a significance funnel plot and assessed whether there was a monotonic relation between standard error and effect size (traditionally referred to as a Spearman rank correlation). Systematically increased effect sizes for studies with higher standard errors (i.e., lower sample size and/or quality of the measurements) indicate that a publication bias is likely.

Frequentist Model. We ran a baseline model without moderators, which returned the overall, mean effect size, followed by moderator models. In the Meta-analysis 1 (narrow-search), models including infants' age, native language, and measurement scale as noninteracting moderators were run. Three separate models were run in order to assess the linear, quadratic, and cubic effects of infants' age. The best fitting model was determined based on the Akaike information criterion (Akaike, 1974), with the model with the lowest Akaike information criterion being considered as the best fit. In the Meta-analysis 2 (broad-search), three moderator models were also run. The first moderator model included infants' age (linear effect, as it has been identified as the best fitting one in Meta-analysis 1), native language, and measurement scale as noninteracting moderators. The second moderator model included design-related methodological moderators, and the third model included vowel-related methodological moderators. Given that not all combinations of moderators were present in the data, we assessed whether planned moderator analyses could be performed, that is, whether we had data for contrasts within the moderators to obtain reliable inferences and whether such contrasts were confounded across moderators (e.g., if British English studies assessed vowel hyperarticulation only on the Hertz scale and Cantonese studies only on logarithmic scales, the analysis of language effects would be confounded with the scale contrast). A comparison was considered valid only when each cell involved contained at least one study (see Table 1 for planned moderators for Meta-analyses 1 and 2).

Bayesian Model. In order to estimate the pooled effect size and credible intervals, a hierarchical Bayesian robust regression model using a Student's *t*-likelihood was fitted to the data. Fat-tailed distributions (here, a student's *t*-distribution) dampen the influence of outliers, incorporating them without allowing them to dominate the inference (Jylänki et al., 2011). Weakly informative priors were chosen so that their influence on the meta-analytic estimates was small, whereas extreme effect sizes were discounted as unlikely (see details and sensitivity checks in the Supplemental Material S2). When moderators could be included (age, some languages, and measurement scales), we assessed whether the model including them was better than the baseline model via Pareto Smoothed Importance Sampling to estimate Leave-One-Out-Cross-Validated error in each model: whether estimated out-of-sample error (estimated error in predicting data points never seen before) was reliably lower in the model that included moderators. Further, we used stacking weights to assess the relative credibility of models: a stacking weight of 1 indicates that this model is estimated to best predict novel data points and no additional information can be gained from the other model, whereas a stacking weight of 0.5 indicates that both models contain relevant and complementary information (Yao et al., 2018). Additionally, between-study heterogeneity was estimated as the population-level variance (standard deviation) in effect size between studies, after adjusting for systematic variance explained by moderators.

Results

Table 2 provides an overview of all studies included in this meta-analysis. Additionally, Table 3 provides descriptive statistics for continuous and categorical moderators.

Meta-Analysis 1 (Narrow-Search)

Data Description

The data were assessed to ensure that contrasts could be meaningfully tested, that is, all relevant cells contained at least one data point. For instance, if for a given language only data on a log scale were present, the effects of scale within that language could not be assessed. The following contrasts could be meaningfully tested:

- Age assessed separately within U.S. English, Australian English, and Mandarin with outcomes in Hertz.
- Age assessed separately within U.S. English and Cantonese with outcomes on a logarithmic scale.
- Language assessed separately for studies conducted in Alur, U.S. English, Australian English, British English, Japanese, Kiswahili, Lugbara, Mandarin, and Swedish with outcomes in Hertz.
- Language assessed separately for studies conducted in U.S. English, Australian English, Cantonese, and Mandarin with outcomes on a logarithmic scale.
- Measurement scales tested separately within U.S. and Australian English, as well as Mandarin.

See details in Table 4 and age distributions across the different cells in Figure 4.

Publication Bias

An analysis of sensitivity to publication bias indicates that if significant results were even just 1.52 times more likely to be published than nonsignificant results, the corrected credible intervals of the meta-analytic estimate would include 0, as depicted in

Table 3
Descriptive Statistics for Moderators

Moderator	Meta-analysis 1			Meta-analysis 2		
	M (*SD*)	Range	*k*	*M* (*SD*)	Range	*k*
Infants' age (days)	313.21 (161.65)	92–730.56	42 (100%)	320.47 (170.19)	91.32–776.22	75 (93%)
Native language						
Alur (Uganda)			1 (2%)			1 (1%)
U.S. English			11 (26%)			18 (22%)
Australian English			13 (30%)			21 (26%)
British English			2 (4%)			2 (2%)
Canadian English			0 (0%)			2 (2%)
Cantonese			6 (14%)			6 (7%)
Danish			0 (0%)			3 (3%)
Dutch			0 (0%)			2 (2%)
German			0 (0%)			6 (7%)
Hungarian			0 (0%)			3 (3%)
Japanese			1 (2%)			4 (5%)
Kiswahili (Uganda)			1 (2%)			1 (1%)
Lugbara (Uganda)			1 (2%)			1 (1%)
Mandarin			4 (9%)			4 (5%)
Norwegian			0 (0%)			6 (7%)
Swedish			1 (2%)			1 (1%)
Mixed			1 (2%)			1 (1%)
Measurement scale						
Hertz			30 (71%)			44 (55%)
Logarithmic			12 (28)			38 (47%)
Recording order						
IDS first						30 (37%)
ADS first						5 (6%)
Not fixed-counterbalanced						25 (31%)
Not fixed-randomized						7 (8%)
Not fixed-ordering principle unspecified						1 (1%)
Ordering principle unspecified						11 (13%)
Elicitation stimulus in IDS						
None						1 (1%)
Object without written label						32 (40%)
Object with written label						11 (13%)
Picture without written label						2 (2%)
Object and picture without written label						12 (15%)
Written verbiage						16 (20%)
Memorized verbiage						4 (5%)
Unspecified						4 (5%)
Elicitation stimulus in ADS						
None						6 (7%)
Object without written label						18 (22%)
Object with written label						10 (12%)
Picture without written label						2 (2%)
Object and picture without written label						8 (10%)
Written verbiage						16 (20%)
Memorized verbiage						4 (5%)
Not specified if IDS stimuli were retained						13 (16%)
Unspecified						5 (6%)
Recording context						
Unconstrained						51 (63%)
Reading						16 (20%)
Singing						1 (1%)
Repetition of sentence/phrase/text provided by experimenter						0 (0%)
Interlocutor in IDS						
Own child						79 (98%)
Other child						2 (2%)
Pictured child						0 (0%)
Absent/imaginary child						1 (1%)
Interlocutor in ADS						
Adult experimenter						67 (83%)
Adult unspecified						3 (3%)
Unspecified						4 (5%)
Absent interlocutor						8 (10%)

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 3 (*continued*)

Moderator	Meta-analysis 1 *M* (*SD*)	Meta-analysis 1 Range	Meta-analysis 1 *k*	Meta-analysis 2 *M* (*SD*)	Meta-analysis 2 Range	Meta-analysis 2 *k*
Method of formant measurements						
Automatic only						30 (37%)
Automatic—outliers excluded						9 (11%)
Automatic—outliers human checked						3 (3%)
Automatic—human reliability						1 (1%)
Automatic—human verification						5 (6%)
Automatic—human verification and reliability						1 (1%)
Human—primary coder						21 (26%)
Unspecified						12 (15%)
Session duration (in minutes)						
0–4						15 (18%)
5–10						21 (26%)
11–30						4 (5%)
More than 30						0
Unspecified						42 (52%)
Vowel type						
Only corner vowel						65 (83%)
Corner vowels and other vowels						10 (12%)
Vowel number				3.32 (0.89)	3–9	75 (93%)
3						61 (76%)
4						14 (17%)
7						1 (1%)
9						1 (1%)
Unspecified						5 (6%)
Target-word types						
One target						35 (43%)
Several targets						31 (38%)
All content words						3 (3%)
All words analyzed in focus utterance						1 (1%)
All words						7 (8%)
Unspecified						5 (6%)
Stress						
Stressed						56 (70%)
Both stressed and unstressed						12 (15%)
Unspecified						13 (16%)

Note. k = number of effect sizes included in meta-analysis, percentage of total number of studies is shown in brackets. IDS = infant-directed speech; ADS = adult-directed speech.

Figure 5. The literature suggests that significant results in developmental psychology are 4.7 times more likely to be published than nonaffirmative ones (cf. Mathur & VanderWeele, 2020). In other words, were we to assume a realistic publication bias and correct our estimate for that, our estimated meta-analytic effect size would not be significantly different from 0. In the extreme, were the estimated effect size to be inferred from nonaffirmative studies only, we would expect an effect practically indistinguishable from 0: 0.01, 95% CIs [−0.33, 0.36]. Further, Spearman rank correlation also indicates weak evidence for a small relation between higher effect sizes and standard error (which would indicate plausible publication bias): estimate: 0.03, $SE = 0.03$, 95% CIs [−0.02, 0.08], ER = 4.93. Taken together, these results suggest strong caution in interpreting meta-analytic effects, as they are quite likely inflated by publication bias, and one cannot exclude the underlying effect size to be actually 0. It should be noted that the publication bias for Meta-analysis 1 was assessed only for the published studies, as there were no unpublished studies in the study sample.

Table 4
Number of Effect Sizes According to Native Language and Measurement Scale

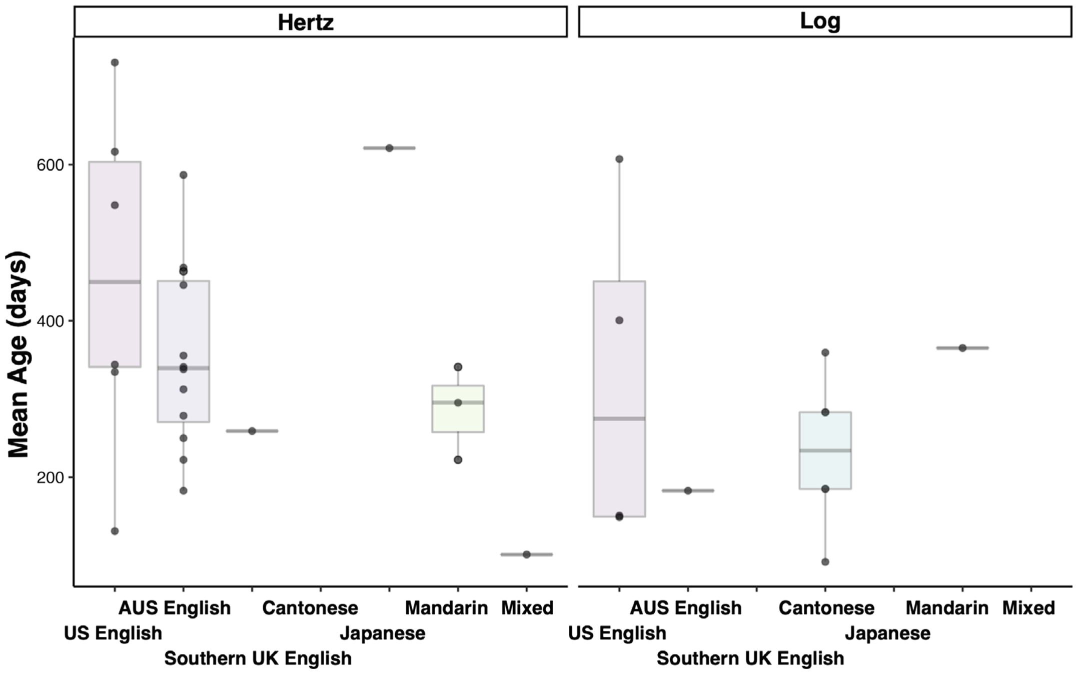
Language	Hertz	Logarithmic
Alur (Uganda)	1	0
U.S. English	7	4
Australian English	12	1
British English	2	0
Cantonese	0	6
Japanese	1	0
Kiswahili (Uganda)	1	0
Lugbara (Uganda)	1	0
Mandarin	3	1
Swedish	1	0
Mixed	1	0

The Baseline Model

Both the Frequentist and Bayesian baseline models indicated reliable vowel hyperarticulation, albeit slightly smaller in the latter approach (Frequentist: Hedges' $g = 0.69$, $SE = 0.16$, 95% CIs [0.37, 1.02]; Bayesian: Hedges' $g = 0.63$, $SE = 0.15$, 95% CIs [0.33, 0.92]). According to Cohen's (1988) criteria for effect size estimates, this pooled result indicates a medium to large effect. A standardized mean

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 4
Age Distributions by Language and Measurement Scale



Note. In order to better qualify our analysis and interpretation of the impact of age, language, and measurement scale on vowel articulation, we represent the average age of the infants for each experiment separately by language and measurement scale. Cases with a broad distribution of ages can be more meaningfully analyzed and interpreted; cases with only one age cannot be meaningfully assessed for age effects, for instance. AUS English = Australian English. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

difference of this size implies that approximately 70% of vowel articulations in IDS will be more acoustically extreme than in ADS. Both approaches indicated substantial heterogeneity in the estimates, that is, variance unexplained by the model as a result of significant differences across the studies. The results show a large between-study heterogeneity, Frequentist: $Q(41) = 62.96$, $p < .05$; Bayesian: 0.62, 95% CIs [0.41, 0.87], small additional heterogeneities between samples (Bayesian: 0.09, 95% CIs [0.00, 0.28]), and small additional heterogeneities between experiments (Bayesian: 0.07, 95% CIs [0.00, 0.23]).

Moderator Model With Infants' Age, Native Language, and Scale as Moderators

Both Frequentist and Bayesian approaches indicate no evidence suggesting that hyperarticulation changes with age (linear models provided the best fit against polynomial and monotonic changes), and the linear slope over age was not reliably different from 0 (see Table 5 and Figure 6). Contrary to the Frequentist approach, Bayesian analyses found cross-measurement scale and cross-linguistic variability in the effects. Specifically, there were reliable differences of measurement scale by language (ERs for the differences >10; see Table 6). Australian English showed systematically larger effects for studies using a logarithmic scale than for those using Hertz (1.19, 95% CIs [−0.20, 2.39], ER = 12.99). Mandarin, on the contrary, showed systematically larger effects for studies using Hertz (0.91, 95% CIs [−0.13, 1.87], ER = 13.08), and U.S. English showed anecdotal evidence of larger effects for studies using Hertz (0.32, 95% CIs [−0.49, 1.16], ER = 3.01). Finally, with regard to native language, there is evidence for vowel hyperarticulation in all languages but Cantonese and Lugbara, which show weak-to-moderate evidence of hypoarticulation across available scales within each language (see Table 6). There was also a reliable cross-linguistic variation (only comparing within available scales), with Australian English showing stronger hyperarticulation on the logarithmic scales than the three other languages with data available, namely U.S. English (1.08, 95% CIs [−0.38, 2.43], ER = 9.20), Cantonese (1.77, 95% CIs [0.29, 3.15], ER = 35.36), and Mandarin (1.07, 95% CIs [−0.50, 2.58], ER = 7.44; see Table 7). On the contrary, Australian English showed lower hyperarticulation on the Hertz scale than those same three languages (U.S. English: −0.42, 95% CIs [−1.13, 0.30], ER = 4.85; British English: −0.75, 95% CIs [−1.78, 0.29], *ER* = 7.66; and Mandarin: −1.02, 95% CIs [−1.78, −0.25], ER = 47.78), with little to no evidence of differences from Japanese (−0.23, 95% CIs [−1.08, 0.69], ER = 2.09; see Table 6), Swedish (0.01, 95% CIs [−0.96, 1.01], ER = 1.03), Alur (0.00, 95% CIs [−0.98, 0.94], ER = 1.03), Kiswahili (0.31, 95% CIs [−0.61, 1.24], ER = 2.46), or Lugbara (0.57, 95% CIs [−0.54, 1.62], ER = 4.17).

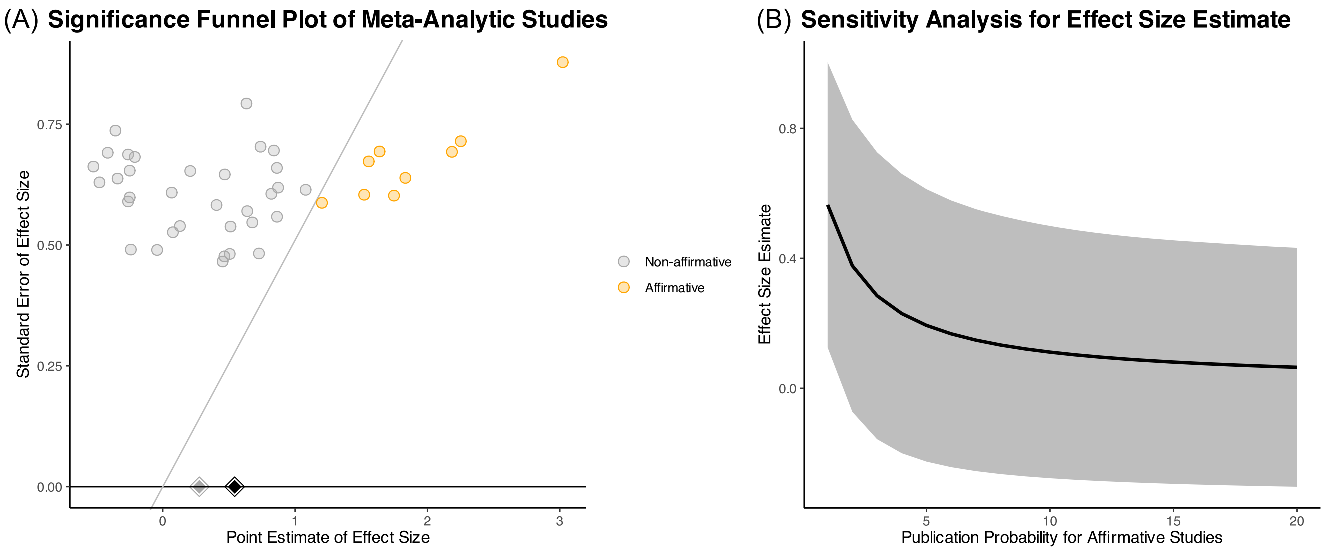
Results

Meta-Analysis 2 (Broad-Search)

Data Description. Inspection of the data (see Table 8) suggested that the following contrasts could be meaningfully assessed:

Figure 5

Panel A: Significance Funnel Plot; Panel B: Sensitivity Plot for Publication Bias



Note. Panel A (left): The *x*-axis represents the estimated average effect size of each experiment, and the *y*-axis represents their standard error (the higher the standard error, the smaller the sample size and/or the larger the heterogeneity between participants, thus indicating less reliable estimates). The gray line separates studies with statistically significant results in the expected direction (affirmative results, on the right) from those with null results or significant but unexpected results (nonaffirmative results, on the left). The dark gray diamond on the *x*-axis represents the pooled estimate across all studies in the meta-analysis. The light gray diamond on the *x*-axis indicates the corrected meta-analytic estimate across studies, where the correction assumes the worst-case publication bias (i.e., estimates the average effect size including only the nonaffirmative results). Panel B (right) presents a more nuanced perspective on the impact of possible levels of publication bias, according to a sensitivity plot. The *x*-axis indicates the severity of the publication bias in terms of relative odds of publishing a significant result in the expected direction against a nonaffirmative result. 10 indicates that it is 10 times easier to publish an affirmative result. The *y*-axis reports the estimated meta-analytic effect size if we correct the estimate according to the publication bias on the *x*-axis. The higher the assumed publication bias, the lower the corrected estimate is. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

- Age for U.S. English, Australian English, British English, Cantonese, Danish, Dutch, German, Hungarian, Japanese, and Mandarin.
- Language for Alur, U.S. English, Australian English, British English, Danish, Japanese, Kiswahili, Lugbara, Mandarin, Norwegian, and Swedish studies with outcomes in Hertz.
- Language for U.S. English, Australian English, Canadian English, Cantonese, Dutch, German, Hungarian, Japanese, and Mandarin with outcomes on a logarithmic scale.
- Measurement scales within American and Australian English, Japanese, and Mandarin.

Publication Bias. An analysis of sensitivity to publication bias indicates that if significant results were 1.65 times more likely to be published than nonsignificant results, the corrected credible intervals of the meta-analytic estimate would include 0, as depicted in Figure 7. This publication bias is much lower than the publication bias of 4.7 estimated for developmental psychology (Mathur & VanderWeele, 2021). In other words, were we to assume a realistic publication bias and correct our estimate for that, our estimated meta-analytic effect size would most likely not be significantly

Table 5

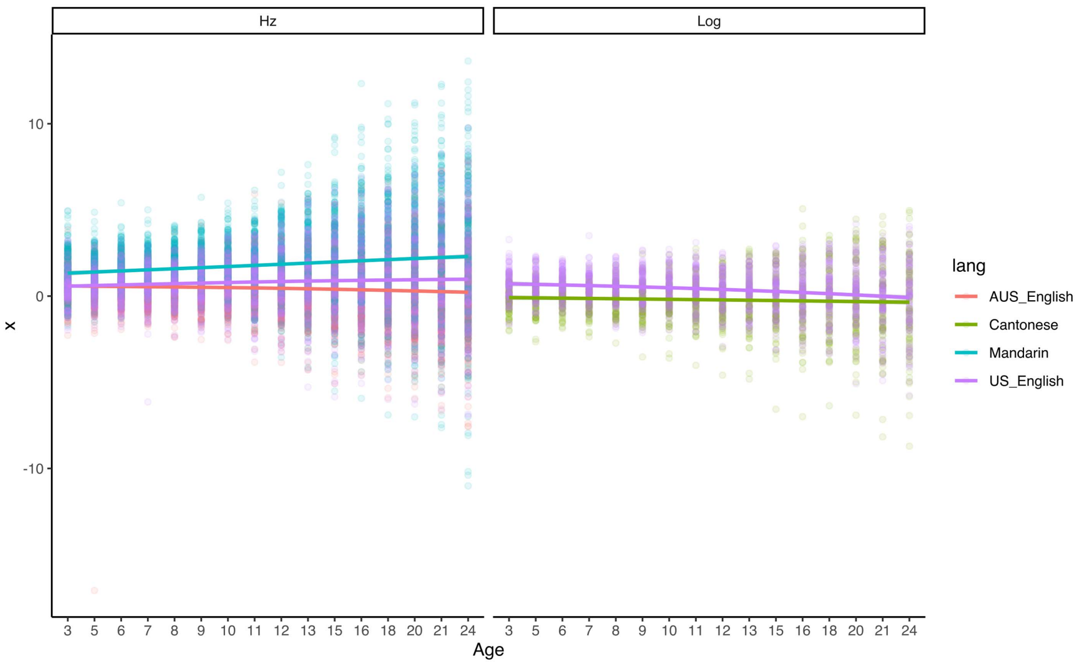
Impact of Age on Estimated Effect Size by Native Language and Measurement Scale in Meta-Analysis 1

		Measurement scale							
		Hertz				Logarithmic			
Language	Age (months)	ES	95% CIs	ER	ER01	ES	95% CIs	ER	ER01
U.S. English	4, 11, 13, 18, 19, 20, 24	.04	[−.21, .28]	.58	6.84	−.03	[−.39, .31]	1.34	5.71
Australian English	6, 9, 11, 15, 19	−.02	[−.26, .20]	1.24	7.57	NA	NA	NA	NA
Cantonese	3, 6, 9, 11	NA	NA	NA	NA	.00	[−.26, .26]	.96	7.60
Mandarin	7, 11, 12	.07	[−.33, .52]	.65	4.38	NA	NA	NA	NA

Note. CIs = confidence intervals; ES = effect size; ER = evidence ratio; NA = not applicable.

Figure 6

The Estimated Relation Between Vowel Articulation Effects of IDS (y-Axis) and Age (x-Axis) for Languages With More Than One Age Point in the Data Set, Separately by Scale



Note. Hertz is presented in the left panel; logarithmic scales are presented in the right panel. Each dot represents a random sample from the posterior distributions. Note the high uncertainty of the predictions. IDS = infant-directed speech; AUS English = Australian English. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

different from 0. However, Spearman rank correlation did not show any correlation between effect sizes and standard error (0.01, 95% CIs [−0.01, 0.03], ER = 2.2). Given the results of publication bias in Meta-analysis 1 that demonstrated that effects were inflated by publication bias, the Meta-analysis 2 results suggest that including more and more diverse studies mitigates but does not completely solve the problem. To sum up, although there was mixed evidence of publication bias in the Meta-analysis 2, caution in interpretation is needed.

The Baseline Model

Both the Frequentist and Bayesian baseline models indicated a reliable vowel hyperarticulation (Frequentist: Hedges' $g = 0.41$,

Table 6

Estimated Effect Sizes by Native Language as Moderated by Measurement Scale in Meta-Analysis 1

	Measurement scale								
	Hertz and logarithmic			Hertz			Logarithmic		
Language	ES	95% CIs	ER	ES	95% CIs	ER	ES	95% CIs	ER
Alur (Uganda)	NA	NA	NA	.33	[−.51, 1.18]	2.99	NA	NA	NA
U.S. English	.61	[.16, 1.04]	57.82	.76	[.19, 1.31]	57.28	.46	[−.20, 1.08]	8.66
British English	NA	NA	NA	1.09	[.13, 1.99]	31.26	NA	NA	NA
Australian English	.94	[.26, 1.60]	65.67	.34	[−.11, .82]	8.17	1.54	[.27, 2.72]	30.25
Cantonese	NA	NA	NA	NA	NA	NA	−.23	[−.80, .36]	3.56
Japanese	NA	NA	NA	.57	[−.23, 1.27]	8.62	NA	NA	NA
Kiswahili (Uganda)	NA	NA	NA	.03	[−.81, .81]	1.10	NA	NA	NA
Lugbara	NA	NA	NA	−.24	[−1.23, .77]	.52	NA	NA	NA
Mandarin	.91	[.41, 1.39]	284.71	1.36	[.70, 1.94]	332.33	.47	[−.40, 1.27]	5.37
Swedish	NA	NA	NA	.32	[−.55, 1.17]	2.87	NA	NA	NA

Note. CIs = confidence intervals; ES = effect size; ER = evidence ratio; NA = not applicable.

Table 7

Differences in Effect Size Between Studies Relying on Hertz Estimates of Vowel Articulation and Studies Relying on Logarithmic Estimates

Language	Estimate	*SE*	95% CI *LL*	95% CI *UL*	ER	Credibility
Total	0.00	0.38	−.62	.62	0.96	.49
U.S. English	.30	0.50	−.54	1.14	2.79	.74
Australian English	−1.20	0.81	−2.49	.19	12.62	.93
Mandarin	.89	0.65	−.15	1.99	11.12	.92

Note. Estimates indicate standardized effects. For example, −0.12 indicates that the average effect was 0.12 *SD* lower in the studies reporting Hertz than in those reporting logarithmic estimates. CI = confidence interval; *SE* = standard error; *LL* = lower limit; *UL* = upper limit; ER = evidence ratio.

SE = 0.08, 95% CIs [0.25, 0.56]; Bayesian: Hedges' $g = 0.47$, $SE = 0.13$, 95% CIs [0.22, 0.76]) with values of Hedges' g indicating a medium to large effect according to Cohen's (1988) criteria. Furthermore, both approaches showed substantial heterogeneity in the estimates, that is, variance unexplained by the model as a result of significant differences across the studies. Specifically, the results show a large between-study heterogeneity, Frequentist: $Q(77) = 119.99$, $p < .01$; Bayesian: 0.63, 95% CIs [0.36, 0.90], and small additional heterogeneities between samples (Bayesian: 0.19, 95% CIs [0.01, 0.52]) and between experiments (Bayesian: 0.05, 95% CIs [0.00, 0.16]).

Moderator Models

Models With Infants' Age, Native Language, and Scale as Moderators. The Frequentist approach did not find evidence for the effects of native language, infants' age, or measurement scale. Similar results were obtained via Bayesian approach, with no evidence for the effects of these moderators.

Table 8

Number of Studies According to Native Language, Infants' Age, and Measurement Scale

Language	Infants' age (in months)	Hertz	Logarithmic
Alur (Uganda)	3	1	0
U.S. English	3, 4, 6, 11, 13, 18, 19, 20, 24	6	11
Australian English	6, 7, 8, 9, 10, 11, 14, 15, 19	14	5
British English	5, 8	2	0
Canadian English	4	0	2
Cantonese	3, 6, 9, 11	0	6
Danish	4, 16	3	0
Dutch	11, 15	0	2
German	5, 6, 9	0	6
Hungarian	4, 16, 25	0	3
Japanese	7, 20	3	1
Kiswahili (Uganda)	5	1	0
Lugbara (Uganda)	3	1	0
Mandarin	9, 7, 11, 12	3	1
Mixed	3	1	0
Norwegian	5	1	0
Swedish	12	1	0

Note. Column "Age" depicts the specific infants' age(s) (in months) at which data were available.

Models With Design-Related Methodological Factors as Moderators. Both Frequentist and Bayesian approaches indicate ambiguous results due to the underlying confounding between conditions. For example, of the 14 studies (32 experiments) that used objects without a written label as stimuli in IDS (factor: elicitation stimulus in IDS), all had unconstrained recording context (factor: recording context; see Supplemental Material S4 for contingency tables).

Models With Vowel-Related Methodological Factors as Moderators. Both Frequentist and Bayesian models indicated that the model including vowel-related methodological moderators is better than the baseline model, Frequentist method: *Q*-test for moderators $Q(6) = 31.16$, $p < .0001$; Bayesian model: stacking weight = .64. However, we refrain from interpreting it due to the sparsity and confoundedness of conditions. For example, of the 27 studies (61 experiments) investigating three vowels (factor: vowel number), 26 (60 experiments) included the corner vowels (factor: vowel type; see Supplemental Material S4 for contingency tables).

Discussion

The existing literature provides conflicting evidence regarding the presence of vowel hyperarticulation in IDS, which refers to caregivers' propensity to produce acoustically exaggerated vowels when speaking to infants. Thus, we conducted a systematic review and meta-analysis to assess the robustness of vowel hyperarticulation in IDS across studies and its moderators. Additionally, we aimed to identify the common practices and pitfalls in the field to better guide future studies.

Our results demonstrate strong evidence for the presence of vowel hyperarticulation in IDS as compared to ADS. Regarding the potential moderators, our results suggest cross-linguistic and cross-measurement scale effects. Furthermore, the results of Meta-analysis 2 (broad-search) suggest confoundedness of data resulting in data sparsity, making the interpretation of potential moderator effects more challenging.

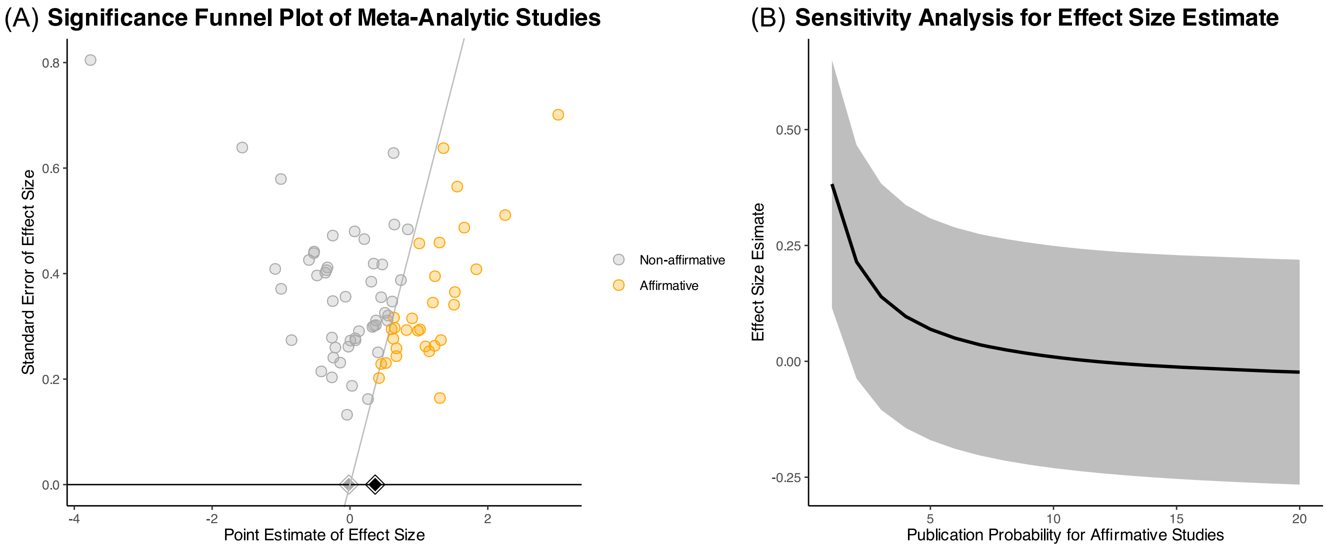
Concurrent with the above goals, we considered the impact of the sampling strategy and analysis method on the meta-analytic outcomes. First, we analyzed both the methodologically consistent subset of studies following Kuhl et al. (1997) study (Meta-analysis 1), as well as a methodologically heterogeneous set of all relevant studies on the topic (Meta-analysis 2; e.g., Lipsey & Wilson, 1993). After assessing effect sizes in a baseline model without moderators, we analyzed how vowel hyperarticulation may be impacted by moderators in both study samples. Second, we combined the dominant Frequentist analysis approach with Bayesian analyses. As our Bayesian approach is more conservative and allows us to more intuitively test contrasts using only the nonconfounded conditions, the discussion presented here will be primarily based on Bayesian findings.

Results for the baseline model across study samples (Meta-analyses 1 and 2) and analytic approaches (Frequentist and Bayesian) indicate the likely presence of vowel hyperarticulation in IDS (Meta-analysis 1: Frequentist effect size estimate = 0.69, Bayesian effect size estimate = 0.63; Meta-analysis 2: Frequentist effect size estimate = 0.41, Bayesian effect size estimate = 0.47). Models including moderators showed less consistent results, likely due to regularizing priors and robust regression methods in the Bayesian

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 7

Panel A: Significance Funnel Plot; Panel B: Sensitivity Plot for Publication Bias



Note. Panel A (left): The *x*-axis represents the estimated average effect size of each experiment, and the *y*-axis represents their standard error (the higher the standard error, the smaller the sample size and/or the larger the heterogeneity between participants, thus indicating less reliable estimates). The gray line separates studies with statistically significant results in the expected direction (affirmative results, on the right) from those with null results or significant but unexpected results (nonaffirmative results, on the left). The dark gray diamond on the *x*-axis represents the pooled estimate across all studies in the meta-analysis. The light gray diamond on the *x*-axis indicates the corrected meta-analytic estimate across studies, where the correction assumes the worst-case publication bias (i.e., estimates the average effect size including only the nonaffirmative results). Panel B (right) presents a more nuanced perspective on the impact of possible levels of publication bias, according to a sensitivity plot. The *x*-axis indicates the severity of the publication bias in terms of relative odds of publishing a significant result in the expected direction against a nonaffirmative result. 10 indicates that it is 10 times easier to publish an affirmative result. The *y*-axis reports the estimated meta-analytic effect size if we correct the estimate according to the publication bias on the *x*-axis. The higher the assumed publication bias, the lower the corrected estimate is. See the online article for the color version of this figure.

approach. Converging evidence for the impact of moderators is thus sparse but is discussed in more detail further below.

Infant-Related Moderators

The results demonstrated cross-linguistic variability in the degree of vowel hyperarticulation in IDS. However, no effects of infants' age were observed. Furthermore, the results indicate the paucity of IDS studies with at-risk infants.

Native Language

Regarding the effect of native language on vowel hyperarticulation in IDS, our results suggest a vowel hypoarticulation in Cantonese. Our study included another tone language, Mandarin, which showed evidence for vowel hyperarticulation in IDS. We did not anticipate such contradictory effects across tone languages and are at present not able to satisfactorily explain these results with reference to the relation of vowel articulation to tone (e.g., Mandarin: Shaw et al., 2016; Vietnamese: Han, 1966; Taiwanese Chinese: Zee, 1980; Ningbo Chinese: Hu, 2004; Thai: Gandour, 1977). In Mandarin, rising and falling tones result in more extreme articulation for /a/ and /i/ (mediated by the effect of tone height on tongue-body position and subsequently formant frequencies; Erickson et al., 2004; Shaw et al., 2016; Torng, 2000). This might leave less scope for vowel hyperarticulation in IDS in words with rising and falling tones compared to words with level and falling-rising tones. Results from previous Mandarin IDS studies are not in line with this speculation, given that vowel hyperarticulation was documented in stimuli with all four tones (Liu et al., 2009), as well as in stimuli with only rising tones (Tang et al., 2017). In Cantonese IDS, by contrast, stimuli with falling tones elicited hypoarticulation (Rattanasone et al., 2013). For now, we can only suggest that studies on vowel hyperarticulation need to continue controlling for tone and that there is a scope for exploring the tone-vowel relationship in IDS.

Another factor that might explain our contrasting findings for Mandarin and Cantonese is vowel inventory size. Although Kuhl et al. (1997) did not observe differences in vowel hyperarticulation in languages with different vowel inventory sizes (American English, Russian, and Swedish), it is nevertheless viable that languages with a smaller inventory size might have a smaller vowel space in ADS (Al-Tamimi & Ferragne, 2005; Bradlow, 1995), thus providing more scope for vowel hyperarticulation in IDS. This is in accordance with our demonstration of vowel hyperarticulation in Mandarin, which has only six vowels (Duanmu, 2007), and vowel hypoarticulation in Cantonese, which has a relatively large vowel inventory of 18 vowels (Barrie, 2003). Note that this is not the largest vowel inventory size in our sample, with Danish being the one with the largest vowel inventory (37 vowels, Stokes et al., 2012).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

No other cross-linguistic differences in hyperarticulation were found, which might be caused by the limited diversity of languages, as well as the scarcity and small size of samples for the less represented languages in this meta-analysis. Our sample includes 10 different languages from four language families (Indo-European, Sino-Tibetan, Japonic/Japanese-Ryukyuan, Nilo-Saharan) with the languages from the Indo-European group being dominant. Hence, more cross-linguistic and cross-cultural research is needed, especially considering the reliance of the field on a limited number of languages and cultures affecting the generalizability of research findings on language acquisition generally (Blasi et al., 2022; Christiansen et al., 2022; Deffner et al., 2022; Kidd & Garcia, 2022) and other aspects of IDS more specifically (Cristia et al., 2019; Farran et al., 2019; Shneidman & Goldin-Meadow, 2012; Weber et al., 2017).

Risk Status

Our literature search revealed a limited number of vowel hyperarticulation studies with infants who are not developing typically (six out of 35 studies). It is thus not possible to reliably conclude if the degree of vowel hyperarticulation is different in IDS for these populations. Extending vowel hyperarticulation research to lesser studied populations would help characterize the input these children receive and also shed light on the mechanisms that drive hyperarticulation. For example, if speakers would speak most clearly to infants who need it most, hyperarticulation could be expected to be stronger in multilingual infants (who need to distinguish more vowel categories) and in infants with or at risk for sensory and cognitive difficulties (who may have difficulties developing vowel categories). The few relevant studies in our meta-analysis do not provide support for these hypotheses. Another population to consider is bilingual infants. These were not the focus of our study due to the limited number of IDS studies with bilingual infants. Regardless, the current state of literature indicates that only one study directly compared hyperarticulation to monolingual and bilingual infants and did not find significant group differences (Kalashnikova & Carreiras, 2022). Furthermore, IDS is, if anything, spoken less clearly to infants with (potential) deficits in auditory processing (infants at risk for developmental dyslexia, Kalashnikova et al., 2018, 2020) or to infants unable to hear mothers' speech (Lam & Kitamura, 2012). More research will be needed to firmly refute the hypothesis that hyperarticulation reflects speakers' intent to speak clearly in response to their infant's needs. Such research will also need to consider an alternative account of hyperarticulation as a response to the (covert) cues that parents receive from infants or unconscious phonetic convergence or accommodation (Lam & Kitamura, 2012; Pardo, 2006).

Design-Related Methodological Factors

The results of the Meta-analysis 1 (narrow-search) indicated reliable differences in measurement scale by language, particularly for Australian English and Mandarin. These will be discussed in more detail below. With regards to the inclusion of other design-related methodological moderators in Meta-analysis 2 (broad-search), we encountered the issue of the dependence between the coding categories resulting in data sparsity (e.g., studies sharing a type of elicitation stimulus tended to use similar recording contexts).

Measurement Scale

Vowel hyperarticulation research has been employing a mix of Hertz and logarithmic scales, but only a few studies (three out of 35, Andruski et al., 1999; Burkinshaw, 2020; Kuhl et al., 1997) report results on both scales. Due to the different nature of these scales (linear vs. logarithmic), hyperarticulation values on different scales cannot be directly compared, even after a transformation to effect sizes (see the Introduction section for more details). The use of different scales was thus a major contributor to the data sparsity in this meta-analysis, limiting the number of contrasts that could be meaningfully tested. For example, vowel hyperarticulation was reported on a single scale for the majority of languages, limiting our possibilities for assessing cross-linguistic differences. Kuhl et al.'s (1997) expressed hyperarticulation on a logarithmic scale under the assumption that this would best capture infants' processing of auditory input, a justification that has been explicitly adopted by several other studies (Andruski et al., 1999; Ashby, 2004; Audibert & Falk, 2018; Benders, 2013; Burkinshaw, 2020; E. B. Burnham et al., 2015; Cox, Dideriksen, et al., 2023; Gergely et al., 2017; Lam & Kitamura, 2010; Rattanasone et al., 2013; Tang et al., 2017; Weirich & Simpson, 2019; Wieland et al., 2015). In contrast, no study employing the linear Hertz scale has provided a motivation for doing so.

One of the goals of this meta-analysis was to assess whether the choice of measurement scale might affect the effect size of vowel hyperarticulation in IDS. Only data from Australian English, American English, Mandarin, and Japanese allowed within-language comparisons of scale. Of these, we only found a scale effect for Australian English and Mandarin. Australian English displayed more hyperarticulation on a logarithmic compared to Hertz scale, whereas Mandarin displayed more hyperarticulation on Hertz scale as opposed to logarithmic. As logarithmic scales accentuate changes in lower frequencies, we expect larger hyperarticulation on logarithmic scales, such as observed for Australian English, to stem from lowering of the formants. However, per-vowel changes (overall formants raising) observed in the figures of Australian English IDS studies (D. Burnham et al., 2002; Kalashnikova & Burnham, 2018; Lam & Kitamura, 2012; Xu et al., 2013) do not reveal an unusual shift to generally lower formant frequencies in that language (thus not supporting the hypothesized relationship between formant lowering and larger hyperarticulation effects on the logarithmic scales). On the other hand, the observed vowel hyperarticulation on Hertz scale in Mandarin might stem from raising of the formants, which is supported by inspection of per-vowel changes in the figures of Mandarin IDS studies (Liu et al., 2003, 2009). However, this is contradicted by the observation of formant raising alongside vowel hyperarticulation in a Mandarin IDS study employing the logarithmic (Bark) scale (Tang et al., 2017). It should be noted that even though we have more data on Australian English compared to other languages and availability of data on both Hertz and logarithmic scales in Mandarin studies, the data are not evenly distributed across scales (Australian English provided 12 effect sizes for the Hertz scale and one effect size for the logarithmic scale, whereas Mandarin provided three effect sizes for the Hertz and one effect size for the logarithmic scale in the Meta-analysis 1 [narrow-search], where the scale effects for Australian English and Mandarin were found). Also, it should be noted that some unidentified factors related to the laboratory might affect vowel hyperarticulation and consequently the cross-measurement scale results discussed here. In the case of Australian English, all six

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

studies (13 experiments) were conducted in the same laboratory, making it less likely that laboratory factors influenced the results. On the other hand, the case of Mandarin is mixed because two studies (three experiments) were conducted in the same lab, whereas one study (one experiment) was conducted in another lab. Thus, there is a possibility of effects due to unidentified laboratory factors. Nonetheless, regarding the measurement scale, we recommend that future (cross-linguistic) studies should ideally report on both scales so that we can get a better idea of the impact of this seemingly mundane choice on effect size estimates for hyperarticulation.

Vowel-Related Methodological Factors

The Bayesian method indicated that vowel-related methodological factors as a group affect vowel hyperarticulation in IDS, but without pinpointing any factor as a specific contributor. This lack of precision is due to dependence between the coding categories and, consequently, data sparsity (for instance, studies that investigated three vowels tended to choose the corner vowels; see Supplemental Material S4). Because of this, we refrain from further interpretation and recommend that future studies address the possible effect of the vowel-related methodological factors.

Considering the observed overall impact of vowel-related methodological factors on hyperarticulation, we suggest the direct manipulation of vowel-related methodological moderators in future studies as an interesting opportunity to directly test several hypotheses about IDS. First, if hyperarticulation in IDS is even stronger when people refer to (near-) minimal-pair words (e.g., Wang et al., 2018), an effect observed in adult-to-adult lab speech (Baese-Berk & Goldrick, 2009), this would provide evidence that caregivers consider lexical contractiveness when speaking to their infant. Hence, this will support the hypothesis of “audience design” (production of certain linguistic features to make speech more intelligible). Second, if vowel hyperarticulation is only true for corner vowels but not others (i.e., noncorner vowels), that has implications for the underlying production mechanisms as well as the learnability of the system as a whole (see Eaves et al., 2016, for more on this topic). We hope that the inability to answer such big questions about the nature of IDS even with the present large set of published studies provides an incentive for more targeted research.

Beyond Vowel Hyperarticulation

Vowel hyperarticulation assessment is based on $F1$ and $F2$ measurements in three (or more) vowels, which are combined in one aggregate measure of vowel space per register. The comparison between these aggregate measures abstracts away from, and thus obscures, a part of infants’ input, namely the exact patterns of $F1$ and $F2$ shifts in IDS compared to ADS. Of the studies included in this meta-analysis, six studies (seven experiments) reported per-vowel statistical analyses of $F1$ and $F2$, three studies (four experiments) reported statistical analyses of $F1$ and $F2$ averaged across vowels, 19 studies (43 experiments) provided per-vowel visualizations (but not statistical analyses), and two studies (four experiments) reported only hyper-scores without any information on individual vowels. This currently limits our potential to gain meta-analytic insight into these shift patterns and thus some of the mechanisms leading to hyperarticulation and its associations with language outcomes. For example, vowel space expansion as a result of prominent $F1$ raising in /a/ might result from wider mouth opening (Chong et al., 2003; Green et al., 2010; Stern, 1974), which could benefit language acquisition directly by enhancing visual cues to speech sounds (Erber, 1974; Rosenblum et al., 1996; Summerfield & Assmann, 1989) and indirectly by grabbing attention (Biringen, 1987; Cohn & Elmore, 1988). Expansion as a result of $F2$ raising in /i/ and /a/ might result from smiling (Benders, 2013; Chong et al., 2003; Stern, 1974), which may indirectly contribute to language acquisition through the resulting affective speech quality (Benders, 2016; Singh et al., 2002) and general developmental benefits (Pearson et al., 2011). Addressing such questions needs to start from observations beyond vowel hyperarticulation in IDS, including the per-vowel changes in IDS giving rise to the hyperarticulation.

Publication Practices in Vowel Hyperarticulation Research

Publication Bias

We found evidence for a publication bias in the Meta-analysis 1 (narrow-search; only for the published studies), whereas evidence from the Meta-analysis 2 (broad-search) is inconclusive. While the former calls for caution in interpreting the results, the latter suggests that including a larger and more diverse sample might alleviate publication bias but does not completely solve the problem.

Examination of the chronological timeline of studies might provide more insights into publication tendencies in the field. The timeline of studies in Meta-analysis 1 (narrow-search) and Meta-analysis 2 (broad-search; see Supplemental Figure S2.2) shows that the first three studies published after Kuhl et al. (1997) study used the same methodology. The first study using different methods appeared in 2004 and remained unpublished (Ashby, 2004). Hence, it is possible that initial studies could only successfully reach publication if the methodology and findings followed the field’s expectations (i.e., real publication bias). As more research was done from 2004 onwards, the scope of study widened to include somewhat different issues (e.g., noncorner vowels, different elicitation contexts, at-risk populations). While one might speculate that the associated changes to methodologies introduced a larger range of results, including in nonexpected directions, this speculation is not supported by the absence of clear design-related methodological effects (apart from the effect of measurement scale in the Meta-analysis 1, which is less relevant to this speculation). Rather, we suspect that the field’s broadening curiosity coincided with openness to “deviant” results. Consequently, more studies reporting an absence of vowel hyperarticulation or even hypoarticulation were published.

Sample Size

Meta-analyses can inform future studies on recommended sample sizes to be able to make correct statistical inferences. Hence, we used G*Power (Faul et al., 2007) to estimate required sample sizes for an α of .05 under the assumptions of (a) one measure per participant per condition (IDS and ADS), (b) Gaussian likelihood, and (c) no measurement error. These assumptions correspond to assumptions in the majority of (if not all) vowel hyperarticulation studies. Table 9 presents the estimated necessary sample sizes for both 80% ($\beta = 0.2$) and 95% ($\beta = 0.05$) power to detect the underlying effect sizes

Table 9
Results of Power Analysis

Power	One-tailed test (Meta-analysis 1)	Two-tailed test (Meta-analysis 1)	One-tailed test (Meta-analysis 2)	Two-tailed test (Meta-analysis 2)	Two repeated measures (Meta-analysis 1)	Two repeated measures (Meta-analysis 2)
80% ($\beta = 0.2$)	18	22	30	38	8	12
95% ($\beta = 0.05$)	29	35	51	61	12	18

Note. Results of power analysis to estimate the necessary number of participants to achieve 80% and 95% power to detect the effect size of 0.63 as per Bayesian results in Meta-analysis 1 (Narrow-search) and 0.47 as per Bayesian results in Meta-analysis 2 (Broad-search). We used one-tailed and two-tailed *t* tests to calculate the difference between two dependent means (matched pairs, IDS/ADS), whereas for repeated measures, ANOVA (within factors) was used. IDS = infant-directed speech; ADS = adult-directed speech; ANOVA = analysis of variance.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

obtained in the more conservative Bayesian analyses. Results suggested that the required sample sizes are 1.3–2.7 (for 80% power) and 2.1–4.4 (for 95% power) times higher than the average sample size of studies included in the present study (average sample size around 14 participants). Furthermore, the sample size of included studies ranged from three to 45 participants per experiment, with only five studies (six experiments) having over 25 participants (the sample size necessary for 80% power to observe the effect size obtained in Bayesian meta-analysis 2 (broad-search meta-analysis); Bohn, 2013; Cox, Dideriksen, et al., 2023; Cristia & Seidl, 2014; Kuhl et al., 1997; Rosslund et al., 2024). Hence, these power analyses suggest that the majority of vowel hyperarticulation studies might suffer from all problems associated with a low sample size (Button et al., 2013; Fraley & Vazire, 2014), leading to the recommendation of increasing the sample size in future studies. This might be challenging due to the time-consuming process of hand-annotating vowel boundaries and measuring formants. However, this difficulty might be overcome by employing at least partially automated analyses. These were utilized by 19 studies (42 experiments) in our meta-analytic sample, and our results did not suggest or refute any reliable difference in results depending on the degree of safeguards added to the automatic formant analysis. As formants will always be measured with a degree of error (Whalen et al., 2022) and studies with no measurement error are generally a myth, more participants, care in measurement practices to limit measurement error, and explicit inclusion of measurement error estimates should be employed. Additionally, the use of repeated measures might lower the required sample size, as long as we can ensure the smaller sample size is still representative of the broad population investigated (see Table 9 for the required sample sizes for 95% and 80% power with the assumptions of two repeated measures and the correlation between measures of 0.5).

Limitations and Directions for Future Research

Although our results suggest the likely presence of vowel hyperarticulation in IDS, they also pinpoint relevant methodological and theoretical future directions. First, as our study demonstrates, studies on vowel hyperarticulation in IDS have lower sample sizes than expected for well-powered studies with (roughly) median effect sizes. Hence, we recommend that future studies increase the sample size. Next, we recommend explicit manipulation and modeling of sources of heterogeneity in future studies. One of the sources of heterogeneity might be the measurement scale used for formants measurement. Considering our observations regarding measurement scale (see the Measurement Scale section), we recommend the following: (a) studies should consistently report vowel hyperarticulation (and other vowel-related values) on both the linear and a logarithmic scale to facilitate comparison across studies; and (b) the field needs to develop a better idea of whether and, if so, how each scale is relevant to different research questions about hyperarticulation. Furthermore, although our result demonstrated that vowel-related methodological factors as a group might affect vowel hyperarticulation, due to data sparsity we could not identify the specific factors affecting vowel hyperarticulation. Thus, future studies should carefully consider how their methodological decisions regarding vowels might influence the observation of vowel hyperarticulation. Regarding the design-related methodological factors, one way to overcome the potential effects of different methodological choices might be an assessment of vowel hyperarticulation in naturalistic observation from day-long home recordings. Although this might be more costly and time-consuming compared to traditional lab observations, in recent years many developmental psycholinguistics researchers have chosen this method, though not yet so much in the context of assessing vowel hyperarticulation, due to its stronger generalizability and objectivity (e.g., Hilton et al., 2022).

The generalizability of the results from this meta-analysis is subject to certain limitations. First, the predominance of participants from “Minority World Countries” (Alam, 2008), especially Australia and the United States in the present study, makes our findings’ generalizability to the majority of the world population an open question. Second, as our search and sample only included mothers’ speech to their babies, these results may not generalize to other speakers, such as fathers, siblings, or grandparents, all of whom have been observed to adopt an IDS register (Dunn & Kendrick, 1982; Fernald et al., 1989; Loukatou et al., 2022; Pancsofar & Vernon-Feagans, 2006). As the number of studies into paternal hyperarticulation to infants is expanding (Benders et al., 2021; Ferjan Ramírez, 2022; Kondaurova et al., 2023; Rosslund et al., 2022, 2024; Schwarz et al., 2024), future work in particular could assess how hyperarticulation is impacted by speaker gender and relation to the child. Another limitation of our meta-analysis is language bias due to English-based search, which might lead to the omittance of studies published in languages other than English. We recommend that future studies try to overcome this linguistic limitation. Furthermore, although our goal was to provide a comprehensive assessment of vowel hyperarticulation in IDS by examining the potential effects of various mostly theoretically motivated moderators, data sparsity and the inclusion of a large number of moderators led to confounded contrasts, and more direct investigation of the moderators in novel studies is needed to identify their contribution.

Conclusions

Our results support the hypothesis that IDS, overall, is characterized by a greater vowel space compared to ADS. While the existence of a publication bias cannot be excluded, considering a larger and more diverse sample of studies appears to alleviate it. It has been near-impossible to pinpoint specific moderators, due to data sparsity, given the many different languages and methodological choices across studies. Hence, we recommend more cross-linguistic research, consideration of methodological decisions, and targeted investigation of those factors that could shed light on the mechanisms leading to vowel hyperarticulation in IDS. We hope that the recommendations and future directions raised by the present study bring us closer to answering central questions, including the mechanisms leading to vowel hyperarticulation, as well as how this property of the infant social-communicative environment may facilitate language acquisition.

References

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

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070–3078. https://doi.org/10.1121/1.4982246

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Al-Tamimi, J. E., & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. *9th European Conference on Speech, Communication and Technology (Interspeech 2005)*. Newcastle University.

Alam, S. (2008). Majority world: Challenging the West's rhetoric of democracy. *Amerasia Journal*, *34*(1), 88–98. https://doi.org/10.17953/amer.34.1.l3176027k4q614v5

*Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). The acoustics of vowels in Japanese women's speech to infants and adults. *Proceedings of the 14th international congress on phonetic sciences* (Vol. 3, pp. 2177–2179). Regents of the University of California.

*Ashby, S. (2004). *Is infant-directed speech hyperspeech?: An acoustical analysis of speech to infants and other accommodative speech styles* [Unpublished doctoral dissertation]. University College Dublin.

*Audibert, N., & Falk, S. (2018, June). Vowel space and f0 characteristics of infant-directed singing and speech. *Proceedings of the 19th international conference on speech prosody* (Vol. 153, p. 157). ISCA.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554. https://doi.org/10.1080/01690960802299378

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22. https://doi.org/10.1006/jmla.1999.2667

Bard, E. G., Aylett, M. P., Trueswell, J., & Tanenhaus, M. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173–191). MIT Press.

Barrie, M. (2003). *Contrast in Cantonese vowels* (Toronto working papers in linguistics, 20).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

*Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*(4), 847–862. https://doi.org/10.1016/j.infbeh.2013.09.001

Benders, T. (2016). Emotion-related explanations of the vowel variability in infant-directed speech. In C. Carignan & M. D. Tyler (Eds.), *Australasian international conference on speech science and technology* (pp. 233–236). Australasian Speech Science and Technology Association.

Benders, T., StGeorge, J., & Fletcher, R. (2021). Infant-directed speech by Dutch fathers: Increased pitch variability within and across utterances. *Language Learning and Development*, *17*(3), 292–325. https://doi.org/10.1080/15475441.2021.1876698

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother–child speech. *Journal of Child Language*, *11*(3), 557–578. https://doi.org/10.1017/S030500090000595X

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language*, *11*(3), 515–522. https://doi.org/10.1017/S0305000900005924

Berry, J. (1953). *Some statistical aspects of conversational speech. Communication theory*. Butterworth.

Biringen, Z. C. (1987). Infant attention to facial expressions and facial motion. *The Journal of Genetic Psychology*, *148*(1), 127–133. https://doi.org/10.1080/00221325.1987.9914543

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

*Bohn, O. S. (2013, June). Acoustic characteristics of Danish infant directed speech. *Proceedings of meetings on acoustics ICA2013* (Vol. 19, No. 1, p. 060055). Acoustical Society of America. https://doi.org/10.1121/1.4798488

Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, *14*(4), 325–337. https://doi.org/10.1016/0167-6393(94)90026-4

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916–1924. https://doi.org/10.1121/1.412064

Bradlow, A. R., Kraus, N., & Hayes, E. (2000). Speaking clearly for learning-disabled children: Sentence perception in noise. *The Journal of the Acoustical Society of America*, *108*(5), Article 2603. https://doi.org/10.1121/1.4743692

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3–4), 255–272. https://doi.org/10.1016/S0167-6393(96)00063-5

Brookman, R., Kalashnikova, M., Conti, J., Xu Rattanasone, N., Grant, K. A., Demuth, K., & Burnham, D. (2020). Depression and anxiety in the postnatal period: An examination of infants' home language environment, vocalizations, and expressive language abilities. *Child Development*, *91*(6), e1211–e1230. https://doi.org/10.1111/cdev.13421

*Burkinshaw, K. D. (2020). *Vowel space, variability, and lexical context in infant speech perception*. University of Calgary.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

*Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), Article 1435. https://doi.org/10.1126/science.1069587

*Burnham, E. B., Wieland, E. A., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., & Dilley, L. C. (2015). Phonetic modification of vowel

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

space in storybook speech to infants up to 2 years of age. *Journal of Speech, Language, and Hearing Research*, *58*(2), 241–253. https://doi.org/10.1044/2015_JSLHR-S-13-0205

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*(1–2), 39–54. https://doi.org/10.1016/0167-6393(94)90039-6

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Castellanos, A., Benedí, J. M., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication*, *20*(1–2), 23–35. https://doi.org/10.1016/S0167-6393(96)00042-8

Chong, S. C. F., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, *12*(3), 211–232. https://doi.org/10.1002/icd.286

Christiansen, M. H., Contreras Kallens, P., & Trecca, F. (2022). Toward a comparative approach to language acquisition. *Current Directions in Psychological Science*, *31*(2), 131–138. https://doi.org/10.1177/09637214211049229

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cohn, J. F., & Elmore, M. (1988). Effect of contingent changes in mothers' affective expression on the organization of behavior in 3-month-old infants. *Infant Behavior and Development*, *11*(4), 493–505. https://doi.org/10.1016/0163-6383(88)90008-2

Cox, C. M. M., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, *7*(1), 114–133. https://doi.org/10.1038/s41562-022-01452-1

*Cox, C. M. M., Dideriksen, C., Keren-Portnoy, T., Roepstorff, A., Christiansen, M. H., & Fusaroli, R. (2023). Infant-directed speech does not always involve exaggerated vowel distinctions: Evidence from Danish. *Child Development*, *94*(6), 1672–1696. https://doi.org/10.1111/cdev.13950

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974

*Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge. https://doi.org/10.4324/9780203807002

de Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks–Evidence from intonation analysis using GToBI. *Journal of Phonetics*, *48*, 29–44. https://doi.org/10.1016/j.wocn.2014.10.008

Deffner, D., Rohrer, J. M., & McElreath, R. (2022). A causal framework for cross-cultural generalizability. *Advances in Methods and Practices in Psychological Science*, *5*(3), Article 25152459221106366. https://doi.org/10.1177/25152459221106366

Dellwo, V., Leemann, A., & Kolly, M. J. (2015). The recognition of read and spontaneous speech in local vernacular: The case of Zurich German. *Journal of Phonetics*, *48*, 13–28. https://doi.org/10.1016/j.wocn.2014.10.011

Dilley, L. C., Lehet, M., Wieland, E. A., Arjmandi, M. K., Kondaurova, M., Wang, Y., Reed, J., Svirsky, M., Houston, D., & Bergeson, T. (2020). Individual differences in mothers' spontaneous infant-directed speech predict language attainment in children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2453–2467. https://doi.org/10.1044/2020_JSLHR-19-00229

*Dilley, L. C., Wieland, E. A., Lehet, M. I., Arjmandi, M. K., Houston, D. M., & Bergeson, T. R. (2015). 2pSC15. Quality and quantity of infant-directed speech by maternal caregivers predicts later speech-language outcomes in children with cochlear implants. *Science*, *6*, Article 3.

Dissen, Y., Goldberger, J., & Keshet, J. (2019). Formant estimation and tracking: A deep learning approach. *The Journal of the Acoustical Society of America*, *145*(2), 642–653. https://doi.org/10.1121/1.5088048

Dodane, C., & Al-Tamimi, J. (2007). An acoustic comparison of vowel systems in adult-directed-speech and child-directed-speech: Evidence from French, English & Japanese. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*. Newcastle University.

*Donnellan, E., Donnelly, E., Atim, S., Benjamin, A., Biroch, H., Buryn-Weitzel, J. C., Graham, K. E., Hoffman, M., Holden, E., Jurua, M., Knapper, C. V., Lahiff, N. J., Marshall, S., Patel, H., Patricia, J., Tusiime, F., Wilke, C., & Slocombe, K. E. (2020). *A cross cultural comparison of Infant Directed Speech in British and Ugandan mothers* [Manuscript in preparation].

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press. https://doi.org/10.1093/oso/9780199215782.001.0001

Dunn, J., & Kendrick, C. (1982). The speech of two- and three-year-olds to infant siblings: 'Baby talk' and the context of communication. *Journal of Child Language*, *9*(3), 579–595. https://doi.org/10.1017/S030500090000492X

Eaves, B. S., Jr., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, *123*(6), 758–771. https://doi.org/10.1037/rev0000031

*Englund, K., & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development*, *15*(2), 139–160. https://doi.org/10.1002/icd.445

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Applied Psycholinguistics*, *39*(1), 67–87. https://doi.org/10.1017/S0142716417000480

Erber, N. P. (1974). Visual perception of speech by deaf children: Recent developments and continuing needs. *The Journal of Speech and Hearing Disorders*, *39*(2), 178–185. https://doi.org/10.1044/jshd.3902.178

Erickson, D., Iwata, R., Endo, M., & Fujino, A. (2004, March). *Effect of tone height on jaw and tongue articulation in Mandarin Chinese* [Paper presentation]. International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Ernestus, M., Hanique, I., & Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics*, *48*, 60–75. https://doi.org/10.1016/j.wocn.2014.08.001

Farran, L. K., Yoo, H., Lee, C. C., Bowman, D. D., & Oller, D. K. (2019). Temporal coordination in mother–infant vocal interaction: A cross-cultural comparison. *Frontiers in Psychology*, *10*, Article 2374. https://doi.org/10.3389/fpsyg.2019.02374

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*(1), 259–271. https://doi.org/10.1121/1.1482078

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1241–1255. https://doi.org/10.1044/1092-4388(2007/087)

Ferjan Ramírez, N. (2022). Fathers' infant-directed speech and its effects on child language development. *Language and Linguistics Compass*, *16*(1), Article e12448. https://doi.org/10.1111/lnc3.12448

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, *9*(10), Article e109019. https://doi.org/10.1371/journal.pone.0109019

Gandour, J. (1977). On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica*, *34*(1), 54–65. https://doi.org/10.1159/000259869

Gasparini, L., Tsuji, S., & Bergmann, C. (2022). Ten easy steps to conducting transparent, reproducible meta-analyses for infant researchers. *Infancy*, *27*(4), 736–764. https://doi.org/10.1111/infa.12470

*Gergely, A., Faragó, T., Galambos, Á., & Topál, J. (2017). Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: An acoustic analysis. *Scientific Reports*, *7*(1), Article 13739. https://doi.org/10.1038/s41598-017-13883-2

Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1529–1542. https://doi.org/10.1044/1092-4388(2010/09-0005)

Han, M. S. (1966). *Studies in the phonology of Asian languages IV: Vietnamese vowels*. Acoustics Phonetics Research Laboratory, University of Southern California.

Han, M. S., de Jong, N. H., Kager, R., & Bertolini, A. B. (2018). Infant-directed speech is not always slower: Cross-linguistic evidence from Dutch and Mandarin Chinese. *Proceedings of the 42nd annual Boston University conference on language development* (Vol. 2, No. 1, pp. 331–344). Cascadilla Press.

*Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language*, *44*(5), 1140–1162. https://doi.org/10.1017/S0305000916000520

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139–2152. https://doi.org/10.1121/1.3623753

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*, *116*(5), 3108–3118. https://doi.org/10.1121/1.1806826

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, *6*(11), 1545–1556. https://doi.org/10.1038/s41562-022-01410-x

Hu, F. (2004, March). *Tonal effect on vowel articulation in a tone language* [Paper presentation]. International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Jylänki, P., Vanhatalo, J., & Vehtari, A. (2011). Robust Gaussian process regression with a student-t likelihood. *Journal of Machine Learning Research*, *12*(11), 3227–3257.

*Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language*, *45*(5), 1035–1053. https://doi.org/10.1017/S0305000917000629

*Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society Open Science*, *4*(8), Article 170306. https://doi.org/10.1098/rsos.170306

Kalashnikova, M., & Carreiras, M. (2022). Input quality and speech perception development in bilingual infants' first year of life. *Child Development*, *93*(1), e32–e46. https://doi.org/10.1111/cdev.13686

*Kalashnikova, M., Goswami, U., & Burnham, D. (2018). Mothers speak differently to infants at-risk for dyslexia. *Developmental Science*, *21*(1), Article e12487. https://doi.org/10.1111/desc.12487

*Kalashnikova, M., Goswami, U., & Burnham, D. (2020). Infant-directed speech to infants at risk for dyslexia: A novel cross-dyad design. *Infancy*, *25*(3), 286–303. https://doi.org/10.1111/infa.12329

Kaplan, P. S., Danko, C. M., Cejka, A. M., & Everhart, K. D. (2015). Maternal depression and the learning-promoting effects of infant-directed speech: Roles of maternal sensitivity, depression diagnosis, and speech acoustic cues. *Infant Behavior and Development*, *41*, 52–63. https://doi.org/10.1016/j.infbeh.2015.06.011

Kavanaugh, R. D., & Jirkovsky, A. M. (1982). Parental speech to young children: A longitudinal analysis. *Merrill-Palmer Quarterly*, *28*(2), 297–311.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110. https://doi.org/10.1207/S15327078IN0401_5

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392. https://doi.org/10.1016/S0163-6383(02)00086-3

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. American Psychological Association. https://doi.org/10.1037/10693-000

Knoll, M., & Scharrer, L. (2007). *Acoustic and affective comparisons of natural and imaginary infant-, foreigner-and adult-directed speech* [Conference session]. Eighth Annual Conference of the International Speech Communication Association, Antwerp, Belgium.

Kondaurova, M. V., Bergeson, T. R., & Dilley, L. C. (2012). Effects of deafness on acoustic characteristics of American English tense/lax vowels in maternal speech to infants. *The Journal of the Acoustical Society of America*, *132*(2), 1039–1049. https://doi.org/10.1121/1.4728169

Kondaurova, M. V., VanDam, M., Zheng, Q., & Welikson, B. (2023). Fathers' unmodulated prosody in child-directed speech. *The Journal of the Acoustical Society of America*, *154*(6), 3556–3567. https://doi.org/10.1121/10.0022571

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. https://doi.org/10.1121/1.1635842

*Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Thomson Wadsworth.

Ladefoged, P., & Maddieson, I. (1996). *Recording the phonetic structures of endangered languages* (UCLA working papers in phonetics, 1–7).

Lam, C., & Kitamura, C. (2010). Maternal interactions with a hearing and hearing-impaired twin: Similarities and differences in speech input, interaction quality, and word production. *Journal of Speech, Language, and Hearing Research*, *53*(3), 543–555. https://doi.org/10.1044/1092-4388(2010/09-0126)

*Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science*, *15*(2), 212–221. https://doi.org/10.1111/j.1467-7687.2011.01118.x

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Lenth, R. V. (2016). Least-Squares means: the R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist*, *48*(12), 1181–1209. https://doi.org/10.1037/0003-066X.48.12.1181

*Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1–F10. https://doi.org/10.1111/1467-7687.00275

*Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. *Journal of Child Language*, *36*(4), 909–922. https://doi.org/10.1017/S030500090800929X

Lorge, I., & Katsos, N. (2019). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approaches to Bilingualism*, *9*(3), 376–397. https://doi.org/10.1075/lab.16054.lor

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2022). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, *49*(6), 1173–1192. https://doi.org/10.1017/S0305000921000623

Lovcevic, I., Benders, T., Fusaroli, R., & Tsuji, S. (2020). *Vowel hyperarticulation in infant-directed speech—A meta-analysis*. https://osf.io/n29mr

Lovcevic, I., Benders, T., Tsuji, S., Fusaroli, R., Dideriksen, C., & Fowler, E. (2021). *Vowel hyperarticulation in infant-directed speech—A meta-analysis-Step 2*. https://osf.io/3jmxs

Lovcevic, I., Burnham, D., & Kalashnikova, M. (2024). Infants' lexical processing: Independent contributions of attentional and clarity cues. *Language Learning and Development*, *20*(1), 1–18. https://doi.org/10.1080/15475441.2022.2149402

*Lovcevic, I., Kalashnikova, M., & Burnham, D. (2020). Acoustic features of infant-directed speech to infants with hearing loss. *The Journal of the Acoustical Society of America*, *148*(6), 3399–3416. https://doi.org/10.1121/10.0002641

*Marklund, E., & Gustavsson, L. (2020). The dynamics of vowel hypo-and hyperarticulation in Swedish infant-directed speech to 12-month-olds. *Frontiers in Communication*, *5*, Article 523768. https://doi.org/10.3389/fcomm.2020.523768

Marklund, U., Marklund, E., & Gustavsson, L. (2021). Relationship between parent vowel hyperarticulation in infant-directed speech and infant phonetic complexity on the level of conversational turns. *Frontiers in Psychology*, *12*, Article 688242. https://doi.org/10.3389/fpsyg.2021.688242

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Applied Statistics*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

Mathur, M. B., & VanderWeele, T. J. (2021). Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers. *Research Synthesis Methods*, *12*(2), 176–191. https://doi.org/10.1002/jrsm.1464

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, *129*(2), 362–378. https://doi.org/10.1016/j.cognition.2013.07.015

*Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, *166*, 84–93. https://doi.org/10.1016/j.cognition.2017.05.003

Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, *96*(1), 40–55. https://doi.org/10.1121/1.410492

Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, *139*(3), 1272–1281. https://doi.org/10.1121/1.4944634

Nearey, T. M., Assmann, P. F., & Hillenbrand, J. M. (2002). Evaluation of a strategy for automatic formant tracking. *The Journal of the Acoustical Society of America*, *112*(5), Article 2323. https://doi.org/10.1121/1.4779372

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), Article 89. https://doi.org/10.1186/s13643-021-01626-4

Pancsofar, N., & Vernon-Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology*, *27*(6), 571–587. https://doi.org/10.1016/j.appdev.2006.08.003

*Panneton, R., Cristia, A., Taylor, C., & Christine, M. O. O. N. (2024). Positive valence contributes to hyperarticulation in maternal speech to infants and puppies. *Journal of Child Language*, *51*(5), 1230–1240. https://doi.org/10.1017/S0305000923000296

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pearson, R. M., Heron, J., Melotti, R., Joinson, C., Stein, A., Ramchandani, P. G., & Evans, J. (2011). The association between observed non-verbal maternal responses at 12 months and later infant development at 18 months and IQ at 4 years: A longitudinal study. *Infant Behavior and Development*, *34*(4), 525–533. https://doi.org/10.1016/j.infbeh.2011.07.003

Peter, V., Kalashnikova, M., Santos, A., & Burnham, D. (2016). Mature neural responses to infant-directed speech but not adult-directed speech in pre-verbal infants. *Scientific Reports*, *6*(1), Article 34273. https://doi.org/10.1038/srep34273

Piazza, G., Martin, C. D., & Kalashnikova, M. (2022). The acoustic features and didactic function of foreigner-directed speech: A scoping review. *Journal of Speech, Language, and Hearing Research*, *65*(8), 2896–2918. https://doi.org/10.1044/2022_JSLHR-21-00609

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

*Rattanasone, N. X., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, *41*(5), 332–343. https://doi.org/10.1016/j.wocn.2013.06.001

Rojczyk, A. (2019). Quality and duration of unstressed vowels in Polish. *Lingua*, *217*, 80–89. https://doi.org/10.1016/j.lingua.2018.10.012

Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research*, *39*(6), 1159–1170. https://doi.org/10.1044/jshr.3906.1159

*Rosslund, A., Hagelund, S., Mayor, J., & Kartushina, N. (2024). Mothers' and fathers' infant-directed speech have similar acoustic properties, but these are not associated with direct or indirect measures of word comprehension in 8-month-old infants. *Journal of Child Language*, *51*(6), 1424–1449. https://doi.org/10.1017/S0305000923000557

*Rosslund, A., Mayor, J., Óturai, G., & Kartushina, N. (2022). Parents' hyper-pitch and low vowel category variability in infant-directed speech are associated with 18-month-old toddlers' expressive vocabulary. *Language Development Research*, *2*(1), 223–267. https://doi.org/10.34842/2022.0547

Schwarz, I. C., Marklund, E., Marklund, U., Gustavsson, L., & Lam-Cassettari, C. (2024). Affect in infant-directed speech of Swedish-speaking mothers and fathers to 3-, 6-, 9-, and 12-month-old infants. *Language Learning and Development*, *20*(2), 145–157. https://doi.org/10.1080/15475441.2023.2239801

Shaw, J. A., Chen, W. R., Proctor, M. I., & Derrick, D. (2016). Influences of tone on vowel articulation in Mandarin Chinese. *Journal of Speech, Language, and Hearing Research*, *59*(6), S1566–S1574. https://doi.org/10.1044/2015_JSLHR-S-15-0031

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

*Shochi, T., Sekiyama, K., Lees, N., Boyce, M., Göcke, R., & Burnham, D. (2009). Auditory-visual infant directed speech in Japanese and English. *Proceedings of International Conference on Audio-Visual Speech Processing* (pp. 107–112).

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*(3), 365–394. https://doi.org/10.1207/S15327078IN0303_5

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3), 1677–1688. https://doi.org/10.1121/1.2000788

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, *128*(1), 389–400. https://doi.org/10.1121/1.3419786

*Starling, G. (2019). *Vowel production in infant-directed speech: An assessment of hyperarticulation and distributional learning* [Unpublished doctoral dissertation]. The University of Edinburgh.

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviours. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187–213). Wiley.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15. https://doi.org/10.1017/S0305000900005092

Stokes, S. F., Bleses, D., Basbøll, H., & Lambertsen, C. (2012). Statistical learning in emerging lexicons: The case of Danish. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1265–1273. https://doi.org/10.1044/1092-4388(2012/10-0291)

Summerfield, Q., & Assmann, P. F. (1989). Auditory enhancement and the perception of concurrent vowels. *Perception & Psychophysics*, *45*(6), 529–536. https://doi.org/10.3758/BF03208060

Swanson, L. A., Leonard, L. B., & Gandour, J. (1992). Vowel duration in mothers' speech to young children. *Journal of Speech and Hearing Research*, *35*(3), 617–625. https://doi.org/10.1044/jshr.3503.617

*Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, *142*(2), 493–503. https://doi.org/10.1121/1.4995998

Torng, P. C. (2000). *Supralaryngeal articulator movements and laryngeal control in Mandarin Chinese tonal production*. University of Illinois at Urbana-Champaign.

*Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication*, *49*(1), 2–7. https://doi.org/10.1016/j.specom.2006.10.003

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Volkmann, J., Stevens, S. S., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, *8*(3), Article 208. https://doi.org/10.1121/1.1901999

Wang, Y., Houston, D. M., Seidl, A., Frühholz, S., & Belin, P. (2018). Acoustic properties of infant-directed speech. In S. Frühholz, & P. Belin (Eds.), *The Oxford handbook of voice perception* (Vol. *1*, 93–116).

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*, *88*(5), 1513–1526. https://doi.org/10.1111/cdev.12882

*Weirich, M., & Simpson, A. (2019). Effects of gender, parental role, and time on infant-and adult-directed read and spontaneous speech. *Journal of Speech, Language, and Hearing Research*, *62*(11), 4001–4014. https://doi.org/10.1044/2019_JSLHR-S-19-0047

Whalen, D. H., Chen, W. R., Shadle, C. H., & Fulop, S. A. (2022). Formants are easy to measure; resonances, not so much: Lessons from Klatt (1986). *The Journal of the Acoustical Society of America*, *152*(2), 933–941. https://doi.org/10.1121/10.0013410

*Wieland, E. A., Burnham, E. B., Kondaurova, M., Bergeson, T. R., & Dilley, L. C. (2015). Vowel space characteristics of speech directed to children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*, *58*(2), 254–267. https://doi.org/10.1044/2015_JSLHR-S-13-0250

*Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog-and infant-directed speech. *Anthrozoos*, *26*(3), 373–380. https://doi.org/10.2752/175303713X13697429463592

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917–1007. https://doi.org/10.1214/17-BA1091

Zee, E. (1980). Tone and vowel quality. *Journal of Phonetics*, *8*(3), 247–258. https://doi.org/10.1016/S0095-4470(19)31474-3

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Tusler, L., & Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science*, *14*(3), 566–581. https://doi.org/10.1111/j.1467-7687.2010.01004.x

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *The Journal of the Acoustical Society of America*, *33*(2), Article 248. https://doi.org/10.1121/1.1908630

Received August 8, 2023
Revision received March 7, 2025
Accepted April 7, 2025 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.